Response to Referees

Composite Sharpening by Vortex Symmetrization and Normalization of Tropical Cyclones

Andrina Caratsch, Sylvaine Ferrachat, and Ulrike Lohmann

May 26, 2026

Dear editor and referees,

We would like to thank the editor for overseeing the processing of our manuscript and the referees for their careful and constructive evaluation of the revised version. In the following, we provide detailed responses to all comments. This response document is structured so that each suggestion, question, and remark from the reviewers is reproduced in lavender boxes, followed by our corresponding reply. Quotations from the revised manuscript appear in *italic font*. Along with the revised manuscript, we have also submitted a version with highlighted tracked changes.

Best regards,
Andrina Caratsch, Sylvaine Ferrachat, and Ulrike Lohmann.

Response to Astrid Kerkweg (CEC1)

Box 0.0

Dear authors,
as GMD requires the permanent archiving of the model code versions used in the publication, please do not only cite the webpage icon-model.org but also cite the DOI which is assigned to the official ICON releases. ICON partnership (DWD; MPI-M; DKRZ; KIT; C2SM) (2024). ICON release 2024.10. World Data Center for Climate (WDCC) at DKRZ. https://doi.org/10.35089/WDCC/IconRelease2024.10
Greetings, Astrid Kerkweg (GMD Executive Editor)

Thanks for indicating this. We adjusted the referencing in the methods section and the code availability section.

Line 95:
For the development and demonstration of the SyNC composites, the North Atlantic TC season 2005 was simulated using the non-hydrostatic numerical weather prediction and climate model ICON (Zängl et al. (2015), ICON partnership (DWD; MPI- M; DKRZ; KIT; C2SM) (2024). ICON release 2024.10. World Data Center for Climate (WDCC) at DKRZ) in limited area mode.

Line 576:
The open-source ICON model code version 2024.10 used for our simulations can be obtained at https://doi.org/10.35089/WDCC/IconRelease2024.10 (ICON partnership (DWD; MPI-M; DKRZ; KIT; C2SM) 2024, World Data Center for Climate (WDCC) at DKRZ) or https://icon-model.org/.

Response to Wojciech W. Grabowski (EC1)

Box 0.1

Dear Authors. before I accept the paper I have the following technical request. Throughout the text, you use the word "resolution" many times. But in some places, it really means "horizontal grid length". For instance, lines 7, 44, 45, etc. It is correct when "resolution" is used in a general sense, like for instance in line 20 and in several other places. However, when you specifically mean grid length (like in line 7: "...5 km horizontal resolutions."), the word "resolution" needs to be replaced by "horizontal grid length". Please do the search for the word "resolution" and replace it with the phrase "horizontal grid length" in all places where a specific number is given (like line 7 in the abstract. Thank you very much for your cooperation. WG.

Thank you for pointing this out. We have revised the wording in the manuscript accordingly. Adjustments were made in line 7, 18, 44, 46, 48, 49, 100, 156, 182, 310, 465, 502 and 551.

1 Response to referee #1

1.1 Comments to the Author

Box 1.1

In this paper, the authors introduce a new methodology meant to improve the compositing approaches often used in the study of cyclones' structure. Their suggestion is to normalize the cyclones according to their size before averaging their structure. Their method appears as a relevant improvement of the usual centered-only approach, especially in the era of km-scale modelling when models are expected to better resolve cyclones' structures (which, surprisingly, is not mentioned as a motivation?). Their approach is rigorous and the improvement they bring is clearly demonstrated. The subject of the paper is also very well suited for the journal, so I think the study deserves to be published in GMD. However, the manuscript needs important revisions, as detailed below.

We thank the referee for the constructive comments and address the concerns below. We also thank the referee for highlighting the relevance of the method for kilometer-scale modeling. In response, we have added a sentence to the abstract and introduction section to emphasize this point.

Line 20:
As numerical models continue to improve in resolution and representation of mesoscale features and vortex variability, TC misalignment in composites will likely become an increasing challenge.

Line 77:
As numerical models achieve finer resolutions and more accurately capture mesoscale features and vortex variability, the challenge of TC misalignment in composites will likely become increasingly relevant.

1.2 Major comments

Box 1.2

As detailed in the specific comments below, I think some of the structure of the manuscript should be revised to better serve the main purpose of the paper, i.e. to demonstrate the improvement brought by your new method. The part on statistical comparison between the observed climatology and the simulations (3.2) should, in my opinion, be substantially reduced. Moreover, an outline and more intermediary intros/transition paragraph would help better follow through the manuscript by always knowing what is the purpose of the section the reader is going through within the manuscript.

Thanks for the feedback! To improve readability, we added a brief manuscript outline in the introduction that prepares the reader for the organization of the paper, the content of each section, and its purpose.

Line 85:
This manuscript first presents the methodology (Sect. 2) including the ICON model setup and TC tracking. A detailed technical description of the SyNC processing steps is presented in Sect. 2.6. Section 3 validates the simulated TC dataset against best track data to estimate its realism and representativeness for other TC datasets thereby indicating the broader applicability of SyNC to other datasets. To assess the SyNC approach, two case studies illustrate the effects of vortex symmetrization and normalization on individual vortices (Sect. 4). Section 5 analyzes the structural evolution of simulated and observed TCs exploiting the fine detection of simulated RMW, Rout, their AIs, and eyewall tilt by SyNC. Finally, Sect. 6 compares the SyNC composites with composites conducted by only aligning the TC centers to evaluate the performance of SyNC and identify its benefits and limitations.

In addition, each subsection states its main purpose at the beginning. We further shortened the validation section (in particular Sections 3.1 and 3.2) by reducing the number of figures and amount of text to focus on the most relevant validation results.

Box 1.3

The question of model resolution is not explicitly treated in the manuscript, and I think it should:

- In my opinion, the advent of km-scale models is the main interest of this method: traditionnal 100-km-or-so models could not resolve cyclone structure well enough for the details of the compositing method to matter. But we expect 10-km-or-less models to better resolve the structure of the storms, and this is why having more rigorous methods like yours is important. This opportunity could even be extended to the apparition of more systematic high-resolution observations such as EarthCare. They might provide enough observationnal references that we can apply the same methods to.
- Along with this lacking motivation, there is a confusion/lack of clarity in the paper between the stakes of 25km models and those of km-scale. You could better highlight the different expectations for the two classes of models.
- L. 121, Bourdin et al. 2024 used 200-25km resolution simulations, whereas Judt et al. 2021 used km-scale simulation. Which of those does the statement L. 120-121 applies to?
- L. 264-265: Again, Bourdin et al. 2024 highlighted the challenge of resolving storm intensity for models with resolution 25km or coarser, which is much more coarse than your own data.
- This is also an important distinction because for models with $\approx$ 25 km resolution or coarser, it can be challenging to define the R17 (your R_{out}), since wind intensities are largely underestimated. Can you comment on the applicability of your method to coarser data?

We thank the referee for highlighting this unclear point and agree that additional precision is necessary. In response, we have added a more detailed discussion of horizontal resolutions and their ability to resolve tropical cyclones.

Line 42:
To which extent vortex structures and mesoscale features are represented in numerical models depends strongly on their horizontal resolution: In models with resolutions of 25 - 250 km, mesoscale precipitation features for example are not resolved, leading to radially smoother and more homogeneous precipitation patterns (Sena et al., 2024; Zhang et al., 2021). Using a higher resolution of 20 km and finer allows for a better representation of mesoscale convection, thereby increasing spatial variability in precipitation and bringing simulated precipitation fields closer to observations (Chen et al., 2007). A similar resolution dependence is found for tangential winds: at coarse resolution (100 km), the wind field is relatively smooth, whereas pronounced horizontal variability in wind emerges at resolutions around 10 km or finer (Manganello et al., 2012; Judt et al., 2021).

