Response to Reviewer 1

In their study, Yu et al consider forecast biases of extratropical cyclones depending on the degree of diabatic heating attributed to them. Studying these heating- dependent biases is of importance, particularly as the cyclones with strong diabatic heating exhibit significant biases. The text is generally well written, and the overall discussion can easily be followed. Furthermore, the figures are of high quality, and they clearly support the statements in the text. Still, there are some aspects that deserve further consideration. At some places, the statements remains relatively descriptive, or speculative. It remains also not so clear how the heat-dependent biases 'only' reflect the latitudinal distribution of weak- and strong-heating cyclones; or a link is speculated about dry intrusions (DI) and/or warm-conveyor belts (WCB), but the statements remains rather qualitative.

In summary, I think that the study is of great interest to the research community, and it also fits perfectly into the scope of Weather and Climate Dynamics. To make it publishable, however, some major revisions are needed, as outlined below in detail.

We are thankful for the constructive comments from the reviewer. We hope that all concerns have been duly addressed in the revised version of this paper.

Comments by the reviewer are in **bold**, followed by our replies. Figures from the original manuscript are referred to following the manuscript's order, while new figures included in this document are labelled as Fig. AR# (Author Response).

Major concerns

1) **Diabatic influence of surface fluxes: In section 2.3, it is written how the diabatic influence on a cyclone is quantified. Essentially, this is done by taking from ERA5 the diabatic tendencies due to all parameterizations, except for radiative tendencies. From the text it becomes *not* clear how this is done, whereby I assume that the ERA5 field for all diabatic tendencies is taken and the AER5 of radiative heating is subtracted from it. The authors should write more explicitly which fields are available from ERA5 short-term forecasts (?), and describe in greater detail which diabatic processes are included (or not). Furthermore, the reader should be informed why radiative tendencies are excluded and whether**

they are expected to have a big impact. Please make also clear whether diabatic surface heat fluxes, latent and sensible ones, are included or not. To my knowledge the diabatic tendencies from ERA5 do not explicitly include them. A short discussion on free-tropospheric diabatic tendencies and the ones originating from the surface would be helpful.

We thank the reviewer for pointing this out. We have clarified the data sources and calculation methods in the revised manuscript (Section 2.3).

Specifically, we address your questions as follows:

Data: The diabatic tendencies are derived from the short-term forecasts of the ERA5 reanalysis (interpolated from model levels to pressure levels). Temperature tendencies due to physical parameterisations are accumulated variables, where we use the same procedure as Tsopouridis et al. (2021); Weijenborg and Spengler (2020). Specifically, they are accumulated ± 3 hours around the respective timestep. We calculate the diabatic heating, excluding radiation, by taking the temperature tendencies due to all physical parameterisations and subtracting the radiative heating tendencies.

Exclusion of radiation: Following Papritz and Spengler (2015), radiative tendencies are excluded because they primarily act as a slow background process. Removing them allows us to isolate the rapid, non-radiative diabatic heating (mainly latent heat release) that rapidly enhances local baroclinicity and drives the core dynamics of cyclone development. This focus aligns with our goal of evaluating biases in 12-h short-term forecasts.

Surface and free troposphere: We do not include surface heat fluxes in our metric. Even though diabatic heating from boundary layer turbulent mixing is included, given that our vertical integral is restricted to the troposphere (700–300 hPa), the heating is dominated by latent heat release. Surface latent heat fluxes can play a crucial indirect role by supplying the boundary layer moisture that subsequently condenses and releases heat.

We have explicitly added these details to Section 2.3 to make the methodology more transparent.

2) **Snapshot diabatic tendencies vs. accumulated ones: The cyclones are categorized into the ones with strong diabatic heating and weak diabatic heating *at the time step of maximum intensification* during the cyclones' life cycle. This is certainly a valid approach,**

however, it makes me also wonder how the *accumulation* of diabatic effects/tendencies prior to this time step contributes to the forecast biases. I agree with the authors that the strongest diabatic effect might be expected to coincide with the time step of maximum intensification, but still it could be possible that in time span from genesis to maximum intensification substantial diabatic tendencies are accumulated. If so, taking the tendencies at this single time step of maximum intensification could be misleading. I don't expect the authors to redo their analysis based on accumulated tendencies. But they should carefully discuss why it is appropriate to only take on snapshot to do the classification. Possibly, the can compare the categorization for a subsample with snapshot vs. accumulated tendencies.

The reviewer raises a valid physical point regarding the accumulation of diabatic effects. However, diabatic heating in cyclones is not evenly distributed but highly localised in time and space. Furthermore, it has been demonstrated that intense latent heat release is concentrated during the rapid deepening phase (e.g., Binder et al. (2016); Madonna et al. (2014)). Furthermore, accumulating heating within a moving spatial domain (e.g., a 750 km radius around the cyclone centre) over time averages over distinct air masses. A rigorous accumulation would require a Lagrangian trajectory tracking, which is beyond the scope of our composite analysis.

3) **Latitude effect: The authors clearly state that strong-heating and weak-heating cyclones occur, on average, at distinctly different latitude. As expected, strong- heating cyclones are predominantly found at lower latitude, where moisture availability is higher. Given this difference in latitude, one wonders however if the differences/biases of the two heating categories are because of these heating effects, or if it is just a latitudinal effect. Would it be possible to get subcategories of cyclones having their time step of maximum intensification at rather similar latitudes, but still differing substantially in their diabatic heating? Or, similarly, how do cyclones compare with similar diabatic heating but occurring at different latitudes? I think the study would benefit if the latitude effect is somewhat more carefully discussed.**

The reviewer raises a valid concern. It is a highly relevant question whether the observed forecast biases are mainly due to the diabatic heating itself or only reflect a latitudinal background effect, given the dependency of moisture availability on latitude.

