https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

A global kilometre-scale tropical cyclone inner-core vector wind field dataset from CYGNSS observations

Xinhai Han[1,2,3], Xiaohui Li[2], Jingsong Yang[2,3,4], Hanyue Ni[2], Zeyi Niu[1], and Wei Huang[1]

[1]Shanghai Typhoon Institute, and Key Laboratory of Numerical Modeling for Tropical Cyclone of the China Meteorological Administration, Shanghai 200030, China
[2]State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China
[3]School of Oceanography, Shanghai Jiao Tong University, Shanghai 200030, China
[4]Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China

Correspondence: Xiaohui Li (lixiaohui1991@live.cn) and Jingsong Yang (jsyang@sio.org.cn)

Abstract. Tropical cyclone (TC) inner-core vector wind fields are essential for intensity forecasting, storm surge prediction, and structural climatology. The Cyclone Global Navigation Satellite System (CYGNSS) constellation provides dense temporal sampling and L-band precipitation-penetrating capability over the tropical belt, but its sparse, scalar wind speed retrievals have not been fully assimilated to produce kilometre-scale TC inner-core vector wind fields with global multi-basin coverage. This paper presents the QiFeng-CYGNSS dataset, which combines CYGNSS observations with a physics-guided score-based diffusion assimilation framework to reconstruct spatially complete 10 m vector wind fields from sparse scalar wind speed observations. The dataset covers 249 TCs across six active global basins during January 2020–September 2022, providing 1.5 km resolution vector wind fields at every IBTrACS reporting time with available CYGNSS coverage (4955 snapshots in total), accompanied by observation metadata and pixel-level ensemble uncertainty estimates for 138 major-hurricane snapshots. Independent validation against spaceborne C-band synthetic aperture radar, airborne Tail Doppler Radar, and GPS dropsondes indicates that the reconstructions represent TC inner-core structures at kilometre scales, reducing the absolute $V_{\max}$ bias relative to ERA5 and CCMP by $\sim$79 % and $\sim$75 %, respectively, on the full sample. The dataset is freely available from Zenodo at https://doi.org/10.5281/zenodo.20046109 (Han et al., 2026b).

1 Introduction

TC inner-core 10 m vector wind fields are fundamental to intensity forecasting (Kaplan et al., 2010), storm surge prediction (Resio et al., 2009), TC–ocean interaction studies, and boundary layer dynamics (Kepert, 2001). The spatial structure of the wind field directly determines storm surge amplitude (Resio et al., 2009), sea surface cooling intensity (Vincent et al., 2012), and the environmental energy exchange that modulates rapid intensification (Rios-Berrios and Torn, 2017); inner-core radial profiles also dominate TC size climatology (Knaff et al., 2014). However, no existing observational product jointly delivers kilometre-scale resolution, vector wind components, and global multi-basin coverage of the TC inner core. The underlying trade-offs among resolution, variable type, and spatiotemporal coverage are detailed in Table 1.

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

The CYGNSS constellation has a relatively high revisit rate (Ruf et al., 2016), but its orbital coverage is stochastic, and the raw retrievals provide only sparse scalar wind speeds without directional information (Ruf et al., 2019). Recent deep learning approaches have demonstrated the potential of generative models for TC wind field reconstruction from sparse observations: generative adversarial networks for reconstructing SAR wind speeds within TCs (Han et al., 2023), adversarial learning for multi-level wind-speed structure reconstruction from dropsonde profiles (Han et al., 2025), and physics-guided score-based diffusion for three-dimensional TC structure reconstruction from multi-variable sparse inputs (Han et al., 2026a). However, SAR-based methods do not address wind direction reconstruction (SAR itself does not independently retrieve direction), while dropsonde- and multi-variable-based methods rely on input observations that carry intrinsic directional information; none has tackled the fundamentally harder problem of reconstructing vector wind fields from purely scalar, direction-free observations. Therefore, how to reconstruct kilometre-scale spatially continuous vector wind fields from these incomplete scalar observations, while reliably identifying moments of insufficient observation coverage, remains a significant technical challenge.

Table 1 compares the existing TC surface wind products quantitatively. ERA5 (Hersbach et al., 2020) and CCMP (Atlas et al., 2011; Mears et al., 2022) offer long-term global vector wind coverage with high temporal availability (hourly and 6-hourly, respectively), though their 0.25° resolution smooths inner-core structures and underestimates $V_{\max}$ for major hurricanes by 40–60 % relative to IBTrACS (cf. Sect. 4.3). The CYGNSS L3 Storm-Centric Grid (SCG) product (Mayers et al., 2023) exploits CYGNSS's precipitation-penetrating capability to provide 0.1° storm-relative wind speed fields, improving resolution relative to reanalyses; however, it reports scalar wind speed only (no direction) and retains gaps from orbital sparsity. SAR wind fields (Mouche et al., 2019; Combot et al., 2020) provide the highest spatial resolution (500 m–3 km) among satellite products, though each TC is revisited only episodically. Liu et al. (2025) applied a globally consistent parametric correction to ERA5, extending TC wind field coverage to all basins, though the 0.25° resolution and parametric model assumptions limit inner-core structure representation. Aircraft reconnaissance (dropsondes (Franklin et al., 2003), TDR (Rogers et al., 2012), SFMR (Uhlhorn et al., 2007)) provides high-accuracy local observations; its basin coverage is limited to the North Atlantic and Eastern Pacific (Aberson et al., 2006), with dropsondes yielding sparse one-dimensional profiles and TDR composites exhibiting spatiotemporal discontinuities.

To fill this observational gap, this paper presents the QiFeng-CYGNSS dataset. This product exploits the dense temporal sampling of CYGNSS (median revisit interval ∼2.8 h within the tropical belt; Ruf et al., 2016, 2019) and its L-band precipitation-penetrating capability. A physics-guided score-based diffusion assimilation framework (Han et al., 2026c) reconstructs complete two-dimensional vector wind fields from the sparse CYGNSS scalar observations. The dataset covers January 2020–September 2022, spanning 249 TCs at every International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al., 2010) reporting time (primarily 6-hourly at 00/06/12/18 UTC, with some agencies providing 3-hourly reports). Reporting times at which no CYGNSS observation falls within the spatiotemporal matching window (±3 h, ±192 km) are excluded from reconstruction, yielding 4955 reconstructed snapshots at 1.5 km resolution (256×256 pixels, 384 km × 384 km domain), each providing 10 m u/v wind components. Each record includes an Observation Coverage Sufficiency (OCS) quality flag (Sect. 2.3); among these, 1960 snapshots (39.6 %) passing OCS screening have reliable reconstruction quality. Training

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

data (Hurricane Weather Research and Forecasting model (HWRF) 2023–2024) and reconstruction period (January 2020–September 2022) are strictly separated in time to exclude data leakage.

Independent validation on the OCS-pass subset yields pixel-level RMSE of 5.58 m s^{-1} against C-band SAR (47 matched cases) and 6.9 m s^{-1} against airborne TDR (23 matched cases), with reduced $V_{\max}$ bias relative to ERA5 and CCMP for moderate-to-strong TCs (see Sect. 4). This dataset provides systematic kilometre-scale vector wind field observations for basins lacking aircraft reconnaissance (Western Pacific (WP), North Indian (NI), South Indian (SI), South Pacific (SP)), supporting storm surge numerical modelling, TC structural climatology, and data-driven forecast model training applications.

Table 1. Comparison of existing global/near-global TC surface wind field products with the QiFeng-CYGNSS dataset presented in this study. "Coverage" indicates spatial completeness of each individual TC snapshot. "Direction" indicates whether wind direction (u, v) is reported.

Product	Resolution	Direction	Coverage	Temporal	Basins	Period
ERA5 (Hersbach et al., 2020)	0.25° (∼25 km)	Yes	Full (global)	Hourly	Global	1940–
CCMP v3.1 (Mears et al., 2022)	0.25° (∼25 km)	Yes	Ocean only	6-hourly	Global	1993–
L3 SCG v1.0 (Mayers et al., 2023)	0.1° (∼11 km)	No	Sparse (obs-only)	6-hourly	NA, EP	2018–
Liu et al. (Liu et al., 2025)	0.25° (∼25 km)	Yes	Full (global)	Hourly	Global	1993–2022
C-band SAR[a] (Mouche et al., 2019)	500 m–3 km	No[b]	Partial (swath)	Episodic	Global	2014–
SFMR (airborne) (Uhlhorn et al., 2007)	Along-track	No	1-D profile	Episodic	NA, EP	1998–
QiFeng-CYGNSS	1.5 km	Yes	Full (384 km domain)	Per IBTrACS time[c]	All 6 basins	2020–2022[d]

[a] Sentinel-1A/B, RADARSAT-2, and RADARSAT Constellation Mission (RCM); both NOAA STAR and CyclObs maintain operational TC SAR wind product archives. [b] SAR does not independently retrieve wind direction; operational SAR wind products use direction from numerical models. [c] Reconstruction is attempted at every IBTrACS reporting time (primarily 6-hourly); only snapshots passing the Observation Coverage Sufficiency (OCS) criterion (∼40 %) are recommended for quantitative use. [d] Specifically January 2020–September 2022 (the data acquisition window). NA = North Atlantic; EP = Eastern Pacific.

