

# Seven reanalyses, seven answers: ocean product choice can flip tropical-cyclone intensity conclusions

**Ze Zhang**

National University of Defense Technology

**Letian Chen**

National University of Defense Technology

**Bin Cheng**

chengbin13@nudt.edu.cn

National University of Defense Technology

**Senliang Bao**

National University of Defense Technology

**Huizan Wang**

National University of Defense Technology

**Weimin Zhang**

National University of Defense Technology

**Xiaohui Zhou**

National University of Defense Technology

---

## Article

**Keywords:** Tropical cyclones, Ocean reanalysis, Sea surface temperature, Tropical cyclone heat potential, Potential intensity, Argo

**Posted Date:** October 5th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-11120720/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** No competing interests reported.

---

# Seven reanalyses, seven answers: ocean product choice can flip tropical-cyclone intensity conclusions

Ze Zhang<sup>1</sup>, Letian Chen<sup>1</sup>, Bin Cheng<sup>1\*</sup>, Senliang Bao<sup>1</sup>,  
Huizan Wang<sup>1</sup>, Weimin Zhang<sup>1\*</sup>, Xiaohui Zhou<sup>1</sup>

<sup>1</sup>College of Meteorology and Oceanography, National University of  
Defense Technology, Changsha, China.

\*Corresponding author(s). E-mail(s): [chengbin13@nudt.edu.cn](mailto:chengbin13@nudt.edu.cn);  
[weiminzhang@nudt.edu.cn](mailto:weiminzhang@nudt.edu.cn);

## Abstract

How ocean-reanalysis choice changes inferred tropical cyclone (TC) intensity is rarely quantified. We evaluated seven products along 2,441 TCs (1993–2020) against storm-time drifter SST and gridded Argo on the five-day pre-storm environment, propagating their differences through the maximum and dynamic potential intensity (MPI, DPI). Surface biases are basin-dependent—warm GLORYS (Mercator global reanalysis), cold HYCOM (Hybrid Coordinate Ocean Model). The layers decouple: the coldest surface product overestimates tropical cyclone heat potential most ( $+3.4 \text{ kJ cm}^{-2}$ ), and the product ranking reorders from MPI to DPI, widening the spread from 1.78 to  $2.69 \text{ m s}^{-1}$ . By storm strength, the disagreement peaks at category 5, where two products significantly underestimate TCHP (up to  $6.3 \text{ kJ cm}^{-2}$ ). The errors are non-stationary: ORAS5 flips its bias sign across 2012, and the only reanalysis intensity trend significant at the nominal level, HYCOM’s DPI decline ( $-0.94 \text{ m s}^{-1} \text{ decade}^{-1}$ ), is a step between 2010 and 2012 that coincides with the product’s own production transition—a caution for trend work built on any single reanalysis. Product choice is a structural, time-varying uncertainty—single-dataset studies may carry erroneous conclusions.

**Keywords:** Tropical cyclones, Ocean reanalysis, Sea surface temperature, Tropical cyclone heat potential, Potential intensity, Argo

# 047 1 Introduction

048  
049 The intensity of tropical cyclones couples ocean thermodynamics to societal risk and  
050 has been the focus of some of the most-cited debates in climate science: observational  
051 analyses report increases in the proportion and destructiveness of intense storms [1–3],  
052 rapid intensification has become more frequent in several basins [4, 5], and assessments  
053 continue to attribute intensity changes with increasing confidence [6]. Intensity theory  
054 holds that the maximum intensity a storm can reach is set by sea surface temperature  
055 (SST) together with the atmospheric thermodynamic profile through the Clausius–  
056 Clapeyron relation [7, 8], and that the subsurface temperature structure further  
057 modulates the intensity attained through wind-driven mixing [9, 10]. Errors in the  
058 ocean fields therefore enter TC intensity research before any scientific interpretation  
059 begins.

060 Those fields are almost always taken from global ocean reanalyses. We evaluated  
061 seven of them over their common 1993–2020 coverage at daily and  $0.08^\circ$ – $0.25^\circ$  horizontal  
062 resolution: the eddy-resolving GLORYS12V1 (hereafter GLORYS;  $1/12^\circ$ ) [11], the  
063 Copernicus Marine Service ensemble GREP and three of its members evaluated  
064 individually—C-GLORS, FOAM and the  $0.25^\circ$  Mercator member GLORYS-GREP  
065 [12]—the ECMWF product ORAS5 [13], and the independent high-resolution HYCOM—  
066 five underlying systems sampled as seven products, a framing we justify in Methods.  
067 They differ in assimilation scheme, model physics, vertical discretization and reference  
068 climatology [14]; they are the mainstream ocean reanalyses now used for TC research,  
069 yet the community selects among them without a systematic basis. Existing evaluations  
070 leave most of the relevant space unexplored: regional assessments examine one basin  
071 or one dataset [15–17], single-metric comparisons address basin-mean fields without  
072 carrying differences through to intensity [18–21], and the work that connects the  
073 ocean to intensity uses single products [22–27] that treat the ocean field as given. No  
074 assessment has shown that a product’s bias is independent of storm intensity or storm  
075 properties, propagated inter-product differences through the chain that connects SST  
076 to estimated storm intensity (MPI and DPI), validated both layers independently  
077 under TC sampling, or followed how the biases drift; which product suits a given basin,  
078 storm class and purpose, and how large the resulting uncertainty is, therefore remain  
079 open questions.

080 Two thermal regimes bracket the ocean’s control on TC intensity. The WP warm  
081 pool [28, 29] keeps SST above  $28^\circ\text{C}$  year-round over a deep warm layer and high  
082 TCHP, and nurtures the strongest typhoons [30, 31]. The eastern equatorial Pacific  
083 cold tongue [32, 33] is an upwelling-maintained band of cold water with a shallow  
084 thermocline; typhoons never cross it, but the cold water extending northward from  
085 it—the cold-tongue influence zone (CTIZ;  $100$ – $140^\circ\text{W}$ ,  $10$ – $20^\circ\text{N}$ )—caps the intensity  
086 of EP tropical cyclones [34]. Over a deep reservoir a storm is barely affected by mixing-  
087 induced cooling, whereas over a shallow thermocline the attainable intensity collapses  
088 as soon as mixing begins [25]; we sample the warm-pool and CTIZ boxes throughout.

089 A further question is how surface errors propagate. MPI converts SST and the  
090 atmospheric profile into the theoretical maximum intensity a storm can attain from  
091 its local thermodynamic environment—an upper bound that sees only the sea surface  
092 [7, 35]. DPI adds the ocean’s vertical structure: it replaces the surface temperature

with a depth-averaged temperature over the wind-driven mixing length, so that the estimate also responds to the subsurface ocean [9]. The two measures thus bracket the attainable intensity from the surface and from the mixed layer, and their difference measures how much ocean stratification lowers the intensity a storm can attain; both now have direct observational support [26, 27]. A product whose SST bias is small could still misrepresent TCHP or D26, and a product that is accurate today could drift tomorrow, which bears directly on trend work [36–39]. Neither the propagation of SST biases along the MPI→DPI chain nor the stationarity of reanalysis errors has been quantified for the TC environment. We therefore ask how the SST biases of the seven reanalyses are structured across basins, thermal regimes and storm classes, how they propagate along the MPI→DPI chain, and whether they are stationary. We answer with an event-level composite analysis of the ocean state along 2,441 tropical cyclones over 1993–2020, using GDP drifter SST as the surface reference and gridded Argo below it. We deliberately avoid satellite SST as a reference: the reanalyses assimilate different satellite analyses that are not independent of the products and differ among themselves by several tenths of a degree [40]. We close with a product-selection guide. The answers are not subtle: the product ranking reorders between MPI and DPI, the disagreement peaks in the strongest-storm environments, and the one significant reanalysis intensity trend is a production-era step in a single product.

## 2 Results

### 2.1 Reanalysis SST biases: a three-tier, basin-dependent structure

The departure of the seven products from the storm-time drifter reference is visible at a glance (Fig. 1a,g): GLORYS sits systematically warm, HYCOM strongly cold, its largest cool biases over WP and NI. Averaged globally (95% block-bootstrap intervals; basin-resolved values in Fig. 2a), the products separate into three tiers (with non-overlapping intervals on the end members): warm—GLORYS  $+0.092$  °C [0.072, 0.112], ORAS5  $+0.046$ , GLORYS-GREP  $+0.020$ ; neutral—GREP  $-0.006$ , C-GLORS  $-0.012$ ; cold—FOAM  $-0.077$ , HYCOM  $-0.150$  [ $-0.173$ ,  $-0.127$ ]. Opposite bias signs therefore separate product families.

The basin breakdown sharpens the picture (Table 1). The warm-pool WP spans  $0.36$  °C, from  $+0.115$  °C (GLORYS) to  $-0.247$  °C (HYCOM); NA tips the other way, six of seven products significantly cold and only GLORYS neutral; and SI is the mirror image, five of seven significantly warm (GLORYS  $+0.198$ , ORAS5  $+0.184$ ) with the remaining two near zero (FOAM  $-0.054$ , HYCOM  $-0.026$ ). The EP column carries the cold-tongue corridor, where surface disagreement is small but the DPI disagreement is largest: FOAM and HYCOM are significantly cold there ( $-0.086$  and  $-0.081$  °C), and in NI HYCOM reaches  $-0.299$  °C. The sign of the bias is therefore a property of the basin as much as of the product.

The biases are defined on the drifter-matched subset, which does not always rank the products as the full sample does (in EP, the full-track HYCOM-minus-GLORYS

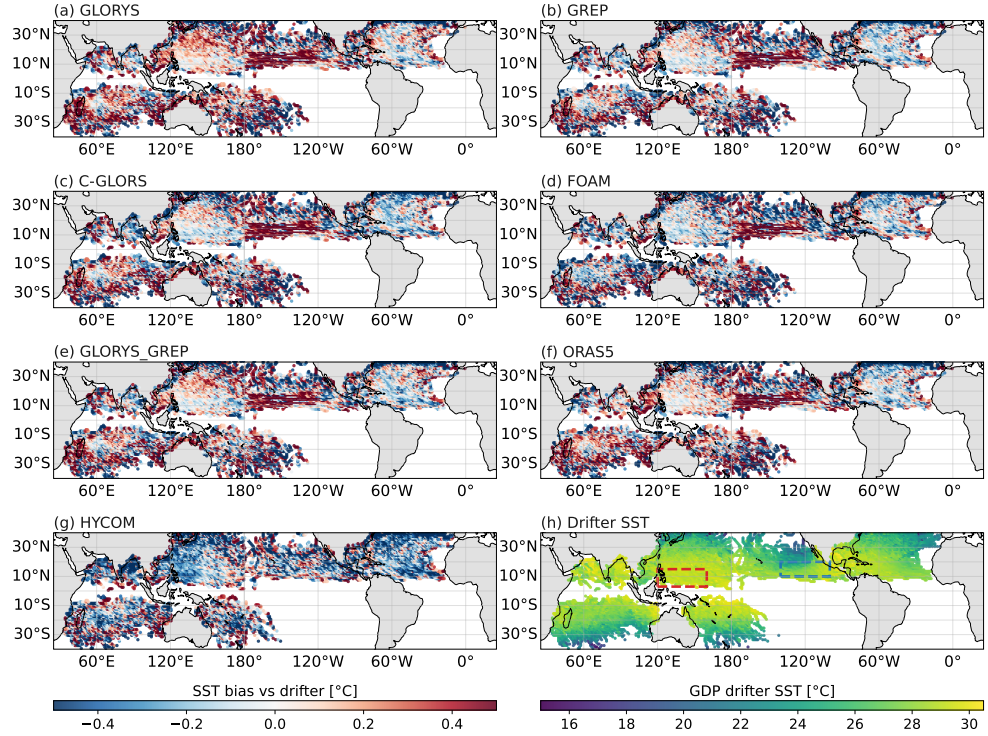


**Fig. 1** Event-level SST bias under tropical-cyclone sampling, 1993–2020 (product minus GDP drifter box-mean SST). (a)–(g) Global  $1^\circ \times 1^\circ$  composite maps (centered at  $155^\circ\text{W}$ ) for the seven products. (h) GDP drifter SST at the matched TC positions (280,108 unique matches) with the warm-pool (red;  $120\text{--}160^\circ\text{E}$ ,  $3\text{--}15^\circ\text{N}$ ) and cold-tongue corridor (blue;  $100\text{--}140^\circ\text{W}$ ,  $10\text{--}20^\circ\text{N}$ ) boxes marked. The maps are Pacific-centered ( $155^\circ\text{W}$ ) so that the western and eastern Pacific basins appear side by side.

difference is  $+0.18^\circ\text{C}$ ); the two sets are reported separately throughout. Neither rapid-intensification status nor translation speed changes a product's bias sign, only its magnitude (Supplementary Fig. 2).