The statement in lines 120–121 (*"The wind–pressure relationship of TCs is poorly represented in numerical models. . . "*) refers to both climate and storm-resolving models, as biases in the wind–pressure relationship can be present in both model classes. We have further elaborated on this statement by discussing the possible causes of these model biases.

Line 151:
In climate models with horizontal resolutions of 20–200 km, the simulated wind–pressure relationship improves with increasing resolution by better resolving the vortex leading to a more realistic representation of its structure and the development of intense tropical cyclones (Cat 3–5; Bourdin et al. (2024); Manganello et al. (2012); Davis (2018)). Storm-resolving models operating at approximately 2–10 km resolution further improve simulated storm structure and intensity compared to coarser-resolution models, although model-dependent biases remain (Chen et al., 2007; Judt et al., 2021; Baker et al., 2024).

Moreover, we adjusted the discussion of the too high frequency of intense storms in our model by comparing our results with an ICON configuration at comparable resolution. Based on this comparison, the statement in lines 264–265 was removed.

Line 307: *This imbalance between weak and intense storm frequency could partially stem from the model setup. This imbalance between weak and intense storm frequency and the overall low number of simulated TCs could partially stem from the model setup, as coupled ocean–atmosphere ICON simulations run globally at a convection-permitting resolution of 5* km *reproduce the observed annual TC counts more accurately (Baker et al., 2024).*

We estimate that the application of SyNC is only beneficial for datasets with sufficiently fine resolution to resolve mesoscale features and vortex variability. For coarse-resolution data (approximately 100 km), vortex structures are relatively smooth and therefore the application of SyNC is not necessary. To clarify the range of applicability, we have added the following statements, which are motivated by the above-cited discussion on tropical cyclones and model resolution (line 42).

Line 18:
SyNC is particularly beneficial for analyzing mesoscale features using high-resolution data (10 - 20 km or finer) capable of resolving theses features as well as the structural variability of TCs.

Line 498:
Based on the performance of SyNC evaluated in Sect. 6, we identify potential for its application to both observational and simulated TC datasets that resolve mesoscale features and TC structural variability, based on previous studies likely requiring resolutions of 10 - 20 km or finer (Sena et al., 2024; Zhang et al., 2021; Manganello et al., 2012; Judt et al., 2021; Chen et al., 2007)

Box 1.4

In the introduction, you mention extra-tropical cyclones (ETCs), but you only apply your method to TCs. Is your method relevant for ETCs as well?

We thank the referee for this comment and have removed all references to ETCs, as the method applies only to axisymmetric cyclones.

Box 1.5

Might be out-of-scope, but I would be curious to see what you method yields for precipitations?

Given that precipitation is a key field, we have added Figure 9 to demonstrate the impact of SyNC on precipitation.

Line 398:
Surface precipitation rates (Fig. 9) are also better preserved by SyNC indicating maximum rates 10 $\mathrm{mm\,h^{-1}}$ *higher than in the centered-only approach.*

1.3 Minor comments

Box 1.6

L. 42: Which "size" are you talking about?

We shifted the size definition (R17) to appear directly after this sentence.

Line 50: *Indeed, observed TCs reveal various vortex sizes and are rarely geometrically axisymmetric vortices. The vortex size is often defined as the radius where the wind field reaches 17* $\mathrm{m\,s^{-1}}$ *(R17, Kepert (2010); Chan and Chan (2012); Judt et al. (2021)).*

Box 1.7

L. 47-65: I assume the role of this paragraph is to introduce the "small-scale features of interest". I find it lacks clarity as it is very dense in information and it is not immediately clear where you are going with it. I suggest reframing it, especially since having a clear idea of what features you are interested in is essential.

We agree that a clear definition of “small-scale features” is helpful. We therefore replaced the term “small-scale features” with “mesoscale features” throughout the manuscript, as this term explicitly refers to spatial scales ranging from a few kilometers to a few hundred kilometers.

Box 1.8

L.75: I suggest providing an outline for your paper at the end of the introduction to let the reader know what to expect.

Thanks for this comment. We have addressed this point in Box 1.2.

Box 1.9

If your simulations start July 1st and you remove the first two weeks, it means you are counting TCs from July 15th only. In 2005, 4 observed tropical storms, including Cindy (Cat. 1) Dennis (Cat. 4) occured before July 15th. In general, many seasons see at least one storm before July 15th. This will impact in particular in figure 2, the number of TCs, the intensity distribution (early-season storms are usually weaker), the month of genesis, as well as potentially the genesis location. Can you justify your choice of timing? I expect this might have been due to other constraints and/or cannot easily be changed. I therefore recommend you crop the HURDAT data to exclude the storms before July 15th for a fair comparison, and acknowledge this limitation.

We agree that this caused an inconsistency. The simulation period was chosen to balance computational and storage costs. Consequently, we adjusted the filtering applied to the HURDAT2 data following the referee's suggestions. This leads to only minor changes in Figure 3, while the overall conclusions of the study remain unchanged.

Line 266:
In addition, only storms that developed between 15 July and 1 December were considered, consistent with the chosen simulation period.

Box 1.10

L.83: I assume you are using the ECMWF IFS HRES analysis? If yes, please precise so, as "ECMWF IFS HRES" could mean the forecast. If not, please explain what you are using.

We thank the referee for this comment and have corrected it.

Line 99:
Initial and six-hourly boundary conditions were taken from ECMWF IFS HRES analysis product.

Box 1.11

Section 2.2: The TC tracking section is not clear enough. I understand you do not want to include an extensive description of the scheme, but even without lengthening it, you could improve clarity. You also mention a sensitivity test, but I do not understand if it was done in Enz et al. or in your own study. And which thresholds did you keep from this sensitivity study? Also, given some of your analysis is sensitive to when the tracker starts/ends the tracks, can you give some details about how you method compares to HURDAT to this respect?

We thank for the feedback. We have reformulated the TC tracker description and addressed the definition of track start and end point (Sect. 2.3).

Line 126:
TCs were identified using the multi-parameter TC tracking algorithm of Enz et al. (2023). The algorithm employs three typical TC characteristics: a surface pressure depression, a cyclonic wind field and a warm core located in the upper troposphere. These characteristics are detected by the algorithm at each time step using four parameters, each evaluated across multiple threshold values (shown in parentheses):

1. *A local sea level pressure minimum within a given distance ($p_{min,dist}$ [50, 100, 150] in km).*
2. *High relative vorticity detected at ≈ 2.5 km altitude (ζ_{min} [1e-6, 1e-5] in s^{-1}).*
3. *A temperature anomaly at 300 hPa (ΔT_{core} [0.5, 0.75, 1.0, 1.25, 1.5] in K) within a given distance (T_{dist} [50, 100, 200, 300, 400] in km).*

Thresholds for the parameters follow the original implementation of Enz et al. (2023) and were selected based on scientific literature and physical feasibility. In total, 150 threshold combinations are possible, whereof weak / strong TCs fulfill a low / high number of threshold combinations, respectively. A system is classified as a tropical depression once at least 10 % of the threshold combinations are met, which defines the start and end points of the detected TC tracks. The TC center is identified by the sea level pressure minimum. Once a TC center is identified at time step t_0*, it is linked to the nearest TC center found at* t_{-1}*, whereby a maximum translation velocity threshold of 25* $\mathrm{m\,s^{-1}}$ *is applied. As shown by Enz et al. (2023), the tracker identifies systems once they reach tropical depression strength. Track termination is primarily associated with the loss of the warm core or excessively high translation speeds in the extratropics (for more detail see Enz et al. (2023)).*

The relationship to the HURDAT2 data is addressed in the validation of the wind–pressure relationship.