Figure 1 illustrates the differences between the QiFeng-CYGNSS product, the existing CYGNSS-derived product (L3 SCG), and the reanalysis product (CCMP) for the same TC snapshot. QiFeng provides a spatially complete kilometre-scale vector wind field in which the eye and asymmetric eyewall structures are discernible, contrasting with the sparse SCG (extensive missing values) and the resolution-limited CCMP. For this case (IBTrACS $V_{\max}$ = 90 kt ≈ 46.3 m s^{-1}), QiFeng's reconstructed $V_{\max}$ is close to the best-track value, whereas CCMP underestimates the intensity. QiFeng's performance is bounded by the sufficiency of CYGNSS observational coverage; for high-intensity TCs, $V_{\max}$ may be underestimated because of L-band GNSS-R signal sensitivity saturation at high wind speeds (see Known Limitations in Sect. 5.5).

The OCS flag is determined entirely from CYGNSS observation attributes prior to reconstruction, without reference to any external dataset (details in Sect. 2.3).

Figure 2 illustrates how QiFeng-CYGNSS tracks inner-core structural evolution during the intensification of Hurricane Ian (2022). At the three OCS-pass IBTrACS reporting times shown, IBTrACS $V_{\max}$ intensified from 25.7 m s^{-1} (50 kt) to 54.0 m s^{-1} (105 kt); QiFeng's reconstructed $V_{\max}$ at these OCS-pass moments correspondingly increased from 35.6 m s^{-1} to

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

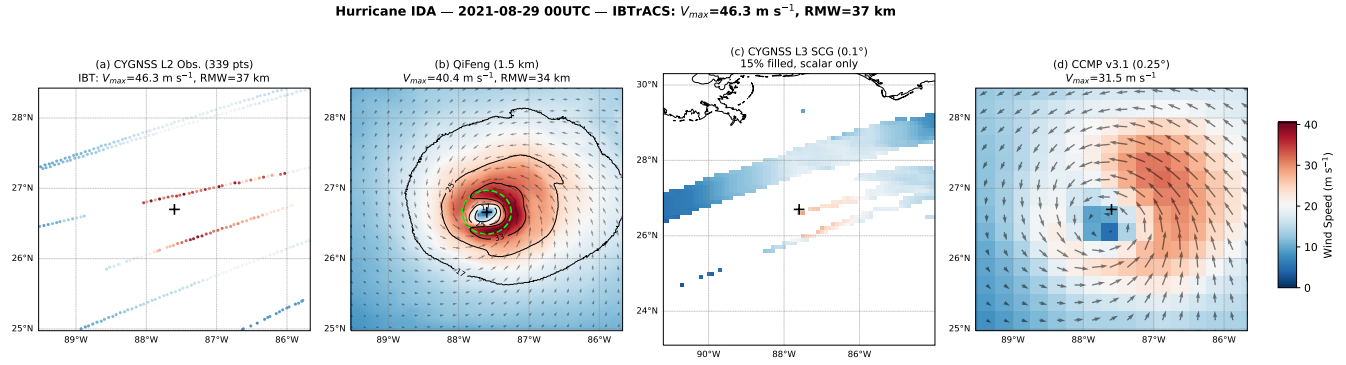

Figure 1. Multi-source wind field comparison for Hurricane Ida (2021-08-29 00 UTC, IBTrACS $V_{\max} = 46.3\,\mathrm{m\,s^{-1}}$ (90 kt), USA-agency radius of maximum wind (RMW) = 37 km, Category 2). (a) CYGNSS L2 Young Seas with Limited Fetch (YSLF) specular reflection points (339 points within ±192 km and ±3 h of the TC centre) used as input; (b) QiFeng-CYGNSS reconstruction (1.5 km, full vector field with wind direction arrows); (c) CYGNSS L3 Storm-Centric Grid product (0.1°, scalar wind speed only, extensive missing data due to orbital sparsity); (d) CCMP v3.1 (0.25°, vector field). QiFeng depicts the TC eye, asymmetric eyewall, and inner-core wind direction structure from the same CYGNSS observations, which the other two products cannot provide.

51.6 m s^{-1}, tracking the overall intensification trend. At the earliest time step (50 kt, tropical-storm intensity), QiFeng overestimates $V_{\max}$ relative to IBTrACS (35.6 vs. 25.7 m s^{-1}), reflecting the diffusion prior's tendency to generate climatological-mean structures that exceed weak-TC intensities (see intensity-stratified performance in Sect. 4.3). QiFeng provides reliable reconstruction only when CYGNSS observation coverage satisfies the OCS criterion; about 55 % of IBTrACS reporting times passed OCS screening for Ian during this period.

Ian subsequently continued intensifying to Category 4 (IBTrACS $V_{\max} = 61.7\,\mathrm{m\,s^{-1}} \approx 120$ kt, 28 September 06 UTC), while QiFeng's reconstructed $V_{\max}$ at that time was 51.4 m s^{-1} and did not fully capture the intensity. Several factors contribute to this underestimation. CYGNSS orbits may not have sufficiently sampled the eyewall maximum wind region at that time, leaving the inner core under-constrained. Even where the eyewall is observed, CYGNSS YSLF retrievals show rapidly increasing uncertainty at high wind speeds: the L-band GNSS-R signal begins to lose sensitivity above 30 m s^{-1} and approaches saturation above 50 m s^{-1}, with retrieval errors growing from ∼2 m s^{-1} (below 30 m s^{-1}) to 8–13 m s^{-1} (above 50 m s^{-1}) (Ruf et al., 2019, 2024; Han et al., 2026c), so the reported wind speed may be biased low. The diffusion prior partially compensates for this low bias by learning the statistical distribution of TC intensities (Han et al., 2026c), but under the combined effects of sparse observations and increased retrieval uncertainty the reconstruction still underestimates extreme intensities (see Known Limitations in Sect. 5.5). CCMP, limited by its 0.25° resolution, cannot adequately resolve inner-core fine-scale structures at these time steps, with its $V_{\max}$ consistently below 20 m s^{-1}.

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

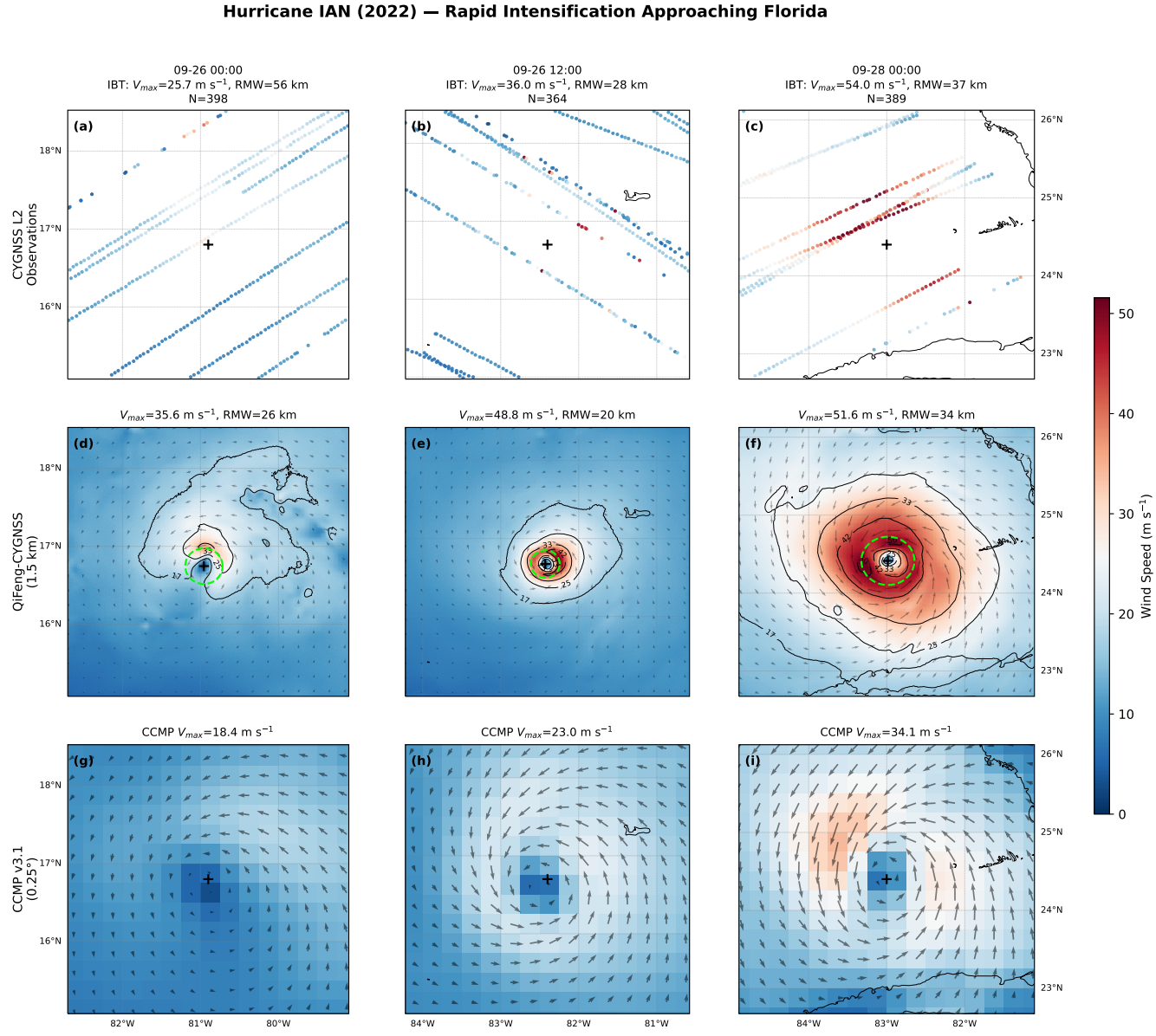

Figure 2. Time-series comparison of Hurricane Ian (2022) during intensification. Row 1: CYGNSS L2 observations used as input; Row 2: QiFeng-CYGNSS vector wind field (wind speed + direction arrows); Row 3: CCMP v3.1 wind speed (0.25°). Three OCS-pass time steps are shown (left to right): 26 September 00 UTC, 26 September 12 UTC, and 28 September 00 UTC. IBTrACS $V_{\max}$ for each step: 25.7 m s^{-1} (50 kt), 36.0 m s^{-1} (70 kt), 54.0 m s^{-1} (105 kt). QiFeng reconstructed $V_{\max}$: 35.6, 48.8, 51.6 m s^{-1}. The "+" marker indicates the IBTrACS TC centre. QiFeng represents the intensification and eyewall development that CCMP does not resolve due to its coarse resolution. Note: these are not consecutive 6-hourly steps—intervening time steps did not pass OCS screening.

