Author's response for

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

Comments by reviewer 3

This manuscript introduces the Reanalysis-Based Global Tropical Cyclone Tracks Dataset for the Twentieth Century (RGTracks-20C), a significant undertaking that aims to reconstruct global TC tracks from 1850–2014. While RGTracks-20C represents a valuable contribution to the field, several major concerns regarding its independence, validation, and methodological limitations must be addressed before it can be considered for acceptance.

Authors' response: We sincerely thank you for your time and effort in providing such valuable and constructive comments on our manuscript. We agree that the initial version did not clearly define the dataset's limitations and intended use, and we acknowledge that some of our earlier wording may have been misleading.

Following your suggestions, we have substantially reorganized the manuscript structure and refined the positioning of the RGTracks-20C dataset. The revised version now explicitly emphasizes its role as a reconstructed supplement to the historical record, specifically designed to augment sparse observational data prior to the satellite era. We have also revised the wording throughout to clarify that this dataset should not be used as a standalone benchmark for analyzing long-term TC trends.

We hope these fundamental revisions adequately address your concerns regarding the dataset's validation, scope, and methodological limitations.

Point (1): The core issue is that the 20th Century Reanalysis (20CRv3) assimilates observed TC-associated data from IBTrACS. Consequently, RGTracks-20C is not fully independent of the best-track data against which it is validated. This circularity undermines the dataset's utility for the early period (pre-1940s), as the reanalysis is inherently informed by the very observations it is meant to supplement. As noted in prior literature (e.g., Emanuel, 2024), this dependency means that 20CRv3 cannot be expected to realistically capture TC activity in data-sparse periods. Without robust uncertainty quantification for these early years, the dataset risks providing a misleading foundation for analyzing long-term trends in response to climate change.

Authors' response: Thank you for your thoughtful comments and for bringing this important issue to our attention.

We agree that IBTrACS central pressure data are included among the observations assimilated in the production of 20CRv3. Specifically, 20CRv3 assimilates surface pressure reports from ISPD 4.7 (Compo et al., 2019), which include station observations, marine observations, and TC-related pressure data from IBTrACS (Knapp et al., 2010; Slivinski et al., 2021). The tropical cyclone-like structures present in the reanalysis therefore represent the dynamical response of the model to pressure observation forcing, rather than being directly constrained by full best-track information. Previous studies have also indicated that the assimilation of pressure observations in 20CRv3 may improve the representation of TC structure, which can in turn facilitate TC detection in the reanalysis (Slivinski et al., 2019, 2021; Chand et al., 2022). Nonetheless, we acknowledge that 20CRv3 and IBTrACS are not fully independent, and this has been explicitly stated in the revised manuscript (Lines 122–131, 483–486 and 619–623).

More importantly, we should clarify the intended scope of this study. The current version of RGTracks-20C, constructed using the OWZ (Objective Wade-Zeithaml) algorithm based on the 20CRv3 ensemble mean, is designed primarily to reconstruct historical TC tracks. It aims to augment the limited observational record during earlier periods characterized by sparse or incomplete observations, rather than to serve as a standalone dataset for detecting long-term TC trends in response to climate change. This application context is fundamentally different from the trend-detection problem critiqued by Emanuel (2024), and accordingly imposes different requirements on dataset independence.

Following your suggestion, we have revised the manuscript to make this scope explicit and to reduce the risk of misleading interpretation. In particular, we have rewritten the Abstract, Introduction, Results and Discussion, and Conclusion to clarify the data's utility. Furthermore, in Section 4, we have added specific statements to advise readers that this dataset should be used as a supplementary reconstruction of historical tracks rather than for long-term trend analysis. We hope that these revisions have addressed your concern more clearly.