Line 284:
The TC tracker likely struggles to detect the weak and very early stages of TCs, when the storm's central pressure is close to the ambient pressure (Enz et al., 2023). As a result, these stages are missed in our analysis.

Box 1.12

Section 2.3: I think it should be re-structured, as it has three somewhat disconnected parts: The filtering of detected tracks, the introduction of HURDAT as a reference, and the time normalization. I would merge the first with the previous section, isolate the second, and maybe keep the third for the paragraph where the corresponding results are discussed.

Thanks for the comment. We split this section into individual subsections to facilitate identification of the applied methods: 2.1 Model setup, 2.2 ensemble generation, 2.3 TC tracker, 2.4 TC filtering, 2.5 time normalization, 2.6 SyNC, 2.7 asymmetry index, 2.8 statistical significance and 2.9 best track dataset.

Box 1.13

Section 2.4: Why not use a polar grid for the initial projection, which would make the subse- quent computations easier, and why re-project the data onto a cartesian grid at the very end, since the composites are azimuthal averages?

We initially considered using a polar grid but ultimately decided to adopt a Cartesian grid. The grid is intended to be suitable not only for analyzing azimuthally averaged cross sections but also for examining plan views (see newly added Figure 9). In plan-view representations, the use of a polar grid entails several drawbacks:
1) Area distortion: grid cells increase in size with radius, causing outer regions to visually dominate and potentially bias the interpretation of magnitudes. For tropical cyclones, where the inner core plays a key role in storm development, this is not desirable.
2) Distance interpretation: The straight coordinate system of a Cartesian grid facilitates the interpretation of distances compared to the curved distances inherent to a polar grid.

Box 1.14

L. 159: I do not understand what this sentence means.

The sentence was moved and revised for clarity.

Line 183:
Additionally, the model output is reduced to a 1000 km ×1000 km square around the TC center to speed up the post-processing. This defines the extent of the data and determines Rmax.

Box 1.15

L. 160: Why is it important to get rid of the eyewall tilt? Is it not a structural feature we may want to keep?

We agree that the eyewall tilt is a structurally relevant feature of TCs and that, with a detailed detection of the RMW, it can be readily diagnosed. To illustrate this capability, we have added the evolution of the eyewall tilt to Fig. 7 and included a corresponding discussion of its temporal evolution.

Line 374:
Cat_{pres} 3–5 TCs also exhibit a reduction in eyewall tilt during intensification and significantly smaller eyewall tilt than weaker TCs at maximum intensity (Fig. 7h, i). These results are consistent with previous observational and modeling studies that found smaller eyewall tilts in more intense TCs (Shea and Gray, 1973; Sanabia et al., 2014; Ohno et al., 2016).

However, a consistent alignment of updrafts and eyewall clouds throughout the depth of the atmosphere is only possible when the eyewall tilt is straightened. This alignment enables a coherent analysis of updrafts, upper-level outflow and super-gradient winds, as well as the investigation of cloud microphysical processes from the warm phase to the ice phase, thereby facilitating a more robust interpretation of the origin and vertical distribution of latent heating within the eyewall. The benefit of the eyewall alignment is demonstrated in Fig. 8, 10 and 11. It contributes to both the preservation of mesoscale features and a reduction in within-group variability (see Sect. 6.1 and 6.2).

Line 395:
Temperature tendencies of the cloud microphysical scheme reveal approximately 10 $\mathrm{K\,h^{-1}}$ higher latent heating within the eyewall (Fig. 8e). Additionally, stronger cooling (evaporative cooling of rain and ice melting), is more distinctly positioned on the inner and outer flanks of the eyewall updraft. Cloud water and precipitation (Fig. D1e and h) also show extended maxima within the eyewall. ... This highlights the improved representation of microphysical signals and precipitation achieved by vertically aligning the eyewalls throughout the depth of the TCs, with implications for physical understanding and TC risk assessments.

Line 409:
It is worth mentioning that due to the large variability of eyewall radii in the weak group, secondary peaks in updraft (Fig. D2a) or latent heating (Fig. 11a) may be misinterpreted as features of outer rainbands. However, the SyNC composite clarifies that these are indeed eyewall features.

Line 432:
Second, the eyewall alignment can reduce the data spread within a TC group, thereby improving the statistical power of the statistical test (Krzywinski and Altman, 2014).

Box 1.16

L. 167: If R_{out} is capped to 700 km and composites are cropped beyond R_{out}, why output the data on a 1000 x 1000 km grid in the first place?

Keeping data outside R_{out} enables the quantification of environmental background conditions. For instance, the environmental background flow is calculated in the 600–800 km radial range and is included in Fig. 5 and 6.

Box 1.17

L. 168: What happends when R_{out} is found within 30 km of RMW?

The manuscript has been clarified accordingly.

Line 199:
In case of a narrow R_{out}-RMW ring, Rout is increased until a minimum distance of 30 km between RMW and R_{out} is fulfilled to avoid extensive extrapolation of data points in case of a narrow R_{out}-RMW ring.

Box 1.18

Table 1 could be an opportunity to schematically detail better the different steps of thection 2.4, which is fairly intricate.

This is indeed the purpose of Fig 1. To strengthen its connection to Table 1, we have added references to the corresponding steps in Table 1 for each subpanel in the figure caption.

Box 1.19

Figure 1: If you have 10 detection sectors, why does the figure only show 6 RMW and 7 R_{out}.

Figure 1 is intended as a general illustration of the method. The number of detection sectors is user-selectable and depends on the underlying grid resolution. For visualization purposes, only seven markers are shown for the radius of maximum wind (RMW) and R_{out}.

Box 1.20

L. 223-224: You say it is more meaningful to compare wind and pressure at the same level, yet you test for winds at 8 km (upper troposphere) vs surface pressure. Your approach is confusing.

We thank the referee for this comment. We address this comment below Box 1.22.

Box 1.21

Figure 2: I am surprised that the wind speed values do not decrease more with height, when we know that the strongest TC winds are found near the surface. Can you comment on this?

Tangential wind speeds typically increase above the surface due to the reduced influence of surface friction. In both numerical models and observations, the maximum tangential winds are therefore often found a few hundred meters above the surface (Bui et al., 2009; Ji and Qiao, 2023; Smith and Montgomery, 2023).

Box 1.22

Section 3.1 : Does it make sense to compare absolute values of wind at 10m (in HURDAT) and at different model levels? Computing the MSE score this way means you will favor most the level that have average wind values close to the observed 10m winds, rather than the one that has the wind-pressure relationship shape closer to your observation target, isn't it?
I don't understand the role of your wind-pressure analysis within the manuscript.

We acknowledge the limited physical meaning of a direct comparison when central pressure are not linked with 10-m winds. We therefore have reduced the comparison to near-surface wind levels (Fig. 2) and reformulated this section to clarify its purpose and explicitly state the limitation of the model configuration in representing 10 m winds.