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

2 Data and methods

The complete technical details of the reconstruction framework are described in Han et al. (2026c). This section documents the data processing steps and key algorithmic elements necessary for understanding the dataset generation and for reproducing the results.

2.1 Input data

The observational input for the QiFeng-CYGNSS dataset is the CYGNSS Level-2 Climate Data Record (CDR) v3.2 Young Seas with Limited Fetch (YSLF) wind speed variable. The CYGNSS constellation consists of eight microsatellites that use Global Navigation Satellite System Reflectometry (GNSS-R) to measure ocean surface wind speed. The YSLF algorithm is optimised for the young sea surface conditions within TC inner cores, providing better wind speeds than the standard Fully Developed Seas retrieval at high wind speeds ($>30\,\mathrm{m\,s^{-1}}$) (Ruf et al., 2019, 2024). This variable is also used by the CYGNSS L3 Storm-Centric Grid product (Mayers et al., 2023). Each CYGNSS specular reflection point provides a single instantaneous scalar wind speed observation (footprint approximately 25×25 km), without any directional information. YSLF nevertheless exhibits substantial L-band GNSS-R signal sensitivity loss at high wind speeds: retrieval uncertainty rises from $\sim2\,\mathrm{m\,s^{-1}}$ below $30\,\mathrm{m\,s^{-1}}$ to 8–13 $\mathrm{m\,s^{-1}}$ above $50\,\mathrm{m\,s^{-1}}$ as the GNSS-R sea-surface scattering cross-section saturates (Ruf et al., 2019, 2024). The QiFeng framework partially mitigates this issue by treating CYGNSS observations as likelihood constraints (rather than direct assignment), combined with the diffusion prior's statistical distribution of high-wind regions (Han et al., 2026c).

For each reporting time of every TC in IBTrACS (standard 6-hourly intervals, with some agencies providing 3-hourly reports), we use the IBTrACS best-track position and CYGNSS L2 orbital data to collect all YSLF specular reflection points within ±192 km of the TC centre and a ±3 h time window. TC motion correction is applied to eliminate trailing effects caused by TC translation, and inter-orbit quality control is performed to remove anomalous orbits, forming the observational constraint set for each snapshot. All snapshots undergo reconstruction, but reconstruction results are considered reliable only when observations satisfy the OCS criterion (specific criteria described in Sect. 2.3). Therefore, each snapshot in the dataset is labelled with a `meets_ocs` quality flag for user screening.

2.2 Reconstruction framework

The reconstruction framework consists of the following key elements:

1. **Diffusion prior**: A score-based generative model trained in two stages. The first stage pre-trains on ∼55 000 TC snapshots from a 30-year (1979–2009) NICAM global cloud-resolving simulation archive (Matsuoka et al., 2023), providing broad cross-basin and multi-intensity coverage at ∼14 km native resolution interpolated to the common 1.5 km grid. The second stage fine-tunes on HWRF 1.5 km TC inner-core u_{10}/v_{10} analysis fields. The HWRF training set is drawn from TCs during 2023–2024, strictly separated by storm name and time from the January 2020–September 2022 product period. The model thereby learns the joint probability distribution $p(\mathbf{u},\mathbf{v})$ of physically consistent TC wind field struc-

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

tures, including characteristic eyewall shapes, radial profiles, and asymmetric features. The complete two-stage training procedure is detailed in Han et al. (2026c).

2. **Nonlinear observation operator**: $H(\mathbf{u}) = \sqrt{u^2 + v^2}$. At each reverse diffusion step, likelihood gradients from the scalar CYGNSS observations are injected into the reconstruction process using:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t) \approx -\frac{1}{2\sigma_y^2} \nabla_{\mathbf{x}_t} \|H(\hat{\mathbf{x}}_0(\mathbf{x}_t)) - \mathbf{y}\|^2 \quad (1)$$

where $\mathbf{y}$ represents the CYGNSS wind speed observations, $\mathbf{x}_t$ is the noisy state at diffusion step t, and $\hat{\mathbf{x}}_0$ is the clean-state estimate derived from the score function. Eq. (1) gives the schematic Gaussian form. The actual implementation uses an observation-wise effective variance and a Huber robust likelihood; the complete formulation is detailed in Han et al. (2026c).

3. **TC boundary layer physical constraints**: Three gradient-based constraints are imposed during the reverse diffusion sampling process: (a) low-level divergence suppression (limiting $|\nabla \cdot \mathbf{u}|$ to be consistent with observed TC boundary layer divergence magnitudes), (b) bounded inflow angle ($10°$–$35°$, encompassing the climatological range of TC boundary layer inflow angles; Zhang and Uhlhorn, 2012), and (c) positive cyclonic vorticity (ensuring physically correct rotational sense). These constraints collectively resolve the directional ambiguity inherent in scalar-to-vector reconstruction.

2.3 Observation Coverage Sufficiency (OCS) criterion

A composite criterion based on three metrics is applied to flag snapshots with sufficient observational constraints for reliable reconstruction: (a) number of independent observations ($n_{\text{obs}} \geq 300$), (b) inner-core spatial coverage (at least 7 super-grid cells containing observations within 100 km of the TC centre), and (c) azimuthal coverage fraction (≥ 5 out of 8 azimuthal octants containing observations, i.e. az_cov ≥ 0.625). This criterion is determined entirely from CYGNSS observation attributes, without reference to any external dataset or to the reconstructed wind field itself. The thresholds were selected based on sensitivity analysis showing a distinct inflection in $V_{\max}$ correlation at $n_{\text{obs}} \approx 300$ (Han et al., 2026c).

2.4 Ensemble uncertainty estimation

For each snapshot, $N = 16$ ensemble members are generated by varying the initial noise seed of the reverse diffusion process. The pixel-wise wind speed ensemble standard deviation serves as the uncertainty estimate. In observed regions, ensemble spread reflects observation noise propagation; in data-void regions, spread reflects the inherent multi-modality of the prior distribution.

3 Data records

The dataset is available through Zenodo (Han et al., 2026b). The total size is approximately 2 GB. Data are provided in NetCDF-4 format (CF-1.8 convention), with one file per TC containing all time steps of reconstructed wind fields, observation information, OCS quality flags, and (where applicable) ensemble uncertainty estimates.

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

3.1 File organisation

Data are organised by storm. Each TC has one NetCDF file, named `{STORM_NAME}_{YEAR}.nc` (e.g. `IAN_2022.nc`). Within each file, the `time` dimension indexes all IBTrACS reporting time snapshots for that TC (including both OCS-pass and OCS-fail). Additionally, ensemble uncertainty estimates for 138 sampled major hurricane snapshots (Category 3 and above; 70 OCS-pass and 68 OCS-fail; sampling details in Sect. 4.6) are provided in a separate file `ensemble_uncertainty.nc`, independent of the per-TC files. Users may filter the ensemble subset to OCS-pass cases by cross-referencing the `meets_ocs` flag in the matching per-TC NetCDF file (matched by `storm_name` and `time`).

3.2 Variables

Table 2 lists the variables contained in each per-TC NetCDF file.

Table 2. Variables contained in each per-TC NetCDF file of the QiFeng-CYGNSS dataset.

Variable	Dimensions	Description
`u10`	(time, y, x)	Reconstructed 10 m eastward wind component ($m\,s^{-1}$).
`v10`	(time, y, x)	Reconstructed 10 m northward wind component ($m\,s^{-1}$).
`time`	(time,)	Snapshot analysis time (ISO 8601, dtype: datetime64).
`center_lat`	(time,)	IBTrACS-interpolated TC centre latitude (°N).
`center_lon`	(time,)	IBTrACS-interpolated TC centre longitude (°E).
`ibt_vmax`	(time,)	IBTrACS best-track V_{max} (knots).
`n_obs`	(time,)	Number of CYGNSS observation points used.
`meets_ocs`	(time,)	OCS quality flag (boolean): `True` = OCS-pass.
`obs_wind_speed`	(time, n_obs_max)	CYGNSS observed wind speed at each point ($m\,s^{-1}$).
`obs_pixel_i`	(time, n_obs_max)	Observation pixel row index.
`obs_pixel_j`	(time, n_obs_max)	Observation pixel column index.

Global attributes include: `storm_name`, `basin`, `year`, `grid_resolution_km` (1.5), `domain_size_km` (384). The separate `ensemble_uncertainty.nc` file contains variable `ws_spread` (time, y, x): pixel-level ensemble standard deviation of wind speed across 16 members ($m\,s^{-1}$), available for 138 sampled major-hurricane snapshots (70 OCS-pass and 68 OCS-fail; sampling details in Sect. 4.6). Users may filter to OCS-pass cases via the `meets_ocs` flag in the matching per-TC NetCDF file.

