Supplementary Information for

A Reanalysis-Based Global Tropical Cyclone Tracks Dataset for the Twentieth Century (RGTracks-20C)

Guiling Ye[1,2,3], Jeremy Cheuk-Hin Leung[2,4]*, Wenjie Dong[1,3]*, Jianjun Xu[5], Weijing Li[6], Weihong Qian[7], Hoiio Kong[4], Banglin Zhang[2,8]

Affiliations

[1] School of Atmospheric Sciences, Key Laboratory of Tropical Atmosphere-Ocean System, Ministry of Education, Sun Yat-sen University, Zhuhai, China

[2] Hunan Institute of Advanced Technology, Changsha, China

[3] Southern Marine Science and Engineering Guangdong Laboratory, Zhuhai, China

[4] Faculty of Data Science, City University of Macau, Macau, China

[5] CMA-GDOU Joint Laboratory for Marine Meteorology, South China Sea Institute of Marine Meteorology, Guangdong Ocean University, Zhanjiang, China

[6] National Climate Center, China Meteorological Administration, Beijing, China

[7] Department of Atmospheric and Oceanic Sciences, Peking University, Beijing, China

[8] College of Atmospheric Science, Lanzhou University, Lanzhou, China

# These authors contributed equally to this work.

Corresponding author(s): Jeremy Cheuk-Hin Leung (chleung@pku.edu.cn); Wenjie Dong (dongwj3@mail.sysu.edu.cn)

Contents of this file

Introduction

This supporting information provides supplementary texts, figures, and tables cited in the main text.

Table S1 | The probability of detection (POD) and false alarm rate (FAR) of the global TCs detected by different trackers in the fifth generation ECMWF reanalysis (ERA5) and 20CRv3. POD (unit: %) and FAR (unit: %) of TCs detected by different trackers in the latest high-resolution ERA5 reanalysis by (Accarino et al., (\2023) (green background), (Bourdin et al., 2022) (green background), and RGTracks-20C (orange background).

	Hybrid[1]	**CNRM[2]**	**TRACK[2]**	**UZ-ERA5[2]**	**OWZ-ERA5[2]**	**UZ-20CRv3**	**OWZ-20CRv3**
POD (%)	71.49	72.77	74.37	71.54	71.75	67.62	76.56
FAR (%)	23	8.62	17.19	3.37	17.43	7.19	15.21

S1. Hits, Misses, and FAs of TCs detected by the OWZ and UZ trackers

The distribution of the number of hits is mainly concentrated below 960 $\hbar Pa$ (Category 3 storms), with the peak occurrence at Category 0-1 storms. As the intensity increases, the number of hits decreases, especially for Category 4 storms and above (Figs. S1a–b). This finding suggests that the 20CRv3 tends to underestimate the observed intensity of TC. The wind-pressure relationships further illustrate this point, as shown in Fig. 2a, independent of the tracker. By contrast, the distributions of misses and false alarms (FAs) are strongly biased toward weak category 0 storms. The UZ tracker shows more misses, whereas the OWZ tracker exhibits more FAs. This suggests that the OWZ tracker is more effective in detecting weaker storms, which presents an advantage for the low-resolution 20CRv3 dataset. The TC duration of hits varies with trackers, the UZ tracker has a relatively short duration compared to the OWZ tracker (Figs. S1c–d). The missed cases correspond to short-lived TCs. It is worth noting that the FAs of the OWZ tracker can last up to 20 days, which is consistent with the findings of (Bell et al., 2018) and (Bourdin et al., 2022). According to (Bell et al., 2018), these FAs may correspond to TCs or their developmental phases that were not recorded or were removed from IBTrACS because they did not reach the tropical storm category. Therefore, it is likely that some of the FAs in this study also correspond to weak storms that were excluded from IBTrACS. The missed tracks consist of weak and short-lived TCs, which is closely related to the limitations of 20CRv3 in simulating these types of TCs (Hodges et al., 2017), as well as its inability to meet the threshold of tracker.

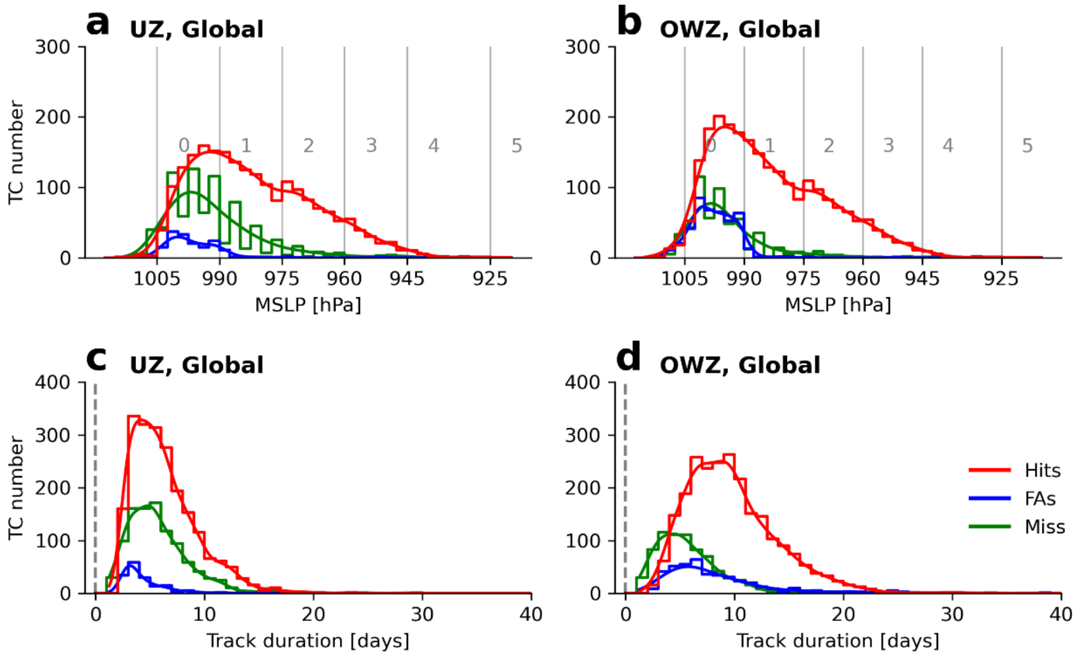

Fig. S1 | Properties of Hits, Misses, and False Alarms tracks for TC Trackers. a–b, Histograms of Hits (red), Misses (green), and FAs (blue) TC cases against TC intensity (minimum sea level pressure (SLP_{min})， unit: *hPa*, with the storm categories as defined according to '*TC classification' shown with vertical gray lines) detected by the UZ (**a**) and OWZ (**b**) trackers. **c–d,** same as **a–b**, except for histograms of Hits (red), Misses (green), and FAs (blue) TC cases against TC duration (unit: days).

S2. Comparison between RGTracks-20C and IBTrACS

S2.1 POD and FAR for individual basins

Both trackers are capable of capturing TCs that are observed globally and in most basins (Fig. 3). While the overall performances of both tracking algorithms are satisfactory, we note some differences in individual basins. There are more misses and lower POD in the Northeast Pacific (ENP). At the same time, the UZ tracker shows that the number of missed TCs exceeds the number of TC hits in the North Atlantic (NATL), and the POD value only reaches 54%. It is also worth noting that the two trackers show relatively higher FAs in the NI, corresponding to the highest FAR, compared to the others. As shown in Figs. S2a–h, the missed TCs in the ENP and NATL basins are primarily those that are shorter-lived and weaker in the observational records, which may not be fully captured in the low-resolution 20CR product. In the NI, a considerable number of FAs were detected (Figs. S2i–l), which may be attributed to issues related to observational records and data processing (more discussion in Text S2.2).