Point (2): The validation presented is insufficient to establish confidence in the dataset's representation of TC behavior over time. The manuscript shows substantial inconsistencies in the trends of TC counts and intensity compared to observations. To address this, a more rigorous comparison of interannual and long-term trends—particularly for intense TCs, which may be better recorded in early observations—is essential. Furthermore, the study does not account for known heterogeneities in the observed record. For instance, the adjustment for TC intensity since 1945 discussed by Emanuel (2005, Nature) is not applied. Directly comparing the constructed TC intensity data to the raw IBTrACS data without such adjustments introduces a risk of misinterpretation.

Authors' response: We thank the reviewer for this important comment.

As noted in the above response, the primary purpose of the first version of RGTracks-20C (constructed using the OWZ algorithm) is to augment the historical tropical cyclone observations, rather than to detect long-term trends in tropical cyclone activity's response to climate change. Based on this objective, we selected the period 1979–2014 as the main validation period, owing to the higher accuracy of satellite-era observational data, which provides a reliable reference baseline. As demonstrated in Figs. 4 to 6 of the revised manuscript, RGTracks-20C exhibits similar climatological characteristics to observations globally and across most ocean basins during this period. This performance in representing TC climatology and interannual variability during the well-observed satellite era supports the overall reliability of the generated dataset. Furthermore, as shown in Figs. 8 and 9, RGTracks-20C meets the core objective of this study by demonstrating its capability in representing the track and intensity of historical tropical cyclones. Figure 8 illustrates that the dataset captures the spatial evolution of past TC tracks with high fidelity. Importantly, for many early-period tropical cyclones where observations are missing either track or intensity information, RGTracks-20C provides new insights, and even identifies TCs that occurred but were not recorded in IBTrACS, further demonstrating its value as a complementary reconstruction product. Given this specific focus on reconstruction rather than climate trend detection, the intensity adjustments for heterogeneities discussed by Emanuel (2005) have not been applied and are considered outside the current scope of this study.

We agree with your suggestion that strong tropical cyclones, being more frequently recorded in historical observations, provide a more robust basis for validation when prioritized. Previous studies (e.g., Hodges et al., 2017; Bourdin et al., 2022) also documented that reanalysis data and detection algorithms reproduce intense TCs more effectively than weaker ones. Following your suggestion, we conducted similar validation analyses for TCs of Category 1 and above. Results show that the agreement between RGTracks-20C and IBTrACS improves markedly (Table R1 and Fig. R1). Specifically, the POD increased from 0.77 to 0.95 (Table R1). For the 1979–2014 period, the interannual correlation for TC counts improved from 0.68 to 0.78, and for duration from 0.63 to 0.80 (Fig. R1). These results, which align with the conclusions of Hodges et al. (2017), demonstrate that RGTracks-20C is more reliable in capturing the climatology and interannual variability of intense TCs.

While we acknowledge the challenges in detecting weaker systems, the core objective of this paper remains the reconstruction of historical TC tracks to augment the sparse observational record. The performance in representing intense TCs confirms that RGTracks-20C provides a new and valuable resource for historical analysis.

Table R1: The Probability of Detection (POD, unit: %) for UZ (a) and OWZ (b). "TS+" represents tropical storm intensity and above, and "Cat1+" represents Category 1 intensity and above.

