



Assessing the representation of tropical cyclones in ERA5 with the CNRM tracker

William Dulac^{1,2} · Julien Cattiaux¹ · Fabrice Chauvin¹ · Stella Bourdin³ · Sébastien Fromang³

Received: 27 January 2023 / Accepted: 20 July 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

The ERA5 dataset from the European Center for Medium-Range Weather Forecasts is the first global reanalysis to reach a horizontal resolution of 31 km and thus provides a unique opportunity to look at tropical cyclones (TC), and in particular at the 3D fields associated with observed TCs. To that end, a specifically calibrated TC tracking scheme is applied on ERA5 along with a track pairing algorithm to match the detected tracks with the IBTrACS catalog in order to investigate how well TCs are represented in the reanalysis. After tuning of the tracking scheme and the application of a dynamic mid-latitude system filtering technique, it is shown that the majority of IBTrACS TCs are detected in ERA5 and that the amount of false alarms is kept reasonably low in most regions. By comparing detected tracks with their IBTrACS counterparts, it is found that TC intensity is still strongly underestimated in ERA5 but that the minimum sea-level pressure distribution is better represented than maximum wind speed. The comparison between the life cycles from both datasets highlights key differences between ERA5 and the best-track catalog, showing in particular that the delay with which TCs from ERA5 reach their peak intensity compared to IBTrACS increases significantly with real TC intensity increase. Finally, the internal structure of TCs in the reanalysis for each intensity class are analyzed and reveal distinct intensification patterns up to Category 3.

Keywords Reanalysis · TC tracking · Internal structure · Life cycle · Composites

1 Introduction

Atmospheric reanalyses have been developed continuously since Kalnay et al. (1996) by a multitude of meteorological institutes throughout the world. While they have benefited from numerous technical improvements since then, the operating principle has remained unchanged: by combining past short-range weather forecasts with observations through an unchanging data assimilation scheme over the period being analyzed, reanalyses offer a physically and dynamically consistent global estimate of the state of the atmosphere at each time step. Such a product is of great value for the study of

the atmospheric and climate system as it allows for direct comparison with the outputs from climate models. In fact, reanalyses are often used as reference for model evaluation, in conjunction with direct observations (Cesana et al. 2015; Voldoire et al. 2019; Fasullo 2020). Because of the good spatial and temporal homogeneity they offer, reanalysis products are particularly useful for the study of tropical cyclones (TCs) at climate scales and can be used to circumvent the shortcomings of best-track catalogs (Schreck et al. 2014) caused in particular by differences in operational methodology among reporting agencies. But the major strength of atmospheric reanalysis for the study of TCs is that they offer the only mean of analyzing the internal 3D structure of present day TCs as well as the large scale environment around them. This is of utmost importance for the study of TCs in climate simulations. Reanalysis are therefore used to bridge the gap between observations and simulations for validating simulated TCs (Murakami and Sugi 2010; Bell et al. 2013; Rathmann et al. 2014). For that reason, the question of TC representation in reanalysis datasets is of great importance.

Yet, TCs in reanalysis suffer from coarse resolutions (Walsh et al. 2007) as well as from a lack of in-situ

✉ William Dulac
william.dulac@meteo.fr

¹ Centre National de Recherches Météorologiques, CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

² Université Toulouse III - Paul Sabatier, Toulouse, France

³ Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ-Université Paris-Saclay, Gif-sur-Yvette, France

observations in certain areas. Consequently, some reanalyses have implemented TC-specific assimilation techniques to improve TC representation. Both the version 2 of the Modern-Era Retrospective Analysis for Research and Applications (MERRA-2, Gelaro et al. 2017) and the Climate Forecast System Reanalysis (CFSR, Saha et al. 2014) make use of a vortex relocation technique which consists in moving the position of the vortex closer to its best-track location or artificially inserting a vortex into the model's first-guess if absent or too weak. Reanalyses produced by the Japanese Meteorological Agency—namely the Japanese 25-year and 55-year Reanalyses (JRA-25, Onogi et al. 2007; Ebata et al. 2011, JRA-55)—retrieve wind profile data from best-track catalogs and assimilate them as dropwindsonde data (Hatsushika et al. 2006). Both these techniques have been shown to be beneficial to the placement of TC in reanalyses and for TC intensity (Schenkel and Hart 2012; Murakami 2014; Hodges et al. 2017).

ERA5 (Hersbach et al. 2020) is the fifth and latest atmospheric reanalysis from the European Center for Medium-Range Weather Forecasts (ECMWF), succeeding to ERA-Interim (Dee et al. 2011) and ERA-40 (Uppala et al. 2005). Currently extending from 1959 to the present day (still ongoing, and is planned to extend back to 1940), ERA5 uses a wide range of observations from satellite data to ground stations, instrumented buoys and reconnaissance aircraft data through the 4DVAR data assimilation scheme and produces hourly fields with a 0.28125° horizontal resolution and 137 vertical levels up to 0.01 hPa. The 4DVAR data assimilation scheme has been shown to perform better than 3DVAR in areas where observations are sparse (Whitaker et al. 2009) and to have better predictive skill for TCs (Tiwari and Kumar 2022). Lastly, The high resolution in ERA5 is also believed to improve TC representation compared to ERA-Interim (Hersbach et al. 2020). Additionally, the ERA5 back extension (BE)—covering the 1959–1978 period—assimilates IBTrACS 6-hourly pressure reports to improve TC intensity (Bell et al. 2021). However, in a preliminary version of ERA5 BE, this unexpectedly resulted in some systems being too intense, firstly because too much weight were inadvertently given to these observations during the production of ERA5 BE, and secondly because the assimilation system would sometimes move the pressure minimum away from the observation, resulting in a minimum deeper than the reported one. The issue has been addressed in the final revision of ERA5 BE. IBTrACS pressure reports were nevertheless not assimilated for the main product that starts from 1979 onwards, but the emergence of new sources of observations at that time (in-situ and satellites) is believed to compensate for this.

In this paper, we apply the CNRM TC Tracking Scheme (Chauvin et al. 2006) to ERA5 to study TC representation in the reanalysis. First, we conduct a sensitivity analysis of our

tracking scheme to its various detection thresholds by pairing detected tracks from the ERA5 to the IBTrACS historical best-track catalog in order to maximize the probability of detection while limiting the false alarm rate. With our tuned tracking scheme, we then focus on TCs that were successfully matched with IBTrACS and look at several of their characteristics including the wind-pressure relationship, comparative life cycles as well as composited internal structure. Section 2 describes the processing done on both ERA5 and IBTrACS, the TC tracking scheme as well as the track pairing algorithm. Section 3.1 presents the performances of the tuned tracking scheme over ERA5 while Sects. 3.2, 3.3 and 3.4 present different aspects of the TC representation in the reanalysis. The tracking scheme's sensitivity analysis is presented in Appendix A. Lastly, Sect. 4 discusses the results and provides concluding thoughts.

2 Data and methods

2.1 Datasets

We use the ERA5 reanalysis from ECMWF with its native horizontal grid resolution of 0.28125° and interpolated to 0.25° . The CNRM tracking algorithm (Chauvin et al. 2006) is applied on ERA5 6-hourly fields to produce TC track data containing longitude/latitude coordinates, sea-level pressure, vorticity and maximum wind speed. NOAA's International Best Track Archive for Climate Stewardship [IBTrACS, Knapp et al. (2010)] was then used as the reference dataset. IBTrACS merges storm track data from agencies all around the world and as such it requires careful processing, particularly for wind and sea-level pressure data. Here, we use the same IBTrACS version (v04) and post-processing procedure for IBTrACS data as in Bourdin et al. (2022). In particular:

- We retrieve surface wind and sea-level pressure data by prioritizing, for each basin, data provided by the WMO agency responsible for that location. For WMO values coming from a U.S agency, we convert them to 10 min winds using a 1.12 correction factor. If the responsible WMO agency is lacking wind data, we attempt the following:
 - We average the data available among the following centers: Tokyo, Réunion, BoM, Nadi and Wellington;
 - If still missing, we use the wind data from USA agencies using a 1.12 correction factor to convert them from 1 to 10 mins winds;
 - If still missing, we use values from the Chinese Meteorological Administration (CMA) values using

- a 1.08 correction factor to convert values to 10 mins winds;
- The remaining missing values are classified as such.
- We remove tracks referenced as 'spurs' in the metadata;
- We only keep 6-hourly data (rather than 3-hourly) and remove tracks shorter than 24 h;
- We only retain cyclonic seasons from 1981 to 2019 in both the northern and southern hemispheres (39 seasons).

Recovering missing SLP values is done similarly as for wind data, except no conversion is needed and therefore we average the available data from Tokyo, Réunion, BoM, Nadi, Wellington, USA and CMA.

2.2 Tracking scheme

We use the CNRM Tropical Cyclone Tracking Scheme (Chauvin et al. 2006) to detect and track tropical cyclones in the ERA5 reanalysis. The tracking is done separately for each hemisphere up to 60° latitude and operates in three steps which are described below.

2.2.1 Detection

Detection of candidate points is first based on a relative 850 hPa vorticity threshold (*VOR*). A local sea-level pressure minimum is then sought within a $10^\circ \times 10^\circ$ box around the candidate which is then considered to be the system's center.

We then compute the characteristic size of the system based on either the sea-level pressure gradient maxima or the radius of maximal tangential wind—whichever one is the smallest—which we call the *Reference Disk Radius* (RDR). This allows us to separate the presumed cyclone from its environment, which is defined as a disk with a radius of three times the RDR, delimiting a surface that is nine times larger than the TC itself. Separating the presumed TC from its environment allows to check for temperature anomaly which is calculated as the difference between the temperature inside the circle delimited by the RDR and the temperature inside the environment (RDR excluded). More specifically, the detection process tests for the following:

- Temperature anomaly: The sum of temperature anomalies at 700, 500 and 300 hPa is to be greater than a threshold hereafter noted *TANOM*;
- Surface wind speed: the 10 m maximum surface wind speed within the RDR is to be greater than a threshold hereafter noted *RES*;
- Warm upper core: the 300–850 hPa temperature anomaly gradient is to be greater than a threshold hereafter noted

PT. This criterion defines to what extent the core should be warmer in the upper levels than in lower levels;

- Stronger wind in lower levels: the 300–850 hPa mean wind speed gradient within the RDR is to be greater than a threshold hereafter noted *PW*. This criterion defines to what extent the wind speed in upper levels should be weaker than that in lower levels, as a result of the thermal wind balance applied to the warm core.