Line 270:
To evaluate the ICON-simulated TCs, the relationship between simulated wind and central pressure is compared to observations in Fig. 2. The simulated wind-pressure relationship displays a bias near the model surface (Fig. 2a-b), where most simulated 10 m winds are weaker than observed winds for a given central pressure. This suggests a model bias that may be related to the surface drag parameterization. As discussed in Sect. 2.4, such behavior is commonly found in numerical models across a range of horizontal resolutions (Bao et al., 2012; Bourdin et al., 2024; Judt et al., 2021; Knutson et al., 2015; Reed et al., 2015; Bao et al., 2012; Manganello et al., 2012). To minimize the influence of surface parameterization, a model level is identified where the wind-pressure relationship aligns well with observations for two reasons: First, it provides a wind-based intensity estimate for simulated TCs that is consistent with simulated pressure-based estimates (see Sect. 3.2), exploiting the higher confidence in the simulated central pressure. Second, it enables a comparison of wind magnitudes and wind field structures between model and observations (see Sect. 3.3) while accepting and excluding model biases near the surface. Indeed, the fit improves at higher altitudes until 0.13 km (Fig. 2c), where the best fits is located with a MSE of only 9 $\mathrm{m^2\,s^{-2}}$*. Above 0.13 km, TC winds appear too strong for a given central pressure. Therefore, the 0.13 km level is selected for further comparisons between observed and simulated wind fields in following sections, as it offers the best alignment with observations while also being rather close to the surface.*

The mean square error (MSE) calculations are deliberately designed such that the simulated central pressure is fixed, under the assumption that the model represents this quantity reliably. A model level is then selected at which the wind magnitudes match observations for a given central pressure. This approach allows us to compare TC wind magnitudes (Fig. 3) and vortex structures (Fig. 4) with observations, based on the model's ability to simulate central pressure.

The vortex structure is affected by surface-related biases. However, this does not necessarily imply that the entire vortex structure is misrepresented in the model. The question we therefore address is how realistically the model simulates vortex structures throughout the atmosphere, since the vortex structure at all vertical levels is subsequently used in SyNC. Consequently, the structural comparison targets a level at which surface effects are minimized, that is, a level where the wind–pressure relationship agrees with observations. Nevertheless, we agree that a fully like-to-like comparison would be preferable. In principle, comparing both near-surface (10 m) and above-surface winds for observed and simulated tropical cyclones would allow a more precise quantification of surface-related biases and their implications on vortex structures. However, observational data at levels above the surface are not available in HURDAT2.

Box 1.23

Section 3.2 constitutes a very tricky statistical exercise: First, you are working with a small amount of data, and second, more importantly, it is not obvious that it makes sense to compare the 1980-2024 climatology to an ensemble representative of the very extreme 2005 season. As shown by my numerous comments on this section, it raises a lot of questions. I would refrain from making a precise statistical comparison of two ensembles that are not equivalent, especially since I don't think you need it within the manuscript (which is meant to demonstrate the improvement brought by your compositing method, not the ability of the ICON model to simulate TCs). Assuming this analysis plays the role of a validation of the TCs simulated by the model, I would recommend reducing a lot the complexity of this section to give only essential keypoints with respect to validating the model's TCs: For example, the low number of simulated TCs is obvious as soon as you mention that it is within the 1-7 TCs/season range, compared to an average 7 in the climatology, and 16 in 2005.

We agree that comparing ensemble members with either a climatology or a single season is challenging. Nevertheless, both provide reference points, to some extent, for evaluating the simulated TCs. We have therefore added a few sentences to clarify the extent to which such comparisons are meaningful and how they should be interpreted.

Line 121:

Since the model is only constrained by the 2005 large-scale environment at the lateral domain boundaries, each member has some degree of freedom to evolve into its own internal state within the domain. Consequently, the simulations represent eight plausible realizations of the 2005 season rather than its exact reproduction. The simulated TC tracks therefore diverge from observed tracks (Fig. A1).

Line 290:

As discussed in Sect. 2.2, the simulated dataset represents plausible realizations of the highly active 2005 hurricane season. Consequently, the simulated activity is not expected to exactly reproduce the observed 2005 season nor to correspond to the climatological mean.

Box 1.24

L. 249-250: I see as many grey as black arrows in fig. 1 for the comparison of the climatology to the simulated ensemble. Why do you say "predominantly insignificant differences"? also, shoudl all metric be considered equals to that respect?

The figure has been updated and the statement has been removed.

Box 1.25

L. 251: You cannot compare the seasonnality of the two ensembles since one starts earlier than the other. I will also impact you counts and intensity statistics (see comment above).

Thanks for this comment. We have addressed this point in Box 1.9.

Box 1.26

L. 255-256: The lower termination latitudes could also be due to the tracker having a propensity to stop detecting the tracks earlier than HURDAT would.

The simulated TC tracks (newly added in Fig. A1) indicate, that the TC tracker is able to track them even after extratropical transition. Consequently, it seems like the limited model domain causes the termination latitude to be relatively low for simulated TCs.

Line 519:

Despite this, there is a good latitudinal agreement between simulated and observed locations of genesis, maximum intensity, and termination evident in Fig. A2a–d, although it is strongly influenced by the model domain.

Box 1.27

Figure 3: The counts between panels (b) and (c) do not match. In panel (c) I count 3 x 2 = 6 major TCs in observed 2005, but in panel (b) you indicate 5. Also, I was confused for a while before I noticed the mention burried in the caption that squares in panel (c) account for more than 1. Maybe use a different shape to warn the reader these are not the same? Plus, you should precise which categories does "major" correspond to (I assumed 3-5).

This issue arises from rounding the number of TCs within each bin to integer values. While this introduces only minor errors for large sample sizes (such as the HURDAT2 climatology), it can indeed be misleading for smaller sample sizes (the 2005 season and the simulated TCs). Consequently, we have removed the scaling for the 2005 season and the simulated TCs. The scaling for the climatology is retained, and an explanatory note has been added to the figure caption in Fig. 3.

Box 1.28

Do you have an idea why the model produces so few TCs?

Unfortunately, despite extensive sensitivity tests involving horizontal and vertical resolutions, turbulence parameterizations and convection representation, we could not identify a model setup that yielded a higher number of TCs.

Line 307:
This imbalance between weak and intense storm frequency and the overall low number of simulated TCs could partially stem from the model setup, as coupled ocean–atmosphere ICON simulations run globally at a convection-permitting resolution of 5 km *reproduce the observed annual TC counts more accurately (Baker et al., 2024).*

Box 1.29

Figure 4: Showing the x-axis grid lines would help with comparing the distributions across the two panels.

We adjusted Fig. 4.

Box 1.30

Figure 5: Why are the dots not equally spaced in azimuth? I thought you divided the space in 10 sectors of equals size.

Within each of the 10 sectors, the grid point closest to the detection criterion is identified. The detected grid cells do not need to be equally distributed in space; however, there is always exactly one detected grid cell per detection sector. The markers shown in Fig. 5 and 6 indicate the exact locations where these grid cells are detected.

Box 1.31

Figures 5 and 6: The colormap/palette could be better chosen to better identify the small scale features and how they change with the transformation. You don't really need the negative side of the colormap, so you could use a non-divergent colormap with more hues so that the details of the circulation are clearer.

Negative values are necessary for Fig. C1, and the colormap has therefore been chosen accordingly. To improve the visualization, we adjusted the colorbar limits of Fig. 5, 6 C1 and C2 to better highlight variability in the tangential wind fields.

Box 1.32

L.307: I expected the AI would be minimum at the lifetime maximum intensity, as I thought the structure would be less symmetric in the early stages when the cyclone is still forming, and the late stages when it is likely being disrupted by vertical wind shear or land interaction. Can you discuss this more? Was this result expected based on the literature? If my naive expectation is wrong, why? And if it's reasonable, why do your result differ?

We have rephrased the section to summarize it more clearly. Generally, the AI of eyewall is closer linked to intensity than AI of size: *"This is in agreement with previous studies (Persing et al., 2013; Martinez et al., 2022), which found that eyewall symmetrization and associated convective organization is beneficial for TC intensification"* (line 373). Regarding size AI, *"Li and Tang (2025) found that weaker TCs typically exhibit larger size asymmetries, often as a consequence of stronger VWS"* (line 363). However, the evolution of size AI is less clear: *"Storm size AIs remain relatively constant throughout the life cycle in both observations and simulations (Fig 7c). This is consistent with the results of Li and Tang (2025), who found that size AI during storm intensification can increase, decrease, or remain constant."* (line 352).