3.3 Grid specification

- Grid size: 256 × 256 pixels
- Pixel resolution: 1.5 km
- Domain extent: 384 km × 384 km, centred on the IBTrACS-interpolated TC centre
- Centre pixel: (128, 128) corresponds to the TC centre

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

– Coordinate system: storm-centred Cartesian; row index increases southward, column index increases eastward

– Geographic coordinates: can be computed from `center_lat`/`center_lon` plus pixel offsets (see Eqs. 2 and 3 in Sect. 5)

3.4 Dataset summary

Table 3 summarises the key statistics of the QiFeng-CYGNSS dataset. The dataset contains a total of 4955 TC snapshots (249 independent TCs), of which 1960 (39.6 %) pass OCS quality screening. All 4955 snapshots are released, each containing the `meets_ocs` quality flag. Users are recommended to use the OCS-pass subset for quantitative analyses; OCS-fail snapshots have lower reconstruction quality (due to insufficient CYGNSS observation coverage) but may still be valuable for qualitative reference or large-sample statistics. Figure 3 shows the geographic distribution of all snapshots, coloured by intensity category. The dataset covers all six active tropical cyclone basins globally; for the full sample of 4955 snapshots the South Indian (SI, 1300 snapshots) and Western Pacific (WP, 1244) basins are the most sampled, followed by the Eastern Pacific (EP, 880) and North Atlantic (NA, 840), with the South Pacific (SP, 465) and North Indian (NI, 226) basins contributing fewer samples. Restricting to the full OCS-pass subset (1960 snapshots), the Western Pacific (WP, 494) and South Indian (SI, 459) basins are the most sampled, followed by the North Atlantic (NA, 382), Eastern Pacific (EP, 373), South Pacific (SP, ∼200) and North Indian (NI, ∼54) basins. Note that the basin-stratified validation in Sect. 4.4 (Table 7) reports slightly smaller per-basin counts (summing to $N = 1779$) because that table further excludes OCS-pass snapshots that lack a valid IBTrACS $V_{\max}$ entry needed for paired comparison.

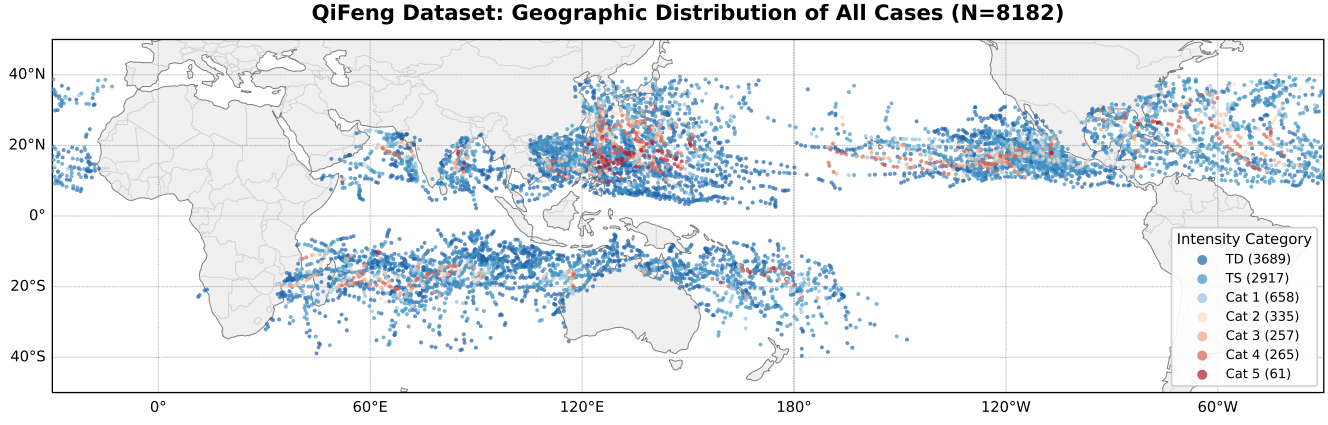

Figure 3. Geographic distribution of all 4955 TC snapshots in the QiFeng-CYGNSS dataset (January 2020–September 2022), coloured by Saffir–Simpson intensity category. The dataset covers all six active tropical cyclone basins globally. Marker colours range from dark blue (tropical depression, TD) to red (Category 5), with sample counts per category indicated in the legend.

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

Table 3. QiFeng-CYGNSS dataset summary statistics (January 2020–September 2022).

Metric	Value
Total TC snapshots (all released)	4955
Independent TCs	249
Active basins covered	6 (WP, EP, NA, NI, SI, SP)
Temporal coverage	January 2020–September 2022
Reporting times	Per IBTrACS report (primarily 6-hourly)
Time window per snapshot	±3 h (motion-corrected)
Snapshots passing OCS	1960 (39.6 %)
Mean CYGNSS observations per snapshot (all)	∼290 (median 250)
Spatial resolution	1.5 km
Domain size	384 km × 384 km (256×256 pixels)
Output variables	u_{10}, v_{10} (10 m wind components, m s^{-1})
Ensemble uncertainty cases	138 (major hurricanes, stratified sample, ≤60 per category; 70 OCS-pass + 68 OCS-fail)
Ensemble members per case	16
File format	NetCDF-4 (CF-1.8)

4 Technical validation

4.1 Independent validation against C-band SAR

To assess the performance of the data product in depicting TC spatial structures, we perform independent validation using all available C-band SAR TC wind field observations during January 2020–September 2022. The SAR wind fields are from the NOAA/NESDIS Center for Satellite Applications and Research (STAR) tropical cyclone SAR wind product (Mouche et al., 2019), which combines acquisitions from Sentinel-1A/B, RADARSAT-2, and the RADARSAT Constellation Mission (RCM) at 500 m pixel spacing with effective wind retrieval resolution of ∼1–3 km, fully independent of CYGNSS (different satellite platforms, different physical observables—C-band radar backscatter vs. L-band GNSS bistatic forward scatter, different retrieval algorithms).

For each SAR observation, we match the nearest QiFeng reconstruction by storm name and observation time (tolerance ±2 h) and additionally restrict matches to snapshots that satisfy the strict OCS criterion of Sect. 2.3, so the SAR validation here is reported on the OCS-pass subset. To eliminate spatial alignment errors caused by TC motion, QiFeng and SAR wind fields independently detect the eye location (local wind speed minimum region) and are aligned in storm-relative coordinates before pixel-level comparison. SAR wind fields undergo quality control prior to matching (Mouche et al., 2019; Combot et al., 2020): (1) edge noise removal (erosion of 10 pixels); (2) land contamination buffering (15-pixel buffer around NaN clusters); (3) sub-

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

swath boundary artefact detection and removal; (4) minimum valid coverage threshold ($\geq$20 % valid pixels). Comparisons are restricted to radius $\leq$150 km and wind speed $\geq$5 m s^{-1}.

This procedure yields 47 matched QiFeng–SAR case pairs on the OCS-pass subset (out of 108 time-matched pairs before OCS screening). After eye-centre alignment, 500 000 pixel pairs are randomly sampled from the complete set for aggregated statistics. Figure 4 presents the results: the pixel-level scatter plot (Fig. 4a) yields root-mean-square error (RMSE) = 5.58 m s^{-1}, mean absolute error (MAE) = 4.00 m s^{-1}, Bias = -0.16 m s^{-1}, $R = 0.728$; the near-zero bias indicates no systematic overestimation or underestimation in the aggregated statistics. The per-case RMSE distribution (Fig. 4b) has a median of 4.88 m s^{-1} (mean 5.21 m s^{-1}), with most cases falling within the 3–8 m s^{-1} range. The per-case MAE distribution (Fig. 4c) has a median of 3.55 m s^{-1} (mean 4.03 m s^{-1}). Table 4 further summarises these statistics.

Table 4. Summary of independent SAR pixel-level validation statistics (January 2020–September 2022, OCS-pass subset, 47 matched cases, eye-aligned, radius $\leq$150 km, wind speed $\geq$5 m s^{-1}).

Metric	Value
Total matched cases (OCS-pass)	47
Randomly sampled pixel pairs	500 000
Pixel-level RMSE	5.58 m s^{-1}
Pixel-level MAE	4.00 m s^{-1}
Pixel-level Bias (QiFeng $-$ SAR)	-0.16 m s^{-1}
Pixel-level R	0.728
Per-case RMSE median (mean)	4.88 (5.21) m s^{-1}
Per-case MAE median (mean)	3.55 (4.03) m s^{-1}

4.2 Best-track intensity validation

We compare the $V_{\max}$ and the radius of maximum wind (RMW) of each QiFeng-CYGNSS snapshot against IBTrACS best-track values, covering the full sample (Table 5). ERA5 and CCMP are included as baselines to quantify the improvement relative to 0.25° reanalysis/blended products.