	RGTracks-20C	UZ
TS+	0.77	0.68
Cat1+	0.95	0.87

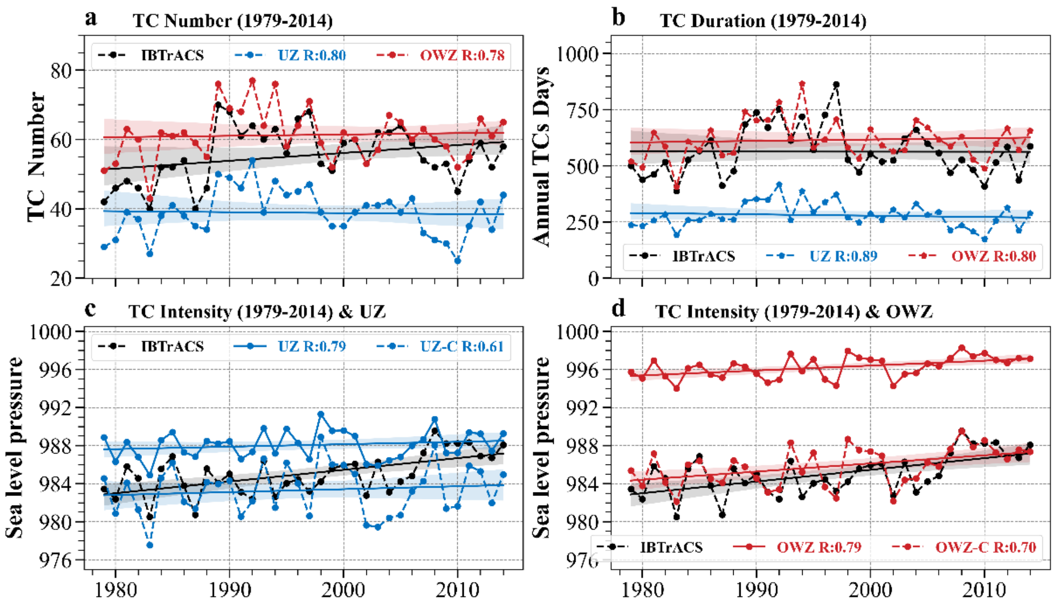

Figure R1: Time series of globally TC activities from IBTrACS, RGTracks-20C (OWZ) and UZ based results during the periods 1979-2014. TC activities are from the IBTrACS and RGTracks-20C (red) and UZ tracker (blue). a, TC number. b, TC days (unit: days). c, TC intensity in SLP_{min} (unit: *hPa*) in IBTrACS (black) and UZ-based before (blue solid line) and after (blue dotted line) bias correction. d, same as (c), except for TC intensity in SLP_{min} (unit: *hPa*) in IBTrACS (black) and RGTracks-20C before (red solid line) and after (red dotted line) bias correction. Shaded areas are the two-sided interval of the linear trend at the 95% confidence level. Straight lines are the linear regression. The correlation coefficients (R) between from IBTrACS and RGTracks-20C are marked in the figure legends. All correlation coefficients are statistically significant at the 99% confidence level.

Point (3): The dataset exhibits notable regional limitations, with lower detection skill in the Eastern North Pacific and North Indian basins, and to some extent in the North Atlantic. While the authors attribute this to biases in IBTrACS, it also suggests inherent limitations in the tracking algorithms or the reanalysis quality in these regions that are not fully resolved. The trade-offs between the two tracking algorithms are also a concern. The UZ tracker appears to miss weaker, shorter-lived TCs, while the OWZ tracker has a higher false alarm rate. The use of globally uniform thresholds may be sub-optimal, as TC characteristics and environmental conditions vary significantly across basins. A discussion on the potential for basin-specific tuning would strengthen the methodological justification.

Authors' response: We sincerely thank you for these constructive comments.

We agree that the lower skill in some basins should not be attributed solely to limitations in IBTrACS. In the revised manuscript, we have clarified this point and now state more explicitly that the basin-dependent discrepancies likely reflect a combination of factors, including uncertainties in the observational record, limitations of the tracking algorithms, and the ability of 20CRv3 to represent TCs in these regions (see revised manuscript, Sects. 3 and 5).

We also agree that the trade-offs between the two tracking algorithms need to be presented more clearly. To address this, we added a new subsection, Sect. 3.5, "Comparison of the OWZ and UZ trackers", in which we summarize the main differences between the two trackers in terms of POD/FAR, climatology, and interannual variability. In the revised manuscript, we now make clear that the first version of RGTracks-20C is constructed primarily from the OWZ tracker, while the UZ tracker is retained as a methodological reference and sensitivity check (see revised manuscript, Sect. 3).