For the present study, we used $\mathbf{VOR} = 15 \cdot 10^{-5} \text{ s}^{-1}$; $\mathbf{TANOM} = 1 \text{ K}$; $\mathbf{RES} = 5 \text{ m.s}^{-1}$; $\mathbf{PT} = -1 \text{ K}$ and $\mathbf{PW} = 5 \text{ m.s}^{-1}$. These values were selected based on the results of the sensitivity study presented in Appendix A

2.2.2 Linking

Linking points together is the process of joining the individual points resulting from the detection process into tracks. In the CNRM TC Tracking Scheme, this is achieved by an iterative process that selects the candidates that maximize the smoothness of the overall trajectory. Originally based on the methodology from Hodges (1994), our tracking scheme maximizes smoothness by combining three quality parameters that each minimize a distance measurement while also factoring the maximum distance a TC is capable of covering in a single time step.

Knowing the position at time t , we select the next position from the candidates at time $t + 1$ as follows: The first parameter measures the euclidean distance between the position at time t and the candidate at $t + 1$. The second parameter measures the euclidean distance from the forward extrapolated position from the track at $t - 1$ to the candidate at $t + 1$. Finally, the third parameter measures the distance from the backward extrapolated position from the track at $t + 2$ to the position at time t . By iteratively evaluating the total quality index for each candidate and at each time step, we ensure that the resulting track shows the least possible amount of inconsistencies, such as sharp turns and fragmented patterns.

2.2.3 Relaxation

In order to avoid the issue where a single TC track might be split into several parts due to one of the parameter falling below its threshold – thereby leading to an incorrect TC count—the tracking scheme performs what is called the relaxation process.

During relaxation, all detection thresholds are lifted and the tracking scheme completes tracks both forward and backward by searching for a vorticity maxima within a 500 km radius disk centered around the advected position of the system computed from the two previous time steps. This process goes on in both directions until the vorticity maxima falls below a newly defined threshold, referred to

as the relaxation level. Because several detected tracks can – prior to the relaxation process – belong to the same TC, relaxing all detected tracks can result in overlapping duplicate trajectories. For that reason, a final step in which such duplicates are identified and removed is applied. This is done by eliminating relaxed tracks that have more than five steps in common with an already existing track. Throughout this study, the relaxation level is set at $25 \cdot 10^{-5} \text{ s}^{-1}$. While the relaxation level is also a tunable threshold, it does not affect the number of detected TCs like the parameters described in Sect. 2.2 do but rather the length of the detected tracks.

2.2.4 Post-processing

Post-processing techniques can be used to retrospectively filter out extra-tropical cyclones and other mid-latitude lows that might have passed the detection tests. In this study, we use the thermal wind warm core diagnosis developed in Hart (2003) and tested in Bourdin et al. (2022), hereafter referred to as the VTU. This method uses the V_U^T parameter from Hart's phase space—defined as the vertical derivative of the isobaric height perturbation within the 600–300 hPa tropospheric layer—to diagnose the presence of a warm upper core. Using the implementation from Bourdin et al. (2022), the V_U^T parameter is computed as follow:

$$V_U^T = P_{\text{mid}} \frac{\Delta Z|_{600 \text{ hPa}} - \Delta Z|_{300 \text{ hPa}}}{\Delta P}$$

where $\Delta Z|_p$ is the maximum isobaric height perturbation within a 500 km radius at the pressure level p , and $P_{\text{mid}} = (600 + 300)/2 = 450 \text{ hPa}$. The V_U^T parameter is computed for each point of the relaxed tracks. A given track is removed from the dataset if it presents less than six time steps where V_U^T is strictly positive, or if not all time steps verify $V_U^T > 0$ for tracks that last less than 36 h. We stress that the VTU post-processing is not used in the sensitivity study presented in Appendix A.

2.3 Track pairing method

We use a track pairing algorithm to pair detected tracks from ERA5 with those from the IBTrACS observational dataset. This allows us not only to provide detailed analysis of the performances of our tracking scheme over ERA5, but it also allows us to better evaluate TCs representation in the reanalysis as we can focus on tropical cyclones that have been successfully matched with IBTrACS, thereby eliminating possible biases that could come from spurious detections, i.e. systems incorrectly classified as tropical cyclones.

The track pairing algorithm works by searching for tracks in IBTrACS that match both spatially and temporally to the detected systems in the reanalysis. More specifically, the

algorithm starts by assigning each detected track a geographic basin. Appendix B Supplementary Figure B1 presents the basins boundaries, which are adapted from Knutson et al. (2020) so that the separation between the Easter Pacific basin (EPac) and the North Atlantic basin (NAtl) follows the American coast. For each detected track in a given basin, we start by searching for potential matches in IBTrACS by selecting tracks that have at least one time stamp in common (temporal overlapping). Then for each candidate track, we compute the great-circle distances between the detected track's positions and the candidate's at equal time stamps. A detected position is considered to be found in IBTrACS if the great-circle distance between ERA5 and IBTrACS for that time step is less than 300 km. Thus for each IBTrACS candidate we compute the coverage ratio defined as the amount of detected time steps located less than 300 km from the candidate divided by the candidate's length. The candidate with the highest coverage ratio is selected and the detected track is considered successfully paired with a track from IBTrACS.

Note that in these conditions, a single time step satisfying the distance threshold can be enough to consider that the entirety of the ERA5 track is indeed found in IBTrACS. However, as it is highly improbable that two TC systems evolve at the same time at less than 300 km from each other, we assume that if a 'match' is found for one time step, then it should correspond to the same TC. This process produces basin specific data that may then be aggregated with other regions to produce pseudo-global information. We point out that we systematically exclude the North Indian region from our analyses due to reliability issues within IBTrACS in this region, as well as the South Atlantic basin due to the general lack of cyclonic activity.

With the resulting data, we can evaluate the performance of the tracking scheme by computing both the Probability of Detection (POD) and the False Alarm Rate (FAR), defined as follow:

$$\text{POD} = \frac{N_{\text{Hit}}}{N_{\text{IBTrACS}}} \quad \text{FAR} = \frac{N_{\text{ERA5}} - N_{\text{Hit}}}{N_{\text{ERA5}}}$$

Where N_{IBTrACS} is the number of reported tracks in IBTrACS, N_{ERA5} is the number of detected tracks in ERA5 and N_{Hit} is the number of detected tracks that are paired with IBTrACS.

3 Results

3.1 Detection score

In this section we focus on characterizing the performances of our TC tracking scheme applied to ERA5 over the 1981

to 2019 cyclonic season range. The tracking scheme was run with the detection thresholds from Sect. 2.2, which were obtained from the tuning experiment described in Appendix A and was done for each hemisphere separately from 0° to 60°. The VTU method from Sect. 2.2 was also used to remove extra-tropical tracks and hence further reduce the FAR. Figure 1 presents the integrated FAR and POD over the 1981–2019 period while Table 1 summarizes the yearly FAR and POD time series over this period.

The probability of detection shows disparities between regions. The SInd and WPac both outperform all three other basins with respectively 73 % and 78 %. The NATl basin on the other hand has the lowest POD of all five with only 57 %, followed by the EPac at 61 %. The NATl basin is the region showing the highest decrease in POD due to the VTU filter: All other regions show a POD decrease ranging from 1 to 2 points after applying the filter, but the NATl basin however drops from 68 to 57 %, i.e 11 points (accounting for 71 tracks). This could be explained by the fact that the NATl region in IBTrACS has the most septentrional data records of all basins. Indeed, as the latitude increases, the amount of records where IBTrACS reports a Nature of either 'Not Reported' or 'Extra Tropical' increases, and systems in IBTrACS that correspond to the tracks that are removed from ERA5 through the VTU filter are found to evolve in

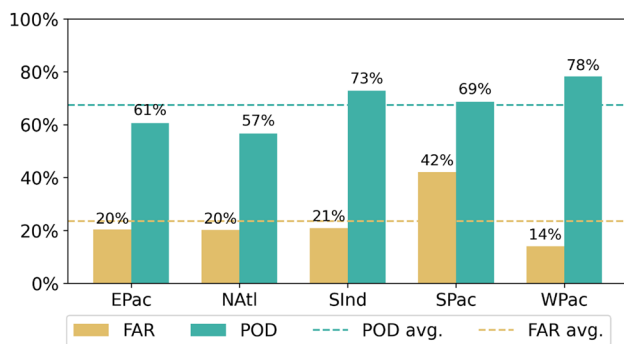


Fig. 1 FAR and POD over the 1981–2019 seasons for all five basins of interest. Dotted lines represent the averages over these regions

higher latitudes than the ones missed the same way in the other regions (Fig. B2 in Supplementary Appendix B). It seems therefore possible that the VTU filter removed tracks that might have lost their tropical nature in IBTrACS and didn't present a strong enough upper tropospheric thermal wind in ERA5 to be kept in the dataset.

In contrast, the FAR is fairly consistent between basins except for the SPac which presents the highest FAR with 42 % while the WPac shows the lowest FAR with 14 %, in addition to having the best POD. Averaging over the five regions, we get a 67 % POD and 24 % FAR. The slight difference in globally averaged FAR we note compared to what Bourdin et al. (2022) found when using the VTU filter on the CNRM Tracking Scheme can probably be explained in large part by the differences in how the VTU post-processing is applied, and also to the fact that we do not include here the North Indian region in the analysis.