Box 1.33

Figure 7: It is difficult to see the faint grey enveloppes. Could you add thinner lines to indicate the quartile values, which would contains the IQR, to make it easier to read?

Using thin lines made the plot visually overcrowded. Instead, we reduced the fading of the interquartile range to improve its visibility.

Box 1.34

Figure 7: Why do you use less interpolation points in the decay than in the intensification stage, whereas you showed previsouly that both stages have in general the same length?

The focus of this study was placed on the intensification phase.

Box 1.35

L. 332: Can you specify how many TCs there are in each group? Could you also comment on how sensitive your composites are on the number of cyclones being aggregated?

We added the number of storms in the text and figure captions. The sample size of the composite analysis is relatively small for classical statistical testing. However, the statistical method employed to assess significance is based on a resampling approach, in which synthetic distributions are constructed by repeatedly exchanging tropical cyclones between the two groups. This procedure evaluates whether the observed differences can be reproduced by random reassignment. Since statistically significant differences are still detected at the 0.05 level, we consider the main results to be robust despite the limited sample size.

Line 381:
For compositing, simulated TCs were split into weak (cat_{pres} 1-3, $n = 15$) and intense (cat_{pres} 4-5, $n = 18$) TC groups allowing to detect systematic differences in the SyNC performance between differently intense TCs.

Box 1.36

Section 6: It is not clear to me at this stage if you aggregate only snapshots of TCs at their lifetime maximum intensity (most usually done, but smaller sample), or all snapshots at all stages (in which case longer-lived cyclone would be over-represented)?

Thanks for the feedback. This was clarified in the manuscript.
Line 388:
A comparison of the two composite methods of the intense group composed at their lifetime maximum intensity can be found in Fig. 8 and Fig. 9.

Box 1.37

L. 360: If the improvement is greater for weak cyclone, why did you chose to show the intense ones in the main body?

We chose to focus on the intense TC group because it represents a more stringent stress test for the composite method, as these storms tend to be structurally more uniform. In addition, intense tropical cyclones are particularly relevant for the assessment of TC-related impacts and risks and are widely studied, making them well represented in the existing literature.

Box 1.38

The conclusion is very good. The abstract should follows the same lines, and probably not mention the not-robust results with respect to the ability of the ICON model to simulate TCs.

We have adjusted the structure of the abstract accordingly, while retaining some ICON validation for consistency.

2 Response to referee #2

2.1 Comments to the Author

Box 2.1

In this manuscript, the authors develop a new compositing framework for tropical cyclones (TCs), termed SYmmetrized-Normalized Cyclone (SyNC) composites, intended to address limitations of traditional center-aligned composite approaches. The authors argue that standard TC composites tend to smooth storm-scale features because storms of varying size and structure are averaged together without accounting for differences in eyewall location or storm extent. To address this, the authors detect the radius of maximum wind (RMW) and the outer storm radius defined by the 17 m s1 wind threshold in multiple azimuthal sectors and at each vertical level, and use these radii to symmetrize and radially normalize individual storms prior to compositing. The method is demonstrated using convection-permitting ICON simulations of the 2005 North Atlantic hurricane season (a very active one, I might add), with ICON-simulated storm characteristics compared against observations. The authors show that SyNC composites retain 'sharper' representations of eyewall- and vortex-relative features, including vertical motion and diabatic heating, compared to traditional center- based composites. They argue that the approach reduces within-group variance and improves the ability to distinguish between storm categories, while noting reduced applicability during early storm stages when coherent vortex structures are not yet established. Overall, the study frames SyNC as a tool for analyzing vortex-relative circulation and microphysical structure in mature TCs, particularly in high-resolution model output. There are a mix of strengths and weaknesses in the manuscript, in my eyes. Some aspects I find particularly appealing are the standardization that removes latitude dependence and the ability for a user to apply different levels of normalization (with SyNC representing an "extreme" normalization, and a simple storm-centered composite being more "moderate," etc.). I also think the ability to retain an asymmetry index/factor is a unique aspect of the framework and provides useful insight into the symmetry of the storm. This could provide insight during a storm's lifetime or across storms with similar characteristics. However, some elements of the normalization strike me as overly aggressive, and I am not fully convinced that the resulting increase in composite sharpness necessarily translates to improved physical insight. In particular, the treatment of vertical structure raises questions about whether important, well-known aspects of TC dynamics (e.g., central vortex tilt) are being suppressed. Given this, I think the paper could be suitable for publication in GMD eventually, but with some required revisions to address the comments below. More details follow, but broadly I believe it would be helpful for the authors to very clearly define the tradeoffs inherent in the SyNC approach, including which physical characteristics of tropical cyclones are preserved versus removed by the normalization procedure. There is also room to more clearly spell out the rationale for specific design choices and to more explicitly link parts of the work to existing literature.

We thank the referee for the careful reading of the manuscript and appreciate the referee's constructive comments. The associated concerns are addressed below.

2.2 Major comments

Box 2.2

The physical tilt that occurs in the eyewall is a known behavior associated with angular momentum conservation (see Stern and Nolan [2009] and references therein). I am not 100% convinced that removing this tilt and vertically aligning the eyewall (Fig. 8) is useful in that context. I am not necessarily saying there is no merit to this approach, but I think the authors need to do a better job justifying the consistent normalization with height when the storm is not expected to be vertically aligned, even when the vortex center at each layer is first computed and then stacked. I suspect this step is fundamentally where much of the composite "sharpening" comes from, but it has not been made clear to me that this sharpening reflects physically useful information.

We have discussed the rationale for the reduction of the eyewall tilt, and the resulting eyewall alignment, in Box 1.15.

Box 2.3

Some decision choices need to be more concretely defined. For example, the 8:1 ratio for R17:RMW appears to be drawn from Chan and Chan [2012], although this is never formally stated. More generally, regarding TC structure between the RMW and the outer radii, I encourage the authors to review and engage with some of Dan Chavas' work on the topic (e.g., Chavas and Knaff [2022], Chavas et al. [2025], and references therein).

We thank the referee for pointing this out. We have clarified our decision to use the 1:8 ratio.

Line 221:

The initial sensitivity analysis (Sect. B) revealed that both RMW and R_{out} vary substantially within the simulated TCs. Near maximum intensity, simulated RMW values are approximately 30–50 km, while R_{out} is about 350 km, yielding RMW / R_{out} ratios between 6 and 11. To account for this structural variability while maintaining a representative normalization, an intermediate ratio of 1 / 8 was selected. Observed North Atlantic TCs exhibit a smaller RMW / R_{out} ratio of 1:5 (see observed radii in Fig. B2), indicating that the choice of ratio is dataset-dependent and may require adjustment when applied to other TC datasets.

Box 2.4

The model verification section (Section 3) feels somewhat tacked on. There are no spatial maps or other visualizations commonly used for model evaluation. The authors do compare some broad statistics in Fig. 3 that provide confidence that the model produces a reasonable spatial distribution of TCs (e.g., via the latitude metrics in the bottom row), but it would be helpful to formalize this evaluation further. I am not familiar with the Enz tracking algorithm, but I assume it has been vetted against reanalysis products and is able to track realistic tropical cyclones. Biases in the tracker may also lead to some biases in statistics.