We further agree that globally uniform thresholds are unlikely to be optimal for all basins, given the regional differences in TC characteristics and environmental conditions. We therefore note that basin-specific tuning is a reasonable and important direction for future development. To respond to your suggestion, we conducted an initial sensitivity experiment for the NATL basin using regionally optimized thresholds with the UZ tracker. We chose UZ for this purpose because its computational efficiency makes it more suitable for ensemble-member testing under the current time constraints. The results show that optimized thresholds improve detection skill (e.g., POD increases from 0.46 to 0.57; Fig. R2), confirming the value of basin-specific calibration. However, we also observe a "ceiling effect": even with optimal thresholds, POD plateaus at approximately 0.60. Analysis of missed cases reveals that the remaining undetected systems are predominantly short-lived, weak TCs (Fig. R3) which account for nearly 50% of observed storms (lifespans of 2-5 days) in the NATL basin (Fig. R4). This suggests that the limitation is structural. Models struggle to reproduce such transient systems, and the UZ algorithm's minimum lifespan criterion inherently constrains its ability to capture them.

At the same time, we chose not to incorporate this regional optimization experiment into the main revised manuscript, because the core objective of the present paper is to introduce and evaluate the first version of RGTracks-20C as currently constructed, rather than to fully optimize basin-specific tracking configurations. We therefore regard the regional tuning experiment as a useful exploratory analysis that supports future development, but as being beyond the central scope of the current paper.

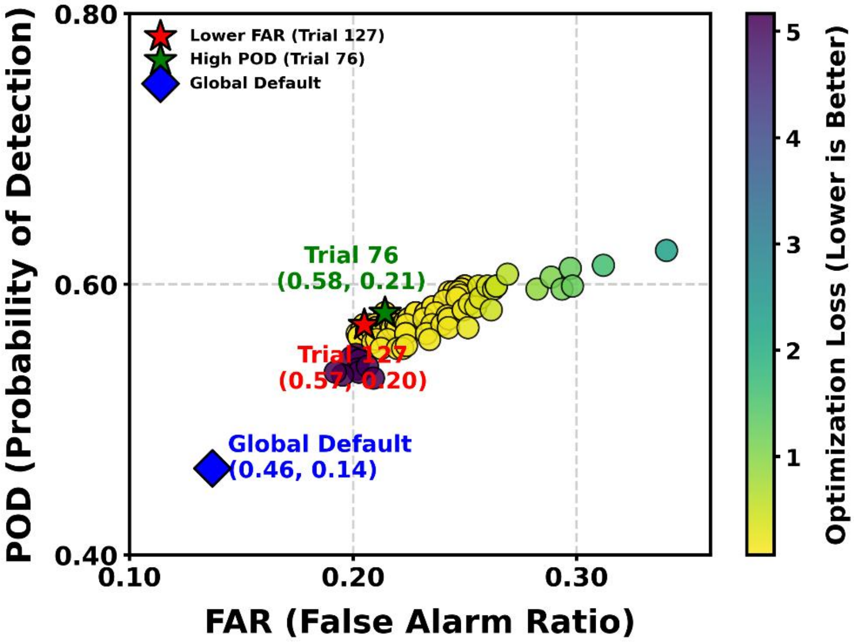

Figure R2: Optimization results of the relevant thresholds for the UZ algorithm based on Bayesian model optimization. Blue diamond represents the global threshold parameters, red stars indicate relatively high POD with relatively low FAR, and green stars represent high POD while ensuring low FAR. Other dots denote post-training results, with a yellowish color indicating a smaller loss function and better performance.