4.3 Intensity-stratified performance

To quantify product performance across different TC intensities, Table 6 compares QiFeng, ERA5, and CCMP stratified by Saffir–Simpson category on the same OCS-pass subset. For weak TCs (tropical depression/tropical storm), ERA5 and CCMP perform comparably to or better than QiFeng, as 0.25° resolution can adequately represent weak inner-core structures. As TC intensity increases, the gap widens: for Category 4–5 TCs, ERA5 and CCMP $V_{\max}$ RMSEs reach 40.2 and 36.1 m s^{-1} respectively, while QiFeng achieves 18.7 m s^{-1} (a 53 % reduction relative to ERA5). QiFeng's positive bias for weak TCs originates from the diffusion prior generating climatological-mean structures that exceed actual tropical depression and weak

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

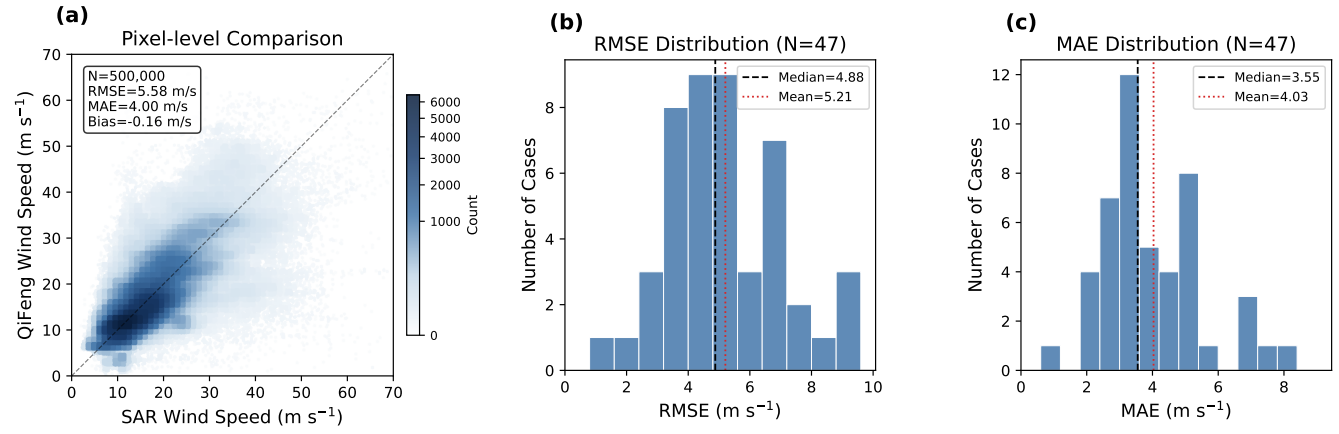

Figure 4. Aggregate pixel-level SAR validation statistics (OCS-pass subset, 47 cases, eye-aligned, radius $\leq$150 km, wind speed $\geq$5 m s^{-1}). (a) Pixel-level scatter plot of QiFeng vs. SAR wind speed ($N = 500\,000$ randomly sampled pairs): RMSE = 5.58 m s^{-1}, MAE = 4.00 m s^{-1}, Bias = -0.16 m s^{-1}, $R = 0.728$. Colours indicate point count density. (b) Per-case RMSE distribution (median = 4.88 m s^{-1}, mean = 5.21 m s^{-1}). (c) Per-case MAE distribution (median = 3.55 m s^{-1}, mean = 4.03 m s^{-1}).

Table 5. TC intensity and structure comparison against IBTrACS best-track (4955 snapshots, January 2020–September 2022). $V_{\max}$ bias: positive = overestimate; RMW bias: positive = overestimate radius. "OCS" = Observation Coverage Sufficiency pass subset (1960 snapshots).

	$V_{\max}$ Bias (m s^{-1})		$V_{\max}$ RMSE (m s^{-1})		RMW Bias (km)	$V_{\max}$ R
Product	All	OCS	All	OCS	All	OCS
ERA5	-7.2	—	12.4	12.7	$+50.7$	0.755
CCMP v3.1	-5.9	—	11.0	11.1	$+49.3$	0.757
QiFeng-CYGNSS	$+1.5$	$+1.1$	10.5	**8.2**	-7.1	**0.796**

On the full sample, QiFeng reduces the absolute $V_{\max}$ bias by $\sim$79 % vs. ERA5 and $\sim$75 % vs. CCMP. QiFeng RMW bias (-7.1 km) is substantially smaller than ERA5/CCMP ($\sim$+50 km), reflecting kilometre-scale resolution. ERA5/CCMP OCS-subset biases are omitted (—) because their values are nearly identical to the full-sample biases (the OCS flag is defined by CYGNSS coverage, not by reanalysis quality); OCS-subset RMSE and R are shown for fair comparison. All three products are evaluated on the same OCS-pass subset for the RMSE (OCS) and R columns.

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

tropical-storm intensities; the negative bias for strong TCs reflects L-band GNSS-R signal sensitivity saturation at high wind speeds (Ruf et al., 2019, 2024; Han et al., 2026c).

Table 6. $V_{\max}$ performance stratified by Saffir–Simpson intensity category (OCS-pass subset, $N = 1779$ valid pairs after excluding snapshots without valid IBTrACS $V_{\max}$). QiFeng, ERA5, and CCMP are evaluated on the identical OCS-pass subset. Units: m s^{-1}.

		QiFeng			ERA5			CCMP		
Category	N	MAE	Bias	RMSE	MAE	Bias	RMSE	MAE	Bias	RMSE
TD	673	4.9	+4.1	6.4	2.4	−1.1	2.9	2.2	−0.7	2.8
TS	757	5.3	+1.8	7.1	6.3	−6.2	7.3	5.5	−4.9	6.6
Cat 1–2	228	8.0	−3.1	9.9	18.5	−18.5	19.4	14.9	−14.8	16.6
Cat 3	57	10.0	−8.8	11.8	28.3	−28.3	28.7	23.1	−23.1	24.2
Cat 4–5	64	16.3	−15.8	18.7	39.5	−39.5	40.2	34.8	−34.8	36.1
All	1779	6.0	+1.1	8.2	8.3	−7.8	12.7	7.1	−6.2	11.1

For Category 1 and above, QiFeng RMSE is consistently lower than ERA5/CCMP. ERA5/CCMP bias for strong TCs is nearly equal to MAE, indicating nearly systematic underestimation. For the overall sample ("All" row), QiFeng's |Bias| (1.1 m s^{-1}) is much smaller than its MAE (6.0 m s^{-1}), indicating a more random error structure in aggregate; however, for Category 4–5 TCs, QiFeng also exhibits near-systematic negative bias (−15.8 m s^{-1}), limited by L-band GNSS-R signal sensitivity saturation at high wind speeds.

4.4 Basin-stratified performance

To evaluate product quality consistency across basins—particularly those without aircraft reconnaissance (WP, NI, SI, SP)—Table 7 presents $V_{\max}$ bias and correlation coefficients stratified by ocean basin.

4.5 Independent airborne validation

In addition to satellite validation, we validate the data product using two types of independent airborne observations: NOAA Hurricane Research Division (HRD) Tail Doppler Radar (TDR) composite wind fields and GPS dropsonde wind speed and direction. TDR provides large-sample gridded validation (47 cases, >750 000 matched points), while dropsondes provide reliable point-by-point validation (24 cases, 78 matched points). Table 8 summarises the validation results; complete validation methodology and scatter plots are presented in Han et al. (2026c).

On the OCS subset, wind speed RMSE against TDR is 6.9 m s^{-1}. The physical constraints reduce directional ambiguity in scalar-to-vector reconstruction, yielding wind direction RMSE of 28.1°, well below the random-direction baseline of ∼104° and consistent with reasonable TC boundary layer flow geometry (Zhang and Uhlhorn, 2012). OCS screening improves results across all validation dimensions, so we recommend filtering by this flag for quantitative use.

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

Table 7. QiFeng-CYGNSS $V_{\max}$ performance stratified by ocean basin (OCS-pass subset, $N = 1779$ valid pairs across 235 TCs). Performance is reported for all six active basins, including those without routine aircraft reconnaissance (WP, NI, SI, SP).

Basin	N **snapshots**	N **TCs**	**Bias** ($\mathrm{m\,s^{-1}}$)	**RMSE** ($\mathrm{m\,s^{-1}}$)	R
NA (North Atlantic)	382	68	-2.9	7.6	0.855
EP (Eastern Pacific)	372	48	$+0.5$	6.9	0.771
WP (Western Pacific)	451	61	$+3.1$	8.8	0.846
NI (North Indian)	45	11	$+0.1$	8.6	0.823
SI (South Indian)	364	44	$+2.8$	8.9	0.738
SP (South Pacific)	165	23	$+2.5$	8.3	0.758
All basins	1779	235	$+1.1$	8.2	0.796

Some TCs traverse multiple basins during their lifetime, and unnamed systems sharing the same IBTrACS designation (`UNNAMED`) across different years/basins are also counted once per basin in this table; hence the sum of per-basin N TCs exceeds the total of 235 unique `tc_name` entries. The product achieves $R > 0.73$ across all basins, with the highest correlations in NA and WP ($R > 0.84$). Southern Hemisphere basins (SI, SP) achieve $R \approx 0.74$–0.76, suggesting cross-basin consistency in product quality.

Table 8. Independent airborne validation of QiFeng-CYGNSS (January 2020–September 2022). TDR: 0.5 km altitude corrected to 10 m using radially dependent Franklin factor, matched within ± 1.5 h. Dropsonde: matched within ± 0.5 h at 10 m altitude. "All" = full matched sample; "OCS" = OCS-pass subset.

Reference	**Subset**	**Cases**	N **points**	**WS RMSE** ($\mathrm{m\,s^{-1}}$)	**WS Bias** ($\mathrm{m\,s^{-1}}$)	**WS** R	**Dir RMSE** (°)
TDR	All	47	754 474	7.5	-3.5	0.702	35.1
TDR	OCS	23	383 803	**6.9**	-3.5	**0.796**	**28.1**
Dropsonde	All	24	78	11.3	-3.5	0.568	—
Dropsonde	OCS	13	51	**8.4**	-2.0	**0.727**	—

OCS screening reduces TDR RMSE from 7.5 to 6.9 $\mathrm{m\,s^{-1}}$ ($-8\,\%$) and improves R from 0.702 to 0.796 ($+13\,\%$). Wind direction RMSE on the OCS subset is 28.1°, well below the theoretical random-direction value ($\sim$104°).