Comparing the simulated tracks with the observed tracks is challenging, as the number of simulated tracks is too small to construct track density maps in the standard manner. *In addition, the simulated tracks do not precisely match the observed trajectories because the model is only forced by the 2005 large-scale environment at the lateral boundaries and no nudging within the domain was applied. Perturbed initial conditions among the ensemble members further contribute to a model spread. Consequently, the simulated TC tracks diverge from observations* (line 518). Nevertheless, we agree that track information is useful for assessing the simulated tropical cyclones. We therefore added the tracks of both simulated and observed TCs, together with a discussion, in Sect. A.

Box 2.5

Regarding the model configuration itself, it may be worth providing more context in the limited-area model (LAM) discussion. Given the size of the domain, I would not expect lateral boundary forcing from ECMWF HRES to strongly constrain the interior solution, although this is not definitively stated. Based on the analysis focusing on the statistical distribution of simulated TCs, it appears that no interior nudging was applied and that each ensemble member developed its own internal meteorology. In addition, the use of prescribed SSTs does not permit cold wakes or other air-sea feedbacks, and this limitation would be worth explicitly noting (there is substantial literature on this issue over the past decade).

We have revised the description of the SSTs and clarified the degrees of freedom of the ensemble members. In addition, we added two statements addressing the limitations of atmosphere-only simulations.

Line 100:

Sea surface temperature was prescribed using daily fields online-interpolated from monthly mean input, itself derived from the ECMWF 6-hourly IFS HRES analysis for 2005.

Line 121:

Since the model is only constrained by the 2005 large-scale environment at the lateral domain boundaries, each member has some degree of freedom to evolve into its own internal state within the domain. Consequently, the simulations represent eight plausible realizations of the 2005 season rather than its exact reproduction. The simulated

TC tracks therefore diverge from observed tracks (Fig.A1).

Line 307:
This imbalance between weak and intense storm frequency and the overall low number of simulated TCs could partially stem from the model setup, as coupled ocean-atmosphere ICON simulations run globally at a convection-permitting resolution of 5 km reproduce the observed annual TC counts more accurately (Baker et al., 2024).

Line 466:
These model biases may be attributable to limitations of the chosen model configuration, such as the absence of atmosphere–ocean coupling.

Box 2.6

Regarding the TC pressure-wind relationship (Section 3.1), I do not think this analysis materially affects the main results, but I found the inclusion of this section somewhat confusing. Pressure-wind relationships are typically defined at the surface and include the correction from gradient wind aloft to the surface through turbulence. They are a useful diagnostic for evaluating models because they help link resolved dynamics with near-surface parameterizations [Chavas et al., 2017, Nardi et al., 2022] and provide information about effective model resolution [Reed et al., 2015]. It is therefore unsurprising that the pressure-wind relationship degrades above a few hundred meters. If winds at or above ≈1 km height were better 'matched' to observations in this framework, it might actually suggest a poorly configured model.

We thank the referee for this feedback. We discuss the modifications regarding the wind–pressure relationship in Box 1.22.

Box 2.7

The authors argue that SyNC composites enhance statistical power (Section 6.2) by reducing within- group variance, enabling the detection of differences between TC groups that would otherwise be missed. While this is a reasonable expectation in principle, I am concerned that the demonstration is somewhat circular. Because the method forces RMW to a normalized radius of 1 by construction, it is unsurprising that variance decreases near the eyewall and that more grid cells pass significance thresholds there. The paper would be strengthened by showing that newly significant regions correspond to physically expected and interpretable contrasts between weak and intense storms, rather than focusing primarily on the number of significant grid cells or p-values at individual locations.

In Sect. 6.2, we aim to highlight impacted regions that can be linked to dynamical or thermodynamical processes relevant for the physical understanding, and we list several examples from the manuscript below. It is important to note that the areas exhibiting statistically significant differences for SyNC but not for the centered-only approach (Fig. 10i,j and 11i,j) extend beyond individual grid cells. The distributions shown for a single grid cell in subpanel k are intended solely to illustrate the mechanism leading to these differences in statistical significance.

Line 426:
While the centered- only composite shows the most prominent differences in radial winds, it misses several mesoscale features that are detected by the SyNC composite (Fig. 10j): Significantly stronger super-gradient outflow above the boundary layer, significantly stronger inflow throughout the vortex and significantly stronger mid-level outflow, potentially induced by rainbands, for intense TCs.

Line 447:
In the temperature tendency due to saturation adjustment (Fig. 11), the centered-only composite method even fails to detect any significant changes (Fig. 11i), whereas the SyNC method reveals significantly enhanced latent heating in the intense storms (Fig. 11j), as expected.

Box 2.8

I understand the motivation for comparing traditional center-based composites to SyNC in later figures, but I wonder whether comparing two model configurations (e.g., different parameterizations or resolutions of ICON, or ICON versus ERA5) would provide a more informative test of the method than stratification purely by intensity. As an aside, while not required for this manuscript, I would have been interested in seeing ERA5 evaluated within this framework.

We deliberately chose a simplified ("dummy") test case with a known outcome in order to focus on evaluating the SyNC method itself. We agree that assessing the performance of SyNC across different horizontal resolutions would be interesting; however, this would be beyond the scope of the present study. We therefore added a more detailed discussion of horizontal resolution, its relevance for resolving mesoscale features, and recommendations for the application of SyNC (see Box 1.3). At a horizontal resolution of 0.25 °, we expect ERA5 to only marginally resolve vortex variability and mesoscale features. Consequently, the benefit of applying SyNC to ERA5 data is likely limited.

2.3 Minor comments

Box 2.9

Composites can be constructed in latitude-longitude space by projecting fields to great-circle distances. This is implicitly what the authors mean when they say "bring TCs to the equator" in Line 27, but it may be worth stating this more explicitly. Note that there is no requirement for longitude to be fixed, as the only geometric distortion arises from converging meridians; a cyclone projected to any longitude will appear identical as long as it is at the equator.

We thank the referee for this comment. We have adjusted the manuscript accordingly.

Line 28:
Some composite methods project each cyclone to the Equator, thereby using great-circle distances to represent cyclone size to make cyclones located at different latitudes more comparable (Vessey et al., 2022)

Box 2.10

Line 226: This is perhaps semantics, but the simulated central pressure is more reliably simulated (compared to surface wind) in lower resolution models (e.g., 25km) due to the effective resolution of the models not being able to support observed RMWs and the associated inner core pressure gradients 2 [Chavas et al., 2017, Hodges et al., 2017, Zarzycki et al., 2021]. In Judt et al. [2021], the pressure-wind errors are likely more tied to surface layer schemes or PBL parameterization since the dynamical cores should (in theory) be able to capture RMWs down around 15-20 km.

We have adjusted the discussion to better distinguish between the limitations of coarse-resolution and higher-resolution models (see Box 1.3).

Box 2.11

Line 283: It may be worth briefly calculating vertical wind shear here. A common approach (e.g., DeMaria and Kaplan [1999]) is to remove the vortex and then compute the shear between two levels (e.g., 850-250 hPa). It would be easy and would add some confirmation to the speculation in '... likely due to wind shear pushing the vortex...'

We have revised Fig. 5 and 6 to include the background flow at 850 hPa, 400 hPa, and 200 hPa, as well as the translation vector and the land mask.

Line 325:
Before the symmetrization and normalization (Fig. 5c), the tangential wind field is shifted into the northeastern quadrant, likely due to the impact of friction over land and the storm translation vector aligning with the tangential winds.

Line 332:
A cat_{pres} 1 storm, illustrated in Fig. 6, has not achieved full structural organization, likely due to strong vertical wind shear and interactions with land

Box 2.12

Line 366: What do the authors mean by "saturation adjustment"? I presume this refers to tendencies associated with parameterized latent heat release in ICON. The linkage between this discussion and the subsequent figures could be clearer in the text.