We conducted a further subsampling analysis by focusing on the latitude band of 45°–55°N. Although the sample sizes differ between the two groups in this band, we applied a Welch's t-test to account for this. Fig. AR1 clearly demonstrates that even when the latitudinal band is confined, the two groups separated by diabatic heating intensity still exhibit distinct different bias patterns. Specifically, the strong heating group maintains a pronounced south-westward displacement bias, whereas the weak heating group only shows a slight westward bias. The areas with statistical significance ($p < 0.05$ indicated by stippling) clearly cover the region of the south-westward bias in the strong heating group. This difference confirms that the diabatic heating acts as an independent factor for the forecast biases, rather than unequal sample sizes, random synoptic noise or latitudinal background effects. We have also added expansions to the discussion in Section 2.3.

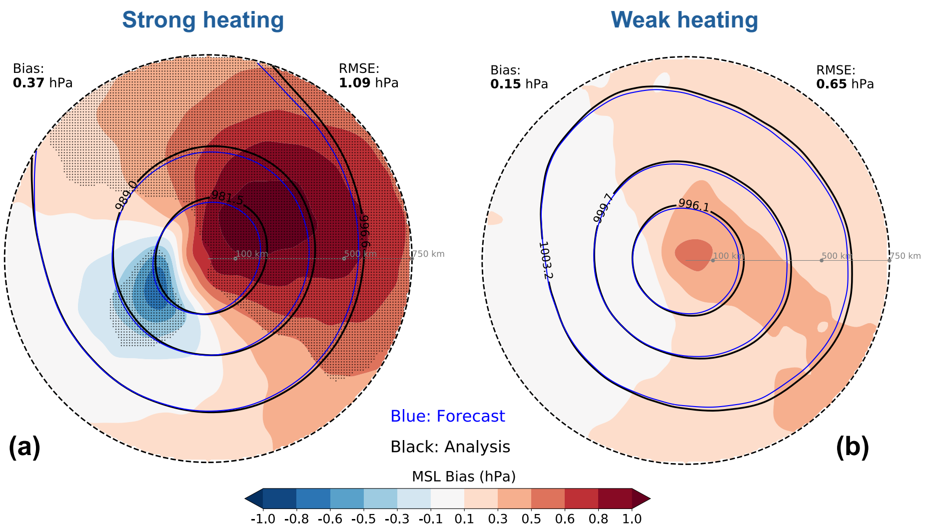

Figure AR1: As in Figure 2, but only for latitude 45°–55°N

4) **Propagation bias for strong-heating cyclones: The authors present in Section2.4 a good way how to get rid of the propagation bias of strong-heating cyclones, and they also introduce the need to**

handle this bias in a clear and didactic way by first showing the 'uncorrected' in Figure 2, before showing the corrected ones in Figure 3. I see that the focus of the study is on structural biases, but it would still be of interest to briefly discuss how diabatic heating could lead to these propagation biases. What is the seasonality of this propagation bias? Which processes could be responsible for the bias? I would appreciate at least a brief discussion on this points.

We thank the reviewer for this highly constructive suggestion. We agree that a brief discussion of how diabatic heating could lead to these propagation biases is valuable for our findings.

We have added a discussion on this topic in the Conclusion section, as follows:

'The pronounced southwestward displacement bias observed in the strong heating group is relevant to the influence of diabatic heating through two processes. First, intense diabatic heating accelerates cyclones (Stoelinga, 1996; Coronel et al., 2015). Thus, the propagation bias suggests a misrepresentation of diabatic heating or its effects in the forecasts, with the bias being more pronounced in the strong heating group than the weak heating group. As the composite cyclones are moving eastward, the westward positional bias manifests a slower movement speed. Second, our results show an underdevelopment of the upper-level ridge exclusively in the strong heating group, which is related to the southward displacement bias (Coronel et al., 2015). The combination of this southward bias and the westward propagation delay explains the distinct southwestward displacement bias in the strong heating cyclones.'

5) **Qualitative/speculative vs. quantitative/confirmed statements: The discussion of the sea-level pressure, moisture, wind and temperature biases remain at a rather qualitative level, which is also reflected in many 'could be' / 'might be related' statement. Some specific examples are: - L163: "The TCWV deficit is thus consistent with insufficient moisture transport in the warm sector, most likely coinciding with warm conveyor belts." - L173: "This bias might be attributable to error sources from both microphysics and dynamics" - L175: "the bias in TCLW could indicate issues with the microphysical scheme" - L197: "Given that these biases for the strong heating group are primarily associated with the warm**

sector, the weaker asymmetry and bias could also be due to weak heating cyclones occurring primarily at higher latitudes (Fig. 1b), limiting the moisture supply, resulting in less pronounced biases." - L213: "This negative bias indicates an overestimation of inflow (convergence), which is most likely associated with an overestimated vertical motion resulting from the frontogenesis bias (Fig. 4d)." - L243: "This overestimation of upper level PV is likely associated with enhanced descending along dry intrusions (DI), which can result in larger cold and dry air advection within the cold sector (Catto and Raveh-Rubin, 2019)." - L247: "The underestimations of both the upper-level PV and geopotential are most likely due to misrepresentations in diabatic heating and are thus consistent with biases presented earlier, such as the underestimation in water vapour transport (Figs. 4a,e,f), wind (Fig. 3c, 6a), and temperature (Fig. 6e)." - L285: "This intensification of frontogenesis is likely driven by the overestimated wind speeds within the CCB/SJ/DI region, which enhance the local kinematic deformation and convergence" - L287: "The intensified ascent can lead to an increase in condensation, thereby resulting in the observed positive bias in total column liquid water" - L289: " This may be exacerbated by limitations of the data assimilation (DA), as liquid water is not directly constrained and the DA relies on indirect adjustments to thermodynamic and kinematic fields." The study would benefit a lot if some of these statements (and others not listed here) can be made more strong, e.g., by applying some extra analysis. I don't think that all 'speculative' statements must become hard facts, but at the moment that the number of these 'vague' statements is too large.