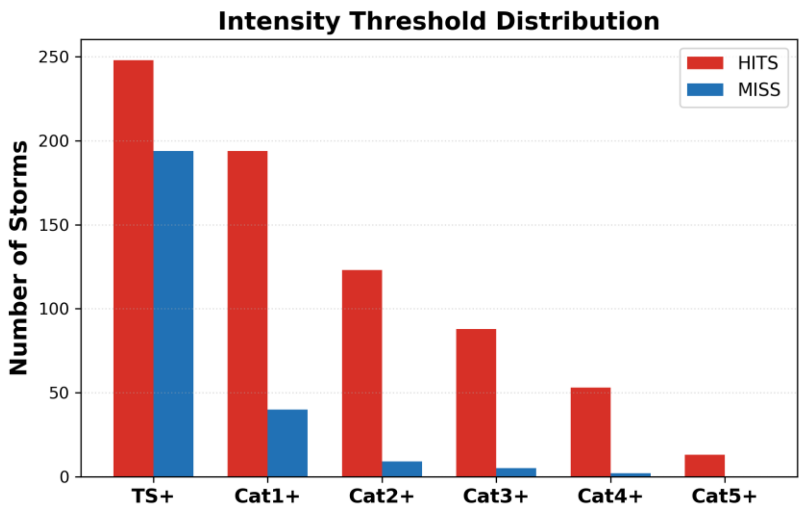

Figure R3: Number of hits and misses for tropical cyclones stratified by intensity category based on Trial 127 (POD = 0.57, FAR = 0.20).