However, these integrated values hide large inter-annual variations as shown by Table 1. Most notably, the SPac basin experiences the strongest variations in both the FAR and POD with a standard deviation of 18.7 % and 21.2 % respectively and also has the highest 95th FAR percentile with 66.6 %. Along with the SInd basin, the SPac also presents a 95th POD percentile of 100 %. Conversely, WPac is the most consistent basin with a standard deviation below 10 % for both metrics, the lowest 95th FAR percentile and highest 5th POD percentile.

Figure 2 presents the probability density map for a detected track for both a successfully matched track and one flagged as false alarm, after application of the VTU filter. In the EPac and the SInd basins, the region of activity are the same for both groups of tracks. On the other hand in the WPac, SPac and the NATl basins, false alarms are seen to evolve predominantly at higher latitudes. False alarm tracks are usually shorter lived and marginally weaker than their counterparts but still pass the detection tests, making it difficult to filter them out dynamically during the tracking process. This is especially true given the results from A, as increasing detection thresholds has the side-effect of strongly penalizing the POD. While some amount of tropical

Table 1 Statistical summary of yearly FAR and POD timeseries from 1981 to 2019

Basin	FAR (%)				POD (%)			
	5 %	95 %	mean	std	5 %	95 %	mean	std
EPac	0.0	38.6	20.4	11.3	30.1	89.1	62.2	17.3
NATl	0.0	43.6	21.4	16.2	28.2	79.1	54.7	16.1
SInd	0.0	44.7	18.7	12.8	46.3	100.0	74.5	18.6
SPac	17.5	66.6	43.9	18.7	34.9	100.0	67.8	21.2
WPac	3.1	27.4	14.1	7.6	64.3	93.1	78.5	9.9
Global	15.4	29.1	21.8	4.6	53.5	79.2	69.2	8.5

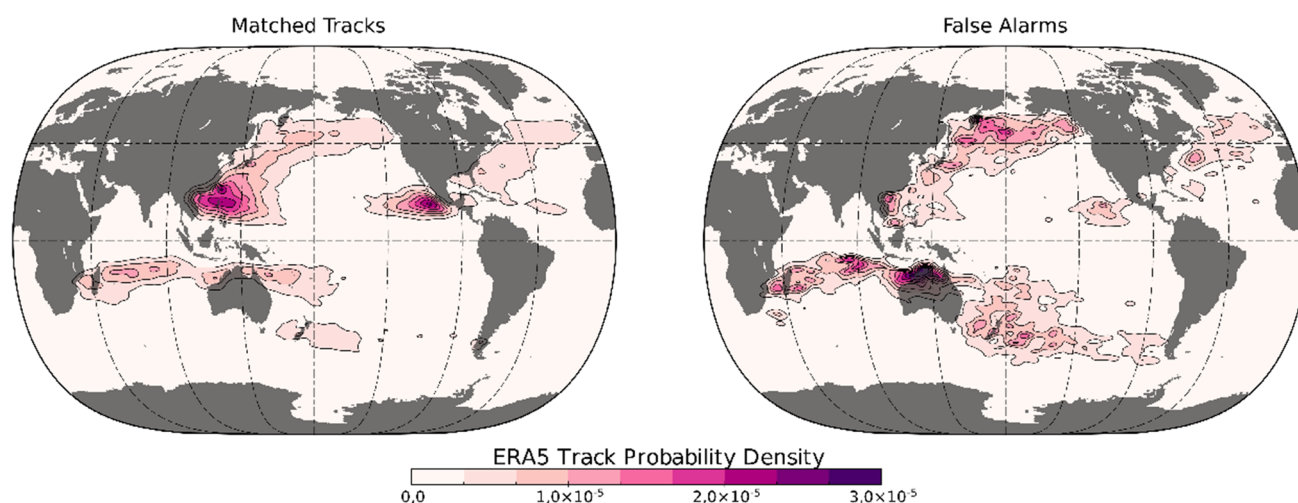


Fig. 2 Probability density map of passage of an ERA5 track for (left) one that is matched to IBTrACS and (right) one that is flagged as false alarm. Computed through kernel density estimation with an ani-

sotropic Gaussian kernel with respectively 3° and 1° zonal and meridional standard deviation over a 0.25° resolution spatial grid

false alarms is likely unavoidable when comparing the output of an objective tracking scheme with a reference dataset that has an inherent part of subjectivity, the false alarm density probability above 45° in Fig. 2 shows that the VTU post-processing fails to remove all mid-latitudes systems. This can be explained by considering two factors: On one hand, the V_U^T parameter is estimated here using only two pressure level within the considered tropospheric layer. It is therefore possible that computing the parameter by fitting the vertical profile of ΔZ to the log of the pressure levels within the layer instead—as suggested by Hart (2003)—would yield a more accurate estimation of the parameter's sign. Secondly, because the VTU method negatively affects the POD in a non-negligible way, the filtering rule needs to express a balance between its efficiency in removing false alarms and the cost in POD. Increasing the required amount of time steps where $V_U^T > 0$ holds ultimately reaches a point of diminishing returns where the loss in POD is too great compared to the improvement in filtering efficiency.

3.2 TC intensity

We focus in this section on how TCs are represented physically and dynamically in ERA5. For this purpose, we filter out detected tracks flagged as false alarms to only keep the ones successfully matched to IBTrACS. Figure 3 represents the wind-pressure relationship (WPR) for the set composed of all temporally and spatially matching points in each track pair, as described in 2.3. Using a $V = a\Delta P^b$ relationship (Atkinson and Holliday 1977) to fit on both groups of points, we get $a = 3.15$ and $b = 0.54$ for ERA5 and $a = 1.4$ and $b = 0.8$ for IBTrACS. Both relationships are therefore vastly different with a faster falling SLP

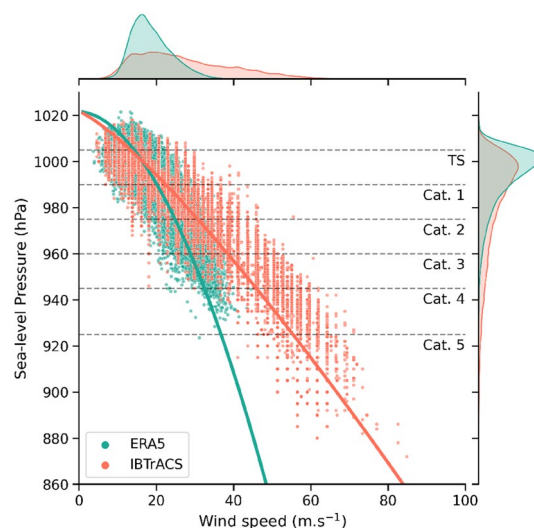


Fig. 3 Pressure-Wind relationship from the subset of matching points in all paired tracks. 58,920 data points in each group. Solid lines show the $V = a\Delta P^b$ fits and the reference SLP at the intercept is set to be greater than the highest SLP record. The marginals represent the density distributions of the data projected on each axis. Horizontal dotted lines represent the pressure-based Saffir-Simpson Hurricane Scale (SSHWS) intensity class thresholds from Klotzbach et al. (2020)

curve in ERA5 compared to IBTrACS. As a result, the wind speeds in ERA5 are considerably lower than in IBTrACS with a 95th percentile of 28.5 m s^{-1} for the reanalysis—which corresponds to a tropical storm on the Saffir-Simpson Hurricane Wind Scale (SSHWS)—compared to 50.5 m s^{-1} for the observational records.

However, the marginal densities on Fig. 3 show that the SLP distribution in ERA5 is better represented than the wind speed and can reach reach category 4 on the revised

SSHPS from Klotzbach et al. (2020) (99.8th percentile) and even reaches category 5 with an absolute minimum of 923 hPa. This difference in how well both variables are represented results in a WPR that is less consistent with IBTrACS and also with ERA-I. Indeed Murakami (2014) compares the WPR diagrams in the northern hemisphere for IBTrACS and six reanalysis products including ERA-I, and finds that the maximum wind speed in ERA5 predecessor extends up to a little more than 30 m s^{-1} (which is only marginally weaker than what we find in ERA5) but also that the minimum SLPs are much less deep, as ERA-I struggles to go below 960 hPa. This comparatively better WPR in ERA-I (despite having weaker TCs) is also found in Hodges et al. (2017).

ERA5	Cat 5	0	0	0	0	0	1	1
	Cat 4	0	0	3	0	6	7	21
	Cat 3	0	1	5	11	30	67	76
	Cat 2	0	3	15	69	89	105	81
	Cat 1	0	16	182	152	131	102	62
	TS	0	269	324	131	119	91	28
	TD	4	43	12	9	2	0	3
		TD	TS	Cat 1	Cat 2	Cat 3	Cat 4	Cat 5
		IBTrACS						

Fig. 4 Confusion matrix between ERA5 and IBTrACS SSHS intensity classes over the five basins of interest. TCs are categorized using the pressure-based SSHS intensity thresholds from Klotzbach et al. (2020)

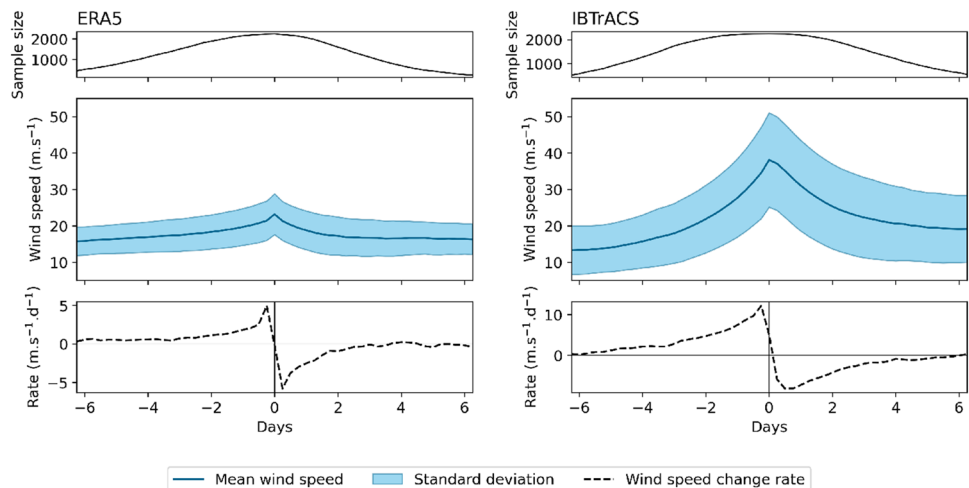
Figure 4 then illustrates the relationship between ERA5 and IBTrACS SSHPS intensity classes by classifying TCs from paired tracks over their entire life cycle. A general agreement between the intensity classes of both datasets is apparent with the majority of entries located below the main diagonal, which is expected as model's horizontal resolutions prevent the accurate representation of TC intensity (Davis 2018). However, the spread in the reanalyzed SSHPS classes—representative of the uncertainty associated to the TC intensity in ERA5—increases with the intensity of the TCs in IBTrACS. For instance, 31 Cat 5 TCs in IBTrACS are reanalyzed into systems rated below Cat 1, amounting to 11.4 % of the total number of Cat 5 IBTrACS TCs represented here.