**Table 1** Event-level SST bias ( $^\circ\text{C}$ ) by basin and product (\* = 95% interval excludes zero).

| Basin | GLORYS  | GREP    | C-GLORS | FOAM    | GLORYS-GREP | ORAS5   | HYCOM   |
|-------|---------|---------|---------|---------|-------------|---------|---------|
| WP    | +0.115* | +0.005  | −0.012  | −0.070* | +0.039*     | +0.061* | −0.247* |
| NA    | −0.032  | −0.114* | −0.150* | −0.092* | −0.112*     | −0.098* | −0.196* |
| EP    | +0.080* | −0.011  | +0.013  | −0.086* | +0.007      | +0.032  | −0.081* |
| NI    | +0.015  | −0.035  | −0.078* | −0.089* | −0.008      | +0.038  | −0.299* |
| SI    | +0.198* | +0.096* | +0.126* | −0.054  | +0.126*     | +0.184* | −0.026  |
| SP    | +0.114* | −0.001  | −0.014  | −0.054  | +0.016      | +0.050  | −0.005  |

The products' ranking flips between the two thermal regimes: the warm pool, which feeds the strongest typhoons, and the cold-tongue corridor, which caps them. GLORYS

pre-TC SST averages 29.35 °C against 29.10 °C for HYCOM in the warm-pool box (+0.25 °C), while in the corridor the ranking reverses (28.06 against 28.22 °C; −0.17 °C) (Fig. 2b,c). The products therefore disagree most where intensity research needs them most.

No TC—not even at tropical-depression strength—crossed the equatorial cold tongue itself over 1993–2020, which is why the comparison is made in its northward corridor; both boxes pass a data-coverage check (91% and 99% of 1° cells). The warm-versus-cold division spans assimilation systems, resolutions and climatologies, and because rapid intensification samples the most energetic storms and slow translation the deepest cold wakes [41, 42], its invariance locates the biases in the products’ representation of the environment.

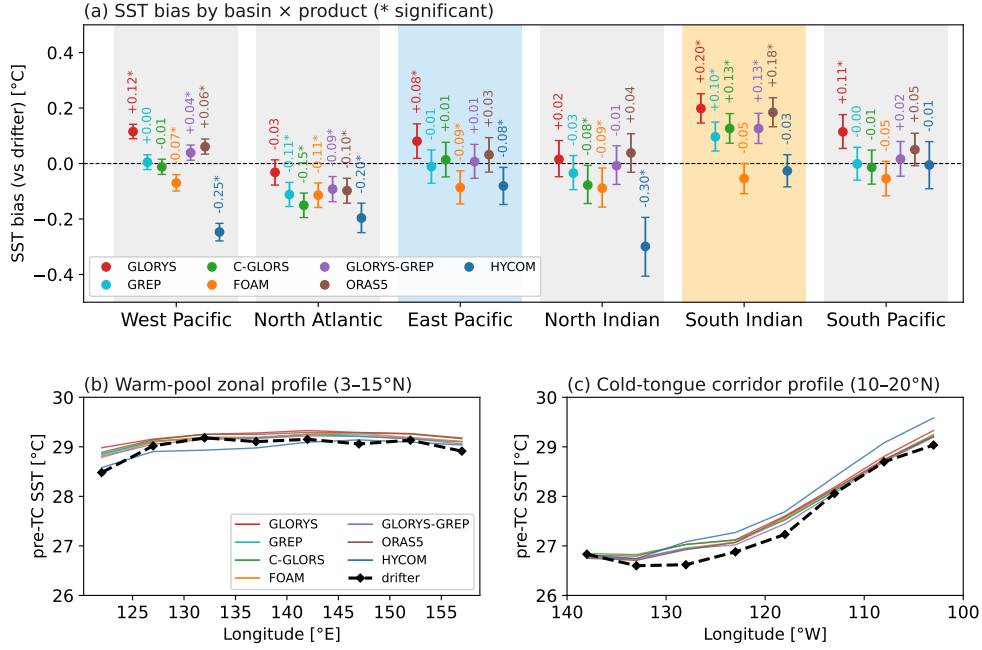


**Fig. 2** Event-level SST bias across basins and pre-TC SST structure in the Pacific, 1993–2020. (a) Event-level SST bias by basin and product (95% confidence intervals; \* = excludes zero; values in Table 1); the SI column is shaded orange and the EP column light blue. (b) Warm-pool (3–15°N) and (c) cold-tongue corridor (10–20°N) zonal profiles of pre-TC SST in 5° longitude bins, seven products plus the GDP drifter reference; panels (b) and (c) use their own y-axis ranges, and the legend in (b) applies to (c).

## 2.2 From surface bias to potential intensity: MPI propagation and the gap to observed intensity

Whether these tenths of a degree survive the step to intensity depends on how steeply MPI responds to SST. Because all seven products share the same ERA5 atmospheric

profiles, MPI is a single function of SST, so Fig. 3a shows that relation with the drifter as the common reference; the same curve, shifted by each product’s SST bias, applies to all. The basin-resolved relation shows that the same  $0.1\text{ }^{\circ}\text{C}$  error produces roughly twice the MPI error in the warm pool as in the cold-tongue regime, because the sensitivity is steep above  $28\text{ }^{\circ}\text{C}$  ( $\approx 9\text{--}14\text{ m s}^{-1}\text{ }^{\circ}\text{C}^{-1}$  in WP) and flat near  $24\text{--}28\text{ }^{\circ}\text{C}$  (EP).

Because MPI responds to SST through  $\text{dMPI}/\text{dSST}$ , these basin biases translate into basin MPI differences of the correct sign and rough magnitude (Fig. 3b); the paired full-sample differences are  $-2.0\text{ m s}^{-1}$  in WP and  $-1.8$  in NI, but  $+2.0$  in EP, where HYCOM’s full-sample SST runs warmer than GLORYS’s (all basins  $p < 0.001$ ). On the drifter-anchored axes the ordering is preserved rather than reversed: GLORYS carries the largest warm offset and HYCOM the largest cold one in both regimes, including the cold-tongue box where the full-sample box means place HYCOM above GLORYS. Which product looks warmest in the cold-tongue corridor therefore depends on how the sample is defined, and nearly every  $\Delta\text{MPI}$  carries the sign of its  $\Delta\text{SST}$ .

On the same events every product differs from GLORYS (paired differences of  $-0.49$  to  $-1.78\text{ m s}^{-1}$ , all  $p < 0.001$ ), and the fixed-environment sensitivity ( $10.8\text{ m s}^{-1}\text{ }^{\circ}\text{C}^{-1}$ ) and the coupled empirical slope ( $7.0$ ) bound the propagation at  $7\text{--}11\text{ m s}^{-1}\text{ }^{\circ}\text{C}^{-1}$ —comparable to best-track intensity uncertainty.

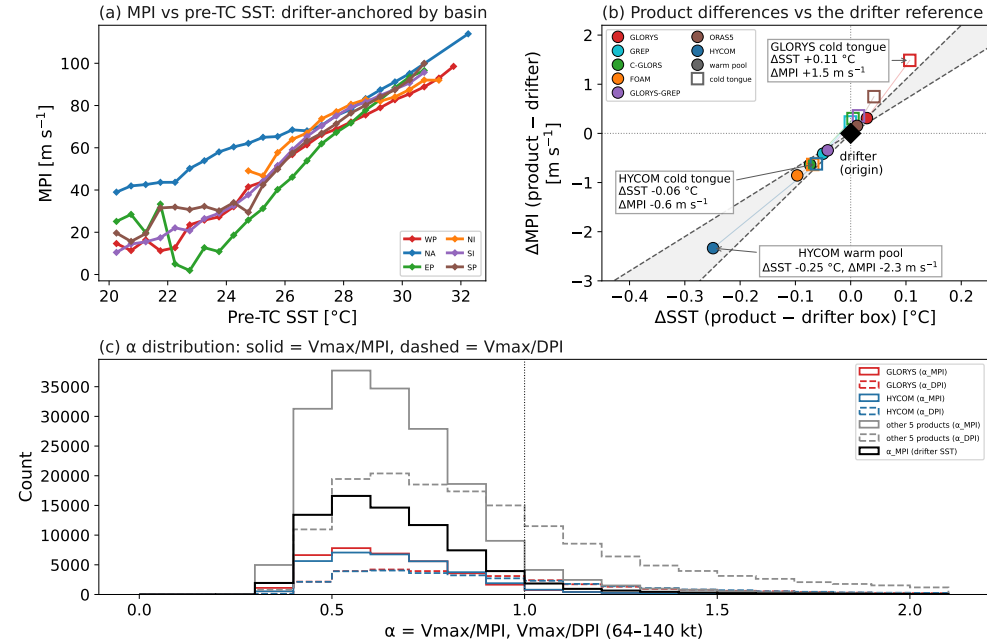


**Fig. 3** Propagation of SST error into potential intensity. (a) Event-level median MPI versus pre-TC SST,  $0.5\text{ }^{\circ}\text{C}$  bins, drifter-anchored curves by basin. (b) Event-mean product-minus-drifter differences: pre-TC box SST ( $\Delta\text{SST}$ ) versus MPI ( $\Delta\text{MPI}$ ) in the warm-pool box (filled circles) and the cold-tongue corridor (open squares), on a common drifter-anchored origin; dashed lines and the shaded band mark the  $7\text{--}11\text{ m s}^{-1}\text{ }^{\circ}\text{C}^{-1}$  sensitivity range, and labeled points mark the two end-states (GLORYS warm, HYCOM cold). (c) Distribution of  $\alpha = V_{\text{max}}/\text{MPI}$  (solid) and  $\alpha = V_{\text{max}}/\text{DPI}$  (dashed); black =  $\alpha$  against MPI with drifter SST.

The ratio  $\alpha = V_{\max}/\text{MPI}$  (Methods) measures how much of MPI a storm attains. Strong storms reach only 0.60–0.62 of it (event-level medians across the seven products; Fig. 3c) and the mean  $|\alpha - 1|$  is 0.37–0.38: observed intensity falls roughly 40% short of the SST-only maximum. The missing ingredient is the vertical temperature structure beneath the storm—surface bias propagates into MPI almost linearly, yet the SST-only maximum leaves roughly 40% of observed intensity unexplained—which motivates the subsurface analysis that follows.

### 2.3 Subsurface structure and the MPI–DPI reordering: the two layers decouple

Checking the same event sample against gridded Argo opens the second layer. Event-level correlations with BOA–Argo are high ( $r = 0.78$ – $0.84$ ; Supplementary Fig. 3), so all products capture the spatial structure of TCHP; the mean biases separate them: GLORYS–GREP ( $+0.9 \text{ kJ cm}^{-2}$ ) sits closest to Argo, followed by FOAM ( $-1.3$ ) and GREP ( $-1.3$ ), while C–GLORS ( $-2.6$ ), ORAS5 ( $+2.3$ ), GLORYS ( $+2.4$ ) and HYCOM ( $+3.4$ , the largest) depart further. Subsurface quality cannot be read off surface quality.

The composite maps and basin panel (Fig. 4a–h) show that HYCOM’s overestimate is itself basin-structured: largest in the SI and SP sectors ( $+7.7$  and  $+5.7 \text{ kJ cm}^{-2}$ ), moderate in WP and EP ( $+3.4$  and  $+2.1$ ), and reversed in NA ( $-0.7$ ); its excess over GLORYS peaks in SP and SI ( $+13.1$  and  $+4.4$ , paired event means; Supplementary Table 2 lists the unpaired values  $+14.6$  and  $+4.6$ ), and the paired composites (Fig. 10a–c) localize it above all in the Indonesian seas. Its D26 is likewise too deep ( $+4.4 \text{ m}$ ) but too shallow in NA ( $-8.5 \text{ m}$ ); the other six products place D26 slightly too shallow (by 1–5.5 m), consistent with their near-zero or negative TCHP biases, and all seven are too deep in the warm pool and too shallow in the cold-tongue corridor (Supplementary Table 3).

The coldest surface product (HYCOM,  $-0.150 \text{ }^{\circ}\text{C}$ ) therefore carries the largest TCHP overestimate ( $+3.4 \text{ kJ cm}^{-2}$ ), while FOAM, the only other cold-surface product, sits in the tight group nearest Argo. Part of the warm products’ positive bias may reflect the reference itself, whose monthly smoothing damps warm cores [34, 43]. Within a single product the two layers are likewise unrelated across basins: GLORYS’s surface bias shows no monotone relation to its subsurface bias.

Whether these differences matter follows from how strongly intensity responds to TCHP. Binning pre-TC TCHP against DPI gives a steep, basin-dependent response that saturates above  $\approx 80 \text{ kJ cm}^{-2}$  (Fig. 5a; slopes  $0.22$ – $0.38 \text{ m s}^{-1}$  per  $\text{kJ cm}^{-2}$ ,  $0.26$  pooled), so each product’s global TCHP bias converts into a diagnostic estimate of its implied DPI shift (Fig. 5b): HYCOM  $+3.4 \text{ kJ cm}^{-2} \rightarrow +0.9 \text{ m s}^{-1}$ , GLORYS  $+2.4 \rightarrow +0.6$ , GREP  $-1.3 \rightarrow -0.3$ . The estimate is statistical—TCHP is not a DPI input, and Argo, with no storm-time profiles, cannot provide one—so the observation anchor stops at TCHP. The physical reason for the response is the regime contrast in ocean mixing: GLORYS event-level pre-TC TCHP correlates negatively with cooling in the warm pool ( $r = -0.13$ ,  $p < 0.001$ ), where the deep warm layer suppresses wind-driven cooling, and positively in the cold-tongue corridor ( $r = +0.28$ ,  $p < 0.001$ ), where the shallow thermocline leaves only a thin warm layer to entrain; the other basins are moderately positive (NA  $+0.17$ , NI  $+0.25$ , SI  $+0.34$ , all  $p < 0.01$ ) and SP ( $+0.15$ ,  $p = 0.011$ ).