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

4.6 Uncertainty characterisation

To quantify pixel-level uncertainty, we drew a stratified random sample of 138 major hurricane snapshots (Category 3 and above, Saffir–Simpson) using a per-category cap of 60: 60 of 96 Category 3 cases, 60 of 113 Category 4 cases, and all 18 Category 5 cases. The pool from which the sample is drawn is selected only on TC intensity and the availability of CYGNSS observations, without applying the strict OCS criterion of Sect. 2.3; the 138 sampled snapshots therefore include 70 OCS-pass cases (35 Category 3, 30 Category 4, 5 Category 5) and 68 OCS-fail cases (25 Category 3, 30 Category 4, 13 Category 5). We deliberately retain the OCS-fail subset because (i) the dataset releases the full 4955 snapshots and OCS-fail reconstructions are still useful for qualitative reference and large-sample statistics, so users benefit from quantitative spread information for both regimes; (ii) the contrast between OCS-pass and OCS-fail spread is a useful diagnostic of how observational sufficiency moderates reconstruction uncertainty. For each sampled snapshot we ran 16-member ensemble diffusion sampling with independent initial-noise seeds and computed the pixel-wise wind-speed ensemble standard deviation. The released `ensemble_uncertainty.nc` can be cross-referenced to the `meets_ocs` flag in the matching per-TC NetCDF file (matched by `storm_name` and `time`); users wishing to restrict to OCS-pass cases may apply this filter directly.

Figure 5 presents the ensemble spread stratified by TC intensity, radial distance, and OCS status. Spread is largest in the inner core ($r < 50$ km, median 5–7 m s^{-1}), decreases to 2–4 m s^{-1} at 50–100 km, and falls to 1–2 m s^{-1} at 100–200 km, consistent with the strong gradients and directional ambiguity in the eyewall region. Within every category and radial band, OCS-fail snapshots show systematically larger spread than OCS-pass snapshots: at 0–50 km the medians are 5.15 vs. 6.47 m s^{-1}, at 50–100 km 2.19 vs. 3.21 m s^{-1}, and at 100–200 km 1.35 vs. 1.84 m s^{-1} (pass vs. fail). Domain-averaged spread also rises in the OCS-fail subset: Category 3 pass/fail = $1.77/2.28$ m s^{-1}, Category 4 pass/fail = $1.73/2.30$ m s^{-1}, Category 5 pass/fail = $1.74/2.57$ m s^{-1}. The pooled by-category means are 1.98 ± 0.45 (Cat 3), 2.02 ± 0.50 (Cat 4) and 2.34 ± 0.56 m s^{-1} (Cat 5), reflecting the joint contribution of intensity and OCS-fail share within each category.

Within each snapshot, observation density modulates the spread. Mean spread in observed grid cells (about 1.5 m s^{-1} overall) is consistently lower than in unobserved cells (about 2.0 m s^{-1} overall), with a gap/observed ratio of 1.3–1.4$\times$ in both subsets (Fig. 6b: 1.29$\times$ for OCS-pass, 1.35$\times$ for OCS-fail). Across the 138 cases, the number of CYGNSS observations is negatively correlated with domain-averaged spread ($r = -0.64$, $p < 10^{-17}$; Fig. 6a). The correlation is much stronger within the OCS-fail subset ($r = -0.66$) than within OCS-pass ($r = -0.25$): once the OCS criterion is met, additional observations bring diminishing reductions in uncertainty. This is consistent with the OCS criterion identifying a coverage threshold above which the reconstruction is largely observation-constrained.

Ensemble spread reflects only the uncertainty from diffusion sampling stochasticity and does not include observation error, model training bias, or other error sources; it should not be equated with total reconstruction error. We therefore recommend that users (i) use the OCS-pass subset for quantitative analyses (consistent with Sect. 4.2–4.5), and (ii) when working with OCS-fail snapshots for qualitative reference, use `ws_spread` as a per-pixel reliability mask, downweighting or excluding regions where spread exceeds a user-defined threshold.

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

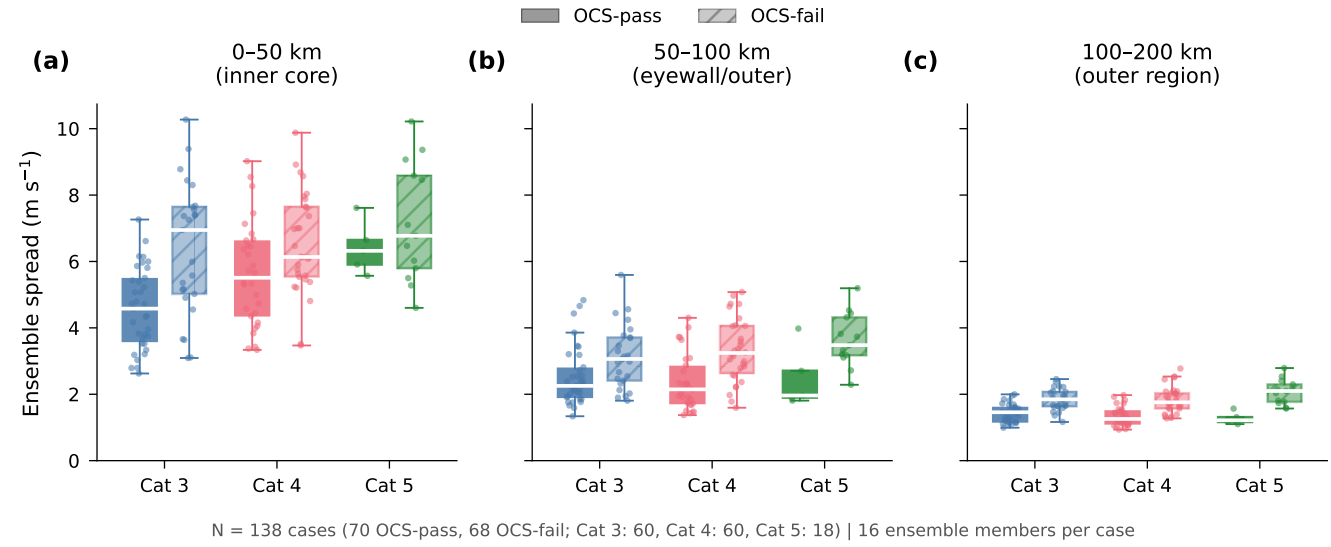

Figure 5. Ensemble spread (standard deviation across 16 members) stratified by TC intensity category, radial distance, and OCS status (pass: solid box; fail: hatched box). (a) Inner core (0–50 km); (b) Mid-range (50–100 km); (c) Outer region (100–200 km). Boxes indicate IQR with white median lines and individual cases are shown as jittered points. Based on 138 major-hurricane cases (60 Category 3, 60 Category 4, 18 Category 5; sample seed = 42), composed of 70 OCS-pass and 68 OCS-fail snapshots.

5 Usage notes

5.1 Reading the data

The following Python code demonstrates how to read and visualise a single snapshot:

```
import xarray as xr, numpy as np

ds = xr.open_dataset('IAN_2022.nc')
t  = '2022-09-28T00:00:00'

# Wind components and speed
u10 = ds['u10'].sel(time=t).values
v10 = ds['v10'].sel(time=t).values
ws  = np.sqrt(u10**2 + v10**2)

# Metadata
vmax = ds['ibt_vmax'].sel(time=t).item()
ocs  = ds['meets_ocs'].sel(time=t).item()
print(f"TC: {ds.attrs['storm_name']}, "
      f"Vmax: {vmax} kt, OCS: {ocs}")
```