The opposite effect also occurs but is much rarer. The confusion matrix is indeed non-triangular and shows 3 % of track pairs as being more intense in ERA5 than in IBTrACS. This effect is the most notable for IBTrACS Tropical Storm (TS) category for which this number reaches 6.4 % and with one system in ERA5 that reaches Cat 3. Finally, while the relative agreement between maximum intensities from Fig. 4 holds when considering the global scale, results can vary for different basins as seen on Appendix B Supplementary Fig. B3. Most notably, the confusion matrix for the Natl region is the most diagonal and shows very little spread while EPac on the contrary presents very little consistency with most systems in ERA5 being classified as TS, independently of the observed intensity.

3.3 Life cycle

Figure 5 compares the average lifecycle between matched TCs from ERA5 and IBTrACS, centered around the maximum wind speed at day 0. The sample sizes (top panels) are of different shape around the center due to ERA5 tracks not necessarily having the same length as their IBTrACS

Fig. 5 Mean TC life cycles in (left) ERA5 and right (IBTrACS) from the set of matching tracks. Life cycles are centered around the time of maximum wind speed and integrated over the five basins of interest. Both upper panels show for each time step the size of the sample used to compute the mean and standard deviation. Lower panels present the derivative of the mean wind speed. The X axis goes from -6.25 to $+6.25$ days with 0 being the maximum intensity. The Y axes on the bottom row plots are not aligned



match, thus indicating that detected tracks in the reanalysis tend to be shorter than in IBTrACS. As also shown by Figs. 3, 5 highlights the low wind intensity in ERA5 with a mean maximum wind speed at day 0 of 23.2 m s^{-1} against 38.1 m s^{-1} in IBTrACS. Furthermore, the relative standard deviation—defined as the ratio of the standard deviation over the sample mean at each time step—is in fact smaller for ERA5 as it ranges from 24 to 31 % against 34 to 50 % for IBTrACS, with day 0 having the lowest relative standard deviation for both datasets.

The shapes of both life cycles are distinguished by a few key differences which are better shown in the bottom panels of Fig. 5, representing the change rate of the mean wind speed. The life cycle in ERA5 is almost linear outside of the $[-2; +2]$ day range, with the intensification phase having a slightly steeper slope than during the weakening phase and also exhibits a sharp spike around day 0 with great symmetry. In IBTrACS however, the intensification phase is curvilinear with an almost linearly increasing change rate before day -2 . The peak around day 0 is not symmetric with the left side showing a sharp intensification until the maximum, followed by a slower decrease and the tipping point occurs slightly later than day 0. This dissymmetry comes solely from the method used to identify the time step of the maximum wind speed; here we select the first occurrence in cases where the maximum wind speed is found on consecutive time steps. Using the last occurrence would yield a mirrored effect, while selecting the central occurrence would even it out. But this subtlety nevertheless illustrates one of the main difference between TCs from the reanalysis and those from the observations, in that TCs in ERA5 appear to be unable to plateau at their maximum intensity like they do in IBTrACS.

Figure 6 represents the distributions of the delays between life cycle from each pair of tracks and for each SSHWS class. A positive lag between two life cycles indicates that the reanalyzed TC reached its maximum intensity later

than in IBTrACS. The mean delay across all categories is of 13.3 h, i.e. slightly greater than two time steps (significantly different from 0). Distributions for each class appear to shift towards higher values with TS presenting a 7.7 h average lag. Starting with Cat 2, the mean lag reaches the overall average and extends to 44.7 h for Cat 5. All intensity groups have a mean statistically significantly different from 0 under a 5 % confidence level two-sided student test. This shift towards higher delays in maximum intensities between ERA5 and IBTrACS could be explained by the lack of rapid intensification in the reanalysis. It should be noted that in cases where the maximum wind speed occurs on several consecutive time steps in IBTrACS—and contrary to the method used for Fig. 5—we select the central occurrence (or the one before if the plateau extends over an even number of time steps) so that we measure the lag with respect to the middle of the maximum intensity plateau. This is done in order to improve the interpretability of these results. The choice of the occurrence to use in such cases acts in fact as an offset to all distributions but has virtually no impact on the placement of each box plot relative to each other and to the overall mean.

3.4 Internal structure

Lastly, Figs. 7 and 8 show the composite azimuthally averaged cross-section of detected TCs in the reanalysis for both the radial and tangential wind as well as the temperature anomaly and for each SSHS intensity class. These composites are aggregated based only on the intensity of the observed track (i.e. IBTrACS) for matched tracks, independently of how intense the TC is in the reanalysis at those times. The amount of averaged time steps for each SSHS category are presented in Table 2. With this methodology, these composite figures therefore show us the average ERA5 representation of an observed TC from a given SSHS category.

The radial wind is seen to move towards the center near the surface, from 1000 hPa up to 900 hPa and then flows outwards radially in the upper part of the atmosphere from 200 hPa up to 100 hPa which demonstrates the low level convergence and the upper-level divergence. Radial wind speed varies from -4.3 to 3.9 m s^{-1} for the TS group and increases up to 9 m s^{-1} for Cat 5 in both directions. The tangential wind speed profile tightens around the core as intensity increases. The maximum wind speed is located at the 900 hPa level and 150 km from the center for TS, within the $15\text{--}18 \text{ m s}^{-1}$ range, and is located at 925 hPa and 100 km from Cat 3 to 5, for which it reaches the $27\text{--}30 \text{ m s}^{-1}$ bin with a maximum of 28.9 m s^{-1} . The temperature anomaly from Fig. 8 exhibits the same tightening around the core as for the tangential wind. This tightening can be quantified in terms of mean radial temperature gradient taken at the pressure level of maximum anomaly and computed until the

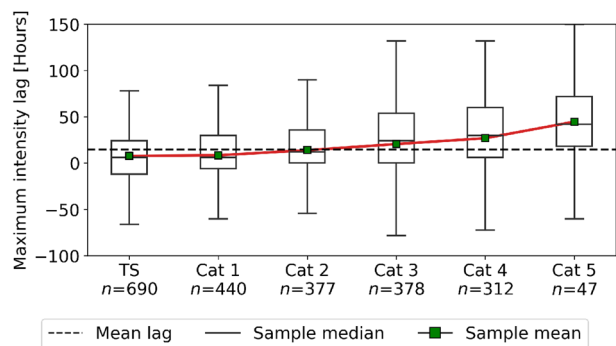


Fig. 6 Distributions of the maximum intensity delays between paired tracks for each SSHS category. Categories are based on IBTrACS maximum wind speed. The delay is relative to IBTrACS such that a positive lag indicates a later intensity peak in ERA5

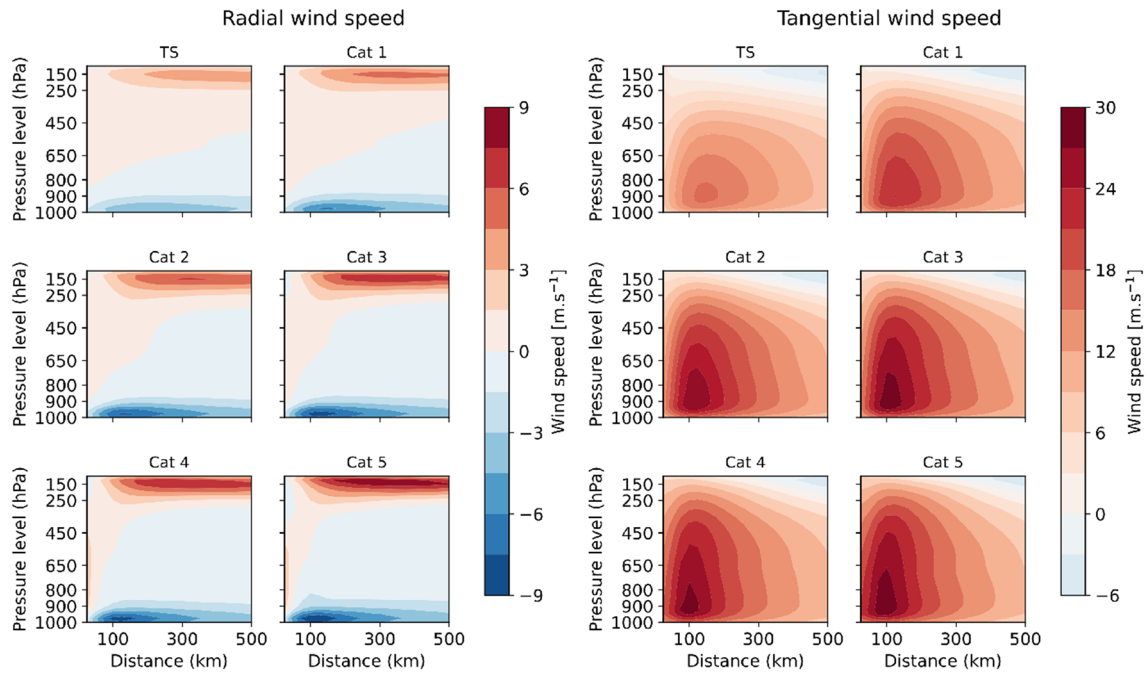


Fig. 7 Radius-height composites cross-sections of azimuthally averaged radial (left) and tangential (right) wind speed of detected TCs in ERA5 sorted by IBTrACS SSHWS categories. Each panel is made from the set of matching time steps where the observed track is at the

given SSHWS intensity. The azimuthal mean is computed within a 41×41 point horizontal box and 27 vertical levels ranging from 1000 to 100 hPa