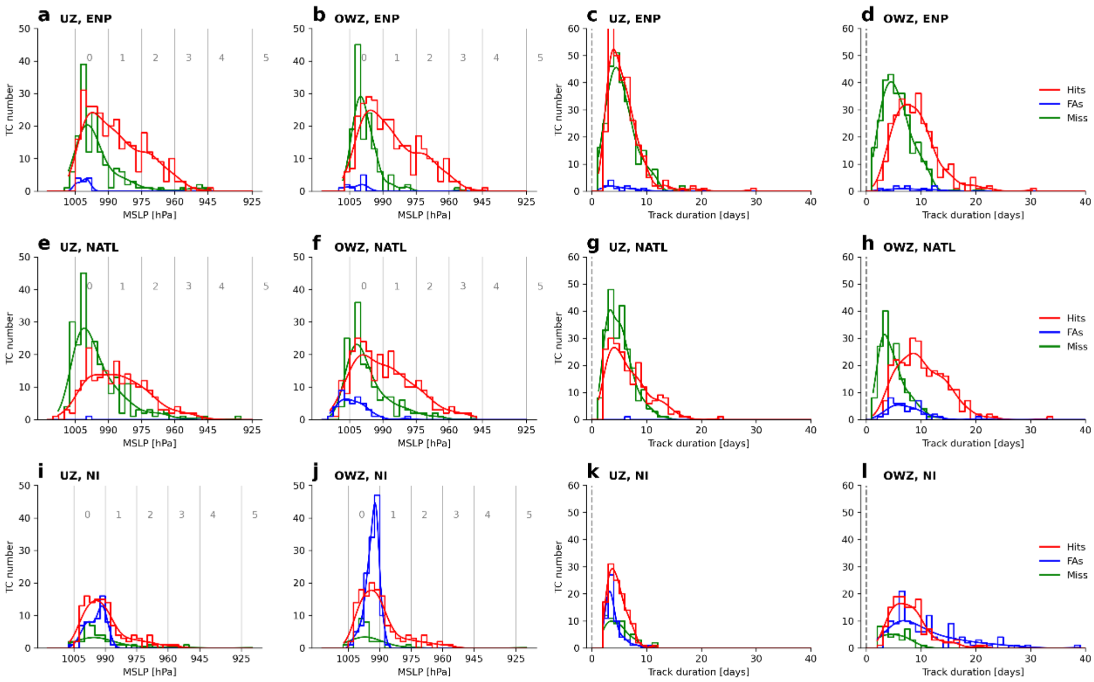

Fig. S2 | As in Fig. S1, but for the ENP, NATL, and NI. a–b, ENP, **e-h**, NATL, and **i-l**, NI. Histograms of TC intensity (SLP_{min}, unit: hPa) for the UZ tracker (**a, e, i**) and OWZ tracker (**b, f, j**), respectively. Histograms of TC duration (unit: days) for the UZ tracker (**c, g, k**) and OWZ tracker (**d, h, l**), respectively.

S2.2 ENP and NI

The RGTracks-20C can reproduce the climatology and interannual variability of observed TC activity globally and in most basins. However, due to uncertainties in the observations and other factors, discrepancies between the RGTracks-20C and IBTrACS exist in specific regions.

In terms of occurring frequency, the RGTracks-20C underestimates the annual average TC numbers compare to IBTrACS in the ENP (Figs. 4g–i). In the NI, RGTracks-20C shows a high FAR (Figs. 3a–b), particularly evident in the OWZ tracker results. Meanwhile, the correlation coefficients fall below 0.6 in the ENP and North Indian (Fig. 8d and Table S2). These results are consistent with a recent publication (Kim et al., 2021). As reported by (Tory et al., 2013a), short-lived TCs and land-based effects may explain the lower correlation in the ENP basin compared to other basins. Meanwhile, in the NI basin, due to the complex geographic conditions, trackers may easily misidentify monsoon depressions as TCs (Raavi and Walsh, 2020; Tory et al., 2013a) (Figs. 3, S2). These, suggesting the limitations of tracking algorithms, may be part of the reasons explaining the inconsistencies between the RGTracks-20C and IBTrACS in the ENP and NI.

On the other hand, we find that the comparisons of TC activity between RGTracks-20C and IBTrACS are sensitive to the starting year of reliable observations (Table S3). For example, in the ENP, the RGTracks-20C severely underestimates TC activity compared to IBTrACS before 1988, which results in a low correlation between the two datasets, as well as a trend with opposite signs compared to observations (Figs. 8, S3a–d, and Tables S2, S4 and S5). In the NI, the IBTrACS record of TC intensity (SLP_{min}) is missing before 1990 (Figs. 8c, S3h–g, and Table S3). With the assistance of Dr. Jennifer Gahtan from NOAA's National Center for Environmental Information, we found that the starting years for the availability of SLP_{min} in the ENP and NI are 1988 and 1990 (Table S3), respectively. In other words, given the high correlation and greater consistency in long-term trend changes between RGTracks-20C and IBTrACS after excluding data in the 1980s (Fig. S3 and Tables S2, S5), we conclude that limitations of the IBTrACS records are the major error source leading to the inconsistencies between the RGTracks-20C and IBTrACS.

Therefore, excluding these two basins increases the correlation coefficients for global TC activity (number, TC days, and intensity) between RGTracks-20C and IBTrACS (Table S6). Specifically, for TC number, the correlation coefficient increases from 0.65 to 0.67 for the UZ tracker and from 0.68 to 0.87 for the OWZ tracker. For TC days, the coefficients for the UZ tracker and OWZ tracker rise from 0.78 to 0.85 and 0.62 to 0.70, respectively. For TC intensity, the correlation coefficients improve from 0.58 to 0.65 for the UZ tracker and from 0.84 to 0.85 for the OWZ tracker.

After excluding these two basins, the long-term trend of global TC activity based on the RGTracks-20C becomes more closely aligned with the IBTrACS (Table S7). In terms of TC days, the results from the UZ tracker are consistent with observations, showing a significant decrease. For intensity, the results from the OWZ tracker align with observations, with a significant increase in intensity.

Table S2 | Correlation coefficients of TC activity (number, days, intensity) in the ENP and NI between IBTrACS and RGTracks-20C. This study initially verified the reliability of RGTracks-20C from 1979-2014. However, considering the lack of observational data, the study periods for ENP and NI were selected from 1988-2014 and 1990-2014, respectively. Asterisks indicate the confidence levels,1 asterisk (*) = 90%, 2 asterisks (**) = 95%, and 3 asterisks (***) = 99%.

Basin	Characteristics	Period	UZ	OWZ
ENP	Number	1979-2014	0.41**	0.35**
		1988-2014	0.77***	0.82***
	Duration	1979-2014	0.65***	0.58***
		1988-2014	0.89***	0.87***
	Intensity	1979-2014	0.01	0.15
		1988-2014	0.74***	0.66***
	Intensity-C	1979-2014	-0.02	0.11
		1988-2014	0.71***	0.67***
NI	Number	1979-2014	0.49***	0.54***
		1990-2014	0.58***	0.61***
	Duration	1979-2014	0.62***	0.34**
		1990-2014	0.63***	0.38*
	Intensity	1979-2014	0.48***	0.36**
		1990-2014	0.69***	0.62***
	Intensity-C	1979-2014	0.55***	0.47***
		1990-2014	0.68***	0.68***

Table S3 | Starting years of TC intensity (*SLP_{min}*) recordings by different agencies across major ocean basins. (Information provided by Dr. Jennifer Gahtan, NOAA's National Center for Environmental Information.)