We thank the reviewer for this constructive observation. We have carefully reviewed the manuscript and agree that using too much phrasing (e.g., "could be", "might be related") weakened the presentation of the results.

Upon reviewing the specific lines you mentioned, as well as other speculative statements throughout the text, we realised that the vast majority of them originate from our interpretations related to the warm conveyor belt (WCB). Because the classic identification of WCBs requires Lagrangian trajectory calculations, it is highly challenging to implement within our current composite framework.

To address this limitation and transform our qualitative speculation into hard, quantitative evidence, we instead use moisture transport axes (MTAs).

As defined by Spensberger et al. (2025), the MTA algorithm explicitly "extracts well-defined maxima in the water vapour transport". As noted in their study, depending on the synoptic context, these coherent moisture filaments are referred to with a variety of names, including atmospheric rivers, warm moist intrusions, and warm conveyor belts.

By applying the same MTA mask as used in the previous study Yu et al. (2025), we quantified the occurrence frequency of these moisture filaments relative to the cyclones for the two different heating groups. This quantitative approach allows us to substantially reduce the speculative content throughout the manuscript. The detailed MTA composite results are presented in our response to Point 6 below.

6) **Processes leading to diabatic heating: In line with point 5), the processes leading to the diabatic heating are not studied extensively. I see that this is *not* the main focus of the study, however, it would benefit from a more thorough process analysis. As a particular example, the diabatic heating and/or cyclone-structure biases are at several places linked to warm-conveyor belts (WCB) and cold- conveyor belts (CCB): - L150: "The strong heating group, on the other hand, exhibits a wind speed underestimation only in the warm sector, while wind speed is overestimated in the cold sector over (Fig. 3c), the region usually associated with the Cold Conveyor Belts (CCBs), dry intrusions (DIs), and sting jets (SJs) (Schultz, 2001; Browning, 1997, 2004). - L164: "The TCWV deficit is thus consistent with insufficient moisture transport in the warm sector, most likely coinciding with warm conveyor belts" - L249: "Overall, these biases are most likely linked to an underestimated strength of the WCB" Since cyclone intensity has previously been related to WCB activity in the cyclone's warm sector, and since WCB are associated to strong diabatic heating due to, e.g., condensational heating, it would be 'excellent' if the authors could 'build' a somewhat stronger link to WCB and other airstreams in a cyclone. Optimally, the authors would link their analysis with a dataset that shows the absence/presence, the intensity and the location of WCB at the time of maximum intensification. Optimally, of course, means that the authors would have such a dataset at hand for all their cyclones. I see that the identification of these WCB datasets is challenging! Still, the authors should consider whether there is a possibility for WCB identification of a sub-**

set of their cyclones and to include this analysis then into the manuscript? Please note also that the following study could be of interest, as it also establish a 'maximum' WCB activity to the time step of maximum cyclone intensification: Heitmann, K., Sprenger, M., Binder, H., Wernli, H., and Joos, H.: Warm conveyor belt characteristics and impacts along the life cycle of extratropical cyclones: case studies and climatological analysis based on ERA5, Weather Clim. Dynam., 5, 537–557, https://doi.org/10.5194/wcd-5-537-2024, 2024. If the authors agree that this is relevant, they might include in their text, and partly base their answer to my concern on this (and similar) publications.

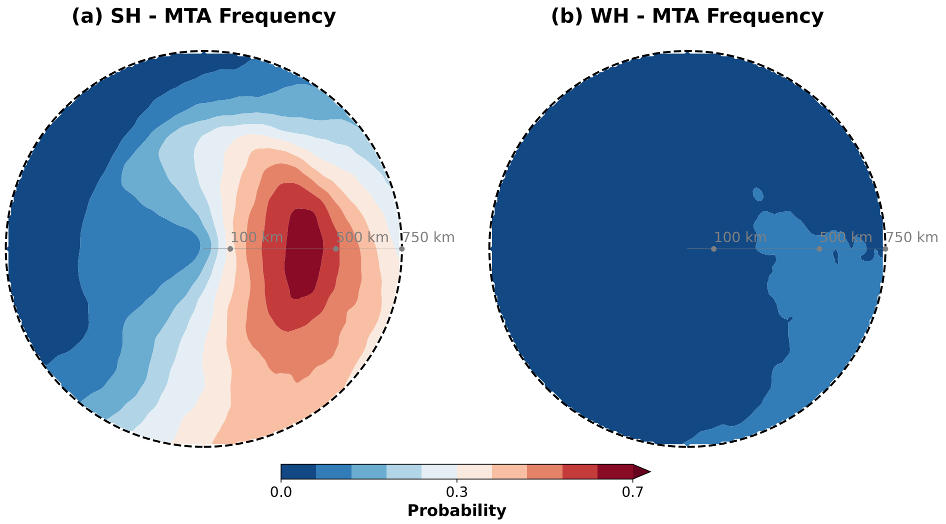

Figure AR2: Cyclone-relative composite spatial frequency of Moist Transport Axis (MTA) for (a) strong heating (SH) group and (b) weak heating (WH) group. A 200-km distance threshold is applied around the MTA line to calculate its spatial frequency.

We sincerely thank the reviewer for this constructive comment. We agree that establishing a stronger link to WCBs significantly strengthens related statements of our study. We are also very grateful for the recommendation of the relevant literature.

As briefly introduced in our response to Point 5, we establish the connection by calculating the occurrence frequency of MTAs for both cyclone groups (Fig. AR2). For the strong heating group, over 70% of the cyclones feature these intense moisture filaments within their warm sectors. In contrast, for the weak heating group, the occurrence frequency is less than 10%. With this clear quantitative proof, we have thoroughly revised the previously speculative phrasing regarding WCBs throughout the manuscript. Furthermore, we agree that Heitmann et al. (2024) is relevant and have added it into our text.