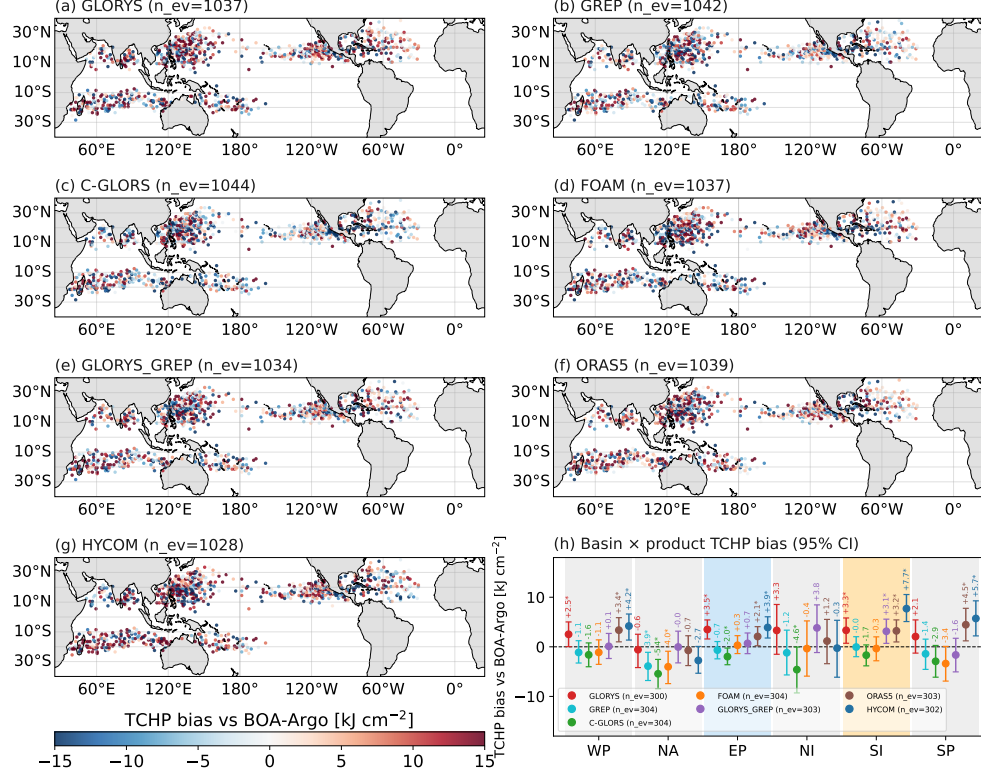


**Fig. 4** TCHP bias against BOA-Argo (maps and basin means). (a)–(g) Global  $1^\circ \times 1^\circ$  composite maps (Pacific-centered at  $155^\circ\text{W}$ , as in Fig. 1), 2004–2019, for the seven products. (h) Event-level TCHP bias by basin (six basins, seven products) with 95% confidence intervals; the SI and EP columns are shaded as in Fig. 2; \* = excludes zero.

The same dichotomy, more weakly, appears in the Argo TCHP on the matched events (WP  $r = -0.10$ ,  $p = 0.081$ ; EP  $r = +0.22$ ,  $p = 0.001$ ; Fig. 5c, open diamonds). The regimes differ correspondingly in their reservoirs ( $\approx 65$  against  $\approx 26 \text{ kJ cm}^{-2}$ ) and in basin-mean cooling: EP is the smallest ( $0.183^\circ\text{C}$ ) against  $0.243\text{--}0.301^\circ\text{C}$  elsewhere (Welch  $p < 0.001$ ). DPI at a given TCHP is nonetheless lower in WP than in NA, NI, or SI, because the warm pool’s warm troposphere lowers MPI for a given SST [7, 36].

The difference between the two estimates,  $\text{MPI} - \text{DPI}$ , is large and product-dependent:  $\text{DPI} < \text{MPI}$  in 92.3% of time steps, by  $9.4 \text{ m s}^{-1}$  on average (GLORYS 11.4, HYCOM 8.7) and most strongly in the EP, with event-weighted basin means  $0.1\text{--}1.3 \text{ m s}^{-1}$  smaller. We call this the mixing correction; it simply measures how far apart the two frameworks sit. Against DPI the  $\alpha$  gap of Fig. 3c closes: strong storms reach 0.81–0.85 of DPI, the mean  $|\alpha - 1|$  falls from 0.37–0.38 to 0.28–0.32, for 82–86% of events DPI lies closer to  $V_{\text{max}}$  than MPI does, and the cross-product range of  $\alpha$  widens from 0.015 to 0.036. The subsurface layer explains observed intensity where the surface saturates: in the warm pool, GLORYS’s  $V_{\text{max}}$  correlates with TCHP at  $+0.32$  ( $p < 0.001$ ) but with SST at only  $+0.06$ , and controlling for either leaves

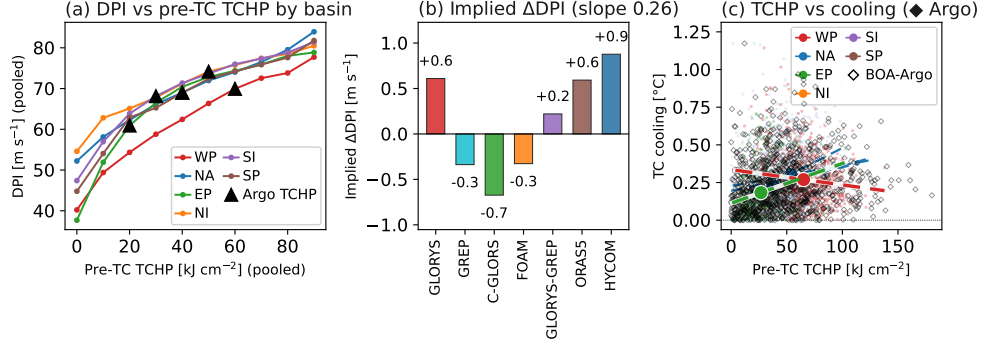


**Fig. 5** Subsurface error propagation into DPI. (a) DPI versus pre-TC TCHP, 10 kJ cm<sup>-2</sup> bin medians pooled over the seven products, by basin; black triangles = BOA-Argo TCHP event means on the same relation. (b) Implied DPI shift per product (global TCHP bias against Argo  $\times$  pooled dDPI/dTCHP slope). (c) Pre-TC TCHP versus TC-induced cooling (GLORYS; cooling = box-mean SST drop pre-TC to storm time), by basin; open diamonds = BOA-Argo TCHP.

TCHP significant (partial  $r = +0.36$ ) and SST not ( $-0.18$ ). Because the products disagree on exactly these fields, the ranking reorders between MPI and DPI (Fig. 6): HYCOM holds the lowest MPI but rises to rank 4 under DPI, GLORYS falls from rank 1 to 7, and the reordering is concentrated at the endpoints (Kendall  $\tau = -0.14$ ; with seven products  $p = 0.77$ , so it rests on paired comparisons). On the 2,433 common events GLORYS's MPI is 1.78 m s<sup>-1</sup> above HYCOM's ( $p = 2 \times 10^{-51}$ ), its DPI 1.51 m s<sup>-1</sup> below ( $p = 5 \times 10^{-28}$ ), and the two mixing corrections differ by 3.29 m s<sup>-1</sup> ( $p < 10^{-300}$ ); across all seven products the spread widens from 1.78 to 2.69 m s<sup>-1</sup>, a 1.5-fold increase (paired event-bootstrap interval of the difference 0.63–1.19 m s<sup>-1</sup>). Nor does the reordering rest on the shared members: restricted to the five independent systems, HYCOM still rises from last to second and GLORYS falls from first to last, and the spread is unchanged (1.78 to 2.69 m s<sup>-1</sup>; Supplementary Fig. 6).

The reordering is mechanical: HYCOM's cold surface bias depresses its MPI while its too-warm, too-deep subsurface thins the mixing correction and lifts its DPI. It also limits how far any single validation generalizes: the Argo-GLORYS D26 consistency reported by Gao et al. [26] does not extend to the product family. The DPI framework [9] and the ocean-coupling intensity index [25] presuppose exactly this subsurface information, and coupled-model intensity forecasts hinge on the fidelity of the ocean initialization [44–47].

## 2.4 The same layers re-sliced by storm strength: the category-5 reversal

Slicing the event sample by storm category answers the surface part first. GREP and C-GLORS never differ from the drifter by more than 0.05 °C at any category, GLORYS is significantly warm throughout (+0.06 to +0.13 °C), FOAM is cool, and HYCOM's cold bias deepens monotonically from  $-0.14$  °C at category 1 to  $-0.32$  °C at category 5 (Fig. 7a). The tiering therefore travels with the product and its basin rather than

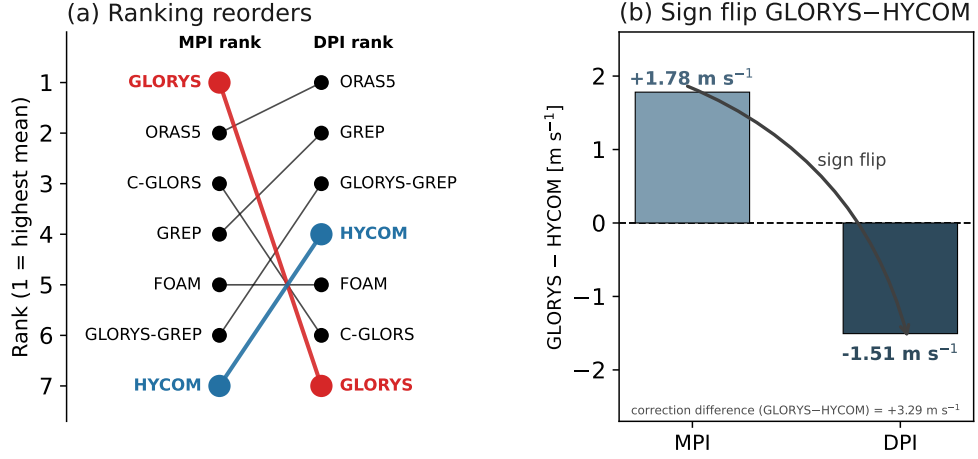


**Fig. 6** MPI and DPI rankings of the seven products. (a) Slope chart of the ranking by event-level mean MPI (left) and DPI (right) on the 2,433 common events. (b) Paired GLORYS – HYCOM differences on common events: MPI, DPI, and the mixing correction (MPI – DPI).

with the storm, extending the insensitivity documented for rapid intensification and translation speed (Supplementary Fig. 2).

Below the surface the same slice tells a different story (Fig. 7b). For categories 1–4, HYCOM significantly overestimates pre-TC TCHP throughout (+4.2 to +6.0  $\text{kJ cm}^{-2}$ ), ORAS5 and GLORYS overestimate it more moderately (+1.8 to +4.5 and +2.1 to +3.4), and GREP, C-GLORS and FOAM stay within  $\pm 2.7 \text{ kJ cm}^{-2}$  of Argo. At category 5 the ordering flips: two of the three products that tracked Argo best in weaker storms—C-GLORS (–6.3) and GREP (–3.8)—now significantly underestimate it, while FOAM (–2.4), GLORYS-GREP (–2.7) and HYCOM (–1.8) drop below zero without significance and GLORYS (+0.6) and ORAS5 (–0.1) stay near zero. The category-5 deficit has a shared and a product-specific part. The all-product mean bias, in which product-specific errors cancel, falls from +1.1 to +2.0  $\text{kJ cm}^{-2}$  in categories 1–4 to –2.4 at category 5; that common term is the part exposed to the reference’s own limitations, since a monthly  $1^\circ$  gridding damps the deep warm cores of the strongest-storm environments—yet degrading the products to the reference’s own  $1^\circ$  and monthly scales moves the all-product category-5 bias from –2.9 to –3.3 and –9.2  $\text{kJ cm}^{-2}$ , deepening rather than creating the common deficit, and leaves the product ordering intact (Supplementary Information). The 6.9  $\text{kJ cm}^{-2}$  product spread at category 5, in contrast, is measured against a common reference and is therefore free of reference-dependent common offsets: C-GLORS and GREP, the two products that tracked Argo most closely in weaker storms, sit 3–7  $\text{kJ cm}^{-2}$  below GLORYS and ORAS5—a differential that reference smoothing cannot create (87–89 matched events per product).