Basin	Agencies							
North Atlantic	HURDAT2	M Chenoweth	DS824	TD9636				
	1979*1	1851	1851	1899				
East Pacific	HURDAT2		DS824					
	1988	0	1949					
Central Pacific	HURDAT2							
	2001							
West Pacific	China	Japan	HKO	JTWC	DS824	TD9636		
	1949	1951	1961	2001	1945	1945		
North Indian Ocean	India	JTWC	DS824					
	1990*2	2001	Mid-19070's					
Southern Hemisphere	La Reunion	Australia	New Zealand,	Nadi	JTWC	DS824	Neumann	TD9636
	1977	1907	1968	1992	2001	1877	1960s	1956

Note:

*1. North Atlantic: 1979 (with prior data given if there was a specific observation) HURDAT2.

*2. North Indian Ocean: 1990 (soon to be 1982) India.

3. Multiple Basins 1945 TD9635.

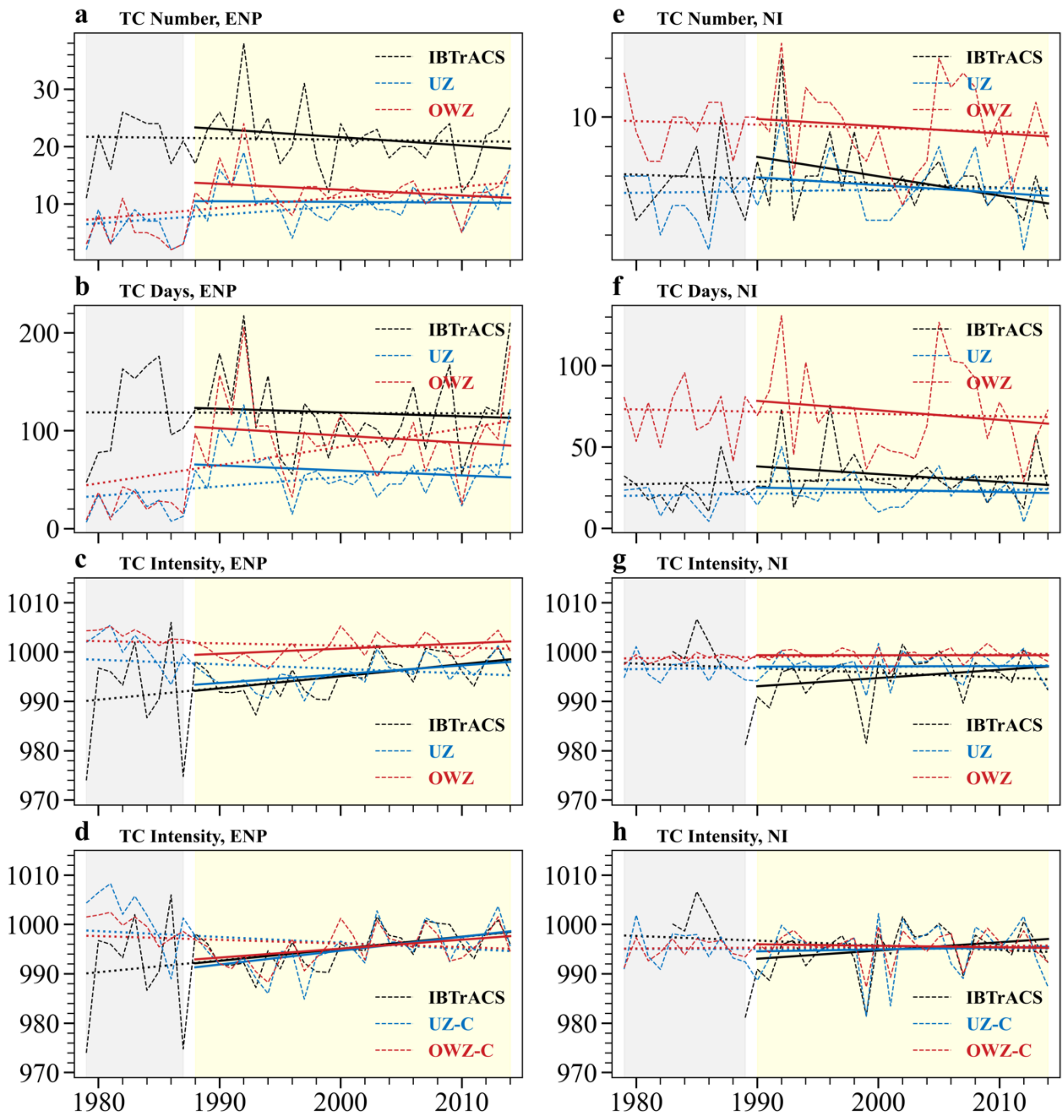

Fig. S3 | As in Fig. 4, but for the ENP and NI. a-d, time series and trends of TC number (unit: $year^{-1}$) **(a)**, days (unit: $day \cdot year^{-1}$) **(b),** and intensity (unit: $hPa \cdot year^{-1}$) before correction **(c)** and after bias correction **(d)**, as recorded by IBTrACS (black) and RGTracks-20C (UZ tracker: blue, OWZ tracker: red) in the ENP. **e-h**, same as **a-d**, but for time series and trends of TC number **(e)**, days **(f)**, and intensity before correction **(g)** and after bias correction **(h)**, as recorded by IBTrACS and RGTracks-20C in the NI.

S2.3 Differences in TC Duration

The duration of TCs also varies with trackers: the OWZ tracker detects TCs with durations that closely resemble the distribution in IBTrACS, while the UZ tracker detects shorter TC lifetime compared to observations in the global and most basins (Figs. 5b, 6b). However, in the NI Basin, the OWZ tracker lasts longer than the IBTrACS. We explain this by comparing the first (Fig. S4a) and last (Fig. S4b) dates of detected and observed tracks, showing that the UZ tracker typically identifies the first point later and the last point earlier than IBTrACS. For the OWZ tracker, it is worth noting that in the NI and NATL basins, the first point detected is earlier than the observation. The OWZ tracker was developed based on the conditions of TC formation and is capable of tracking vorticity perturbations that later develop into genuine TCs (Tory et al., 2013b), an example being the African easterly waves associated with TCs in the NATL[2]. We observe a weak correlation between RGTracks-20C and IBTrACS for TC duration in the NI basin, especially OWZ tracker (0.34, Table S2), which is consistent with the detection of TCs from ERA-Interim based on the OWZ by (Bell et al., 2018) (0.338). They suggest that this weaker correlation may be attributed to a higher FAR in this specific region (detailed discussion in Text S1).