Specific/minor comments:

- **L17: "Our findings highlight the impact of diabatic heating on structural cyclone forecast biases that can guide future model improvements." This stement remains rather vague, and I would either strengthen or remove it from the abstract.**

  We have revised the final paragraph of the abstract.

- **L41: Which instability? Further, is it 'increasing instability' or is it 'reducing stability'?**

  We have rephrased the sentence by using 'reduces the static stability'.

- **L61: The research question remains somewhat vague, and I would appreciate if some (2-3) very concrete research questions are raised and later discussed in the conclusions. At the moment, the reader only learns that a composite view on the bias will be given.**

  We thank the reviewer for pointing this out. To clarify our focus, we have added the specific objectives of this study at the end of the Introduction section. Accordingly, we have also provided direct answers in the concluding section at the end of our study. The added text in the introduction reads as follows:

  'By applying a composite approach, we provide a statistically aggregated view of how different diabatic conditions yield systematic structures in forecast biases. The objective of this study is to identify the link between the intensity of diabatic heating and 12-hour ETC forecast biases. Ultimately, this study helps to understand the shortcomings in representing diabatic moist processes and their related dynamics, which can guide future model improvements.'

- **L74:'tangential' and 'radial' relative to what? This becomes only clear in the result section, where the cyclone-centered composites are shown and discussed. There, the wind decomposition becomes clear, but not at this place.**

  We have provided more detail for clarity. The revised sentence reads: 'In addition, to evaluate the low-level circulation and frontal dynamics, we compute tangential and radial wind components at 925 hPa relative to the cyclone centre, as well as kinematic frontogenesis at 850 hPa ...'

- **L85: "to be identified as a closed system" Please explain in 1-2 sentences what 'closed system' means in this context.**

  We have added a concise explanation in the text as follows: 'All tracked cyclone centres are required to be identified as a closed system at least once within their lifetime, in line with Murray and Simmonds (1991a,b). Specifically, the cyclone centre must be enclosed by at least one closed isobar, ensuring the feature develops into a distinct low-pressure system rather than an open trough.'

- **L104: "all fields are rotated so that motion in all composites is aligned along the x-axis" Okay, but I am not completely convinced that the rotation brings a benefit, and other readers might also wonder. Please 1-2 sentence that motivates this rotation.**

  We have expanded the explanation. As cyclones propagate in various directions, rotation is crucial to prevent implicit smoothing due to directions of propagation during the spatial averaging process. The revised text now reads: 'Following (Catto et al., 2010), cyclone-relative composites were calculated by averaging over all cyclones. During this procedure, all fields are rotated to align the propagation direction of each cyclone along the x axis. As individual cyclones move in diverse directions, this rotation is essential to prevent misalignment of features, such as warm and cold fronts, that would then be smoothed out during the averaging process.'

- **Figure 2: The number of wind vectors could be increased?!**

  Thank you for this comment.

  The main purpose of Figure 2 is to illustrate the wind biases. The wind vectors are included simply to indicate the general wind direction. Increasing their number would clutter the plot and distract the reader from the core bias information without adding new insights. We thus decided to keep the vector density unchanged.

- **Would it make sense to combine Section 3.1 and 3.3? Both are dealing with winds, and at the moment the paper's structure seems not immediately clear to me. On the other hand, the discussion in L140-147 could also be shifted to Section 2.4. Hence, already discuss in Section 2.4 the need for a position correction (propagation bias). If, on the other hand, the propagation bias itself becomes a feature to be further discussed, I think it is find to keep it here.**

  Regarding the combination of Sections 3.1 and 3.3: We carefully considered combining these sections. However, we prefer to maintain the current structure. While Section 3.3 also seems to be dealing with winds, it focuses specifically on the primary and secondary cyclonic circulations. This is tightly related to, and motivated by, the thermodynamic and frontogenetic biases found in Section 3.2. To address your valid point that the structure was not immediately clear, we have added a structurally guiding paragraph at the end of the introduction.

  Regarding the propagation bias: We appreciate the reviewer providing the option to keep this discussion in its current place. We prefer to retain it in Section 3.1 because the propagation bias is one of our key scientific findings. We briefly introduced it in Section 2.4 primarily to provide the necessary context for the re-centring part. We have now added a paragraph discussing this propagation bias in the conclusion section.

- **L184: "Through thermal wind, the resulting enhanced temperature gradient strengthens the vertical shear of the geostrophic wind, consistent with the wind speed overestimation (Fig. 3c)." Be careful in using thermal wind balance to *cause* vertical wind shear. The thermal wind balance is a purely diagnostic relationship between horizontal temperature field and vertical wind shear, and – in my view – it is not valid to build any causality on it: Does wind shear react to temperature gradient, or is it the other way round: does the temperature gradient react to wind shear? I would stress that it is neither nor.**

  The revised sentence now reads: 'In accordance with thermal wind balance, the resulting enhanced temperature gradient corresponds to a stronger vertical shear of the geostrophic wind, consistent with the wind speed overestimation (Fig. 3c).'

- **Figure 3 and 4: Some of the effects are rather small, especially if comparing the strong-heating SLP contours from analysis and forecast. Either the authors could provide some significance test, or they discuss in greater detail the physical relevance of the differences. Note that a difference can be statistically significant, but still physically not relevant (because the difference amplitude is so small). A brief discussion on this would be appreciated.**

  We have added the relevant discussion to the first paragraph of the Conclusion section.

  'This study employs a cyclone-centred composite framework to quantify short-term (12-hour) forecast biases for wintertime (DJF) maritime extratropical cyclones (ETCs) within the ERA5 for the period 1979–2022. To compare the influence of diabatic processes, cyclones are categorised into groups based on strong and weak intensity of domain-averaged diabatic heating. Overall, 12-hour forecasts of North Atlantic extratropical cyclones underestimate cyclone intensity at the time of maximum intensification. Notably, the bias patterns differ distinctly between the strong and weak heating groups. Although the 12-hour forecast biases are of small absolute amplitude, their physical relevance is significant: cyclones with stronger diabatic heating manifest more pronounced systematic biases tied to specific physical processes. Specifically, while forecasts underestimate cyclone intensity near the cyclone centre for the weak heating group, forecasts for the strong heating group not only underestimate cyclone intensity but also feature a propagation bias, manifesting as a southwestward displacement of the cyclone position.'