The consequences follow (Fig. 7c). Event-level MPI rises from  $\approx 72 \text{ m s}^{-1}$  at categories 1–4 to  $\approx 75 \text{ m s}^{-1}$  at category 5, with a cross-product spread that stays between 1.15 and 1.81  $\text{m s}^{-1}$  at every category. DPI spreads the products apart exactly where MPI holds them together: its spread narrows from 2.7 (category 1) to 1.6  $\text{m s}^{-1}$

(category 4) and then doubles to  $3.4 \text{ m s}^{-1}$  at category 5, the same category as the TCHP reversal—the largest cross-product intensity disagreement in this study lives in the strongest-storm environments. HYCOM’s DPI drops to  $61.1 \text{ m s}^{-1}$  at category 5, the lowest of the set, while its MPI (73.8) remains mid-pack.

The ratio view closes the loop (Fig. 7d): the median  $\alpha$  rises with intensity, from 0.49 (category 1) to 0.73 (category 5) against MPI, and from 0.65 to 0.93 against DPI—category-5 storms essentially attain their DPI, so any product error there translates nearly one-for-one into inferred intensity. Part of that attainment is structural, because the mixing length follows the storm’s own intensity wind: with a fixed  $50 \text{ m s}^{-1}$  environment wind the category-5 median falls from 0.93 to 0.87 (Supplementary Information), leaving that conclusion intact. In time, the category-4 environment is where the Argo TCHP trend turns negative ( $-13.5 \text{ kJ cm}^{-2} \text{ decade}^{-1}$ ,  $p = 0.033$ ; window-dependent) against a rising product mean ( $+2.1$ ), and it reappears in the category-5 LMI trends below. In short, product choice matters most exactly where storms are strongest and nearest to their DPI.

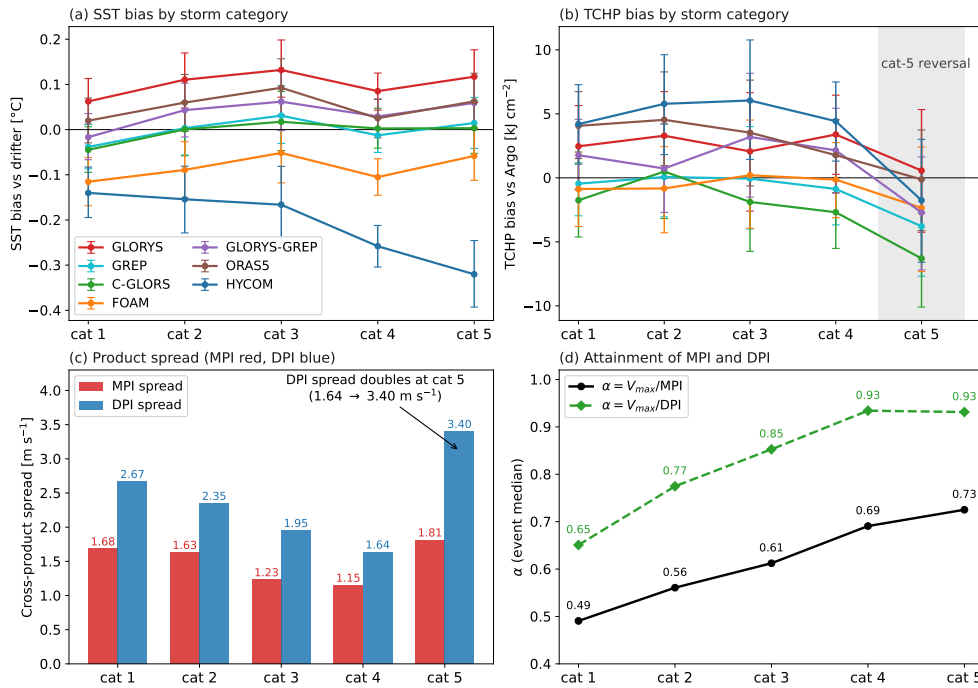


**Fig. 7** Structure of the product disagreement by storm category. (a) Event-level mean SST bias versus the drifter (°C); error bars = 95% confidence intervals. (b) Pre-TC TCHP bias versus Argo ( $\text{kJ cm}^{-2}$ ); the category-5 column is shaded. (c) Cross-product spread (range across the seven products of their event-mean values) of MPI (red) and DPI (blue). (d) Event-median  $\alpha = V_{\max}/\text{MPI}$  (black) and  $\alpha = V_{\max}/\text{DPI}$  (green), on 64–140-kt GLORYS positions. Categories follow the IBTrACS lifetime-maximum intensity ( $n = 402/208/201/361/152$  events; the panel-(d) category-1 median uses 400 events, two lacking defined GLORYS MPI within the 64–140-kt window).

## 2.5 Non-stationary errors and trend inference

Turning to time, the background SST of the TC environment changed little over 1993–2020: every product’s global trend is small and insignificant (Fig. 8a;  $-0.01$  to  $+0.07$   $^{\circ}\text{C}$  decade $^{-1}$ , all  $p > 0.1$ ), below the all-time tropical warming ( $\approx +0.1$   $^{\circ}\text{C}$  decade $^{-1}$ ; 48) because the quantity mixes regional warming with storm-track migration [49]. Basins still differ: WP warms nominally significantly (GLORYS  $+0.151$   $^{\circ}\text{C}$  decade $^{-1}$ ,  $p = 0.014$ ), NI tends the same way but not significantly, and EP is negative. The drifter reference behaves the same way (Supplementary Fig. 5), so the flatness is a property of the observed environment.

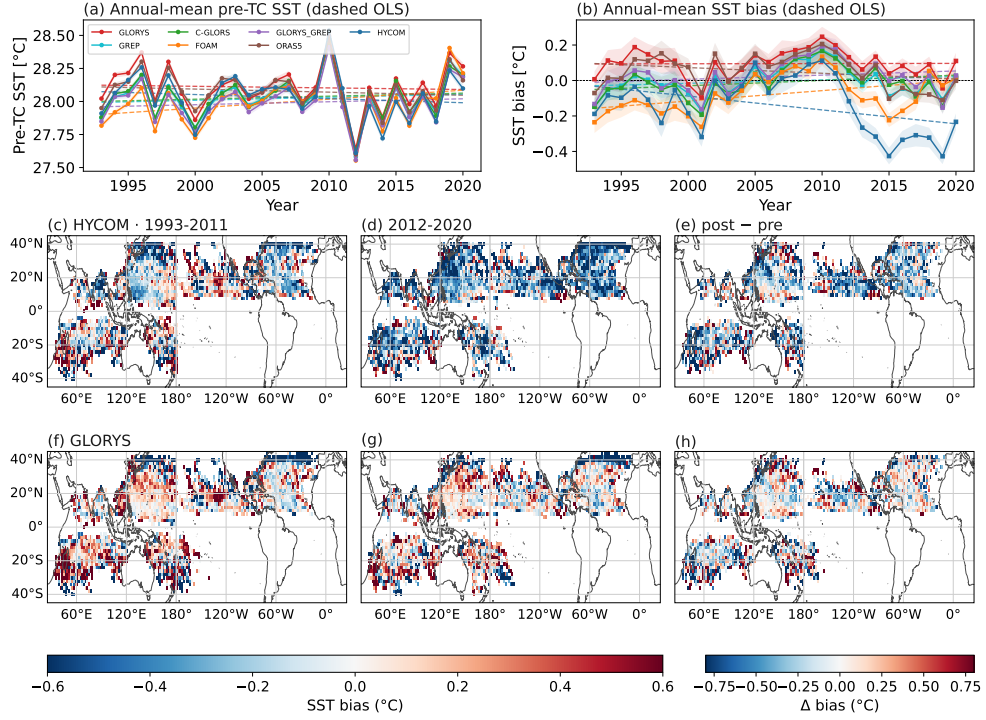


**Fig. 8** Surface non-stationarity of the TC-environment ocean state. (a) Annual-mean pre-TC SST (all 6-h fixes; shading  $\pm 1$  SEM; dashed OLS), 1993–2020. (b) Annual-mean SST bias against the drifter (event-level annual means; dashed OLS). (c)–(h)  $2^{\circ} \times 2^{\circ}$  composites of the event-level SST bias, 1993–2011 (c, f) and 2012–2020 (d, g), and their difference (post – pre, blue = colder; e, h), for HYCOM (c–e) and GLORYS (f–h).

The bias itself drifts, however. Annual-mean SST bias (Fig. 8b) shows nominally significant trends for two products in opposite directions: FOAM warms relative to the reference ( $+0.059$   $^{\circ}\text{C}$  decade $^{-1}$ ,  $p = 0.017$ ) and HYCOM cools ( $-0.078$ ,  $p = 0.033$ ). Splitting the record at 2012 (Fig. 8c–h)—the year of HYCOM’s GOFS 3.1 transition, applied to all products, so a post-hoc split for the rest—localizes the shift: ORAS5 reverses sign, from  $+0.087$   $^{\circ}\text{C}$  (1993–2011) to  $-0.025$   $^{\circ}\text{C}$  (2012–2020) (Welch  $p <$

0.001)—a change of the 2010s rather than of the Argo era. ORAS5’s own documentation places its SST-input switch (HadISST2 to OSTIA) in 2008, not 2012 [13], and we find no documented observing-system change around 2012 that would explain the flip, which we therefore treat as an unexplained inhomogeneity of the product or of its inputs—and HYCOM cools relative to the reference in every basin ( $-0.17$  to  $-0.49$  °C), with a sign reversal in SP ( $+0.08 \rightarrow -0.17$  °C). Decomposing the pooled bias into product and reference terms (Supplementary Fig. 5) shows that over 2012–2020 the product term declines at  $-0.13$  °C decade $^{-1}$  against  $-0.02$  for the reference: the drift is mostly product-side, though with a single surface reference we cannot fully exclude reference drift. In short, the environment is nearly trend-free, but its measurement is not—two products drift and one flips sign.

These drifting fields feed the intensity trends directly. Over 2004–2019, reanalysis TCHP trends on the 1,134 events matched by all seven products and Argo are positive in six of seven products ( $+2.1$  to  $+4.3$  kJ cm $^{-2}$  decade $^{-1}$ , none significant; HYCOM flat at  $-0.1$ ), while the Argo reference on the same events is trend-free ( $-0.7$ ,  $p = 0.79$ ) (Fig. 9a). The warming is therefore fragile.

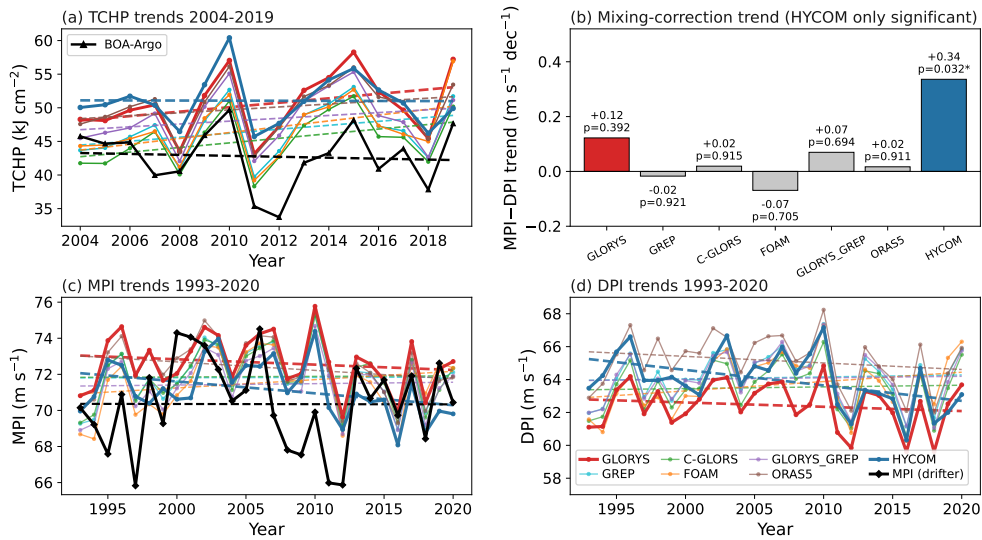


**Fig. 9** Intensity and heat-content trends (event-level annual means; dashed = OLS fits; product colors in all panels follow the legend in (d)). (a) TCHP, 2004–2019, on the 1,134 matched events; black triangles = BOA-Argo. (b) Mixing-correction (MPI – DPI) trends by product (bars annotated with slope and  $p$ ; \* =  $p < 0.05$ ). (c) MPI trends, 1993–2020, with drifter-anchored MPI in black. (d) DPI trends, 1993–2020.

Global MPI trends are small, insignificant, and of mixed sign (Fig. 9c;  $-0.66$  to  $+0.46$  m s $^{-1}$  decade $^{-1}$ ), and a drifter-anchored MPI confirms the flatness ( $-0.01$ ,  $p = 0.99$ ); it departs from the products only as two dips ( $\approx 5$  m s $^{-1}$  in 1997,  $1.5$ – $3$  m s $^{-1}$  over 2007–2012), the intensity-space imprint of the pre-2012 warm bias. That MPI

stays flat while observational studies report increasing intensity [1–3] is consistent with the trend being dominated by the shared ERA5 profile.