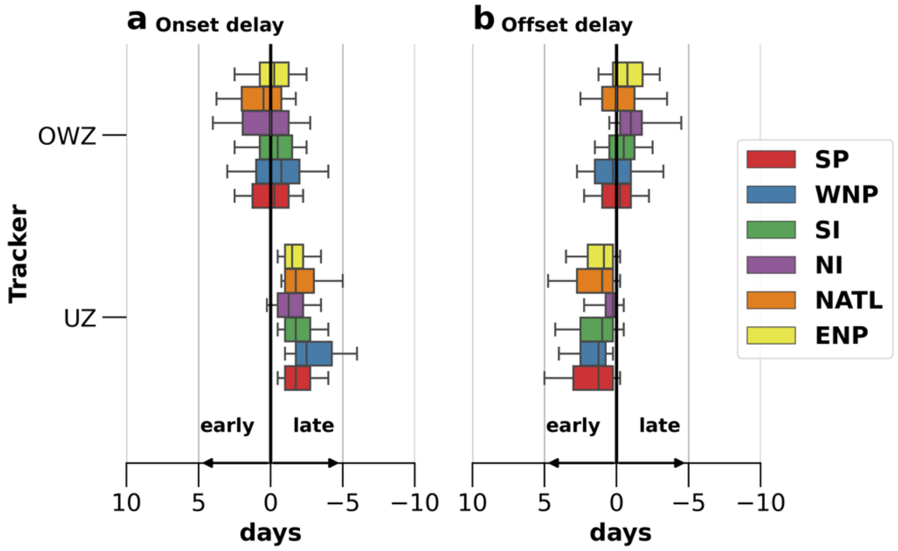

Fig. S4 | Onset and offset delays in TC detection by OWZ and UZ trackers. Onset delay **(a)**: the delay between the first detection by both trackers and the first record in IBTrACS. Offset delay **(b)**: the delay between the last detection by both trackers and the last record in IBTrACS. Colors represent different ocean basins: South Pacific (SP) (red), western North Pacific (WNP) (blue), South Indian (SI) (green), NI (purple), NATL (orange), and ENP (yellow). Box plots indicate the 25th, 50th, and 75th percentiles, whiskers show the 10th and 90th percentiles, and outliers are not shown.

S2.4 TC activity trends

For TC number (Figs. 7a, 8a, and Table S4), based on IBTrACS, the WNP and SP exhibit significant decreasing trends of -0.12 $year^{-1}$ and -0.10 $year^{-1}$, respectively. Our results show that the RGTracks-20C is able to reproduce the decreasing trends in these basins, with the UZ and OWZ trackers both indicating a significant decrease of -0.14 $year^{-1}$ and -0.12 $year^{-1}$ for the WNP. For the SP, the UZ tracker shows a significant decrease (-0.12 $year^{-1}$), while the trend by the OWZ tracker is not statistically significant. In the NATL basin, RGTracks-20C is able to reproduce the significant increasing trend in TC number recorded by IBTrACS (0.24 $year^{-1}$), with the OWZ algorithm (0.14 $year^{-1}$) showing a significant increase consistent with observations. On a global scale, as well as in the NI and SI, the trends observed by IBTrACS are not statistically significant, and RGTracks-20C yields consistent results in these basins. Notably, in the ENP, there are significant differences in the long-term trends between IBTrACS and RGTracks-20C, which is discussed in Text S2.2.

For TC days (Figs. 7b, 8b, and Table S4), the RGTracks-20C reproduces the significant decreasing trend in the SP and the significant increasing trend in the NATL. However, it should be noted that the increasing trend of the UZ algorithm in the NATL does not pass the 90% confidence level. In the WNP, the RGTracks-20C shows agreement with the observed long-term trend, although the trends are not significant. In addition, at the global scale and in the NI and SI, the RGTracks-20C well captures the observed insignificant long-term trends of TC days. On the other hand, in the ENP, there is a significant difference in the slopes of the long-term trends of IBTrACS and RGTracks-20C, which is consistent with the difference in the number of TCs between them (see Text S2.2 for more details).

For TC intensity, the IBTrACS shows a significant increase in TC intensity globally (0.17 $hPa \cdot year^{-1}$) and in the SI (0.27 $hPa \cdot year^{-1}$), and the increasing trend is successfully reproduced by the RGTracks-20C, especially for the OWZ tracker (Figs. 7c–d, 8c, and Table S4). Although the long-term trends given by the UZ tracker are also consistent with the observations, except that they do not reach the significance level. In addition, the RGTracks-20C captures the insignificant trends in TC intensity over the WNP, NATL, NI, and SP. In summary, the RGTracks-20C shows high consistency with observations globally and in most regions. However, in the ENP, IBTrACS shows a significant increasing trend, while RGTracks-20C shows a non-significant decreasing trend (see Text S2.2).

Overall, the RGTracks-20C can generally reproduce the long-term trends of TC activity in IBTrACS globally and over most of the basins. However, some discrepancies are observed in some specific basins, especially the ENP and NI. These inconsistencies may be related to uncertainties in the observations and limitations of tracking algorithms (see Text 2.2). If the evaluation period is restricted to the 1980s, the correlation between RGTracks-20C and IBTrACS in recording TC activity improves significantly (Table S2), and more consistent long-term trend results are obtained (Table S5).

Table S4 | Linear trends in TC activity globally and across six basins, as recorded in IBTrACS and RGTracks-20C. Grey background indicates that the trends between IBTrACS and RGTracks-20C are consistent sign and statistical significance. Asterisks indicate the confidence levels,1 asterisk (*) = 90%, 2 asterisks (**) = 95%, and 3 asterisks (***) = 99%. UZ-C and OWZ-C represent results after intensity bias correction.

		Global	WNP	ENP	NATL	NI	SI	SP
Number ($year^{-1}$)	IBTrACS	**-0.06**	**-0.12***	-0.08	**0.24*****	**0.01**	**-0.04**	**-0.10***
	UZ	**-0.01**	**-0.14***	0.16***	**0.07**	**0.02**	**-0.01**	**-0.12****
	OWZ	0.19	**-0.12***	0.20***	**0.14****	-0.04	0.04	**-0.08**
TC days ($day \cdot year^{-1}$)	IBTrACS	**-2.70**	**-3.64*****	-0.03	**1.93*****	**0.16**	**-0.15**	**-1.02***
	UZ	**-0.29**	**-0.95**	0.97**	**0.59**	**0.11**	**-0.40**	**-0.72***
	OWZ	1.82	**-1.32**	1.86***	**1.38****	-0.14	0.63	**-0.79***
Intensity ($hPa \cdot year^{-1}$)	IBTrACS	**0.17*****	**0.03**	0.24**	**-0.06**	-0.09	**0.27*****	**0.08**
	UZ	**0.03**	**0.01**	-0.09	**-0.06**	0.03	**0.05**	-0.06
	OWZ	**0.05*****	**0.03**	-0.05	**-0.05**	0.03	**0.08*****	**0.03**
	UZ-C	**0.04**	**0.02**	-0.12	**-0.10**	0.01	**0.07**	-0.10
	OWZ-C	**0.09*****	**0.04**	-0.07	**-0.08**	0.02	**0.13*****	**0.05**

Table S5| As in Table S4, but for ENP and NI. This study initially verified the reliability of RGTracks-20C for the period 1979–2014. Due to limited observational data, the study periods for the ENP and NI basins were adjusted to 1988–2014 and 1990–2014, respectively.