- **Section 3.4: I am not completely convinced that the temperature field needs to be discussed separately?! Possibly, it can be combined with the (qualitative) discussion on the vertical-wind forcing?!**

  We agree with the reviewer that the temperature field can be combined with the discussion of other fields. Accordingly, we have integrated the temperature discussion into Section 3.3. This restructuring creates much more cohesive and logically sequential content, as demonstrated in the revised text below: 'These cyclonic circulation biases are associated with the thermal structure through altered temperature advection. The temperature bias for the strong heating group shows a pronounced cold bias in the upper right-hand quadrant. This is consistent with the underestimated cyclonic flow and water vapour transport in the warm sector, thus featuring reduced warm air advection (Figs.3c, 4e, 4f, and 6a). Additionally, the lower-left quadrant also exhibits a notable cold bias. This cold bias is linked to the upper level potential vorticity (PV) biases discussed in the following section.'

References

Binder, H., Boettcher, M., Joos, H., and Wernli, H.: The role of warm conveyor belts for the intensification of extratropical cyclones in Northern Hemisphere winter, Journal of the Atmospheric Sciences, 73, 3997–4020, 2016.

Catto, J. L., Shaffrey, L. C., and Hodges, K. I.: Can climate models capture the structure of extratropical cyclones?, Journal of Climate, 23, 1621–1635, 2010.

Coronel, B., Ricard, D., Rivière, G., and Arbogast, P.: Role of moist processes in the tracks of idealized midlatitude surface cyclones, Journal of the Atmospheric Sciences, 72, 2979–2996, 2015.

Heitmann, K., Sprenger, M., Binder, H., Wernli, H., and Joos, H.: Warm conveyor belt characteristics and impacts along the life cycle of extratropical cyclones: case studies and climatological analysis based on ERA5, Weather and Climate Dynamics, 5, 537–557, 2024.

Madonna, E., Wernli, H., Joos, H., and Martius, O.: Warm conveyor belts in the ERA-Interim dataset (1979–2010). Part I: Climatology and potential vorticity evolution, Journal of climate, 27, 3–26, 2014.

Murray, R. J. and Simmonds, I.: A numerical scheme for tracking cyclone centres from digital data. Part I: Development and operation of the scheme, Australian meteorological magazine, 39, 155–166, 1991a.

Murray, R. J. and Simmonds, I.: A numerical scheme for tracking cyclone centres from digital data. Part II: Application to January and July general circulation model simulations, Australian Meteorological Magazine, 39, 167–180, 1991b.

Papritz, L. and Spengler, T.: Analysis of the slope of isentropic surfaces and its tendencies over the North Atlantic, Quarterly Journal of the Royal Meteorological Society, 141, 3226–3238, 2015.

Spensberger, C., Konstali, K., and Spengler, T.: Moisture transport axes: a unifying definition for tropical moisture exports, atmospheric rivers, and warm moist intrusions, Weather and Climate Dynamics, 6, 431–446, 2025.

Stoelinga, M. T.: A potential vorticity-based study of the role of diabatic heating and friction in a numerically simulated baroclinic cyclone, Monthly weather review, 124, 849–874, 1996.

Tsopouridis, L., Spensberger, C., and Spengler, T.: Characteristics of cyclones following different pathways in the Gulf Stream region, Quarterly Journal of the Royal Meteorological Society, 147, 392–407, 2021.

Weijenborg, C. and Spengler, T.: Diabatic heating as a pathway for cyclone clustering encompassing the extreme storm Dagmar, Geophysical Research Letters, 47, e2019GL085 777, 2020.

Yu, Q., Spensberger, C., Magnusson, L., and Spengler, T.: Forecast Errors Attributed to Synoptic Features, Meteorological Applications, 32, e70 093, 2025.

Response to Reviewer 2

This study by Yu et al. considers short-term (12h) mean forecast errors (biases) in North-Atlantic extratropical cyclones using ERA5 reanalysis data. The authors distinguish between cyclones with strong and weak latent heat release during the time of strongest intensification. The results show a low bias of cyclone intensity, more prominent so for cyclones with strong latent heat release, as well as structural differences between the strong and weak latent-heating groups.

The study has a clear focus and the presentation of results in the text and by the figures is very good. The manuscript is fine from the perspective of reporting observed patterns. Unfortunately, however, it is not clear to me what we learn from these observations. I expand on this issue below.

Before publication, the manuscript needs at least to discuss a fundamental caveat of the design of the study, which is not acknowledged in the current version. While I do not have many comments, I recommend major revisions before potential publication.

We thank the reviewer for the comments. We have revised the manuscript accordingly and hope these changes have improved its readability and quality.

Comments by the reviewers are in **bold**, followed by our replies. Figures from the original manuscript are referred to following the manuscript's order while new figures included in this document are labelled as Figure AR# (Author Response).

Main issue: Conditional verification

1) **The authors consider conditional verification, i.e., they condition their examination of forecast errors on the existence of a cyclone in the analysis. A cyclone constitutes a (strong) anomaly from the climatological mean. Forecasts tend to underestimate analysis anomalies, not because of inherent biases in the forecast system but as an inherent feature of conditioning the verification on analysis anomalies! Supposedly, this effect of conditional verification is well known in the verification community, but less so in academia (Mark Rodwell, personal communication). Unfortunately, I am not aware of a reference in the literature of this effect. The effect seems plausible when considering forecasts that have lost all skill. An average of such forecasts represents a climo state. Forecasts without skill thus evidently underestimate on average any anomaly**

from the climo state when conditioned on anomalies existing in the analysis. Forecasts with less than 100% skill hence exhibit the tendency to underestimate anomaly amplitude in conditional verifications. This effect may be accentuated in the current study by the authors' choice to condition on maximum intensification rate.