No observational analog exists for DPI, because no observing system resolves the TC-time upper-ocean profiles it requires. HYCOM’s global DPI falls at  $-0.94 \text{ m s}^{-1} \text{ decade}^{-1}$  ( $p = 0.013$ ), the only reanalysis intensity trend significant at the nominal level—no trend in either test family survives Benjamini–Hochberg control (Supplementary Information)—and its mixing correction (MPI – DPI) carries the mirror image ( $+0.34$ ,  $p = 0.032$  OLS,  $p = 0.137$  after an AR(1) adjustment; Fig. 9b): its rise is arithmetically implied by the flat MPI series and the robust DPI decline, and so is not independently significant—the signal is subsurface-driven. The fall, however, is a step, not a trend. A Pettitt test locates a single break in 2010 ( $p = 0.004$ , significant even under a Bonferroni correction across the seven products,  $p \times 7 = 0.028$ ): the annual means average  $64.8 \text{ m s}^{-1}$  through 2010 and  $62.5 \text{ m s}^{-1}$  from 2012 onward, no sub-period trend significantly on its own, and SI repeats the pattern (change point 2010,  $p = 0.039$ ; Supplementary Information). The step coincides with the end of HYCOM’s reanalysis-era experiments and the product’s transition to the GOFS 3.1 analysis in late 2012—production inhomogeneity attached to the product, not evidence of a physical decline. No basin-level DPI trend is significant on its own, and removing the 2015 El Niño changes none of this. Trend work on any single reanalysis therefore reads production history, not physics.

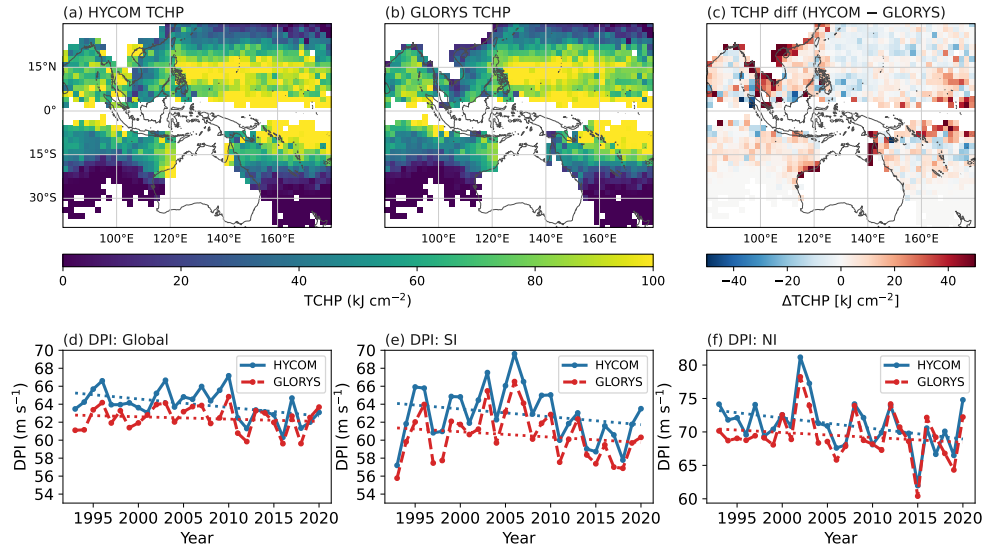


**Fig. 10** Subsurface structure: TCHP differences and HYCOM’s DPI trends. (a)–(c) Pre-TC TCHP composites in the Indian-sector zoom (80°E–180°E, 40°S–30°N): HYCOM (a), GLORYS (b) and HYCOM – GLORYS (c). (d)–(f) HYCOM versus GLORYS DPI time series (event-level annual means; dotted = OLS): (d) Global, (e) SI, (f) NI; (d) and (e) share one y-axis.

Decomposing the MPI trend (Methods) shows that the atmospheric profile controls it: the profile term is nearly identical across products ( $-0.5$ ) because they share one

atmospheric reanalysis, while the SST term spans  $-0.1$  to  $+1.0 \text{ m s}^{-1} \text{ decade}^{-1}$ . That is not an independent constraint, but it agrees with earlier decompositions [36, 38, 39, 50]. At the basin scale HYCOM’s NA MPI trend ( $-0.6$ ,  $p = 0.50$ ) is more negative than GLORYS’s ( $-0.1$ ,  $p = 0.95$ ), but neither is significant, so a single-product assessment still carries sign risk; the DPI drop is significant, subsurface-driven, and concentrated in the 2010–2012 step (Fig. 10). Where the bias sign changes with time, removing a climatological offset cannot recover the trend and may introduce a spurious one.

## 2.6 A product-selection matrix

The selection matrix (Fig. 11; per-basin and per-product values in Supplementary Tables 1 and 2) condenses these results into actionable guidance, expressed against a reference throughout, because absolute values would reproduce basin climatology. The two bias panels (Fig. 11a,b) carry the per-layer selection rules. GLORYS is the natural choice for WP warm-pool surface work and for cold-wake structure, and GREP offers the smallest global surface bias ( $-0.006 \text{ }^{\circ}\text{C}$ ) and is the globally robust baseline: near-zero surface bias, subsurface bias close to Argo, and the ensemble rationale of partial error cancellation. In WP specifically GLORYS-GREP is the TCHP-accurate choice ( $+0.1 \text{ kJ cm}^{-2}$  against Argo, versus GLORYS’s  $+2.5$ ), and FOAM and HYCOM serve as conservative cold-end bounds, with the caveat that after 2012 HYCOM cools relative to the reference in every basin. SI deserves special care: five of seven products are significantly warm at the surface and HYCOM is warm by  $+7.7 \text{ kJ cm}^{-2}$  below it.

For trend work the constraints are stricter: do not use a single-product HYCOM NA MPI series, do not apply a climatological ORAS5 bias correction (its sign changes across 2012), and always report TCHP trends against Argo; GREP or C-GLORS are the robust baselines.

The ratio panels (Fig. 11c,d) say how much the choice matters. Against MPI the event-median  $\alpha$  differs between products by  $0.02$ – $0.04$  within any basin while the basins themselves span  $0.49$ – $0.66$ ; against DPI the spread widens to  $0.02$ – $0.09$  with the product order reshuffling (Fig. 6). In the cold-tongue-influenced EP the DPI maximum sits below observed intensity: six of the seven products return an event-median  $\alpha$  above 1 ( $0.98$ – $1.07$ ; HYCOM just below), because the shallow thermocline makes the mixing correction strongest exactly where it is least constrained (fewer than 1.2% of qualifying events are missing from any product’s cells, so this is not a selection artifact; Methods), and part of the excess is structural—with a fixed  $50 \text{ m s}^{-1}$  environment wind GLORYS’s EP median falls from  $1.03$  to  $0.92$  (Supplementary Information). For DPI-based work the ranking therefore differs, and a product chosen on surface grounds will not automatically suit subsurface-driven applications.

An intensity cut of the same matrices (Fig. 11e,f) condenses the category-5 reversal into selection rules: for categories 1–4 the surface choice carries over (GREP or C-GLORS for both SST and TCHP), but for category-5 storms the TCHP layer should come from GLORYS or ORAS5 rather than from the significantly low GREP and C-GLORS, and HYCOM—the coldest surface product—should be avoided in intensity-critical applications. In short, the choice is settled by the application: surface and weak-storm work tolerate any well-behaved product, category-5 subsurface work does not, and trend work tolerates none.

Product-selection matrices: basin cut (a–d) and storm-category cut (e, f)

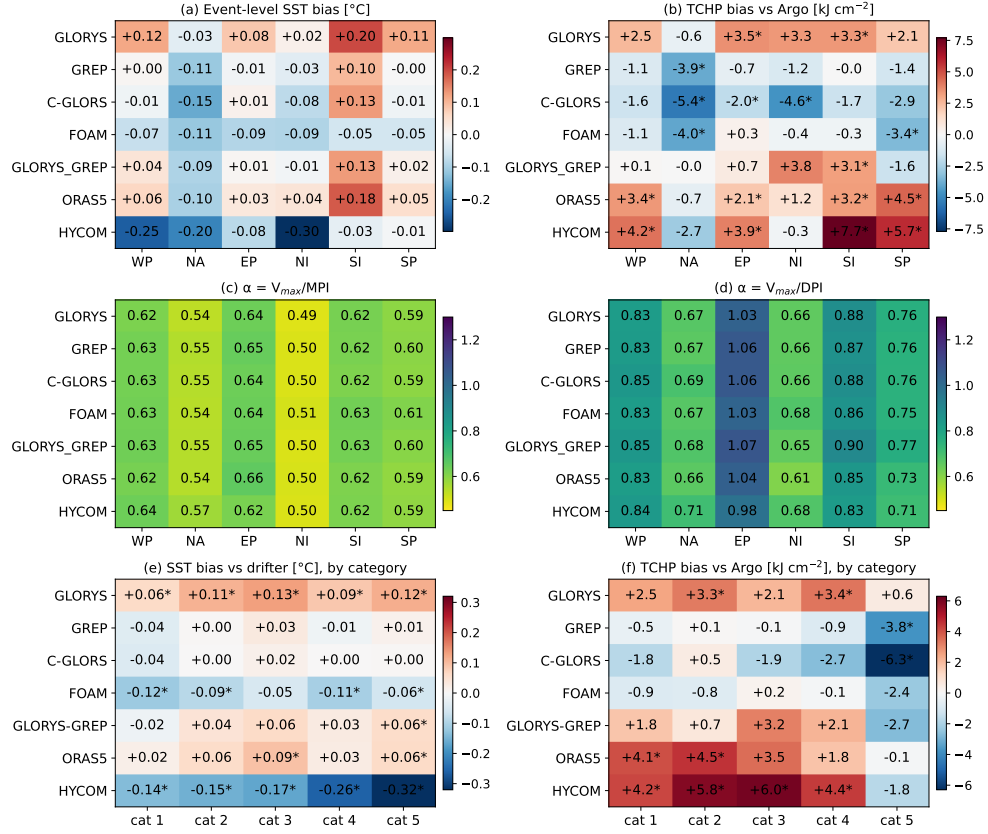


**Fig. 11** Product-selection matrices: basin cut (a)–(d) and storm-category cut (e, f). (a) Event-level mean SST bias versus the GDP drifter reference (°C; blue = cold, red = warm); (b) event-level mean TCHP bias versus gridded Argo (kJ cm<sup>-2</sup>); event-median (c)  $\alpha = V_{max}/MPI$  and (d)  $\alpha = V_{max}/DPI$ ; seven products (rows) by six basins (columns). (e) Event-level mean SST bias versus the drifter and (f) pre-TC TCHP bias versus gridded Argo (kJ cm<sup>-2</sup>), by storm category (columns;  $n = 402/208/201/361/152$  events); asterisks mark cells whose 95% confidence interval excludes zero.

### 3 Discussion

Using GDP drifter SST at the storm instant and gridded Argo on the five-day pre-storm environment, we assessed seven global ocean reanalyses along 2,441 tropical cyclones in 1993–2020 and traced their differences through the MPI→DPI chain. The surface evaluation targets the state drifters sample; MPI, DPI, TCHP and D26 are evaluated on the pre-storm environment that feeds intensity theory. Three results organize what follows.

- *Surface.* SST biases form a warm–neutral–cold three-tier structure whose sign is basin-dependent (GLORYS +0.115 °C against HYCOM –0.247 °C in WP; six of seven products cold in NA, five of seven warm in SI) and reverses between the warm

pool and the cold-tongue corridor (+0.25 versus  $-0.17$  °C); the bias is a property of the product, only weakly modulated by storm characteristics.

- *Two layers.* The layers decouple: the coldest surface product (HYCOM) carries the largest TCHP overestimate ( $+3.4$  kJ cm $^{-2}$ ) and a too-deep D26 ( $+4.4$  m), the products reorder from MPI to DPI (HYCOM 7 $\rightarrow$ 4, GLORYS 1 $\rightarrow$ 7; Kendall  $\tau = -0.14$ ; the reordering persists for the five independent systems behind the seven products, Supplementary Fig. 6), and the spread widens from 1.78 to 2.69 m s $^{-1}$ ; against DPI strong storms reach 0.81–0.85 (versus 0.60–0.62 of MPI).
- *Trends.* The errors are non-stationary: ORAS5 reverses its bias sign across 2012 ( $+0.087 \rightarrow -0.025$  °C,  $p < 0.001$ ), the only reanalysis intensity trend significant at the nominal level—HYCOM’s subsurface-driven DPI decline ( $-0.94$  m s $^{-1}$  decade $^{-1}$ ,  $p = 0.013$ )—is a production-era step across 2010–2012, not a physical trend, and the apparent TCHP rise in six of seven products ( $+2.1$  to  $+4.3$  kJ cm $^{-2}$  decade $^{-1}$ ) is absent from the trend-free Argo reference.