Basin	Characteristics	Period	IBTrACS	UZ	OWZ
ENP	Number	1979-2014	-0.08	0.16*	0.20*
		1988-2014	**-0.04**	**0**	**-0.07**
	Duration	1979-2014	-0.03	0.97*	1.86*
		1988-2014	**-0.39**	**-0.51**	**-0.73**
	Intensity	1979-2014	0.24**	-0.09	-0.05
		1988-2014	**0.24*** **	**0.18*** **	**0.10***
	Intensity-C	1979-2014	0.24**	-0.12	-0.07
		1988-2014	**0.24***	**0.28*** **	**0.18** **
NI	Number	1979-2014	**0.01**	**0.02**	-0.04
		1990-2014	**-0.07**	**-0.04**	**-0.06**
	Duration	1979-2014	**0.16**	**0.11**	-0.14
		1990-2014	**-0.48**	**-0.14**	**-0.58**
	Intensity	1979-2014	-0.09	0.03	0.03
		1990-2014	**0.27**	**0.03**	**0.01**
	Intensity-C	1979-2014	-0.09	0.01	0.02
		1990-2014	**0.27**	**0.06**	-0.01

Table S6 | Correlation coefficients of global TC activity (number, days, intensity) between RGTracks-20C and IBTrACS with the ENP and NI included and excluded, respectively. Asterisks indicate the confidence levels,1 asterisk (*) = 90%, 2 asterisks (**) = 95%, and 3 asterisks (***) = 99%.

Characteristics	Region	UZ	OWZ
Number	ENP & NI included	0.65***	0.68***
	ENP & NI excluded	0.67***	0.87***
Days	ENP & NI included	0.78***	0.63***
	ENP & NI excluded	0.85***	0.70***
Corrected Intensity	ENP & NI included	0.58***	0.84***
	ENP & NI excluded	0.65***	0.85***

Table S7 | As in Table S4, but with ENP and NI basins included and excluded, respectively.

Characteristics	Dataset	ENP & NI included	ENP & NI excluded
Number	IBTrACS	**-0.06**	**0.04**
	UZ	**-0.01**	-0.19**
	OWZ	0.19	**0.08**
Duration	IBTrACS	**-2.70**	**-2.83***
	UZ	**-0.29**	**-1.37***
	OWZ	1.82	0.10
Intensity	IBTrACS	**0.17*****	**0.12*****
	UZ	**0.03**	**0.02**
	OWZ	**0.05*****	**0.05*****
	UZ-C	**0.04**	**0.02**
	OWZ-C	**0.09*****	**0.09*****

S3 Case studies

S3.1 1928 ‘Okeechobee’ hurricane

The 1928 Okeechobee hurricane, the first recorded Category 5 hurricane in U.S. history, caused approximately 2,500 fatalities, making it one of the deadliest natural disasters in Florida's history (Mitchell, 1928). The lifespans and tracks (location) of the hurricane were recorded in IBTrACS (Knapp et al., 2018), beginning at 00:00 on 6 September 1928 and ending at 00:00 on 21 September 1928, lasting 16 days.

The RGTracks-20C almost completely reproduces the lifespans of the hurricane as recorded in IBTrACS, and the OWZ algorithm (06:00 on 5 September 1928 to 00:00 on 20 September 1928) performs particularly well. The detected positions of Okeechobee in RGTracks-20C are highly consistent with those recorded by IBTrACS (Figs. 10a, S5), with the bias within the range of ±1°. Specifically, the UZ tracker shows an average latitude bias of -0.05° and longitude bias of 0.13° for the hurricane (Figs. S5a, c); the OWZ tracker shows an average latitude bias of 0.02°and longitude bias of -0.12° (Figs. S5b, d), which are much smaller than the horizontal resolution of 20CRv3.

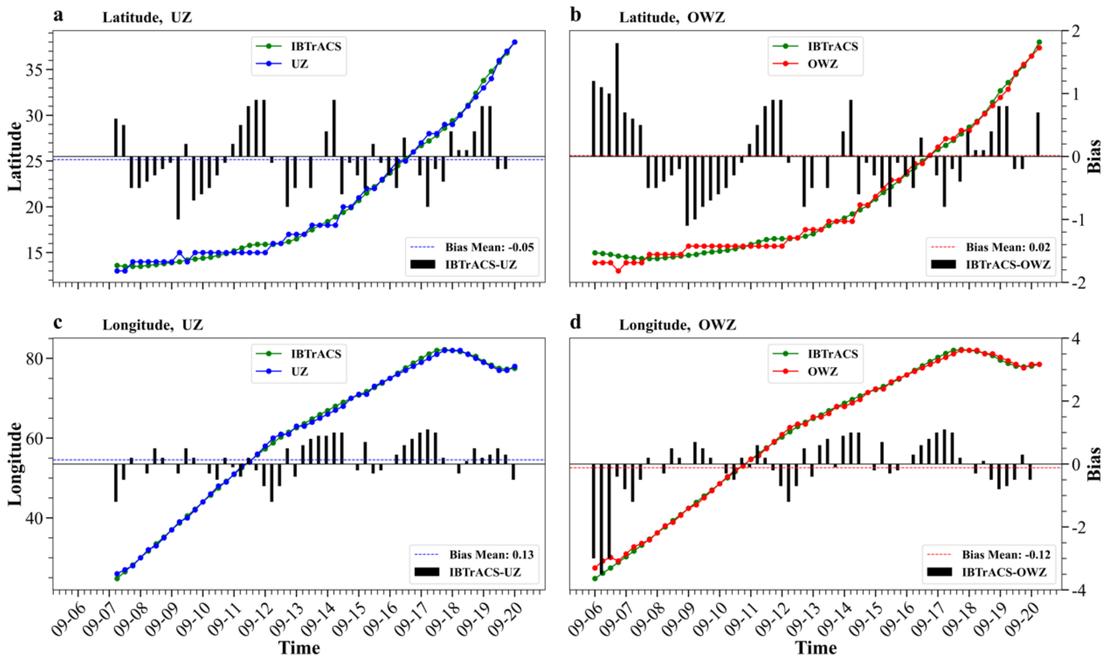

Fig. S5| Position records of Hurricane Okeechobee from IBTrACS and RGTracks-20C, and analysis of positional bias between these two datasets. a-b, latitude (unit: °) records for Hurricane Okeechobee as reported by IBTrACS and the UZ (**a**) and OWZ (**b**) trackers, with corresponding bias values shown in bar and dashed lines represent (UZ: -0.05, OWZ: 0.02). **c-d**, longitude (unit: °) records for Hurricane Okeechobee as reported by IBTrACS (green) and the UZ (blue) (**c**) and OWZ (red) (**d**) trackers, with corresponding bias values shown in bar and dashed lines represent (UZ: 0.13, OWZ: -0.12).

However, it is worth noting that IBTrACS lacks most of the intensity records for this hurricane. Intensity records are mostly concentrated during the hurricane's landfall

(Table S8). Between 17:30 and 18:30 UTC on September 12, the hurricane's center passed over Guadeloupe, where a minimum pressure of 940 *hPa* was recorded, marking the first recorded intensity in IBTrACS. In the RGTracks-20C, the pressure directly obtained from the 20CRv3 was 972 *hPa*, higher than the observed value. However, after applying bias correction, the RGTracks-20C results closely matched the observations (UZ: 955 *hPa*; OWZ: 940 *hPa*).

After leaving Guadeloupe, the hurricane continued moving west-northwestward. On September 13, the 15-mile-wide eye of the hurricane crossed Puerto Rico from southeast to northwest, making landfall near Guayama and exiting between Aguadilla and Isabela (Pérez, O., 1899). During this time, a ship near the southern coast reported a pressure of 931 *hPa* (Landsea et al., 2008a). In the RGTracks-20C (Table S8), the UZ algorithm recorded a pressure of 930 *hPa*, while the OWZ algorithm recorded 920 *hPa*. Since the observed pressure was not at the center of the hurricane, we consider the OWZ estimate of 920 *hPa* to be a more reasonable estimate of the central pressure of the storm.