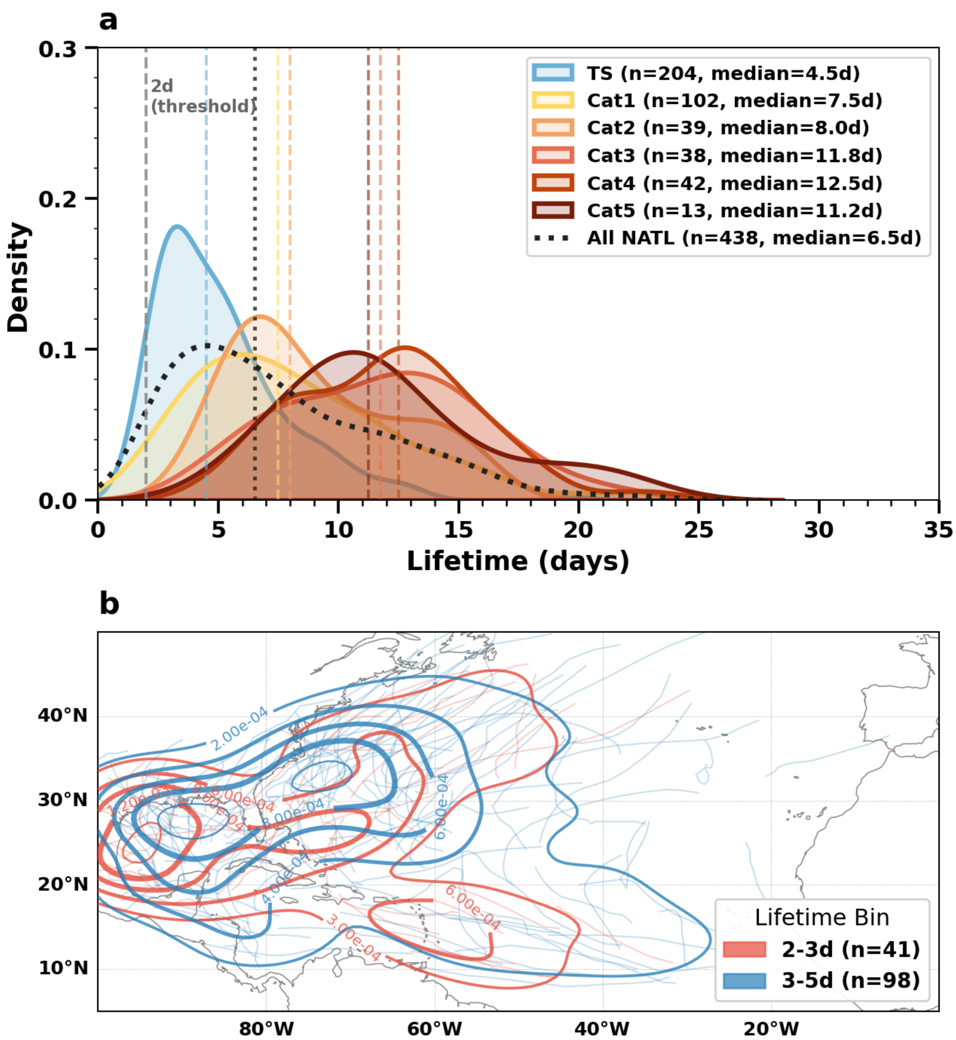

Figure R4: Probability density distribution of tropical cyclone lifespan as a function of intensity category in the North Atlantic (NATL) basin(a). Spatial distribution of TC tracks and the corresponding probability density for short-lived TCs with lifespans of 2–3 days and 3–5 days (b).

Point (4): Validation is heavily focused on the satellite era (1979–2014). Early-period validation relies on limited case studies, which may not be representative.

Authors' response: We thank the reviewer for this important comment.

We selected 1979–2014 as the primary validation period for two reasons: first, observational records during this period are relatively accurate and internally consistent, providing a reliable reference baseline for validation; second, RGTracks-20C (constructed using the OWZ algorithm) is to provide a reconstructed historical TC track dataset to augment the sparse record of pre-satellite observations, rather than to evaluate long-term climatic trends.

Regarding the concern about the representativeness of early-period "validation" based on case studies, we acknowledge that selected cases cannot serve as a rigorous systematic validation. These selected cases serve as illustrative examples to demonstrate the dataset's capacity to provide supplementary track and intensity information for earlier periods when the historical record is sparse or incomplete. As noted in the literature and historical archives, IBTrACS often lacks intensity or track information for many impactful early-period TCs. RGTracks-20C provides relatively complete records for these events, which aligns with the core goal of this version of the dataset, offering new insights into historical periods where observations are absent.

We recognize the importance of systematic early-period validation; however, this remains challenging due to the scarcity and inconsistency of pre-satellite observations. To address your concerns, we have rewritten the relevant sections to ensure more rigorous terminology. Specifically, we have added a new explanation in Sect. 3.6 regarding the selection criteria for these cases, focusing on their historical significance (e.g., significant casualties or economic impact) and their role in demonstrating the dataset's complementary value.

References

Compo, G., Slivinski, L., Whitaker, J., Sardeshmukh, P., McColl, C., Brohan, P., Allan, R., Yin, X., Vose, R., Spencer, L., and others: The international surface pressure databank version 4, 2019.

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

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.

Slivinski, L. C., Compo, G. P., Sardeshmukh, P. D., Whitaker, J. S., McColl, C., Allan, R. J., Brohan, P., Yin, X., Smith, C. A., Spencer, L. J., Vose, R. S., Rohrer, M., Conroy, R. P., Schuster, D. C., Kennedy, J. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Kubota, H., Lee, T. C., Lorrey, A. M., Luterbacher, J., Mock, C. J., Przybylak, R. K., Pudmenzky, C., Slonosky, V. C., Tinz, B., Trewin, B., Wang, X. L., Wilkinson, C., Wood, K., and Wyszyński, P.: An Evaluation of the Performance of the Twentieth Century Reanalysis Version 3, J. Clim., 34, 1417–1438, https://doi.org/10.1175/JCLI-D-20-0505.1, 2021.

Chand, S. S., Walsh, K. J. E., Camargo, S. J., Kossin, J. P., Tory, K. J., Wehner, M. F., Chan, J. C. L., Klotzbach, P. J., Dowdy, A. J., Bell, S. S., Ramsay, H. A., and Murakami, H.: Declining tropical cyclone frequency under global warming, Nat. Clim. Chang., 12, 655–661, https://doi.org/10.1038/s41558-022-01388-4, 2022.

Emanuel, K. Limitations of reanalyses for detecting tropical cyclone trends. Nat. Clim. Chang. 14, 143–

145. https://doi.org/10.1038/s41558-023-01879-y, 2024.

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.