Two design choices condition these numbers. Because drifter and Argo observations are assimilated into some of these reanalyses, the references are not fully independent; they are constructed directly from observations, and shared information would shrink, not create, the reported differences.

On interannual-to-decadal scales our results re-frame the “warming ocean, flat potential” pattern of relative-SST theory [36, 38, 39]: it holds in our products—MPI is mixed-sign and small while decadal intensity change tracks the subsurface ocean [51]—but the ocean side of the pattern is itself product-dependent, since the WP TCHP/D26 rise is not confirmed by Argo. Quantifying the pattern therefore requires fixing product selection first, not the theory.

Over 1993–2020 the observed LMI of category-5 systems rises significantly ( $+1.9$  kt decade $^{-1}$ ,  $p = 0.017$ ) while their composite MPI shows a larger point estimate that is not statistically significant (7-product MPI  $+1.4$  m s $^{-1}$  decade $^{-1} \approx +2.7$  kt; drifter-anchored  $+1.1 \approx +2.1$  kt) and  $\alpha$  at the LMI position is stable; the reconciliation must therefore lie in the (unconfirmed) category-5 MPI drift or in best-track inhomogeneity, which is largest in this category. The slower growth of strong storms that Wang et al. [27] attribute to subsurface mixing is directionally present here too—category-4–5 storms contribute 52% of the all-storm LMI trend in our sample, against their 43%—and because their mixed-layer and OCPI diagnostics rest on GLORYS profiles they inherit the product-dependent error structure quantified here—their category-5 end is exactly where two products fall significantly below Argo in TCHP.

Measured against prior work, this is to our knowledge the first evaluation of a global ocean-reanalysis suite under multi-basin TC sampling that jointly assesses both layers and carries their differences through to MPI and DPI—reordering the product ranking between metrics—and the first to stratify product biases by storm intensity. Surface-only single-basin comparisons [15, 16], single-basin or single-dataset subsurface assessments [17, 26] and single-metric multi-product comparisons [18–20] cannot reveal the family split, the basin and intensity dependence, the decadal drift, or the downstream consequences; the advantage of DPI over MPI [9, 25] is here turned from a directional claim into a quantitative one under product uncertainty.

783 With the error structure quantified—the surface tiers, the surface–subsurface  
784 decoupling, the intensity stratification and the post-2012 drift—the natural next step  
785 is targeted bias correction by basin, intensity class and metric, and Fig. 11 supplies  
786 the criteria for that choice.

787

## 788 4 Methods

789

### 790 4.1 Data

791

#### 792 4.1.1 Tropical cyclone observations

793 TC positions and intensities come from the International Best Track Archive for  
794 Climate Stewardship (IBTrACS v4; 52) at 6-hourly resolution over 1993–2020. We  
795 retain systems with intensity of at least 34 kt at some stage, a duration of at least 24 h,  
796 and all positions equatorward of 40° latitude, which leaves 2,441 TCs and 146,945 six-  
797 hour positions across six basins: the western North Pacific (WP), North Atlantic (NA),  
798 eastern North Pacific (EP), North Indian (NI), South Indian (SI), and South Pacific  
799 (SP). The 1993–2020 window matches the common coverage of all seven products, so  
800 that each event is sampled by every product at the same instants—apart from a few  
801 isolated time steps at which a product lacks an ocean record (see Data Availability).  
802 A small number of rows (186) fall in January 2021, belonging to 2020-season storms  
803 sampled through their full lifetime; these are retained but excluded from the 1993–2020  
804 trend computations. IBTrACS intensity records are inhomogeneous across decades and  
805 basins as observing practice matured; this inhomogeneity affects our intensity-trend  
806 estimates and the  $\alpha$  ratios ( $V_{\max}/\text{MPI}$  and  $V_{\max}/\text{DPI}$ ), and event-level inference with  
807 our intensity screens (the  $\geq 34$ -kt storm criterion and the 64–140-kt  $\alpha$  window) reduces  
808 but does not remove it.

809

#### 810 4.1.2 Ocean reanalyses

811

812 Seven products are evaluated: GLORYS (GLORYS12V1), the 1/12° eddy-resolving  
813 global reanalysis of Mercator Ocean International [11]; GREP, the Copernicus Marine  
814 Service (CMEMS) 0.25° multi-system ensemble-mean product [12], together with three  
815 of its members evaluated individually—C-GLORS (CMCC), FOAM (UK Met Office),  
816 and GLORYS-GREP (GLORYS2V4, the 0.25° Mercator member); ORAS5 (ECMWF;  
817 13), the fourth GREP member; and HYCOM (GLBu0.08), an independent 0.08°  
818 product. Counted by underlying system, the seven products represent five independent  
819 ocean-reanalysis systems (Mercator Ocean, CMCC, the Met Office, ECMWF and  
820 HYCOM); because GREP additionally averages its members, we treat it as the ensemble-  
821 mean product rather than an additional independent system and do not use agreement  
822 with GREP as independent evidence. All headline results hold in the five-system view  
823 (Results; Supplementary Fig. 6). All products share 1993–2020 coverage (HYCOM’s  
824 missing 1993 records were filled from the product’s full archive). The temperature  
825 profiles of all products are extracted and quality-controlled uniformly (temperature  
826 range  $-5$  to  $45$  °C, physical fills removed) before computing SST, TCHP, D26, MPI, and  
827 DPI. Atmospheric profiles for MPI/DPI are taken from the fifth-generation ECMWF  
828

atmospheric reanalysis (ERA5) [53] at 0.25° on pressure levels (temperature, specific humidity, geopotential), 6-hourly at 00/06/12/18 UTC, from 1000 to 70 hPa.

### 4.1.3 Reference observations

The surface reference is Global Drifter Program (GDP) drifter SST [54]—the standard 6-hourly interpolated product, quality-controlled here by a 10–35 °C physical range, a 40° latitude bound and a 5 °C per-6-h jump screen—matched to each TC position (the storm instant  $t_{TC}$ , defined below) by averaging drifter temperatures within  $\pm 3$  h and a 500-km radius (this storm-time mean is denoted  $SST_{drifter}$ ). The radius approximately matches the product’s  $\pm 2^\circ$  ( $4^\circ \times 4^\circ$ ) sampling box ( $\sim 440$  km; the circle and the box average comparable areas), so the reference samples the same local environment; the  $|\text{bias}| \leq 3$  °C screen removes records where a single drifter sampled a wake or eddy the box mean missed. Rain-formed fresh lenses can cool the drifter’s hull measurement below the box mean, but the effect enters every product’s bias with the same sign and cannot create inter-product differences, and the two screens bound its size. Drifters provide the only in situ surface reference that does not share a satellite SST retrieval with the reanalyses. A drifter match exists at 280,108 TC positions from 2,119 events; evaluated for each of the seven products this gives 1,964,813 product-position records (not exactly  $7 \times 280,108$ , because isolated time steps lack a product record), and after the physical screen defined below, 2,117 events remain (2,039–2,116 per product; Fig. 1). The subsurface reference is BOA-Argo, a  $1^\circ \times 1^\circ$  monthly gridded temperature product built from Argo observations (2004–2020, 58 levels to 2000 m; 43), used through 2019 to cross-check TCHP and D26. Validation begins in 2004, when Argo approached global coverage [55], and early-year comparisons must be read with care because the array reached its full global design only around 2007; before 2004 the gridded fields revert toward climatology and cannot serve as an independent reference.

## 4.2 Analysis

### 4.2.1 TC-environment sampling

For each TC position we define the storm instant  $t_{TC}$  (the six-hourly fix) and the pre-TC instant  $t_{pre} = t_{TC} - 5$  d, when the pre-TC environment is sampled (cold wakes typically persist 3–10 days; 41). Reanalysis SST is the  $\pm 2^\circ$  ( $4^\circ \times 4^\circ$ ) box mean around the TC center, taken at both instants:  $SST_{box}(t_{TC})$  for the storm-time state, which enters the surface bias and the TC cooling, and  $SST_{box}(t_{pre})$  for the pre-TC state, which alone feeds MPI, DPI, TCHP, and D26. MPI and DPI use the ERA5 profile at the same pre-TC instant as the ocean state—a deliberate deviation from the convention of sampling the atmospheric profile at the storm instant; by storm time the boundary layer near the storm is modified by the storm itself, so a storm-time profile paired with the five-day-earlier SST collapses CAPE and yields spuriously low MPI (156 instants with SST above 26 °C: below  $20 \text{ m s}^{-1}$ , rising to 32–72  $\text{m s}^{-1}$  with the pre-TC profile; Supplementary Fig. 1). All ocean variables are computed on each product’s own grid and then sampled at the TC location, so inter-product differences are unaffected by

interpolation schemes. The two instants evaluate two distinct error scales: the storm-time surface state that observations sample, and the pre-storm environment that feeds intensity theory; both are quantified because TC research consumes both.

#### 4.2.2 SST bias and its screen

The reanalysis SST bias at a TC position is  $SST_{\text{box}}(t_{\text{TC}}) - SST_{\text{drifter}}$ , both at the storm instant. Records pass a  $|\text{bias}| \leq 3^\circ\text{C}$  physical screen: genuine systematic reanalysis–drifter differences in the tropics reach at most  $1\text{--}2^\circ\text{C}$ , and larger records (1.1% of the total) reflect sampling mismatch (a single drifter sampling a wake or eddy while the box mean held undisturbed water) or sensor anomalies. Because the product is a daily mean and the drifter value instantaneous, the comparison carries a diurnal-cycle term (tropical amplitudes  $0.1\text{--}0.5^\circ\text{C}$ ; 56–59) that is common to all products, so it shifts absolute biases but cannot reorder products. Absolute bias is defined only where a drifter match exists, and the matched subset does not always rank the products as the full sample does (in EP, full-sample HYCOM-minus-GLORYS  $+0.18^\circ\text{C}$  vs the drifter-referenced Table 1 values  $-0.081$  and  $+0.080^\circ\text{C}$ ); we report drifter-referenced biases and full-sample inter-product differences separately, and do not interchange them.

#### 4.2.3 Ocean variables

Tropical cyclone heat potential (TCHP) and the  $26^\circ\text{C}$  isotherm depth (D26) follow Leipper and Volgenau [60]:

$$\text{TCHP} = \rho_0 c_p \int_0^{\text{D26}} \max[T(z) - 26^\circ\text{C}, 0] dz \quad (\text{kJ cm}^{-2}), \quad (1)$$

with  $\rho_0 = 1026 \text{ kg m}^{-3}$  and  $c_p = 4200 \text{ J kg}^{-1} \text{ K}^{-1}$ . TCHP and D26 are taken from the five-day pre-storm profile throughout; the storm-time ocean state enters the analysis through the SST cooling and the DPI mixing correction. Mixed-layer depth (MLD) follows the  $0.5^\circ\text{C}$  criterion relative to the surface [61]. TC cooling is  $SST_{\text{box}}(t_{\text{pre}}) - SST_{\text{box}}(t_{\text{TC}})$ , clipped at zero (basin comparisons use the GLORYS-based cooling).

#### 4.2.4 Maximum potential intensity and its trend decomposition

MPI is computed in the Bister and Emanuel [7] potential-temperature framework with the pyPI implementation [62], using ERA5 pressure-level temperature and specific humidity and each product’s pre-TC box SST:

$$V^2 = \left( \frac{C_k}{C_d} \right) \left( \frac{T_s}{T_0} \right) (\text{CAPE}^* - \text{CAPE}_b). \quad (2)$$

Here  $C_k/C_d$  is the exchange-coefficient ratio for enthalpy and momentum at the sea surface,  $T_s$  the sea surface temperature,  $T_0$  the outflow-layer temperature,  $\text{CAPE}^*$  the saturation convective available potential energy of a parcel lifted from the sea surface along a reversible moist adiabat, and  $\text{CAPE}_b$  the corresponding boundary-layer CAPE;

the SST dependence enters through both CAPE\* and the  $T_s/T_0$  factor, which is why MPI steepens with SST.  $C_k/C_d$  is set to the Bister and Emanuel [35] default of 0.9; inter-product differences, bias propagation, and trends are insensitive to this value, though absolute magnitudes are not (Supplementary Information). All products share the same ERA5 profiles, so the atmospheric term in MPI trends is not independent of the atmospheric reanalysis choice. The fixed-environment sensitivity dMPI/dSST is computed event by event as the central 1 °C difference in SST with the atmospheric profile held fixed, globally averaging 10.8 m s<sup>-1</sup> °C<sup>-1</sup>—the propagation coefficient that carries reanalysis SST error into MPI error, because in an inter-product comparison all products share the same profiles and SST is the only input that varies. The 28-year MPI trend is decomposed into an SST term (the fixed sensitivity times the SST trend) and a profile term (the residual), separating oceanic and atmospheric contributions. Drifter-anchored MPI corrects a product’s MPI by its storm-time surface bias scaled by the product’s own local sensitivity: at each fix with a GDP drifter observation within ±3 h of  $t_{TC}$ ,

$$\text{MPI}_{\text{obs}} = \text{MPI}[\text{SST}_{\text{box}}(t_{\text{pre}})] + \frac{\text{dMPI}}{\text{dSST}} [\text{SST}_{\text{drifter}} - \text{SST}_{\text{box}}(t_{TC})]. \quad (3)$$

The drifter SST is thus a storm-time measurement carried into the pre-TC MPI through Eq. (3), which brings in the drifter reference without recomputing pyPI on drifter-matched profiles (Figs. 3, 9c). The linear correction cannot distinguish product error from a physically real cold wake that the product represents correctly, so the anchored series blends the two; we use it only as an observational reference for the flatness of MPI, not as a corrected product.