After leaving Puerto Rico, the hurricane weakened slightly, recording 941 *hPa* at Isabela. The storm brushed the northern coast of Hispaniola and then moved northwestward, gradually re-intensifying. By September 15, the storm crossed the Bahamas as a strong Category 4 hurricane, passing near Nassau at around 10:00 UTC on September 16. At 00:00 UTC on September 17, the hurricane made landfall in West Palm Beach, southeastern Florida, with a pressure of 929 *hPa* (Landsea et al., 2008a, b), breaking the previous record of 935 *hPa* set by the 1926 Miami hurricane. The RGTracks-20C successfully reproduced the weakening and subsequent intensification trend from Puerto Rico to Florida (Fig. 10 and Table S8), closely matching observations, with UZ recording 925 *hPa* and OWZ recording 915 *hPa* (Figs. 10c, d and Table S8).

Table S8 | Intensities (*SLP_{min}*) of Hurricane Okeechobee recorded in IBTrACS and RGTracks-20C. The World Meteorological Organization (WMO) and the United States of America (USA) agencies in IBTrACS recorded the hurricane intensities, while RGTracks-20C results are represented by UZ, OWZ, UZ-C, and OWZ-C.

UZ and OWZ indicate the SLP_{min} directly obtained from 20CRv3, while UZ-C and OWZ-C represent bias-corrected results based on observational data.

Time	WMO	USA	UZ	OWZ	UZ-C	OWZ-C
1928/9/12 18:00	940	940	971	972	955	940
1928/9/13 18:00	931	931	958	958	930	920
1928/9/13 21:00		936				
1928/9/14 0:00	941	94	958.	958	930	920
1928/9/17 0:00	929	929	955	955	925	915
1928/9/18 6:00	977	977	958	958	982	965
1928/9/18 9:00		976				
1928/9/18 12:00	976	976	985	985	982	965
1928/9/18 15:00		976				
1928/9/18 18:00	977	977	986	986	983	966
1928/9/18 19:00	977	977				
1928/9/20 12:00	1008	1008				
1928/9/20 15:00		1006				
1928/9/20 18:00		1005				

After making landfall, the hurricane moved inland, passing over Lake Okeechobee. The IBTrACS recorded pressures of 976–977 hPa during this phase, although stronger unofficial records were also found, including a pressure reading of 942hPa at a canal point near the lake, and 966hPa in Bartow, north of Lake Okeechobee (Mitchell, 1928). In the RGTracks-20C, the UZ algorithm's intensity estimates were closely aligned with IBTrACS, while the OWZ algorithm was more consistent with the unofficial records, particularly the 966 hPa.

S3.2 1920 ‘1920232N24150’ Typhoon in the WNP

The typhoon labeled ‘1920232N24150’ occurred in the WNP in 1920. According to meteorological records from Kagawa Prefecture of Japan (https://www.shikoku-saigai.com/archives/25443?preurl&query_pref&query_city&query_dis_kind&query_s=%E5%8F%B0%E9%A2%A8+&query_date=18680908-19261225&query_paged=11), the typhoon brought 20 to 110 mm of rainfall to the region between

August 20 and 21. Although the typhoon had a short duration and caused relatively low rainfall, it damaged crops, particularly fruit trees and vegetables, indicating that it made landfall in Japan.

While the IBTrACS provides a record of this typhoon, it includes only part of the storm's tracks, especially missing the tracks during its landfall in Japan (August 20–21). However, the RGTracks-20C successfully reconstructs the typhoon's tracks. To verify the accuracy of the track in the RGTracks-20C, we used the 20CRv3 to generate sea level pressure maps for August 19 to 21, 1920, and compared them with historical weather chart archives (Figs. S6–8).

Historical weather charts provided by the National Institute of Informatics (NII) confirm the presence of a low-pressure system southeast of Japan on August 19 (Fig. S6). By August 20 (Fig. S7), the system was moving northwestward, approaching the Japanese coast, and by August 21 (Fig. S8), it was located over Japan's southern islands, further confirming its landfall.

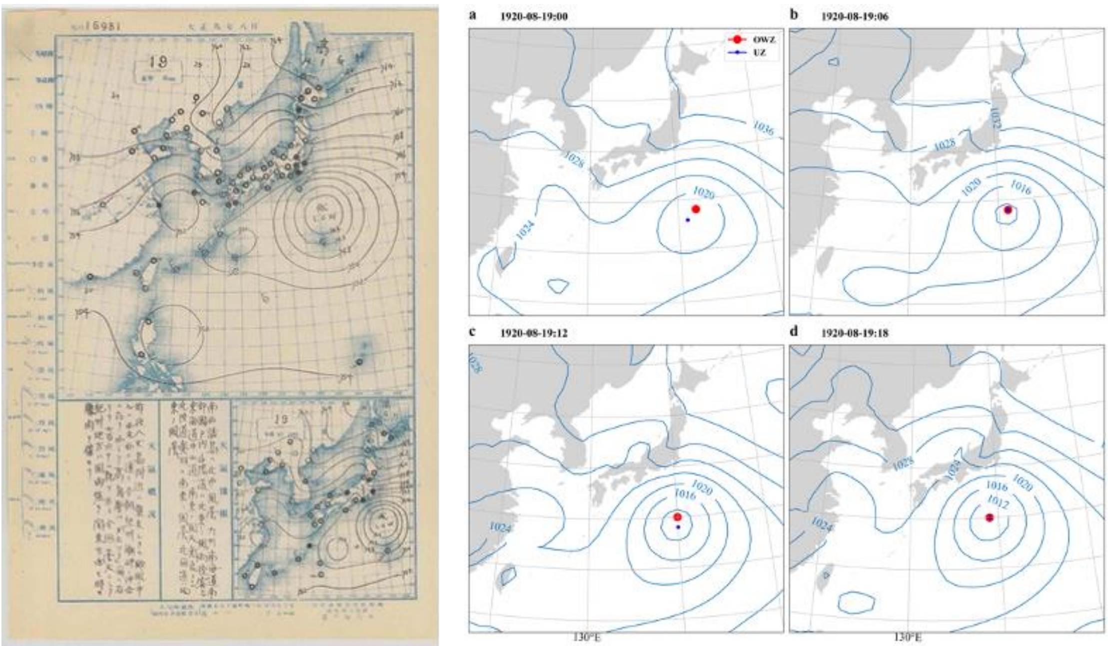

Fig. S6 | Comparison of the historical Asia-Pacific weather map and sea-level pressure from 20CRv3 for the 1920 typhoon '1920232N24150' on August 19, 1920. On the left is a historical Asia-Pacific weather map provided by the National Institute of Informatics (NII). **a-d**, Sea level pressure on August 20, 1920 at 00:00 (**a**), 06:00 (**b**), 12:00 (**c**), and 18:00 (**d**) based on 20CRv3. The blue and red dots indicate the positions of the typhoon as identified by the UZ (blue) and OWZ (red) trackers, respectively. The historical weather chart was created by

NII "Digital Typhoon" based on "Weather Charts" from Japan Meteorological Agency and obtained from https://agora.ex.nii.ac.jp/digital-typhoon/weather-chart/.