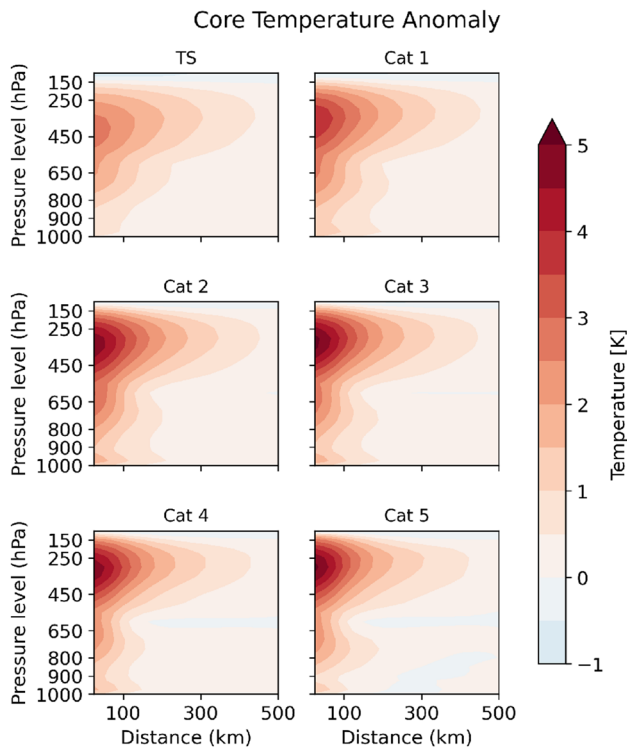


Fig. 8 As in Fig. 7 but for the temperature anomaly. The reference temperature is computed from the four corners of the 41×41 box, i.e. from points 500–700 km away from the center

anomaly decreases by 50 %, thus representing the temperature anomaly decrease rate close to the core. The mean radial temperature gradient decreases with SSHS categories from -0.69 K per 100 km for TS down to -2.31 K per 100 km for Cat 5. Values for the remaining SSHS categories are presented in Table 2. Maximum temperature anomalies range from 2.74 K for the TS group up to 5.16 K for Cat 5. The maximum temperature anomaly increases for each SSHS category except for Cat 4 where it drops from 5.14 K (Cat 3) to 4.87 K (Cat 4). A similar decrease in intensity from Cat 3 to Cat 4 is also noted for the tangential wind speed (Table 2). The fact that this decrease can be seen in both variables is expected because the warm core from a TC is linked to its primary circulation through the gradient and thermal wind balance (Willoughby 1990). It is however interesting to note that the temperature radial gradient does not follow the same pattern as it either increases or remains constant throughout all SSHS categories. Lastly, the height of the maximum temperature anomaly, defined as the level at which maximum temperature anomaly at 25 km from the center is reached, also increases with intensity. The maximum temperature anomaly is located at 400 hPa for TS, 350 hPa for Cat 1 and 2 and 300 hPa for Cat 3 to 5.

The smooth intensity increase with SSHS categories depicted here is however to be nuanced by the fact that these composites are the result of strong averaging—first azimuthally and then over many time steps and different

Table 2 The amount of time steps that are averaged to produce the radius-height composites from Figs. 7 and 8, for each SSHS category. Below are presented general characteristics for the tangential wind speed and core temperature anomaly for all intensity categories

	TS	Cat 1	Cat 2	Cat 3	Cat 4	Cat 5
Count	27,047	10,065	6514	3953	1894	205
Tangential wind speed						
Maximum wind speed (m/s)	16.5	22.6	27	28.7	28.6	29.7
Height of max. wind speed (hPa)	900	900	900	925	925	925
Radius of max. wind speed (km)	150	125	125	100	100	100
Core temperature anomaly						
Maximum anomaly (K)	2.74	3.99	4.89	5.14	4.87	5.16
Height of max. anomaly (hPa)	400	350	350	300	300	300
Radial gradient (K per 100 km)	−0.69	−1.22	−1.73	−1.89	−1.88	−2.31

systems—which tends to smooth out imperfections. Appendix B Supplementary Figure B4 presents the distributions of the maximum tangential wind speed and maximum temperature anomaly for each radius-height cross-section and for each SSHS category and show a large intensity spread in each class, as also pointed out by Fig. 4. For instance, the 5th percentile for the Cat 5 maximum tangential wind speed is 19.6 m s^{-1} , which is halfway between the mean maximum wind speeds from the TS and Cat 1 categories. Conversely, the TS category exhibits a 95 % percentile of 30.8 m s^{-1} , which is stronger than the mean Cat 5 maximum wind speed from Table 2.

Overall, the visual differences in internal structure for all composite variables from Figs. 7 and 8 are the strongest from Cat 1 to 3. This includes both the tightening of the wind vortex with SSHWS category, the tightening and elevation of the warm core and also the strengthening of the radial fluxes from the radial wind composites. Above Cat 3, TCs' internal structures produced by ERA5 become almost indistinguishable, as also highlighted in Table 2. This inability to distinguish TCs above Cat 3 is consistent with ERA5 underestimation of TC intensity in general, as seen from Figs. 3, 4 and 5. Combined with the aforementioned uncertainty in the intensity of reanalyzed TCs, this effect could likely explain the apparent drop in intensity between Cat 3 and Cat 4.

4 Discussion and conclusion

As a preliminary part of our work, we conducted a sensitivity analysis of the CNRM TC Tracking Scheme to its five different detection thresholds with respect to the FAR and POD metrics, which is described in Appendix A. The ultimate purpose of the analysis was to use this information to make an informed choice of threshold values for the complete ERA5 tracking, such that both detection metrics would be optimized to some extent. The analysis provided interesting insights about the tracking scheme with the maximum surface wind and core temperature anomaly

thresholds (*RES* and *TANOM*) having the biggest influence on the tracker's performance. Most strikingly, it showed that lowering *RES* to 10 and 5 m s^{-1} consistently improved FAR and POD compared to *RES*=15. This is irregular since most tracking schemes tend to use wind speed thresholds of at least 15 m s^{-1} (Walsh et al. 2007; Ullrich and Zarzycki 2017, Appendix B). But this is not to say that a 5 m s^{-1} wind speed is considered cyclonic in ERA5. In fact, the *RES* threshold only has a limited impact on the wind speed distribution from the final set of tracks. Indeed, while 10 m s^{-1} can be seen as a more physically acceptable value for the surface wind speed threshold when considering the underestimation of TC intensity in ERA5, it must be noted that the wind speed distributions from the *RES*=5 and *RES*=10 non-paired set of tracks in the Natl basin—all other thresholds held constant—are almost identical (not shown), which is consistent with the observed POD saturation on Fig. 9 at *RES*=10. The mean wind speeds from these two distributions are non-significantly separated by 0.1 m s^{-1} (*p* value = 0.2). The amount of values strictly below 10 m s^{-1} are 5.3 % and 4.9 % for *RES*=5 and *RES*=10, respectively. For *RES*=15, the mean wind speed is only 1 m s^{-1} greater than that of *RES*=10. The reason for the low impact of the *RES* parameter on the measured wind speed distribution is two-fold. First, during the detection process, all detection criteria must be held for at least 4 time steps for the track to be kept, meaning that any track not maintaining a maximum wind speed greater than *RES* for at least 24 h will be discarded. The second reason is that the relaxation process completes tracks forward and backward with no other requirement than a set relative vorticity threshold and is intended to capture the genesis and ending of TCs, where wind speeds are the weakest. This particular aspect results in the lower end of the wind speed distributions to remain about the same, independently of the *RES* parameter. Lowering this threshold therefore mostly affects the number of detected tracks, which is why it acts as a POD upper bound.

The performances of the tracking scheme on ERA5 from 1981 to 2019—presented in Sect. 3.1—show FAR and POD that vary from basin to basin, with a 67 % mean POD and

24 % mean FAR. Bourdin et al. (2022) conducted an inter-comparison analysis of four tracking schemes on ERA5, including our own, for which we provided a set of tracks produced with the same parameters than here and uses a post-processed sub-tropical jet (STJ) diagnostic as cut-off latitude to filter out mid-latitude systems, which is applied over all four tracking schemes. This study places the CNRM TC Tracking Scheme's probability of detection on par with the OWZ (Tory et al. 2013) and UZ (Ullrich et al. 2021) tracking schemes, both in terms of globally averaged POD and inter-regional differences. However, without any form of mid-latitude filtering, both these tracking schemes present considerably less false alarms than the CNRM scheme. Moreover, the STJ filter used on our tracking scheme in Bourdin et al. (2022) offers similar performances to the VTU method in terms of false alarms but has the benefit of preserving the probability of detection, particularly in the NATl basin. However, The VTU filter presents the interesting property of being dependent only on the instantaneous and local state of the atmosphere whereas the STJ filter requires the use of temporal smoothing over large scale fields. This property of the VTU method makes it possible to implement this criterion as an online diagnostic within a tracking algorithm rather than as a post-processing.