#### 4.2.5 Dynamic potential intensity

DPI follows the Balaguru et al. [9] mixing-length method: the average temperature over the wind-driven mixing length  $L$  ( $0-L$ ) is the dynamic temperature  $T_{dy}$ , and  $\text{DPI} = \text{MPI}(T_{dy})$ . The mixing-length inputs follow that method: the friction velocity  $u_*$  is computed from the storm’s own best-track intensity wind with the Donelan et al. [63] high-wind drag coefficient, the mixing time uses  $t = R/U$  with  $R = 50$  km (the mean radius of maximum winds; 9) and  $U$  the translation speed between successive best-track positions, and  $L$  is capped at 300 m; where the archive profile is shallower than the depth the mixing length requires, it is completed by the same bounded linear rule as the operational TCHP/D26 pipeline—a continuation of the profile’s own deepest gradient, no steeper than 0.10 °C m<sup>-1</sup> and capped at 200 m. Here  $\alpha$  is the thermal-stratification parameter of Balaguru et al. [9]: the density gradient across the pycnocline beneath the mixed layer (kg m<sup>-4</sup>, equivalently  $\alpha = \rho_0 N^2/g$  with  $N^2$  the buoyancy-frequency squared), operationalized here as  $\rho_0 \beta_T |dT/dz|$  with  $\beta_T = 2 \times 10^{-4}$  °C<sup>-1</sup> and  $dT/dz$  the mean temperature gradient over the 100 m beneath the mixed layer, so that a more stratified upper ocean shortens the mixing length and thins the mixing correction:

$$L = \text{MLD} + \left( \frac{2\rho_0 u_*^3 t}{\kappa g \alpha} \right)^{1/3}. \quad (4)$$

Here  $\kappa = 0.4$  is the von Kármán constant. Where the mixed-layer mean temperature falls below roughly 25 °C (the exact threshold varies with the product’s profile), the algorithm drives VMAX toward a non-physical value (0), which is recorded as undefined. Event-mean DPI is defined for every event of every product (0 of 2,436–2,440 events undefined), and fewer than 1.2% of qualifying events are absent from any product’s selection-matrix cells, so none of the class-resolved results rests on a filtered subsample. MPI and DPI are evaluated on the same pre-TC baseline, so their difference isolates the ocean-mixing effect.

975

#### 976 4.2.6 Statistics

977

**Inference caliber and uncertainty.** Because successive 6-h positions of a TC are strongly autocorrelated, all inference is event-level: each TC contributes one representative value (intensity taken as the event peak; biases, MPI, TCHP, and cooling averaged within the event). Annual values follow the same caliber for MPI, DPI and the mixing correction (Figs. 9, 10) but use all six-hourly fixes for environmental quantities (Fig. 8, Supplementary Fig. 4). Confidence intervals are 95% event-level block-bootstrap intervals (5,000 resamples; 64), and the cross-product spread of a metric is the range (maximum minus minimum) across the seven products of their product-level event means (globally or within a category), with confidence intervals from an event-level block bootstrap (2,000 resamples, seed 42) that recomputes the range on each resample. Each ratio  $\alpha = V_{\max}/\text{MPI}$  or  $V_{\max}/\text{DPI}$  is computed on 64–140 kt positions: the lower bound restricts the ratio to hurricane-strength environments, where the high-wind drag and mixing formulation behind DPI is calibrated, and the upper bound excludes the few most extreme fixes so that a handful of storms cannot dominate the event medians; each event contributes the median of its position-wise ratios, and the selection-matrix panels of Fig. 11 instead use the  $\geq 64$ -kt event median without the 140-kt cap, so that the matrix covers the full storm record.

**Group comparisons and trend tests.** The tests are chosen for the structure of the data: cross-basin comparisons use Welch’s analysis of variance with Welch post-hoc tests [65], which tolerate unequal sample sizes and variances, together with Benjamini–Hochberg false-discovery-rate control [66]; correlations are event-level Pearson  $r$  with partial correlations separating SST and TCHP. Trends are ordinary least squares on annual values over 1993–2020 with two-tailed t-tests; because the annual residuals are weakly autocorrelated ( $|\rho| \leq 0.34$ ), an AR(1)-adjusted test with Quenouille effective degrees of freedom [67] is reported alongside (the HYCOM DPI trend gives  $p = 0.025$ ; Supplementary Information), and all reported trend tests are placed under Benjamini–Hochberg family control in the Supplementary Information. Non-stationarity is examined with the Pettitt change-point test [68] and by splitting the record at 2012.

**Argo cross-validation.** Each product’s event-level TCHP and D26 are paired with the BOA profile at the nearest grid point for the month of the pre-TC instant over 2004–2019, yielding 7,997 event-level matches (1,150 events, 1,142–1,144 records per product); the TCHP comparisons use the 7,261 records (1,028–1,044 per product) in which both fields have positive TCHP. Because BOA-Argo is a monthly  $1^\circ \times 1^\circ$  gridded product, these differences also carry a representativeness term: the monthly averaging and the nearest-grid-point sampling smooth the storm-scale anomalies that the daily

product fields retain, so part of the reported bias is a scale mismatch between the reference and the products rather than a product error (see Results). On the matched events the reference’s event-to-event standard deviation is 1.02–1.10 times the products’ own, so the smoothing concerns the storm-scale warm cores rather than the event-to-event variability; Results separates its contribution (the all-product common bias) from the reference-independent product spread. Sensitivity tests, including the coarsening and monthly-averaging tests that isolate the reference’s spatial- and temporal-scale contributions to the category-5 deficits, are described in the Supplementary Information.

**Supplementary information.** Supplementary Information is available as a separate file: Supplementary Figures S1–S6 and Supplementary Tables S1–S4 (the per-basin selection matrix, split over Tables S1 and S2 to fit the page, the D26 bias table S3, and the trend-test multiplicity table S4), cited in the text as “Supplementary Fig. 1–6” and “Supplementary Tables 1–4”.

**Acknowledgements.** This research was supported by the National Natural Science Foundation of China Grants (42305014, 42505154).

## Declarations

- **Conflict of interest/Competing interests** The authors declare no competing interests.
- **Ethics approval and consent to participate** Not applicable.
- **Consent for publication** Not applicable.
- **Data availability** All data used here are public. TC best tracks: IBTrACS v4, NOAA National Centers for Environmental Information, <https://www.ncei.noaa.gov/products/international-best-track-archive> [52]. Atmospheric reanalysis: ERA5 pressure levels, Copernicus Climate Data Store, <https://cds.climate.copernicus.eu> [53]. Ocean reanalyses: GLORYS12V1 (1/12°), Copernicus Marine Service product GLOBALMULTIYEAR\_PHY\_001\_030, [https://data.marine.copernicus.eu/product/GLOBALMULTIYEAR\\_PHY\\_001\\_030](https://data.marine.copernicus.eu/product/GLOBALMULTIYEAR_PHY_001_030) [11]; GREP and its members (GLORYS2V4, C-GLORS, FOAM, ORAS5), product GLOBAL\_REANALYSIS\_PHY\_001\_031 [12]; ORAS5, Copernicus Climate Data Store reanalysis-oras5 [13]; HYCOM GLBu0.08, <https://www.hycom.org/dataserver/glb-reanalysis>. Drifter SST (GDP): NOAA AOML, <https://www.aoml.noaa.gov/phod/gdp/> [54]. Gridded Argo temperature (BOA-Argo; 1° monthly fields, 2004–2019 as used here): China Argo Real-time Data Center, <http://www.argo.org.cn> [43]. The potential-intensity code pyPI is available at <https://github.com/dgilford/tcpyPI> [62]. The data that support the findings of this study are openly available in Zenodo at <https://doi.org/10.5281/zenodo.22874096>. The deposit contains 52 pre-aggregated tables (CSV) that provide the data needed to reproduce every figure and every sensitivity test in this paper.
- **Code availability** The code that supports the findings of this study is openly available in the same Zenodo deposit at <https://doi.org/10.5281/zenodo.22874096>. The deposit includes 21 Python scripts (17 figure scripts plus the data-assembly and sensitivity-test pipelines), which together with the data provide the code needed to reproduce every figure in this paper. The results are reproducible using the pipeline described in the repository’s README.

1059 • **Author contribution** Z.Z.: Conceptualization, Methodology, Software, Investiga-  
 1060 tion, Formal analysis, Writing – original draft. L.C.: Investigation, Formal analysis.  
 1061 B.C.: Conceptualization, Software, Writing – review & editing. S.B.: Investigation,  
 1062 Data curation. H.W.: Resources, Writing – review & editing. W.Z.: Supervision,  
 1063 Writing – review & editing. X.Z.: Validation. All authors approved the final  
 1064 manuscript.

1065

## 1066 References

1067

- 1068 [1] Emanuel, K. A. Increasing destructiveness of tropical cyclones over the past 30  
 1069 years. *Nature* **436**, 686–688 (2005).
- 1070 [2] Webster, P. J., Holland, G. J., Curry, J. A. & Chang, H.-R. Changes in tropical  
 1071 cyclone number, duration, and intensity in a warming environment. *Science* **309**,  
 1072 1844–1846 (2005).
- 1073 [3] Kossin, J. P., Knapp, K. R., Olander, T. L. & Velden, C. S. Global increase in  
 1074 major tropical cyclone exceedance probability over the past four decades. *Proc.*  
 1075 *Natl. Acad. Sci. (USA)* **117**, 11975–11980 (2020).
- 1076 [4] Bhatia, K. T. *et al.* Recent increases in tropical cyclone intensification rates. *Nat.*  
 1077 *Commun.* **10**, 635 (2019).
- 1078 [5] Bhatia, K. *et al.* A potential explanation for the global increase in tropical cyclone  
 1079 rapid intensification. *Nat. Commun.* **13**, 6626 (2022).
- 1080 [6] Knutson, T. *et al.* Tropical cyclones and climate change assessment: Part II:  
 1081 Projected response to anthropogenic warming. *Bull. Amer. Meteor. Soc.* **101**,  
 1082 E303–E322 (2020).
- 1083 [7] Bister, M. & Emanuel, K. A. Dissipative heating and hurricane intensity. *Meteor.*  
 1084 *Atmos. Phys.* **65**, 233–240 (1998).
- 1085 [8] Wang, Y. & Wu, C.-C. Current understanding of tropical cyclone structure and  
 1086 intensity changes—a review. *Meteor. Atmos. Phys.* **87**, 257–278 (2004).
- 1087 [9] Balaguru, K. *et al.* Dynamic potential intensity: An improved representation of  
 1088 the ocean’s impact on tropical cyclones. *Geophys. Res. Lett.* **42**, 6739–6746 (2015).
- 1089 [10] Wu, L., Wang, B. & Braun, S. A. Impacts of air–sea interaction on tropical  
 1090 cyclone track and intensity. *Mon. Wea. Rev.* **133**, 3299–3314 (2005).
- 1091 [11] Lellouche, J.-M. *et al.* The Copernicus Global 1/12° Oceanic and Sea Ice  
 1092 GLORYS12 Reanalysis. *Front. Earth Sci.* **9**, 698876 (2021).
- 1093 [12] Storto, A. *et al.* The added value of the multi-system spread information for ocean  
 1094 heat content and steric sea level investigations in the CMEMS GREP ensemble  
 1095