We have provided additional details on the saturation adjustment.

Line 107:
Cloud droplets' growth is implemented by saturation adjustment, whereby updraft-induced water vapor supersaturation is reduced to 100 % by condensation on cloud droplets.

Line 422:
Figures 10 and 11 display the radial wind and temperature tendencies due to saturation adjustment (parametrized cloud droplet growth)...

Box 2.13

The authors discuss extratropical cyclones a few times (in the abstract and introduction). My preference is to perhaps remove or downplay them, particularly in the abstract, since all work here is axisymmetric.

We agree with the referee's comment and removed extratropical cyclones from the manuscript.

Box 2.14

- Line 48: "loosing" should be "losing"
- Line 55: "withing" should be "within"
- Line 80: The domain is listed as "105 °E to 18 °E," but this is the North Atlantic; these should almost certainly be °W, not °E.
- Line 88: "Cloud droplet's growth" should be "Cloud droplets' growth"
- Lines 80 vs. 256: The northern boundary is given as 45 °N in the domain description, but later the text refers to the domain being limited to 55 °N. This inconsistency should be resolved.
- Line 123: 'dynamic core' should be 'dynamical core'
- Line 258: "median central pressure median of 938 hPa"
- Line 264: Might be worth also citing Davis [2018].
- Line 273: R33, R25, and R17 are referenced for validation, but only R17 is formally defined. Given the use of standard TC wind thresholds (34, 50, 64 kt), it would be helpful to clarify these definitions.

We thank the referee for pointing out the typos in the manuscript and corrected all of them.

3 Response to referee #3

Comments to the Author

Box 3.1

Overview: The authors introduce a new technique for compositing tropical cyclone structure and physical processes by symmetrizing vortex structure using 10 distinct detection sectors for the storm's radius of maximum wind and outer size. While preserving information about these different radii to allow for investigations of storm asymmetry, they then produce a normalized, storm-centered, Cartesian analysis that the authors argue sharpens mesoscale structures like supergradient winds and eyewall convective heating. The authors contend that their "SyNC" technique can strengthen the analysis of composite TC structure, particularly as the community continues to approach global storm-resolving modeling. I recommend major revisions. The manuscript is well-written and organized, the figures are particularly engaging and well-designed, and the technique described throughout the manuscript has potential to address important problems in TC structural and climatological research. I have no doubts that it will be a valuable addition to GMD in time, and many of my comments below are more minor in nature, asking for further clarification at points throughout the manuscript. Some of my outstanding questions are more significant, however, surrounding the applicability of SyNC to other modeling frameworks and its ability to reliably capture asymmetric TC processes such as vortex tilt and eyewall replacement cycles. I will outline these major comments below with specific examples, then conclude with a list of line-by-line minor suggestions. Thank you for your effort on a compelling manuscript, apologies for my delay in reviewing, and I look forward to reading the next version!

We thank the referee for the constructive review and address the comments point by point below. Moreover, we thank the referee for linking the method to "mesoscale" features, which aligns perfectly with the intention of SyNC. We therefore replaced the less precise term "small-scale features" with "mesoscale features" throughout the manuscript.

3.1 Major comments

Box 3.2

Applicability for some important mesoscale structural components of TCs: The authors do well to explain how SyNC captures axisymmetric features like supergradient outflow above the boundary layer, diabatic processes in the eyewall, and subsidence within the eye. They also argue that their methodology accounts for structural asymmetries caused by storm motion, vertical wind shear, and vortex tilt. While I would agree that sharpening boundary layer and eyewall processes is useful, I think there are other applications that are important for SyNC to be a "game changer" in the field. This includes:

a) Lines 16-17, Line 356: You discuss challenges of the technique when the vortex does not exhibit a Rankine-like wind profile. You also mention some uncertainty where outer latent heating maxima could be interpreted as eyewall or rainband features. My question: Could SyNC be applied to identify and track secondary eyewall development? I am admittedly unsure if ICON in its current form would explicitly simulate such a process (though the 3 km HAFS has encouragingly done so in recent years). I will suggest that the authors at least try to perform a case study on an eyewall replacement with SyNC, but if this is not possible, at least some discussion/speculation in the Conclusions section on whether or not this can be done would be useful in my view.

b) Lines 159-160: Allowing the RMW to be different at each vertical level, to me, would make eyewall convection that is normally slanted appear more upright in the resulting composite. So I agree that eyewall tilt is accounted for. But would this actually account for vortex tilt? Is the center allowed to vary with height based on a different threshold such as a geopotential minimum or vorticity centroid? See Ryglicki and Hart 2015 for an overview of different center finding techniques aloft, which has since been expanded upon. As written, it is unclear how vortex asymmetries are accounted for outside of the horizontal plane.

We thank the referee for the input and questions.

a) A case study of secondary eyewall formation is beyond the scope of this study. However, secondary eyewalls do form in the ICON configuration used here, and testing SyNC in such cases could therefore be part of future work. We have added a brief discussion and hypotheses regarding this aspect in the conclusions section.

Line 488:
Moreover, a challenge for the eyewall detection arises from secondary eyewalls since within the current implication, only one eyewall can be defined. Secondary eyewalls are only identified as the primary eyewall once they exceed the primary eyewall in tangential wind strength within a given detection sector. This limitation also suggests a potential use of SyNC for investigating the formation of secondary eyewalls. The transition point at which the secondary eyewall becomes dominant may induce a sudden increase in the RMW within individual sectors and may temporarily increase eyewall asymmetry. Accordingly, variations in RMW and/or eyewall AI could be used to identify secondary eyewalls or even eyewall replacement cycles, a hypothesis that could be tested in future work.

b) We consider the eyewall tilt and the overall vortex tilt to be related to some extent. Consequently, aligning the eyewall likely also straightens the vortex to a certain degree. However, to be precise, we have adjusted the wording in the manuscript to consistently refer to this process as eyewall alignment. In addition, we have added a discussion of the current definition of the vortex center and potential improvements through the use of a vertically varying detection method.

Line 137:
The TC center is identified by the sea level pressure minimum.

Line 194:
Note that the cyclone center is defined as the location of the sea level pressure minimum and is assumed not to vary with height. This assumption can introduce errors in radial and tangential wind components in the presence of vortex tilt, particularly at upper levels. However, errors mainly affect radial winds, while tangential winds are only weakly impacted (Ryglicki and Hart, 2015). Therefore, the wind radii detection are not expected to be strongly affected by this simplification.

Line 505:
A different detection method of the vortex center and a vertically varying vortex center could further improve the precision of composites.

Box 3.3

Methodological questions that may need to be expanded upon:

a) Lines 133-134: Are genesis and lysis defined as the first and last time steps identified by your tracking algorithm? I mostly ask because the definition can vary significantly, especially for lysis where powerful extratropical cyclones can be perhaps mislabeled as "decayed" after transition.

b) Lines 153-154: For reproducibility and applicability to different model frameworks – how straightforward is it to alter the resolution of the projected grid, and is it important that the projected grid approximately match the effective resolution of the model you are applying the SyNC technique to? This should be discussed in the manuscript, in my view.

c) Line 268: "While weaker storms are relatively rare". Are they? This may be an artifact of the tracking methodology, where you filter out any TCs that do not achieve a lifetime of maximum intensity of 990 hPa or lower. Bourdin et al. (2024) introduced a "Category 0" (1005-990 hPa) to account for simulated tropical storms, which you may consider including in your analysis while keeping the longevity criterion you use in your tracking. You already acknowledge that SyNC has challenges with weaker TCs, however, so if you choose not to expand your analysis to $>$ 990 hPa TCs, you should at least acknowledge the tracking approach as a factor in the intensity bias.