In Sects. 3.2 to 3.4 we presented our results on the physical representation of TCs in the ERA5 reanalysis. These results highlighted the underestimation of TC induced wind speeds in ERA5. This underestimation is in fact not specific to ERA5 but concerns all reanalyses, even those (to a lesser extent) with the TC-specific assimilation techniques mentioned in Sect. 1, and cannot solely be explained by the coarse grid resolution of these products (Schenkel and Hart 2012; Hodges et al. 2017; Zarzycki et al. 2021). Moreover, we note that the wind-pressure relationship measured in ERA5 (Fig. 3) appears somewhat degraded compared to those found with two predecessors of ERA5: ERA-40 and ERA-Interim (Murakami 2014; Hodges et al. 2017) due to the discrepancy in how well both the wind speeds and SLPs are represented in ERA5. Therefore—and while the exact origin of the underestimation remains unknown at this point—these elements could hint towards an issue with the model's physics.

We also showed in Sect. 3.3 that TC intensification rate prior to peak intensity was greater in the best-track catalog than in ERA5, which is consistent with the results from Schenkel and Hart (2012) obtained on five reanalyses (including ERA-40 and ERA-Interim). However, we find that while all categories of ERA5 TCs reach their peak intensity later than in IBTrACS when looking at the mean lag, the first quartile for the TS and Cat 1 groups below 0 indicate that a considerable amount of TCs peak earlier than in IBTrACS in these categories. This distinction may be the result of differences in methodologies compared to Schenkel

and Hart (2012), mainly with regard to the identification of the time of maximum intensity in IBTrACS, as discussed in Sect. 3.3, but could also partly result from the increase in horizontal resolution in ERA5 compared to the reanalyses used in Schenkel and Hart (2012). Nevertheless, this result is consistent with how ERA5 appears to represent low intensity TCs better than those from higher categories.

Zick and Matyas (2015) analyzed the internal structure of TCs in both the North American Regional Reanalysis (NARR) and CFSR datasets by computing the mean composited radial-height cross section over the ten most intense TCs (according to the reanalyses) and highlighted key structural differences between both. Interestingly, the temperature anomaly cross-section in ERA5 (Fig. 8) has similar features to that from NARR—namely, the distinct upper-tropospheric peak and the secondary, weaker, mid-level warming— while the ERA5 tangential wind speed cross-section (Fig. 7, right panel) shares more similarity with that from CFSR with a vertical distribution of the maximum wind speed. As stated by Zick and Matyas (2015), this vertical distribution of the tangential wind speed is consistent with observational studies based on dropwindsonde and Doppler radar data (Franklin et al. 2003; Stern and Nolan 2011). Furthermore, the positive correlation between the altitude of the maximum temperature anomaly and TC intensity (Table 2) has also been documented by observational studies (Zhang et al. 2015; Gao et al. 2017). More recently, Wang and Jiang (2019) produced a 13 years climatology of TC warm-core structures based on data from the Atmospheric Infrared Sounder onboard the AQUA satellite, from which they derived typical warm-core heights for each SSHS intensity class. We find that compared to the aforementioned study, the warm core in ERA5 do not rise as high, capping at 300 hPa against 150 hPa for the observations. However, we acknowledge that there is no consensus about the link between warm-core elevation and TC intensity to this day (Stern and Nolan 2012). Moreover, it is important to remind the reader that the selected compositing times here are based on the observed intensity rather than on the reanalysis. Bearing in mind that the delay between the maximum intensity of ERA5 and IBTrACS TCs increases with IBTrACS TC intensity, it is possible that replicating this analysis according to the model's intensity would yield slightly different results. For instance, warm core elevation for Cat 4 and 5 could increase, and the decrease in intensity between Cat 3 and Cat 4 would also likely disappear. At any rate, the inability to distinguish TC intensity from their internal structure above Cat 3 in the averaged composites, combined with the large intensity spread associated with reanalyzed TCs within any given SSHS group adds as many extra sources of uncertainty for users wanting to study the internal structure of historical TCs.

Despite some discrepancies in the representation of TCs in ERA5, the reanalysis can still be a valuable tool for the study of tropical cyclones and is capable of producing substantially stronger TCs than ERA-I. It also offers good spatial and temporal homogeneity whereas the differences in detection procedure and maximum wind speed measurement techniques between the meteorological agencies that contribute to IBTrACS produce substantial heterogeneity within the best track catalog (Schreck et al. 2014; Hodges et al. 2017). This is of course to be nuanced by the fact that the quality of observations assimilated in ERA5 also improves with time (Bengtsson et al. 2004; Rienecker et al. 2012) and that not all regions have the same observational coverage. For instance, the Northern Hemisphere benefits from aircraft reconnaissance as well as from a large network of ground stations in addition to satellite observations, the latter being the primary data source in the Southern Hemisphere.

One could consider using bias correction techniques to alleviate bias in wind speed and minimum SLP values. This is however not trivial as most bias correction methods are rank-preserving (e.g. quantile-quantile) and we showed with Fig. 4 that the reanalysis does not preserve the observed rank of TCs. One other strength of atmospheric reanalyses lies within the fact that they allow not only to study the 3D variables associated within the TC itself (as was showcased in this paper), but also the large scale environment in which TCs occur. In both cases, being able to locate precisely TCs in the reanalysis is essential as TC position between a reanalysis and IBTrACS rarely coincide perfectly, hence the need for an objective TC tracking scheme. ERA5 continues on the path of improvement, following in the footsteps of its predecessors, in part due to its increased horizontal resolution, but TC representation would undoubtedly benefit from the use of TC-specific treatments.

Appendix A: tuning the tracker's detection thresholds

In this section we present a sensitivity analysis of the tracking scheme to its detection thresholds in order to identify a set of tracking parameters (see Sect. 2.2) that maximizes the efficiency of detection while limiting the amount of false alarms. In such cases where two or more objectives are to be simultaneously optimized, no single solution can be derived as there is a trade-off between the objectives. One can however aim to approach a solution that is said to be *Pareto efficient* if one of the objectives cannot be improved further without degrading the other (Zitzler and Thiele 1999; Deb 2011).

To find such a solution, we explore the parameters' space by perturbing the tracking parameters so that they can each take three different values, thus forming a set of 243

Table 3 Detection thresholds values tested for the sensitivity analysis

Criterion	Tested values		
<i>VOR</i> Vorticity threshold ($10^{-5} s^{-1}$)	5	10	15
<i>RES</i> Surface wind speed threshold (ms^{-1})	5	10	15
<i>TANOM</i> Temperature anomaly threshold (K)	1	2	3
<i>PT</i> Temperature vertical gradient (K)	-1	-2	-3
<i>PW</i> Wind speed vertical gradient (ms^{-1})	2	5	10

Combinations of these values form a set of 3^5 vectors. The boxed value in each row denotes the value that was selected for the corresponding parameter to conduct the complete ERA5 tracking

combinations of five thresholds, called vectors or solutions. The values of each parameters are presented in Table 3. For each vector, our tracking scheme is applied on both the North Atlantic (NAtl) region—which is the best observed basin—and the South Indian (SInd) basin—which is under Météo-France forecasting responsibility—from season 2008 to 2018. We justify why the conclusions drawn from two basins can be largely be applied to other regions at the end of Section A. We then proceed to pair the detected tracks with IBTrACS, following the methodology from Sect. 2.3 and compute the FAR and POD integrated over the 11 year period.

Figure 9 presents each combination of thresholds in FAR / POD space for both basins, called the objective space. In the objective space, an ideal solution maximizing the POD while minimizing the FAR would be located in the upper left corner of the plots, i.e. POD=100 % and FAR=0 %. However such a solution does not exist here as FAR and POD appear to be related to each other through a non-linear relationship such that improving the POD generally tends to degrade the FAR, and vice versa. In fact, points located along the leftmost of the scatter plots indicate the best trade-off between POD and FAR based on our sample. Choosing a solution among the ones presented here is therefore a subjective choice expressing a balance between detection efficiency and false alarms rate.

These plots also inform us on how each threshold affects the performance of the tracking scheme in terms of these two metrics. For the *VOR*, most vectors are stacked on top of each other, indicating that the vorticity threshold has little impact on the tracker's performance, or that the sampling on this criterion was not selective enough. However, and while the effect cannot be seen visually here, higher *VOR* value appears to be associated with a lower FAR for vectors with high PODs. Indeed, vectors with a POD greater than 70 % show a 1 % relative decrease in FAR with *VOR* set