- reanalysis product. *Climate Dyn.* **53**, 287–312 (2019b). 1105
- [13] Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K. & Mayer, M. The 1106  
ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: 1107  
A description of the system and assessment. *Ocean Sci.* **15**, 779–808 (2019). 1108  
1109
- [14] Storto, A. *et al.* Ocean reanalyses: Recent advances and unsolved challenges. 1110  
*Front. Mar. Sci.* **6**, 418 (2019a). 1111  
1112
- [15] He, H. Sea surface temperature error under tropical cyclone in the western North 1113  
Pacific. *Sci. Rep.* **14**, 26734 (2024). 1114  
1115
- [16] Nakada, E. & Tomita, H. Intercomparison of bias of global sea surface temperature 1116  
products associated with tropical cyclone passages in the western North Pacific. 1117  
*J. Oceanogr.* **81**, 361–380 (2025). 1118  
1119
- [17] Paul, S., Maheswaran, P. A., Satheesan, K. & Kottayil, A. Assessing ocean reanal- 1120  
ysis accuracy for marine extremes in the Indian Ocean using in-situ observations. 1121  
*Climate Dyn.* **63**, 58 (2025). 1122  
1123
- [18] Toyoda, T. *et al.* Intercomparison and validation of the mixed layer depth fields 1124  
of global ocean syntheses. *Climate Dyn.* **49**, 753–773 (2017). 1125  
1126
- [19] Valdivieso, M. *et al.* An assessment of air–sea heat fluxes from ocean and coupled 1127  
reanalyses. *Climate Dyn.* **49**, 983–1008 (2017). 1128  
1129
- [20] Liu, C. *et al.* Assessment of the global ocean heat content and North Atlantic 1130  
heat transport over 1993–2020. *npj Clim. Atmos. Sci.* **7**, 314 (2024). 1131  
1132
- [21] Rahman, R. & Rahaman, H. Evaluation of sea surface temperature from ocean 1133  
reanalysis products over the North Indian Ocean. *Front. Mar. Sci.* **11**, 1461696 1134  
(2024). 1135
- [22] Wada, A., Usui, N. & Sato, K. Relationship of maximum tropical cyclone intensity 1136  
to sea surface temperature and tropical cyclone heat potential in the North Pacific 1137  
Ocean. *J. Geophys. Res.* **117**, D11118 (2012). 1138  
1139
- [23] Chih, C.-H. & Wu, C.-C. Exploratory analysis of upper-ocean heat content and 1140  
sea surface temperature underlying tropical cyclone rapid intensification in the 1141  
western North Pacific. *J. Climate* **33**, 1031–1050 (2020). 1142  
1143
- [24] Balaguru, K., Foltz, G. R., Leung, L. R., Hagos, S. M. & Judi, D. R. On the use of 1144  
ocean dynamic temperature for hurricane intensity forecasting. *Wea. Forecasting* 1145  
**33**, 411–418 (2018). 1146  
1147
- [25] Lin, I.-I. *et al.* An ocean coupling potential intensity index for tropical cyclones. 1148  
*Geophys. Res. Lett.* **40**, 1878–1882 (2013). 1149  
1150

- 1151 [26] Gao, C. *et al.* Crucial role of subsurface ocean variability in tropical cyclone  
1152 genesis. *Nat. Commun.* **16**, 1050 (2025).  
1153
- 1154 [27] Wang, X. *et al.* Strong tropical cyclones have a more gradual increase trend  
1155 through interaction with ocean subsurface. *npj Clim. Atmos. Sci.* (2026). Article  
1156 in press.
- 1157 [28] Wyrski, K. *Some thoughts about the west Pacific warm pool.* Proc. Western Pacific  
1158 International Meeting and Workshop on TOGA COARE, 99–109 (Nouméa, 1989).  
1159
- 1160 [29] Yan, X.-H., Ho, C.-R., Zheng, Q. & Klemas, V. Temperature and size variabilities  
1161 of the western Pacific warm pool. *Science* **258**, 1643–1645 (1992).  
1162
- 1163 [30] Pun, I.-F., Lin, I.-I. & Lo, M.-H. Recent increase in high tropical cyclone heat  
1164 potential area in the Western North Pacific Ocean. *Geophys. Res. Lett.* **40**,  
1165 4680–4684 (2013).  
1166
- 1167 [31] Wada, A. & Chan, J. C. L. Increasing tropical cyclone heat potential in the  
1168 western North Pacific and its influence on the intensity of Typhoon FAXAI and  
1169 HAGIBIS in 2019. *SOLA* **17A**, 29–32 (2021).  
1170
- 1171 [32] Wyrski, K. An estimate of equatorial upwelling in the Pacific. *J. Phys. Oceanogr.*  
1172 **11**, 1205–1214 (1981).  
1173
- 1174 [33] Philander, S. G. H. *El Niño, La Niña, and the Southern Oscillation* (Academic  
1175 Press, 1990). 293 pp.
- 1176 [34] Shay, L. K. & Brewster, J. K. Oceanic heat content variability in the eastern  
1177 Pacific Ocean for hurricane intensity forecasting. *Mon. Wea. Rev.* **138**, 2110–2131  
1178 (2010).  
1179
- 1180 [35] Bister, M. & Emanuel, K. A. Low frequency variability of tropical cyclone potential  
1181 intensity. 1. Interannual to interdecadal variability. *J. Geophys. Res.* **107**, 4801  
1182 (2002).  
1183
- 1184 [36] Wing, A. A., Emanuel, K. & Solomon, S. On the factors affecting trends and  
1185 variability in tropical cyclone potential intensity. *Geophys. Res. Lett.* **42**, 8669–8677  
1186 (2015).  
1187
- 1188 [37] Kossin, J. P. Validating atmospheric reanalysis data using tropical cyclones as  
1189 thermometers. *Bull. Amer. Meteor. Soc.* **96**, 1089–1096 (2015).  
1190
- 1191 [38] Vecchi, G. A. & Soden, B. J. Effect of remote sea surface temperature change on  
1192 tropical cyclone potential intensity. *Nature* **450**, 1066–1070 (2007).  
1193
- 1194 [39] Ramsay, H. A. & Sobel, A. H. Effects of relative and absolute sea surface  
1195 temperature on tropical cyclone potential intensity using a single-column model.  
1196 *J. Climate* **24**, 183–193 (2011).

- [40] Martin, M. *et al.* Group for high resolution sea surface temperature (GHR SST) analysis fields inter-comparisons. Part 1: A GHR SST multi-product ensemble (GMPE). *Deep Sea Research Part II: Topical Studies in Oceanography* **77–80**, 21–30 (2012).
- [41] Price, J. F. Upper ocean response to a hurricane. *J. Phys. Oceanogr.* **11**, 153–175 (1981).
- [42] Vincent, E. M. *et al.* Processes setting the characteristics of sea surface cooling induced by tropical cyclones. *J. Geophys. Res.* **117**, C02020 (2012).
- [43] Li, H. *et al.* Development of a global gridded Argo data set with Barnes successive corrections. *J. Geophys. Res. Oceans* **122**, 866–889 (2017).
- [44] Mogensen, K. S., Magnusson, L. & Bidlot, J.-R. Tropical cyclone sensitivity to ocean coupling in the ECMWF coupled model. *J. Geophys. Res. Oceans* **122**, 4392–4412 (2017).
- [45] Domingues, R. *et al.* Ocean conditions and the intensification of three major Atlantic hurricanes in 2017. *Mon. Wea. Rev.* **149**, 1265–1286 (2021).
- [46] Chiodi, A. M. *et al.* Surface ocean warming near the core of Hurricane Sam and its representation in forecast models. *Front. Mar. Sci.* **10**, 1297974 (2024).
- [47] Gramer, L. J., Steffen, J., Aristizabal Vargas, M. & Kim, H.-S. The impact of coupling a dynamic ocean in the Hurricane Analysis and Forecast System. *Front. Earth Sci.* **12**, 1418016 (2024).
- [48] Bulgin, C. E., Merchant, C. J. & Embury, O. Tendencies, variability and persistence of sea surface temperature anomalies. *Sci. Rep.* **10**, 7986 (2020).
- [49] Kossin, J. P., Emanuel, K. A. & Vecchi, G. A. The poleward migration of the location of tropical cyclone maximum intensity. *Nature* **509**, 349–352 (2014).
- [50] Vecchi, G. A., Fueglistaler, S., Held, I. M., Knutson, T. R. & Zhao, M. Impacts of atmospheric temperature trends on tropical cyclone activity. *J. Climate* **26**, 3877–3891 (2013).
- [51] Mei, W., Xie, S.-P., Primeau, F., McWilliams, J. C. & Pasquero, C. Northwestern Pacific typhoon intensity controlled by changes in ocean temperatures. *Sci. Adv.* **1**, e1500014 (2015).
- [52] Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J. & Neumann, C. J. The International Best Track Archive for Climate Stewardship (IBTrACS). *Bull. Amer. Meteor. Soc.* **91**, 363–376 (2010).
- [53] Hersbach, H. *et al.* The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.* **146**, 1999–2049 (2020).

- 1243 [54] Lumpkin, R. & Pazos, M. in *Measuring surface currents with Surface Velocity*  
1244 *Program drifters: the instrument, its data, and some recent results* Lagrangian  
1245 Analysis and Prediction of Coastal and Ocean Dynamics 39–67 (Cambridge  
1246 University Press, 2007).
- 1247
- 1248 [55] Wong, A. P. S. *et al.* Argo data 1999–2019: Two million temperature-salinity  
1249 profiles and subsurface velocity observations from a global array of profiling floats.  
1250 *Front. Mar. Sci.* **7**, 700 (2020).
- 1251
- 1252 [56] Price, J. F., Weller, R. A. & Pinkel, R. Diurnal cycling: Observations and models  
1253 of the upper ocean response to diurnal heating, cooling, and wind mixing. *J.*  
1254 *Geophys. Res.* **91**, 8411–8427 (1986).
- 1255
- 1256 [57] Fairall, C. W. *et al.* Cool-skin and warm-layer effects on sea surface temperature.  
1257 *J. Geophys. Res.* **101**, 1295–1308 (1996).
- 1258
- 1259 [58] Stuart-Menteth, A. C., Robinson, I. S. & Challenor, P. G. A global study of  
1260 diurnal warming using satellite-derived sea surface temperature. *J. Geophys. Res.*  
1261 **108**, 3155 (2003).
- 1262
- 1263 [59] Kawai, Y. & Wada, A. Diurnal sea surface temperature variation and its impact  
1264 on the atmosphere and ocean: A review. *J. Oceanogr.* **63**, 721–744 (2007).
- 1265
- 1266 [60] Leipper, D. F. & Volgenau, D. Hurricane heat potential of the Gulf of Mexico. *J.*  
1267 *Phys. Oceanogr.* **2**, 218–224 (1972).
- 1268
- 1269 [61] de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A. & Iudicone, D.  
1270 Mixed layer depth over the global ocean: An examination of profile data and a  
1271 profile-based climatology. *J. Geophys. Res.* **109**, C12003 (2004).
- 1272
- 1273 [62] Gilford, D. M. pyPI (v1.3): Tropical cyclone potential intensity calculations in  
1274 Python. *Geosci. Model Dev.* **14**, 2351–2369 (2021).
- 1275
- 1276 [63] Donelan, M. A. *et al.* On the limiting aerodynamic roughness of the ocean in very  
1277 strong winds. *Geophys. Res. Lett.* **31**, L18306 (2004).
- 1278
- 1279 [64] Künsch, H. R. The jackknife and the bootstrap for general stationary observations.  
1280 *Ann. Statist.* **17**, 1217–1241 (1989).
- 1281
- 1282 [65] Welch, B. L. On the comparison of several mean values: an alternative approach.  
1283 *Biometrika* **38**, 330–336 (1951).
- 1284
- 1285 [66] Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and  
1286 powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57**, 289–300  
1287 (1995).
- 1288
- 1289 [67] Quenouille, M. H. *Associated Measurements* (Butterworths, London, 1952).

|                                                                                                                                                             |      |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| [68] Pettitt, A. N. A non-parametric approach to the change-point problem. <i>J. Roy. Statist. Soc. Ser. C (Appl. Statist.)</i> <b>28</b> , 126–135 (1979). | 1289 |
|                                                                                                                                                             | 1290 |
|                                                                                                                                                             | 1291 |
|                                                                                                                                                             | 1292 |
|                                                                                                                                                             | 1293 |
|                                                                                                                                                             | 1294 |
|                                                                                                                                                             | 1295 |
|                                                                                                                                                             | 1296 |
|                                                                                                                                                             | 1297 |
|                                                                                                                                                             | 1298 |
|                                                                                                                                                             | 1299 |
|                                                                                                                                                             | 1300 |
|                                                                                                                                                             | 1301 |
|                                                                                                                                                             | 1302 |
|                                                                                                                                                             | 1303 |
|                                                                                                                                                             | 1304 |
|                                                                                                                                                             | 1305 |
|                                                                                                                                                             | 1306 |
|                                                                                                                                                             | 1307 |
|                                                                                                                                                             | 1308 |
|                                                                                                                                                             | 1309 |
|                                                                                                                                                             | 1310 |
|                                                                                                                                                             | 1311 |
|                                                                                                                                                             | 1312 |
|                                                                                                                                                             | 1313 |
|                                                                                                                                                             | 1314 |
|                                                                                                                                                             | 1315 |
|                                                                                                                                                             | 1316 |
|                                                                                                                                                             | 1317 |
|                                                                                                                                                             | 1318 |
|                                                                                                                                                             | 1319 |
|                                                                                                                                                             | 1320 |
|                                                                                                                                                             | 1321 |
|                                                                                                                                                             | 1322 |
|                                                                                                                                                             | 1323 |
|                                                                                                                                                             | 1324 |
|                                                                                                                                                             | 1325 |
|                                                                                                                                                             | 1326 |
|                                                                                                                                                             | 1327 |
|                                                                                                                                                             | 1328 |
|                                                                                                                                                             | 1329 |
|                                                                                                                                                             | 1330 |
|                                                                                                                                                             | 1331 |
|                                                                                                                                                             | 1332 |
|                                                                                                                                                             | 1333 |
|                                                                                                                                                             | 1334 |

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformation.pdf](#)