A comparison with historical weather charts shows that the low-pressure system in 20CRv3 closely matches the observation records, indicating that 20CRv3 effectively simulated the typhoon's development and landfall in Japan. Thus, RGTracks-20C not only captures the typhoon's tracks but also accurately represents the pre-landfall (Fig. S6), landfall (Fig. S7), and post-landfall phases (Fig. S8). This reconstruction fills the gaps in IBTrACS regarding the missing tracks of the typhoon during this period.

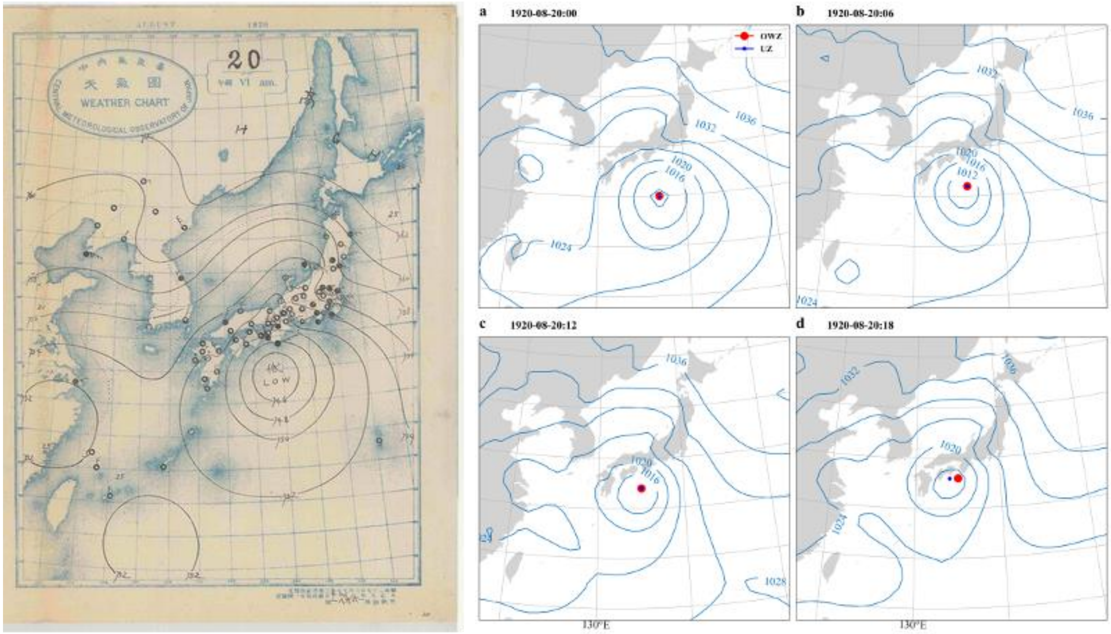

Fig. S7 | As shown in Fig. S6, but for August 20, 1920. The historical weather chart was created by NII "Digital Typhoon" based on "Weather Charts" from Japan Meteorological Agency and obtained from https://agora.ex.nii.ac.jp/digital-typhoon/weather-chart/.

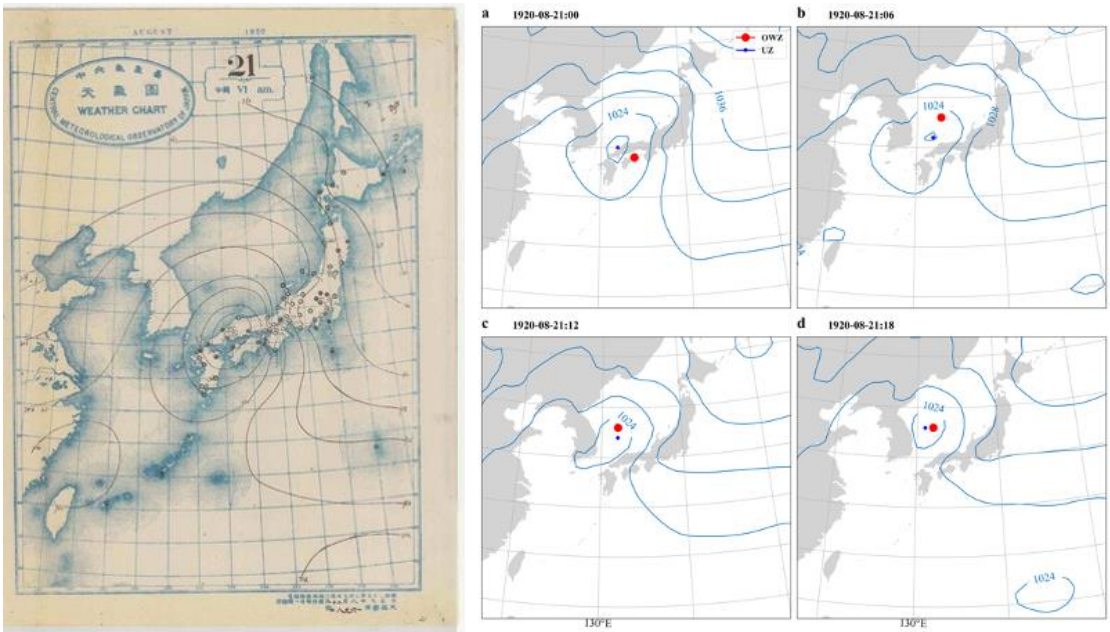

Fig. S8 | As shown in Fig. S6, but for August 21, 1920. The historical weather chart was created by NII "Digital Typhoon" based on "Weather Charts" from Japan Meteorological Agency and obtained from https://agora.ex.nii.ac.jp/digital-typhoon/weather-chart/.

S4 Impacts of assimilating IBTrACS in the 20CRv3

SLP_{min} data provided by IBTrACS were assimilated during the production of 20CRv3, which implies that the IBTrACS data could have an impact on the constructed RGTracks-20C. In other words, the RGTracks-20C and IBTrACS are not fully independent data. For instance, the RGTracks-20C underestimates the TC activity over the ENP basin prior to 1988 because of the incomplete TC intensity records of the IBTrACS (mentioned above in Text S4.2). To understand this discrepancy, we analyze the annual availability of observational data and that assimilated into the 20CRv3, and find a significant increase over time, particularly after 1950 (Fig. S9). Part of the assimilated data came from IBTrACS, exhibiting similar trends and variations (Fig. S9). Although these observations were relatively few, they were crucial for accurately reproducing TCs in the reanalysis. Fig. S10 shows that the RGTracks-20C exhibits trends and variability that are highly similar to those recorded in IBTrACS, particularly after 1950. These suggest that the assimilation of observational data from IBTrACS influences the number of TCs in the 20CRv3 (Wang et al., 2012). This assimilation process likely enhances the accuracy in simulating the structure and intensity of TCs (Slivinski et al., 2019), facilitating detection by TC trackers.