Most of the signal that the authors find is a low bias in intensity. As frequently noted by the authors, most of the other features they discuss are consistent with this low bias. The low bias in turn may be a mere artifact of the authors' conditional verification. Hence, what do we learn in this study about the role of latent heat release in cyclone forecasts? Other than the well-known effect of latent heat release to additionally intensify cyclones. How can we then distinguish between structure bias that is consistent with low intensity bias and additional biases due to latent heat release?

And as a corollary of the above discussion: To what extent are the bias structure differences between the weak and strong heating groups mere reflections of the structure differences between cyclones with weak and strong latent heat release?

We agree with the theoretical premise that conditional verification can lead to an artificial underestimation of anomalies (regression to the mean), particularly when considering medium- to long-range forecasts that begin to lose predictability and regress toward a climatological state (as the reviewer aptly noted: "when considering forecasts that have lost all skill"). However, to minimise this statistical artefact, we specifically focus on the 12-hour forecast (first-guess) and fast-physics processes.

Our rationale is supported by three key points in the NWP literature:

1. For conditional verification: while unconditional global metrics (like RMSE or ACC) evaluate overall performance, they average errors across space and time, thereby hiding localised issues or poor performance in predicting critical extreme events. Conditional evaluations evaluate how well a model simulates specific features, such as extratropical cyclones (Dacre et al., 2026).

2. For forecast skill: at a 12h lead time, models maintain a remarkably high degree of accuracy (Bauer et al., 2015). The atmospheric state has not yet decorrelated into chaos, and the model is not regressing to a climatological mean.

3. About what can be learned from this study: using short-range NWP forecasts to evaluate parametrised "fast-physics" (e.g., diabatic heating, cloud microphysics) is a well-established, efficient, and computationally feasible method to ultimately improve long-term weather and climate predictions (Rodwell and Palmer, 2007; Xie et al., 2012; Klocke and Rodwell, 2014). By systematically evaluating cyclone biases on weak and strong diabatic heating, our study reveals that cyclones with strong heating exhibit distinct biases tied to the representation of specific physical processes. These distinct patterns provide concrete and actionable targets for improving the parameterisation in NWP models.

Finally, the reviewer raises two insightful and interconnected questions: (1) how to separate latent-heat-driven structural biases from general intensity biases, and (2) how to ensure these bias differences are not merely reflections of differing background structures between the two groups.

Regarding the first question, if the biases were merely a general "intensity bias", the resulting error pattern would manifest largely as a symmetric, near-centre bias. This symmetric pattern is indeed what we primarily observe in the weak heating group (although, as weak heating is not zero heating, it still exhibits a slight underestimation of propagation speed). With stronger heating, distinct asymmetric bias patterns are observed in the strong heating group. To address the second question and to isolate the role of latent heat release from background environmental differences, we conducted an additional analysis. We compared the forecast biases of strong and weak heating cyclones restricted to a fixed latitudinal band, thereby accounting for the climatological background state. As shown in Fig. AR1, even when the latitudes are constrained, the two groups continue to exhibit distinctly different bias patterns. Specifically, the strong heating group maintains its pronounced south-westward displacement bias, with the area of statistical significance ($p < 0.05$, indicated by stippling) clearly covering this south-westward region. In contrast, the weak heating group within the same latitudinal band only shows a slight westward bias. This difference confirms that diabatic heating acts as an independent factor for the forecast biases, rather than unequal sample sizes, random synoptic noise, or latitudinal background effects.

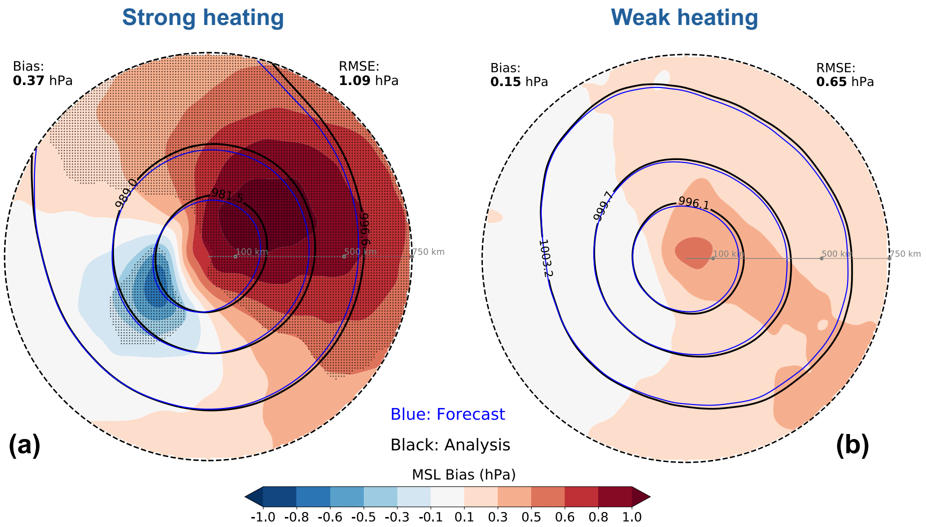

Figure AR1: As in Figure 2, but only for latitude 45°–55°N

Other non-minor issues:

1) **Section 2.1.: It remains unclear to me what the forecast data is that is examined by the authors.**

We have restructured this paragraph in section 2.1 to explicitly state that the forecast data we examine are the short-range forecasts that provide the background fields for the ERA5 4D-Var data assimilation, rather than external operational forecasts, as follows:

'We perform our analysis over the North Atlantic using the European Centre for Medium Range Weather Forecasting (ECMWF) ERA5 reanalysis for the period from 1979 to 2022 at a spatial resolution of 0.5° × 0.5° for DJF (December, January, February) (Hersbach et al., 2020). Following Yu et al. (2025), we calculate 12-hour forecast errors using the analyses and respective 12-hour forecasts. Specifically, the forecast data utilised here serve as the background fields (first-guess trajectories) for the ERA5 4D-Var data assimilation. We select the short-range forecasts initialised at 0600 and 1800 UTC because they provide the background for the subsequent 0900–2100 UTC and 2100–0900 UTC assimilation windows, respectively, ensuring close consistency between the forecasts and analyses. Such close consistency is desirable, as the differences between these internally coupled forecasts and analyses can serve as a robust indicator of the model's physical realism (Rodwell and Palmer, 2007).'