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

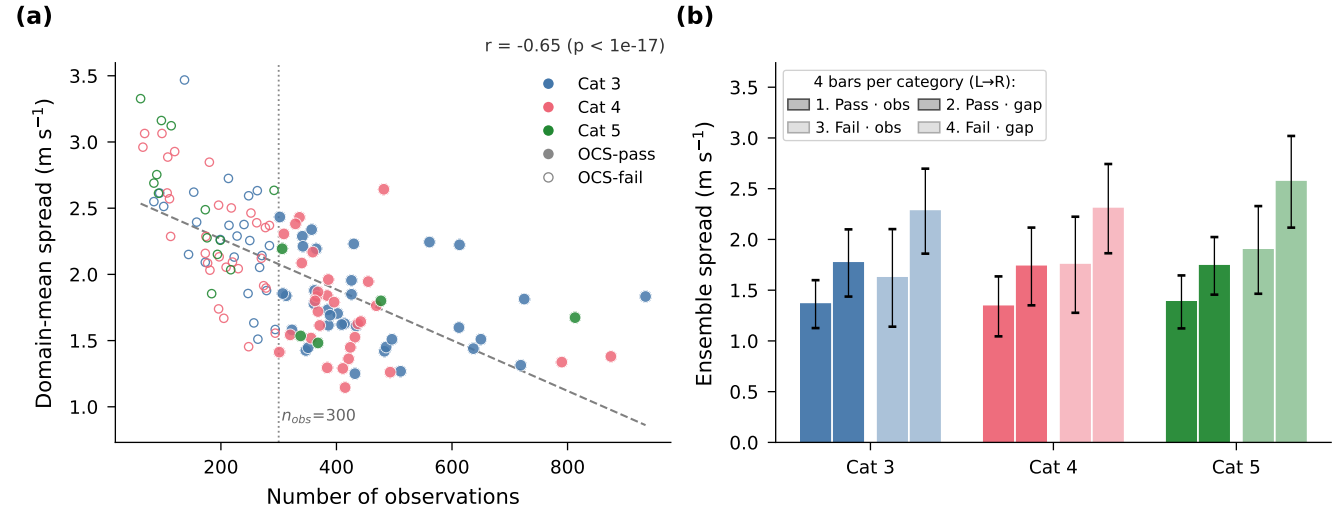

Figure 6. Relationship between observational constraints and ensemble uncertainty, split by OCS status. (a) Number of CYGNSS observations vs. domain-mean ensemble spread, coloured by TC intensity category and shaped by OCS pass/fail (large filled marker: pass; small open marker: fail; pooled $r = -0.64$, $p < 10^{-17}$; vertical dotted line marks the OCS n_{obs} threshold of 300). (b) Ensemble spread in observed ("obs", i.e. grid cells containing at least one CYGNSS observation point) vs. unobserved ("gap") grid cells, by TC intensity category. Each category has four bars (left → right; see legend): (1) Pass · obs, (2) Pass · gap, (3) Fail · obs, (4) Fail · gap. OCS-pass bars are darker (full alpha) and OCS-fail bars are lighter; within each pass/fail pair, the obs bar is on the left and the gap bar on the right. Error bars indicate the case-to-case sample standard deviation within each (category, OCS) subgroup. The gap/observed ratio is nearly category-independent (1.26–1.30× for OCS-pass, 1.32–1.41× for OCS-fail; pooled values 1.29× and 1.35×), demonstrating that observations reduce uncertainty in both regimes by a roughly constant fraction; OCS-fail bars are also higher than OCS-pass bars in absolute terms (and have visibly longer error bars, $\sigma \approx 0.43$–0.48 vs. 0.24–$0.38\,\text{m s}^{-1}$), reflecting the larger residual uncertainty under sparse coverage.

```
# Ensemble uncertainty (if available)
ds_ens = xr.open_dataset(
    'ensemble_uncertainty.nc')
spread = ds_ens['ws_spread'
    ].sel(time=t).values
```

5.2 Geographic coordinate conversion

QiFeng-CYGNSS uses storm-centred Cartesian coordinates. To convert pixel (i, j) to geographic coordinates, use `center_lat` and `center_lon` from the NetCDF file (note that `center_lat` is in degrees, hence the explicit conversion to radians inside

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

the cosine):

$$\text{lat} = \texttt{center\_lat} + (128 - i) \times 1.5/111.32 \tag{2}$$

$$\text{lon} = \texttt{center\_lon} + (j - 128) \times 1.5 \,/\big(111.32\cos(\texttt{center\_lat} \cdot \pi/180)\big) \tag{3}$$

5.3 Quality filtering recommendations

Downstream users are recommended to filter based on the OCS flag:

- **High confidence**: Use only snapshots with `meets_ocs = True` (1960 snapshots, 39.6 % of total).
- **Uncertainty-weighted**: For major hurricane cases with ensemble uncertainty, combine the `ws_spread` field to mask or downweight regions exceeding a user-defined threshold. Typical domain-averaged spread is 2.0 m s^{-1}; inner-core ($r < 50$ km) values may reach 5–7 m s^{-1}.

5.4 Data access tools

A companion code repository is mirrored on the same Zenodo record as the dataset (Han et al., 2026b) and is also available on GitHub at https://github.com/Watanabeyouuu/QiFeng-CYGNSS-dataset-tools (kept up to date for ongoing development). The Zenodo record provides the permanent archive and contains:

- `export_tiff.py`: Batch conversion of NetCDF snapshots to multi-band GeoTIFF (u_{10}, v_{10}, wind speed) with geographic projection (EPSG:4326), suitable for GIS applications and storm surge model forcing.
- `visualize_dataset.ipynb`: Jupyter notebook demonstrating data loading, wind field plotting, TC evolution visualisation, azimuthal-mean radial profiles, and overlay of CYGNSS observations on reconstructed fields.

5.5 Known limitations

Users should be aware of the following limitations:

1. Approximately 60 % of snapshots do not pass OCS criteria due to CYGNSS orbital sparsity; their reconstruction quality is lower, and users should filter by the OCS flag.
2. For very intense TCs ($V_{\max} \gtrsim 50$ m s^{-1}), CYGNSS L-band GNSS-R signal sensitivity saturates, with retrieval error rising to 8–13 m s^{-1} (Ruf et al., 2019, 2024; Han et al., 2026c); this translates into a mean $V_{\max}$ underestimation of about 9 m s^{-1} for Category 3 cases and about 16 m s^{-1} for Category 4–5 cases (cf. Sect. 4.3).
3. The diffusion prior is fine-tuned on HWRF analyses (with prior pre-training on the NICAM cloud-resolving climate archive; see Sect. 2.2) and may inherit HWRF's structural biases (e.g. underrepresentation of extreme asymmetries).

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

4. The dataset's temporal resolution depends on IBTrACS reporting frequency (primarily 6-hourly, some 3-hourly), and cannot capture rapid structural changes within reporting intervals.

5. Although the South Indian basin contributes a sample size comparable to the Western Pacific, the South Pacific basin has substantially fewer snapshots, and Southern Hemisphere basins overall exhibit slightly lower $V_{\max}$ correlation against IBTrACS ($R \approx 0.74$–0.76) than the Northern Hemisphere basins (cf. Table 7).

6. TDR and dropsonde validation is available only in NA/EP; other basins can only be validated indirectly through SAR and IBTrACS.

7. The 16-member ensemble spread is computed for only 138 major hurricane cases (2.8 % of total; 70 OCS-pass and 68 OCS-fail) and reflects only diffusion sampling stochasticity, not total error.

6 Conclusions

The QiFeng-CYGNSS dataset is, to our knowledge, the first globally consistent kilometre-scale vector wind field product for TC inner cores. It complements coarse-resolution reanalyses such as ERA5 and CCMP, which smooth inner-core structure, and episodic high-resolution observations from SAR and aircraft, which lack temporal continuity. The 1.5 km 10 m u/v fields cover 249 TCs (4955 snapshots) across all six active basins during January 2020–September 2022; an OCS quality flag marks the 39.6 % of snapshots suitable for quantitative analysis. Independent validation on the OCS-pass subset against C-band SAR and airborne TDR yields pixel-level wind speed RMSE of 5.58 m s^{-1} (47 cases, $R = 0.73$) and 6.9 m s^{-1} (23 cases, $R = 0.80$) respectively, and wind direction RMSE of 28.1° against TDR. For Category 4–5 TCs, $V_{\max}$ has a low bias of about 16 m s^{-1} (RMSE 19 m s^{-1}) because of L-band GNSS-R signal sensitivity saturation at high wind speeds (Sect. 5.5). We expect the dataset to support storm surge modelling, TC structural climatology, and the training and evaluation of data-driven forecast models, especially in basins without routine aircraft reconnaissance.

7 Data availability

The QiFeng-CYGNSS dataset (v1.0) is deposited on Zenodo under DOI https://doi.org/10.5281/zenodo.20046109 (Han et al., 2026b). The dataset contains 4955 TC snapshots (249 independent TCs, January 2020–September 2022) at IBTrACS reporting times with available CYGNSS coverage (primarily 6-hourly, with some agencies providing 3-hourly reports; reporting times at which no CYGNSS observation falls within the spatiotemporal matching window are excluded from reconstruction), of which 1960 pass OCS quality screening. Data are provided in NetCDF-4 format (CF-1.8 convention), organised as one file per TC (e.g. `IAN_2022.nc`) plus a separate `ensemble_uncertainty.nc` for 138 major hurricane cases. Each file contains reconstructed 10 m u/v wind components, observation metadata, and OCS quality flags. Total size is approximately 2 GB. The dataset is released under the CC BY 4.0 licence. CYGNSS L2 CDR v3.2 data are available from NASA's

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

Physical Oceanography Distributed Active Archive Center (PO.DAAC; https://podaac.jpl.nasa.gov/CYGNSS). IBTrACS best-track data are from NOAA's National Centers for Environmental Information (NCEI; https://www.ncei.noaa.gov/products/international-best-track-archive).

8 Code availability

Custom code for dataset access and visualisation is deposited together with the data on the same Zenodo record (DOI https://doi.org/10.5281/zenodo.20046109 (Han et al., 2026b)), which serves as the permanent archive. The repository contains a GeoTIFF export script (`export_tiff.py`), a Jupyter notebook (`visualize_dataset.ipynb`), and a `README.md` documenting the dataset structure and variable conventions. A development copy of the same scripts is mirrored on GitHub for community contributions: https://github.com/Watanabeyouuu/QiFeng-CYGNSS-dataset-tools. The code is released under the MIT licence. All data processing in the construction of the dataset was performed using Python 3.10 with PyTorch 2.0, xarray, and NumPy.

Author contributions. XH conceived and designed the study, developed the score-based diffusion assimilation algorithm, produced the dataset, performed technical validation analyses, and wrote the manuscript. XL co-supervised the research, coordinated computational resources, and revised the manuscript. JY provided overall scientific guidance and project administration. HN assisted with data processing and quality control. ZN contributed to analysis tool development. WH contributed to project coordination. All authors reviewed and approved the final manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (42530402, 42192561), the Special Project–Original Exploration (Grant 42450254), the NSFC–FDCT Grants (62361166662), the National Key R&D Program of China (2023YFC3503400, 2022YFC3400400), the Project of State Key Laboratory of Satellite Ocean Environment Dynamics (SOEDZZ2527), and the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (311024004, SML2023SP202).