We thanks for the comments.

a) We have added clarification regarding lysis in the manuscript and included the major causes of track termination in the description of the TC tracker. As shown in the newly added track maps (Sect. A), the simulated tracks extend far northward, indicating that extratropical transition can be detected by the tracker. Since the intensification and decay phases are separated by the time of minimum central pressure, the defined "decay phase" is characterized by an overall decrease in intensity. Nevertheless, periods of re-intensification can occur during this phase, although they do not reach the lifetime maximum intensity again. We agree that this represents a limitation of the applied time normalization.

Line 166:
TC genesis and lysis are defined by the start and end points of the detected TC tracks.

b) The projected grid resolution can be easily modified, as it is a selectable parameter in the SyNC settings. We have clarified our motivation for choosing the original model resolution for the projected grid. In addition, the term "effective resolution" was incorrectly used in the manuscript and has therefore been removed.

Line 181:
The projected grid resolution is set to match the model resolution (x= 5 km) to avoid data extrapolation or interpolation, thereby minimizing data modification in this initial step.

c) We agree that this statement was unclear and have therefore reformulated the section. Category 0 storms are not included because, as noted above, SyNC faces challenges when applied to such weak systems. In addition, the tracker has difficulties reliably following storms that remain at very low intensities, resulting in uncertain track representations. We acknowledge this limitation explicitly in Sect. 3.1.

Line 285:
The TC tracker likely struggles to detect the weak and very early stages of TCs, when the storm's central pressure is close to the ambient pressure (Enz et al., 2023). As a result, these stages are missed in our analysis.

Box 3.4

On the discussion of statistical power and "reduced within- group variance": I will repeat that I see the utility of SyNC in sharpening inner-core features. But is reducing variance necessarily a good thing? On one hand, differences between composite means are more likely to be significant and physically meaningful when there is little variance within the sample. However, many researchers may be interested in explicitly quantifying the variability within a composite. Can SyNC do this? You may consider briefly discussing this application in the Conclusions.

We thank the referee for this feedback. SyNC is particularly well suited for feature-to-feature comparisons, as it captures variability among individual features, which can lead to a reduction in apparent variability across composites. If the focus is instead on variability at a fixed distance from the storm center, and less on feature-to-feature alignment, we agree that SyNC is less suitable, since the true radial distance from the TC center is no longer preserved. We added this aspect to the conclusion sections.

Line 483:
SyNC is particularly well suited for feature-to-feature comparisons and it enables the identification of variability among individual features. If the focus is instead on field magnitudes or variability at a fixed distance from the storm center, SyNC is less suitable, as the true radial distance from the TC center is no longer preserved and the identified variability by SyNC does not correspond to the variability at a given radius.

Line 509:
Although the normalization results in the loss of true spatial distances within the vortex, it offers a cyclone-relative framework that facilitates distinguishing between processes occurring within the eyewall and those in the outer regions of the TC.

3.2 Minor comments

Box 3.5

- Line 6: Briefly specify in the abstract how the size of the TC is defined.
- Line 36: Another excellent paper to reference when discussing VWS would be Rios-Berrios et al. (2024) - A Review of the Interactions between Tropical Cyclones and Environmental Vertical Wind Shear
- Line 42: Again, how do you define size here? The radius of 17 m/s winds? The radius of the outermost closed isobar? Some other metric? On that note, add "wind" before "radii" on Line 44, assuming you use R17 as your outer size metric.
- Line 55: Minor typo – correct "withing" to "within" .
- Lines 83-84: Are the monthly mean SSTs you refer to specifically from 2005, or are these averaged across a broader climatology? Specify this, and what dataset SSTs are prescribed from.
- Lines 106-108: Be a bit more specific about the tracking algorithm (though I appreciate that you allude to Enz et al. 2023 for more details). What is the necessary amplitude of the local pressure minimum/vorticity maximum/upper- level temperature anomaly? What specific vertical levels are you looking at for the latter two? Also, I believe Line 107 should be changed from zeta min to zeta max.
- Lines 263-265: I would argue that a 5 km model simulating sub-900 hPa TCs is not very surprising, given that the community has achieved convection-permitting operational NWP resolutions with models like HAFS and NAM that do explicitly simulate Category 5 hurricanes. It may be worth briefly mentioning any known biases that ICON has in simulating TC structure, climatology, and intensity from other recent studies, besides just

noting that Judt et al. (2021) found a better alignment between observed and modeled TC frequency/ACE with 5 km resolution.

• Lines 283-284: Can you verify the direction of the wind shear in this particular example to support this claim? A very simple technique you can use is averaging the 200 and 850 hPa winds in a 200-800 km radial annulus surrounding the TC to calculate the vector difference. Consider including an inset on Figure 5 with the motion and shear direction arrows.

• Figure 7: Consider adding sample size information into this figure or the corresponding text discussion, and reminding the reader in the text that you are classifying TCs according to their lifetime maximum intensity for Figures 7d-g, so that TCs near the beginning or end of their life cycles could vary widely in their intensities within an individual composite.

We thank the referee for the comments and for pointing out the typographical errors. We have revised the manuscript accordingly. In particular:

- We added the size definition to the abstract and moved the definition used in the manuscript to (line 50) to avoid confusion.
- We added an appropriate reference for vertical wind shear.
- We clarified the description of SSTs (see Box 2.5).
- We reformulated the description of the TC tracker (see Box 1.11).
- We agree that it is not surprising that a model with 5 km resolution can simulate intense TCs. We therefore removed this statement and added an additional reference on ICON TC simulations.

Line 307:

This imbalance between weak and intense storm frequency could partially stem from the model configuration. The too low number of simulated TCs could also be related to this configuration, as ICON simulations run globally at convection-permitting resolution (at 5 km) with ocean–atmosphere coupling reproduce observed annual TC counts more accurately (Baker et al., 2024).

- We added the background flow and translation vectors to Fig. 5 and 6 (see Box 2.11).
- We included the sample size in the figure legends and added a statement clarifying the temporal alignment of the evolution.

Line 343:

The temporal evolution is aligned with respect to the lifetime maximum intensity. Consequently, storm intensities at a given normalized time step may vary.

References

Bui, Hai Hoang, Roger K. Smith, Michael T. Montgomery, and Jiayi Peng (Oct. 2009). "Balanced and unbalanced aspects of tropical cyclone intensification". en. In: *Quarterly Journal of the Royal Meteorological Society* 135.644, pp. 1715–1731. ISSN: 0035-9009, 1477-870X. DOI: `10.1002/qj.502`. URL: `https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.502` (visited on 02/20/2026).

Ji, Dong and Fangli Qiao (June 2023). "What Are the Balanced and Unbalanced Dynamics of a Tropical Cyclone?" EN. In: *Journal of the Atmospheric Sciences* 80.7, pp. 1719–1737. ISSN: 0022-4928, 1520-0469. DOI: `10.1175/JAS-D-22-0165.1`. URL: `https://journals.ametsoc.org/view/journals/atsc/80/7/JAS-D-22-0165.1.xml` (visited on 02/20/2026).

Smith, Roger K. and Michael T. Montgomery (2023). "Chapter 6 - Frictional effects". In: *Tropical Cyclones*. Ed. by Roger K. Smith and Michael T. Montgomery. Vol. 4. Developments in Weather and Climate Science. Elsevier, pp. 137–162. ISBN: 978-0-443-13449-4. DOI: `https://doi.org/10.1016/B978-0-44-313449-4.00014-X`. URL: `https://www.sciencedirect.com/science/article/pii/B978044313449400014X`.