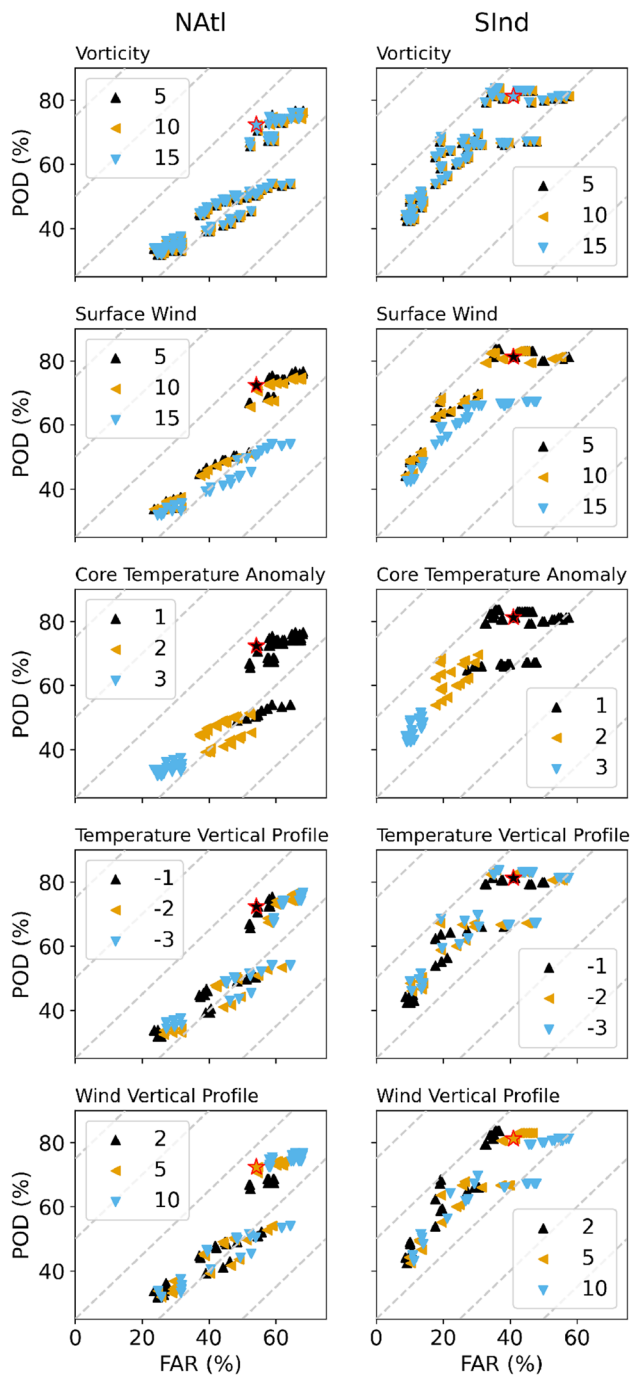


Fig. 9 Objective space representation of each combination of thresholds. The first column shows the results for the NATl basin and the second column those for the SInd basin. Each of the five rows correspond to a detection criterion. Each point (combination of thresholds, a.k.a vector) is colored according to the value taken by said criterion. In each panel, the red-bordered star denotes the selected vector

at $15 \cdot 10^{-5} \text{ s}^{-1}$ compared to $5 \cdot 10^{-5} \text{ s}^{-1}$ in the NATl basin, and a 1.3 % relative decrease for SInd. This tends to show that the vorticity threshold may act as a false alarm filter.

However, increasing the *VOR* too much would inevitably lead to a loss in POD as weaker TCs would fail to meet the criterion. Therefore we considered that the gain from testing with higher values would have been marginal with respect to the cost of conducting new experiments. Moreover, the final set of tracks from the global tracking made with the threshold values from Table 3 show a 0.1 % percentile of the maximum vorticity per matched track of $22 \cdot 10^{-5} \text{ s}^{-1}$ which tends to support this claim.

The surface wind speed threshold (*RES*) on the other hand has a clear impact on performances as it effectively defines an upper bound for the POD. Setting *RES* to 15 m.s^{-1} bounds the POD to 54 % in NATl and 67.3 % in SInd. Boundaries formed by vectors with *RES* values set to 10 and 5 m.s^{-1} are located further up the POD axis and are capped respectively to 74.8 % and 76.7 % for NATl and 83 % and 83.6 % for SInd. The sensitivity to the *RES* parameter shows in fact that the POD saturates at *RES*=10, with *RES*=5 vectors offering only marginally better PODs in the NATl basin, and no apparent change in the SInd basin. Because of the POD gap between *RES*=10 and *RES*=15, it is possible that the saturation value is actually located between 10 and 15 m.s^{-1} .

The temperature anomaly threshold (*TANOM*) stratifies the FAR as each value taken by this parameter corresponds to a certain FAR range with little overlapping. Reducing the threshold leads to an increased FAR, as it allows the tracker to detect cooler systems. However, because of the link between FAR and POD, an increased FAR generally implies a higher POD, making *TANOM*=1 vectors prime choices for finding a good performing solution within our set.

As for the last two detection thresholds that define the strength of vertical profiles for respectively the temperature and horizontal wind speed (*PT* and *PW*), they act by design as filters for extra-tropical cyclones—counting as false alarms in our methodology—by ensuring the presence of respectively a warmer upper core and stronger near-surface winds. Both of these properties are indeed reversed in extra-tropical cyclones, and linked together by the thermal wind relationship. No distinct pattern applicable to both regions and all vectors emerge from the analysis. However, when considering only the group of vectors with the highest POD in each region, increasing these thresholds tends to deteriorate the FAR with marginal benefit to the POD.

As a result, we chose *VOR*=15 for its potential to reduce the amount of false alarms, *RES*=5 for the extended POD upper bound, *TANOM*=1 because of the advantageous location of these vectors in the objective space for the NATl basin and selected a solution from the remaining candidates in our set which we felt constituted a satisfying compromise in both regions—leaving us with *PT*=1 and *PW*=5. The choice of a 5 m.s^{-1} wind speed threshold and its meaning for the CNRM TC Tracking Scheme is

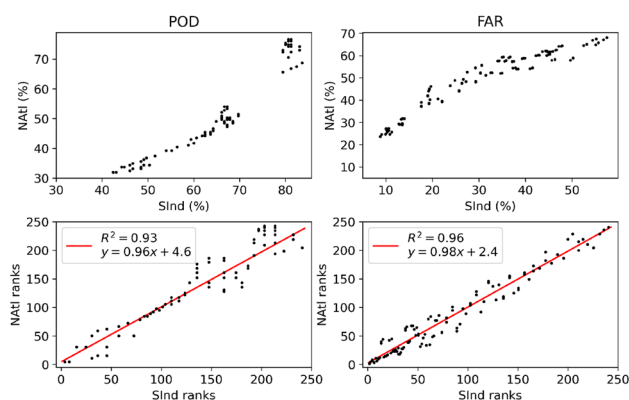


Fig. 10 The left column places each vector POD-wise (upper left) and their respective ranks—from 1 to 243—between both experiments (bottom left). The right column does the same FAR-wise. PODs are ranked in ascending order such that higher POD values have a higher rank and FARs in descending order. The red line on the bottom row panels show the fitted line between NATl ranks and exhibit a 0.97 and 0.94 correlation coefficient for POD and FAR, respectively

further discussed in the Discussion section (Sect. 4). This particular solution exhibits a 54 % FAR and 72.4 % POD in the NATl region, and a 41 % FAR and 81.2 % POD in the SInd basin. The important amount of false alarms in the chosen solution, the majority of which being mid-latitude systems, motivated us to then add the VTU post-processing method described in Sect. 2.2 as a mean of improving performances even further. Applying the VTU method on the selected solution brings down the FAR to respectively 15 % and 27 % for NATl and SInd between 2008 and 2018. This reduction in FAR comes at a cost to the POD in the NATl basin which is lowered to 58 %. However, the POD in the SInd remains unaffected by the use of the additional filter, which is likely due to the fact that there is little TC activity at mid-latitudes in this region.

Finally, it is worth noting that while our test vectors in the SInd region present generally higher PODs in this experiment (14 points more in average) and are arranged differently in the objective space than in the NATl region, the rankings of each vector on FAR and POD scales is approximately preserved between both region, as can be seen in Fig. 10 below.

POD ranks between both basins indeed show a 0.96 correlation coefficient FAR ranks are at 0.98, meaning that any given vector performs about as well in both basins with respect to the other vectors. This gives us confidence in the fact that tuning our algorithm over two basins only can be relevant for all basins.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00382-023-06902-8>.

Author contributions WD carried out the study under the guidance of JC and FC. FC provided the tracking scheme. SB provided the VTU post-processing. WD did the figures and wrote the manuscript. All authors provided critical feedback and contributed to the revisions.

Funding This work was partly supported by the French National program LEFE (Les Enveloppes Fluides et l'Environnement), specifically by the CYPRESSA project.

Data availability The ERA5 data used was obtained from the ESPRI Mesocentre from the IPSL, Polytechnique, but is otherwise available through the Copernicus Climate Change Service (C3S) Climate Data Store (CDS). The IBTrACS catalog is available at <https://www.ncei.noaa.gov/products/international-best-track-archive>. The tracking and track pairing data generated and analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. The authors declare no competing interests.

References

- Atkinson GD, Holliday CR (1977) Tropical cyclone minimum sea level pressure/maximum sustained wind relationship for the Western North Pacific. *Mon Weather Rev* 105(4):421–427. [https://doi.org/10.1175/1520-0493\(1977\)105<0421:TCMSLP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1977)105<0421:TCMSLP>2.0.CO;2)
- Bell R, Strachan J, Vidale PL et al (2013) Response of tropical cyclones to idealized climate change experiments in a global high-resolution coupled general circulation model. *J Clim* 26(20):7966–7980. <https://doi.org/10.1175/JCLI-D-12-00749.1>
- Bell B, Hersbach H, Simmons A et al (2021) The ERA5 global reanalysis: preliminary extension to 1950. *Q J R Meteorol Soc* 147(741):4186–4227. <https://doi.org/10.1002/qj.4174>
- Bengtsson L, Hagemann S, Hodges KI (2004) Can climate trends be calculated from reanalysis data? *J Geophys Res Atmos*. <https://doi.org/10.1029/2004JD004536>
- Bourdin S, Fromang S, Dulac W et al (2022) Intercomparison of four algorithms for detecting tropical cyclones using ERA5. *Geosci Model Dev* 15(17):6759–6786. <https://doi.org/10.5194/gmd-15-6759-2022>
- Cesana G, Waliser DE, Jiang X et al (2015) Multimodel evaluation of cloud phase transition using satellite and reanalysis data. *J Geophys Res Atmos* 120(15):7871–7892. <https://doi.org/10.1002/2014JD022932>
- Chauvin F, Royer JF, Déqué M (2006) Response of hurricane-type vortices to global warming as simulated by ARPEGE-Climat at high resolution. *Clim Dyn* 27(4):377–399. <https://doi.org/10.1007/s00382-006-0135-7>
- Davis CA (2018) Resolving tropical cyclone intensity in models. *Geophys Res Lett* 45(4):2082–2087. <https://doi.org/10.1002/2017GL069666>
- Deb K (2011) Multi-objective optimisation using evolutionary algorithms: an introduction. In: Wang L, Ng AHC, Deb K (eds) *Multi-objective evolutionary optimisation for product design and manufacturing*. Springer, London, pp 3–34. https://doi.org/10.1007/978-0-85729-652-8_1
- Dee DP, Uppala SM, Simmons AJ et al (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137(656):553–597. <https://doi.org/10.1002/qj.828>