2) **Confusion of (unbiased) forecast errors vs. bias and error source vs. amplification: To be clear upfront: I have no doubt that (mis)representation of diabatic processes induces forecast biases. The theme of this study are biases. The introduction, however, also discusses the role of (moist) diabatic processes in the amplification of (mostly unbiased) forecast errors. This discussion can be interpreted such that this contribution to error growth is due to deficiencies in the representation, i.e., the model's parameterizations of the diabatic processes. By no means, however, such deficiencies need to dominate or even make a prominent contribution to error growth. Cyclone amplification with strong latent heat release is a highly nonlinear process and one may expect error growth in such a situation even with perfect representation of latent heat release, as e.g., seen in spread growth in ensembles or in perfect model experiments. Similarly, the introduction contains a discussion of diabatic processes as sources of forecast errors, which states that "forecast biases and errors in extratropical cyclones have increasingly been attributed to deficiencies in the representation of diabatic processes" (L44ff). At least two of the references cited by the authors to this end (Lamberson et al. 2016 and Pickl et al. 2023) conclude the opposite in their abstracts: Lamberson et al. attribute the main differences in the cyclone evolution to initial condition uncertainty in the upstream trough, and so do Pickl et al. (Berman and Torn (2019, 2022(?)) found similar results). In addition, Pickl et al. consider latent heat release in warm conveyor belts as amplifier of forecast uncertainty – as diagnosed in an ensemble – hence this amplification does not rely on "deficiencies in the representation of diabatic processes". In L173 the authors note that "This bias might be attributable to error sources from both microphysics and dynamics." and subsequently attempt to disentangle the distributions. While disentangling is very difficult, the study would benefit at least from a clearer discussion of the background knowledge in the introduction.**

Thanks for pointing this out. WCBs are important for NWP in two main ways: The microphysical processes they involve can have a strong impact on the larger-scale dynamics and they are a major source and magnifier of forecast uncertainty. Because our study focuses on the 12-hour deterministic short-range forecasts to diagnose model deficiencies, it is indeed inappropriate to cite ensemble-based chaotic amplification studies (like Pickl et al. and Lamberson et al.).

We now realise that the unclear content in the introduction makes the main purpose of this study not obvious. This may also be the reason for your first concern regarding whether we are observing predictability limits or actual model biases. We have rewritten this section of the introduction with appropriate citations, as follows:

'Given the aforementioned systematic biases in ETCs forecasts, we extend the study by Yu et al. (2025) by employing a cyclone-centred composite framework to quantify short-term (12-hour) forecast biases for wintertime maritime ETCs categorised into strong and weak diabatic heating. We select the 12-hour forecast lead time, as error growth during this initial period is predominantly dominated by diabatic processes (Baumgart et al., 2019), making a comparison with the respective analysis highly effective in isolating 'fast-physics' errors in NWP models (Xie et al., 2012; Klocke and Rodwell, 2014). We focus on wintertime maritime ETCs, as their intensification and structural evolution are critically influenced by diabatic processes (Hoskins and Hodges, 2002; Joos and Forbes, 2016). By applying a composite approach, we provide a statistically aggregated view of how different diabatic conditions yield systematic structures in forecast biases. The objective of this study is to identify the link between the intensity of diabatic heating and 12-hour ETC forecast biases. Ultimately, this study helps to understand the shortcomings in representing diabatic moist processes and their related dynamics, which can guide future model improvements.'

Minor comments:

- **L25: Is there a distinction between systematic forecast errors and biases? I suggest rephrasing, e.g., leaving out „systematic“**

  We thank the reviewer for pointing this out. We have rephrased the sentence by removing "systematic" to avoid any tautology.

- **L35: I suggest adding a reference to Davis, Stoelinga, and Kuo 1993, which in my view is a seminal paper on quantitative diagnostic of the impact of latent heat release on cyclones.**

  We have added the reference: Davis, Stoelinga, and Kuo (1993).

- **End of introduction: I'd appreciate a brief outline of the paper here.**

  We believe that the structure of the paper is well reflected in the section headings and thus refrain from such an outlining paragraph at the end of the introduction.

- **L105: It is not obvious to me that the distribution of mesoscale features aligns with the direction of motion. Can you provide a brief explanation/ illustration. Is this a heuristic or based on theory/ conceptual understanding?**

  We have expanded the explanation to state that because cyclones propagate in various directions, rotation is crucial to reduce smoothing of spatial features during the averaging process. The revised text now reads: 'Following (Catto et al., 2010), cyclone-relative composites were calculated by averaging over all cyclones. During this procedure, all fields are rotated to align the propagation direction of each cyclone along the x axis. As individual cyclones move in diverse directions, this rotation is essential to prevent misalignment of features, such as warm and cold fronts, that would then be smoothed out during the averaging process.'

- **L123: I do not understand why the *analysis* cyclone needs to be re-centred. Please explain.**

  The analysis cyclone needs to be re-centred because its original centre is identified by a tracking algorithm based on Laplacian of MSLP. This algorithm requires a continuous time series covering the cyclone's full lifecycle.

As it cannot be applied to 12-hour forecasts, we use the MSLP minimum to define the forecast cyclone centre. Furthermore, in developing ETCs, the Laplacian of MSLP is frequently displaced from the MSLP minimum. To ensure a direct comparison between the analysis and the forecast, we hence also re-centre the analysis to its corresponding MSLP minimum.