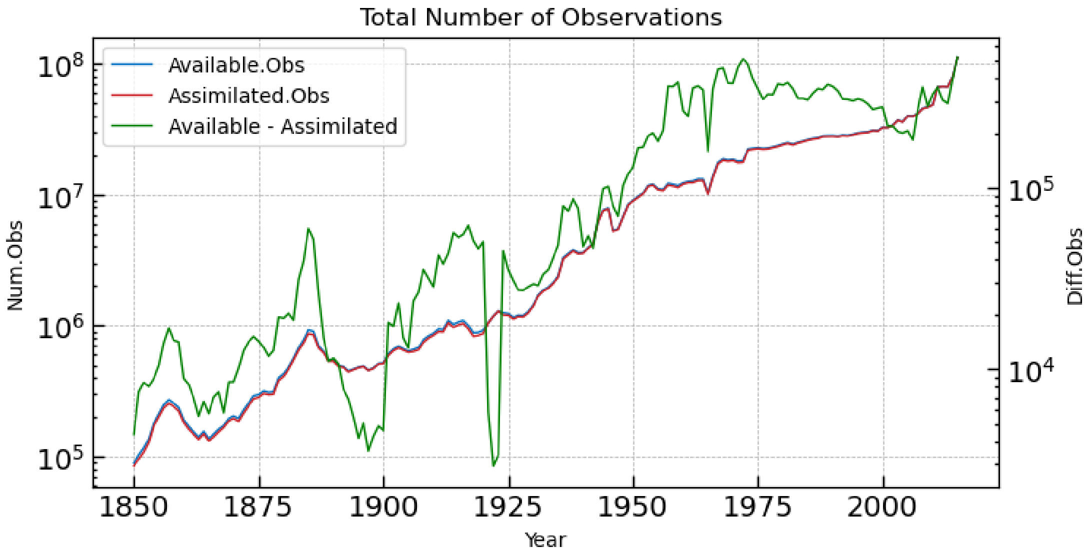

Fig. S9 | Time series of the total number of available and assimilable observations from 1850 to 2015. Blue and red lines represent assimilable and assimilated observations, respectively. And the green line indicates the difference between the available and assimilated observations.

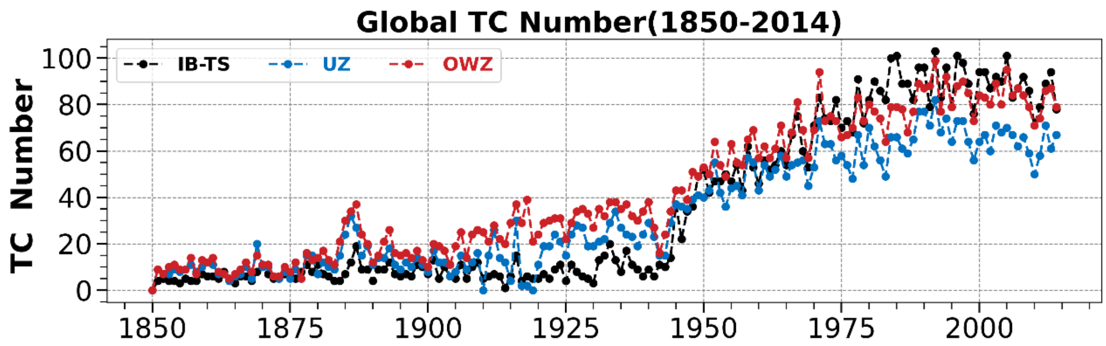

Fig. S10 | Time series of the TC number from 1850 to 2014. The black dotted line represents the IBTrACS, while the blue and red dotted lines represent the results identified by the UZ and OWZ trackers, respectively.

References

Accarino, G., Donno, D., Immorlano, F., Elia, D., and Aloisio, G.: An Ensemble Machine Learning Approach for Tropical Cyclone Localization and Tracking From ERA5 Reanalysis Data, Earth Space Sci., 10, e2023EA003106, https://doi.org/10.1029/2023EA003106, 2023.

Bell, S. S., Chand, S. S., Tory, K. J., and Turville, C.: Statistical Assessment of the OWZ Tropical Cyclone Tracking Scheme in ERA-Interim, J. Clim., 31, 2217–2232, https://doi.org/10.1175/JCLI-D-17-0548.1, 2018.

Bourdin, S., Fromang, S., Dulac, W., Cattiaux, J., and Chauvin, F.: Intercomparison of four algorithms for detecting tropical cyclones using ERA5, Geosci. Model Dev., 15, 6759–6786, https://doi.org/10.5194/gmd-15-6759-2022, 2022.

Hodges, K., Cobb, A., and Vidale, P. L.: How Well Are Tropical Cyclones Represented in Reanalysis Datasets?, J. Clim., 30, 5243–5264, https://doi.org/10.1175/JCLI-D-16-0557.1, 2017.

Kim, H., Lee, M.-I., Kim, S., Lim, Y.-K., Schubert, S. D., and Molod, A. M.: Representation of Tropical Cyclones by the Modern-Era Retrospective Analysis for Research and Applications Version 2, Asia-Pac. J. Atmospheric Sci., 57, 35–49, https://doi.org/10.1007/s13143-019-00169-y, 2021.

Knapp, K. R., Diamond, H. J., Kossin, J. P., Kruk, M. C., Schreck, C. J., and others: International best track archive for climate stewardship (IBTrACS) project, version 4, NOAA Natl. Cent. Environ. Inf., 10, https://doi.org/10.25921/82ty-9e16 (Accessed on 03-AUG-20243), 2018.

Landsea, C., Anderson, C., Bredemeyer, W., Carrasco, C., Charles, N., Chenoweth, M., Clark, G., Dunion, J., Ellis, R., Fernandez-Partagas, J., and others: Documentation of atlantic tropical cyclones changes in HURDAT, Atl. Oceanogr. Meteorol. Lab., 2008a.

Landsea, C. W., Glenn, D. A., Bredemeyer, W., Chenoweth, M., Ellis, R., Gamache, J., Hufstetler, L., Mock, C., Perez, R., Prieto, R., Sánchez-Sesma, J., Thomas, D., and Woolcock, L.: A Reanalysis of the 1911–20 Atlantic Hurricane Database, J. Clim., 21, 2138–2168, https://doi.org/10.1175/2007JCLI1119.1, 2008b.

Mitchell, C. L.: The West Indian hurricane of September 10–20, 1928, Mon. Weather Rev., 56, 347–350, https://doi.org/10.1175/1520-0493(1928)56<347:TWIHOS>2.0.CO;2,1928.

Notes on the Tropical Cyclones of Puerto Rico: https://www.aoml.noaa.gov/hrd/data_sub/perez_21_34.pdf, last access: 15 October 2024.

Raavi, P. H. and Walsh, K. J. E.: Basinwise Statistical Analysis of Factors Limiting Tropical Storm Formation From an Initial Tropical Circulation, J. Geophys. Res. Atmospheres, 125, e2019JD032006, https://doi.org/10.1029/2019JD032006, 2020.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H., Kennedy, J., Spencer, L. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kubota, H., Blancq, F. L., Lee, T.-C., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Moore, G. W. K., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Smith, C. A., Tinz, B., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K., and Wyszyński, P.: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system, Q. J. R. Meteorol. Soc., 145, 2876–2908, https://doi.org/10.1002/qj.3598, 2019.

Tory, K. J., Chand, S. S., McBride, J. L., Ye, H., and Dare, R. A.: Projected Changes in Late-Twenty-First-Century Tropical Cyclone Frequency in 13 Coupled Climate Models from Phase 5 of the Coupled Model Intercomparison Project, J. Clim., 26, 9946–9959, https://doi.org/10.1175/JCLI-D-13-00010.1, 2013a.

Tory, K. J., Dare, R. A., Davidson, N. E., McBride, J. L., and Chand, S. S.: The importance of low-deformation vorticity in tropical cyclone formation, Atmospheric Chem. Phys., 13, 2115–2132, https://doi.org/10.5194/acp-13-2115-2013, 2013b.

Wang, X. L., Feng, Y., and Swail, V.: North Atlantic wave height trends as reconstructed from the 20th century reanalysis, Geophys. Res. Lett., 39, https://doi.org/10.1029/2012GL053381, 2012.