- Ebita A, Kobayashi S, Ota Y et al (2011) The Japanese 55-year reanalysis JRA-55: an interim report. *Sola* 7:149–152. <https://doi.org/10.2151/sola.2011-038>
- Fasullo JT (2020) Evaluating simulated climate patterns from the CMIP archives using satellite and reanalysis datasets using the Climate Model Assessment Tool (CMATv1). *Geosci Model Dev* 13(8):3627–3642. <https://doi.org/10.5194/gmd-13-3627-2020>
- Franklin JL, Black ML, Valde K (2003) GPS dropwindsonde wind profiles in hurricanes and their operational implications. *Weather Forecast* 18(1):32–44. [https://doi.org/10.1175/1520-0434\(2003\)018<0032:GDWPIH>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0032:GDWPIH>2.0.CO;2)
- Gao S, Chen B, Li T et al (2017) AIRS-observed warm core structures of tropical cyclones over the western North Pacific. *Dyn Atmos Oceans* 77:100–106. <https://doi.org/10.1016/j.dynatmoce.2016.12.001>
- Gelaro R, McCarty W, Suárez MJ et al (2017) The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J Clim* 30(14):5419–5454. <https://doi.org/10.1175/JCLI-D-16-0758.1>
- Hart RE (2003) A cyclone phase space derived from thermal wind and thermal asymmetry. *Mon Weather Rev* 131(4):585–616. [https://doi.org/10.1175/1520-0493\(2003\)131<0585:ACPSDF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<0585:ACPSDF>2.0.CO;2)
- Hatsushika H, Tsutsui J, Fiorino M et al (2006) Impact of wind profile retrievals on the analysis of tropical cyclones in the JRA-25 reanalysis. *J Meteorol Soc Jpn Ser II* 84(5):891–905. <https://doi.org/10.2151/jmsj.84.891>
- Hersbach H, Bell B, Berrisford P et al (2020) The ERA5 global reanalysis. *Q J R Meteorol Soc* 146(730):1999–2049. <https://doi.org/10.1002/qj.3803>
- Hodges KI (1994) A general method for tracking analysis and its application to meteorological data. *Mon Weather Rev* 122(11):2573–2586. [https://doi.org/10.1175/1520-0493\(1994\)122<2573:AGMFTA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<2573:AGMFTA>2.0.CO;2)
- Hodges K, Cobb A, Vidale PL (2017) How well are tropical cyclones represented in reanalysis datasets? *J Clim* 30(14):5243–5264. <https://doi.org/10.1175/JCLI-D-16-0557.1>
- Kalnay E, Kanamitsu M, Kistler R et al (1996) The NCEP/NCAR 40-year reanalysis project. *Bull Am Meteor Soc* 77(3):437–472. [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2)
- Klotzbach PJ, Bell MM, Bowen SG et al (2020) Surface pressure a more skillful predictor of normalized hurricane damage than maximum sustained wind. *Bull Am Meteor Soc* 101(6):E830–E846. <https://doi.org/10.1175/BAMS-D-19-0062.1>
- Knapp KR, Kruk MC, Levinson DH et al (2010) The international best track archive for climate stewardship (IBTrACS): unifying tropical cyclone data. *Bull Am Meteor Soc* 91(3):363–376. <https://doi.org/10.1175/2009BAMS2755.1>
- Knutson T, Camargo SJ, Chan JCL et al (2020) Tropical cyclones and climate change assessment: Part II: projected response to anthropogenic warming. *Bull Am Meteor Soc* 101(3):E303–E322. <https://doi.org/10.1175/BAMS-D-18-0194.1>
- Murakami H (2014) Tropical cyclones in reanalysis data sets. *Geophys Res Lett* 41(6):2133–2141. <https://doi.org/10.1002/2014GL059519>
- Murakami H, Sugi M (2010) Effect of model resolution on tropical cyclone climate projections. *Sola* 6:73–76. <https://doi.org/10.2151/sola.2010-019>
- Onogi K, Tsutsui J, Koide H et al (2007) The JRA-25 reanalysis. *J Meteorol Soc Jpn Ser II* 85(3):369–432. <https://doi.org/10.2151/jmsj.85.369>
- Rathmann NM, Yang S, Kaas E (2014) Tropical cyclones in enhanced resolution CMIP5 experiments. *Clim Dyn* 42(3):665–681. <https://doi.org/10.1007/s00382-013-1818-5>
- Rienecker MM, Dee D, Woollen J, et al (2012) Atmospheric Reanalyses-Recent Progress and Prospects for the Future. A Report from a Technical Workshop, April 2010. Tech. Rep. GSFC. TM.6867.2012, NASA, <https://ntrs.nasa.gov/citations/20120014167>, nTRS Author Affiliations: NASA Goddard Space Flight Center, European Centre for Medium-Range Weather Forecasts, I M Systems Group, Colorado Univ., Japan Meteorological Agency, Met Office (Meteorological Office), National Oceanic and Atmospheric Administration NTRS Document ID: 20120014167 NTRS Research Center: Goddard Space Flight Center (GSFC). Accessed 7 Sept 2022
- Saha S, Moorthi S, Wu X et al (2014) The NCEP climate forecast system version 2. *J Clim* 27(6):2185–2208. <https://doi.org/10.1175/JCLI-D-12-00823.1>
- Schenkel BA, Hart RE (2012) An examination of tropical cyclone position, intensity, and intensity life cycle within atmospheric reanalysis datasets. *J Clim* 25(10):3453–3475. <https://doi.org/10.1175/2011JCLI4208.1>
- Schreck CJ, Knapp KR, Kossin JP (2014) The impact of best track discrepancies on global tropical cyclone climatologies using IBTrACS. *Mon Weather Rev* 142(10):3881–3899. <https://doi.org/10.1175/MWR-D-14-00021.1>
- Stern DP, Nolan DS (2011) On the vertical decay rate of the maximum tangential winds in tropical cyclones. *J Atmos Sci* 68(9):2073–2094. <https://doi.org/10.1175/2011JAS3682.1>
- Stern DP, Nolan DS (2012) On the height of the warm core in tropical cyclones. *J Atmos Sci* 69(5):1657–1680. <https://doi.org/10.1175/JAS-D-11-010.1>
- Tiwari G, Kumar P (2022) Predictive skill comparative assessment of WRF 4DVar and 3DVar data assimilation: an Indian Ocean tropical cyclone case study. *Atmos Res* 277(106):288. <https://doi.org/10.1016/j.atmosres.2022.106288>
- Tory KJ, Chand SS, Dare RA et al (2013) The development and assessment of a model-, grid-, and basin-independent tropical cyclone detection scheme. *J Clim* 26(15):5493–5507. <https://doi.org/10.1175/JCLI-D-12-00510.1>
- Ullrich PA, Zarzycki CM (2017) TempestExtremes: a framework for scale-insensitive pointwise feature tracking on unstructured grids. *Geosci Model Dev* 10(3):1069–1090. <https://doi.org/10.5194/gmd-10-1069-2017>
- Ullrich PA, Zarzycki CM, McClenny EE et al (2021) TempestExtremes v2.1: a community framework for feature detection, tracking, and analysis in large datasets. *Geosci Model Dev* 14(8):5023–5048. <https://doi.org/10.5194/gmd-14-5023-2021>
- Uppala SM, Kållberg PW, Simmons AJ, et al (2005) The ERA-40 reanalysis. *Q J R Meteorol Soc* 131(612):2961–3012. <https://doi.org/10.1256/qj.04.176>
- Voldoire A, Saint-Martin D, Sénési S et al (2019) Evaluation of CMIP6 DECK experiments With CNRM-CM6-1. *J Adv Model Earth Syst* 11(7):2177–2213. <https://doi.org/10.1029/2019MS001683>
- Walsh KJE, Fiorino M, Landsea CW et al (2007) Objectively determined resolution-dependent threshold criteria for the detection of tropical cyclones in climate models and reanalyses. *J Clim* 20(10):2307–2314. <https://doi.org/10.1175/JCLI4074.1>
- Wang X, Jiang H (2019) A 13-year global climatology of tropical cyclone warm-core structures from AIRS data. *Mon Weather Rev* 147(3):773–790. <https://doi.org/10.1175/MWR-D-18-0276.1>
- Whitaker JS, Compo GP, Thépaut JN (2009) A comparison of variational and ensemble-based data assimilation systems for reanalysis of sparse observations. *Mon Weather Rev* 137(6):1991–1999. <https://doi.org/10.1175/2008MWR2781.1>
- Willoughby HE (1990) Gradient balance in tropical cyclones. *J Atmos Sci* 47(2):265–274. [https://doi.org/10.1175/1520-0469\(1990\)047<0265:GBITC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1990)047<0265:GBITC>2.0.CO;2)
- Zarzycki CM, Ullrich PA, Reed KA (2021) Metrics for evaluating tropical cyclones in climate data. *J Appl Meteorol Climatol* 60(5):643–660. <https://doi.org/10.1175/JAMC-D-20-0149.1>

- Zhang JA, Nolan DS, Rogers RF et al (2015) Evaluating the impact of improvements in the boundary layer parameterization on hurricane intensity and structure forecasts in HWRF. *Mon Weather Rev* 143(8):3136–3155. <https://doi.org/10.1175/MWR-D-14-00339.1>
- Zick SE, Matyas CJ (2015) Tropical cyclones in the North American Regional Reanalysis: an assessment of spatial biases in location, intensity, and structure. *J Geophys Res Atmos* 120(5):1651–1669. <https://doi.org/10.1002/2014JD022417>
- Zitzler E, Thiele L (1999) Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE Trans Evol Comput* 3(4):257–271. <https://doi.org/10.1109/4235.797969>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.