The CYGNSS Level-2 Climate Data Record (v3.2) wind speed data are available from NASA PO.DAAC. ERA5 reanalysis data are available from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu). CCMP v3.1 ocean surface wind data are available from Remote Sensing Systems (https://www.remss.com/measurements/ccmp/). TDR composite wind products are available from NOAA HRD (https://www.aoml.noaa.gov/ftp/pub/hrd/data/radar/level3/). SAR-derived TC surface wind fields are available from NOAA STAR (https://www.star.nesdis.noaa.gov/socd/mecb/sar/sarwinds_tropical.php).

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

References

Aberson, S. D., Black, M. L., Black, R. A., Burpee, R. W., Cione, J. J., Landsea, C. W., and Marks, F. D.: Thirty Years of Tropical Cyclone Research with the NOAA P-3 Aircraft, Bulletin of the American Meteorological Society, 87, 1039–1055, https://doi.org/10.1175/BAMS-87-8-1039, 2006.

Atlas, R., Hoffman, R. N., Ardizzone, J., Leidner, S. M., Jusem, J. C., Smith, D. K., and Gombos, D.: A Cross-Calibrated, Multiplatform Ocean Surface Wind Velocity Product for Meteorological and Oceanographic Applications, Bulletin of the American Meteorological Society, 92, 157–174, https://doi.org/10.1175/2010BAMS2946.1, 2011.

Combot, C., Mouche, A., Knaff, J., Zhao, Y., Zhao, Y., Vinour, L., Quilfen, Y., and Chapron, B.: Extensive High-Resolution Synthetic Aperture Radar (SAR) Data Analysis of Tropical Cyclones: Comparisons with SFMR Flights and Best Track, Monthly Weather Review, 148, 4545–4563, https://doi.org/10.1175/MWR-D-20-0005.1, 2020.

Franklin, J. L., Black, M. L., and Valde, K.: GPS Dropwindsonde Wind Profiles in Hurricanes and Their Operational Implications, Weather and Forecasting, 18, 32–44, https://doi.org/10.1175/1520-0434(2003)018<0032:GDWPIH>2.0.CO;2, 2003.

Han, X., Li, X., Yang, J., Wang, J., Zheng, G., Ren, L., Chen, P., Fang, H., and Xiao, Q.: Dual-Level Contextual Attention Generative Adversarial Network for Reconstructing SAR Wind Speeds in Tropical Cyclones, Remote Sensing, 15, 2454, 2023.

Han, X., Li, X., Yang, J., Wang, J., Han, G., Tao, W., and Aouf, L.: Tropical Cyclone Multi-Level Wind-Speed Structure Reconstruction from Sparse Dropsonde Data via Adversarial Learning, Geophysical Research Letters, 52, e2024GL110661, https://doi.org/10.1029/2024GL110661, 2025.

Han, X., Li, X., Niu, Z., Yang, J., Han, G., Wang, J., Tao, W., Aouf, L., Peng, S., and Chen, D.: Physics-guided score-based diffusion for 3D reconstruction of tropical cyclones from sparse observations, npj Climate and Atmospheric Science, https://doi.org/10.1038/s41612-026-01413-9, 2026a.

Han, X., Li, X., Yang, J., Ni, H., Niu, Z., and Huang, W.: QiFeng-CYGNSS: A Global Kilometre-Scale Tropical Cyclone Inner-Core Vector Wind Field Dataset (v1.0), https://doi.org/10.5281/zenodo.20046109, CC BY 4.0, 2026b.

Han, X., Li, X., Yang, J., Niu, Z., Han, G., Wang, J., Huang, W., Zheng, Y., Ni, H., Wang, Y., Tao, W., Aouf, L., Peng, S., and Chen, D.: Global kilometre-scale tropical cyclone inner-core vector winds from sparse scalar CYGNSS observations, arXiv preprint arXiv:2605.18477, https://doi.org/10.48550/arXiv.2605.18477, 2026c.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Kaplan, J., DeMaria, M., and Knaff, J. A.: A Revised Tropical Cyclone Rapid Intensification Index for the Atlantic and Eastern North Pacific Basins, Weather and Forecasting, 25, 220–241, https://doi.org/10.1175/2009WAF2222280.1, 2010.

Kepert, J. D.: The Dynamics of Boundary Layer Jets within the Tropical Cyclone Core. Part I: Linear Theory, Journal of the Atmospheric Sciences, 58, 2469–2484, https://doi.org/10.1175/1520-0469(2001)058<2469:TDOBLJ>2.0.CO;2, 2001.

Knaff, J. A., Longmore, S. P., and Molenar, D. A.: An Objective Satellite-Based Tropical Cyclone Size Climatology, Journal of Climate, 27, 455–476, https://doi.org/10.1175/JCLI-D-13-00096.1, 2014.

https://doi.org/10.5194/essd-2026-364
Preprint. Discussion started: 22 May 2026

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., and Neumann, C. J.: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying Tropical Cyclone Data, Bulletin of the American Meteorological Society, 91, 363–376, https://doi.org/10.1175/2009BAMS2755.1, 2010.

Liu, G., Jiang, S., Zheng, M., Lin, S., Kong, Y., and Zhan, P.: A Global ERA5-based Tropical Cyclone Wind Field Dataset Enhanced by Integrated Parametric Correction Methods, Scientific Data, 12, 1429, https://doi.org/10.1038/s41597-025-05789-w, 2025.

Matsuoka, D., Kodama, C., Yamada, Y., and Nakano, M.: Dataset of Tropical Cyclone in a High-Resolution Global Nonhydrostatic Atmospheric Simulation, https://doi.org/10.17632/xtvvkfvycr.1, v1, 2023.

Mayers, D. R., Ruf, C. S., and Warnock, A. M.: CYGNSS Storm-Centric Tropical Cyclone Gridded Wind Speed Product, Journal of Applied Meteorology and Climatology, 62, 329–339, https://doi.org/10.1175/JAMC-D-22-0054.1, 2023.

Mears, C., Lee, T., Ricciardulli, L., Wang, X., and Wentz, F.: Improving the Accuracy of the Cross-Calibrated Multi-Platform (CCMP) Ocean Vector Winds, Remote Sensing, 14, 4230, https://doi.org/10.3390/rs14174230, 2022.

Mouche, A., Chapron, B., Knaff, J., Zhao, Y., Zhang, B., and Combot, C.: Copolarized and Cross-Polarized SAR Measurements for High-Resolution Description of Major Hurricane Wind Structures: Application to Irma Category 5 Hurricane, Journal of Geophysical Research: Oceans, 124, 3905–3922, https://doi.org/10.1029/2019JC015056, 2019.

Resio, D. T., Irish, J., and Cialone, M.: A Surge Response Function Approach to Coastal Hazard Assessment – Part 1: Basic Concepts, Natural Hazards, 51, 163–182, https://doi.org/10.1007/s11069-009-9379-y, 2009.

Rios-Berrios, R. and Torn, R. D.: Climatological Analysis of Tropical Cyclone Intensity Changes under Moderate Vertical Wind Shear, Monthly Weather Review, 145, 1717–1738, https://doi.org/10.1175/MWR-D-16-0350.1, 2017.

Rogers, R., Lorsolo, S., Reasor, P., Gamache, J., and Marks, F.: Multiscale Analysis of Tropical Cyclone Kinematic Structure from Airborne Doppler Radar Composites, Monthly Weather Review, 140, 77–99, https://doi.org/10.1175/MWR-D-10-05075.1, 2012.

Ruf, C., Al-Khaldi, M., Asharaf, S., Balasubramaniam, R., McKague, D., Pascual, D., Russel, A., Twigg, D., and Warnock, A.: Characterization of CYGNSS Ocean Surface Wind Speed Products, Remote Sensing, 16, 4341, https://doi.org/10.3390/rs16224341, 2024.

Ruf, C. S., Atlas, R., Chang, P. S., Clarizia, M. P., Garrison, J. L., Gleason, S., Katzberg, S. J., Jelenak, Z., Johnson, J. T., Majumdar, S. J., O'Brien, A., Posselt, D. J., Ridley, A. J., Rose, R. J., and Zavorotny, V. U.: New Ocean Winds Satellite Mission to Probe Hurricanes and Tropical Convection, Bulletin of the American Meteorological Society, 97, 385–395, https://doi.org/10.1175/BAMS-D-14-00218.1, 2016.

Ruf, C. S., Gleason, S., and McKague, D. S.: Assessment of CYGNSS Wind Speed Retrieval Uncertainty, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12, 87–97, https://doi.org/10.1109/JSTARS.2018.2825948, 2019.

Uhlhorn, E. W., Black, P. G., Franklin, J. L., Goodberlet, M., Carswell, J., and Goldstein, A. S.: Hurricane Surface Wind Measurements from an Operational Stepped Frequency Microwave Radiometer, Monthly Weather Review, 135, 3070–3085, https://doi.org/10.1175/MWR3454.1, 2007.

Vincent, E. M., Lengaigne, M., Madec, G., Vialard, J., Samson, G., Jourdain, N. C., Menkes, C. E., and Jullien, S.: Processes Setting the Characteristics of Sea Surface Cooling Induced by Tropical Cyclones, Journal of Geophysical Research: Oceans, 117, C02 020, https://doi.org/10.1029/2011JC007396, 2012.

Zhang, J. A. and Uhlhorn, E. W.: Hurricane Sea Surface Inflow Angle and an Observation-Based Parametric Model, Monthly Weather Review, 140, 3587–3605, https://doi.org/10.1175/MWR-D-11-00339.1, 2012.