- **L128ff: More than 20% of the data is lost for the weak heating group but less than 5% for the strong heating group. Why is this difference so large? On a related note: Did the authors consider also an EOF/ PCA approach to better understand the modes of variability of the systematic errors/ biases. I recommend considering such an approach, which can be applied without loss of data.**

  Regarding the difference in data loss: Our composite analysis focuses specifically on the timestep of maximum intensification. At this stage, strong heating cyclones experience intense low-level vortex stretching driven by diabatic heating, which rapidly deepens the local pressure centre directly beneath/adjacent to the heating region. This makes the kinematic centre (Laplacian) and the pressure minimum close to each other.

  In contrast, without this intense localised diabatic deepening, weakly heating cyclones are driven primarily by broader baroclinic forcing. The kinematic centre (tracked along the highly sheared frontal zones) has more chance of displacing from the broader, flatter pressure minimum. When this spatial change exceeds our criteria of a 250 km search radius, the data is excluded. Therefore, the 20% rate is not a loss of valid data but a necessary quality control step to ensure the physical validity of our subsequent composite analysis. We have added a brief explanation to the text: 'A smaller proportion of strong heating cyclones is filtered out, which physically reflects that their kinematic centres and MSLP minima are more consistently co-located.'

  Regarding the use of EOF/PCA: We agree that EOF/PCA is a powerful tool for exploring modes of variability. However, it is not the best tool for this specific analysis. The main reason is as follows:

  Mean bias vs. maximum variance: EOF extracts modes that maximise total variance within the dataset. Our goal, however, is not to find the dominant modes of general forecast variability but specifically to extract the mean bias explicitly tied to a physical forcing (diabatic heating). Composite analysis conditioned on heating distributions directly isolates this mean signal, which is a more appropriate method for our specific research question.

- **L138: I would not say intensity underestimation is restricted. Rather, the maximum of underestimation is in the center, whereas the maximum of underestimation in the strong-heating group is at larger radii, ahead and left of motion.**

  We have rewritten this paragraph. It now reads: 'Cyclone-centred MSLP biases reveal an underestimation of cyclone intensity, with structural differences for the two groups (Fig.3a, b). In the strong heating group, a broad area of positive biases is evident (Fig.3a, shading), indicating that the forecasted cyclones are generally too shallow. This intensity underestimation is also visible in the forecasted MSLP contours (blue) that exhibit a smaller radius compared to the analysis (black). Furthermore, the spatial location of the maximum underestimation in the strong heating group is in the upper right-hand quadrant of the composite centre. In contrast, for the weak heating group, the intensity underestimation is mainly around the cyclone centre (Fig.3b).'

- **L149: I do not understand this potential explanation. Please clarify.**

  We have added more information for clarity, as follows: 'One potential explanation for this overestimation is that the stronger forecasted winds may actually be closer to reality. The incremental 4D-Var assimilation system in ERA5 computes analysis increments at a reduced spatial resolution (Hersbach et al., 2020; ECMWF, 2016). Hence, highly localised sharp wind gradients are often smoothed out. This is consistent with previous findings that marine CCB jets are underestimated in ERA5 (Gentile and Gray, 2023).'

- **L187ff: I do not understand why this is consistent. Please expand your argument for clarity.**

  We have updated this paragraph as follows: 'Given that vertical velocity in the analysis is diagnosed and not directly observed, we employ the kinematic frontogenesis of potential temperature (θ) as a proxy, as it quantifies the dynamic forcing that drives a secondary vertical motion (Sawyer, 1956; Eliassen, 1962). Pronounced frontogenesis at 850 hPa is evident in the upper left- and right-hand quadrants of the cyclone centre (Fig. 4d). The frontogenesis positioned closer to the cyclone centre is associated with a positive bias in frontogenesis near the centre ($\sim 17\%$). As latent heating strongly influences cyclone frontal structure (Martínez-Alvarado et al., 2014), this bias reflects a feedback between the dynamics and diabatic processes. In accordance with thermal wind balance, the resulting enhanced temperature gradient corresponds to a stronger vertical shear of the geostrophic wind, consistent with the wind speed overestimation (Fig.3c). The intensified frontogenesis is also associated with a secondary circulation that produces enhanced vertical motion on the warm side of the front (not shown). This ascent can enhance condensation, resulting in an overestimation of liquid water along the bend back warm front (Fig.4c).'

References

Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, Nature, 525, 47–55, 2015.

Baumgart, M., Ghinassi, P., Wirth, V., Selz, T., Craig, G. C., and Riemer, M.: Quantitative view on the processes governing the upscale error growth up to the planetary scale using a stochastic convection scheme, Monthly Weather Review, 147, 1713–1731, 2019.

Catto, J. L., Shaffrey, L. C., and Hodges, K. I.: Can climate models capture the structure of extratropical cyclones?, Journal of Climate, 23, 1621–1635, 2010.

Dacre, H., Charlton-Perez, A., Driscoll, S., Gray, S., Harvey, B., Harvey, N., Hodges, K., Hunt, K., and Volonté, A.: Northern hemisphere midlatitude cyclone intensity biases in machine learning weather prediction models, Bulletin of the American Meteorological Society, 107, E208–E221, 2026.

ECMWF: IFS Documentation CY41R2 – Part II: Data Assimilation, Tech. rep., European Centre for Medium-Range Weather Forecasts, Reading, UK, URL `https://www.ecmwf.int/en/elibrary/16648-part-ii-data-assimilation`, 2016.

Eliassen, A.: On the vertical circulation in frontal zones, Geofys. publ, 24, 147–160, 1962.

Gentile, E. S. and Gray, S. L.: Attribution of observed extreme marine wind speeds and associated hazards to midlatitude cyclone conveyor belt jets near the British Isles, International Journal of Climatology, 43, 2735–2753, 2023.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, 2020.

Hoskins, B. J. and Hodges, K. I.: New perspectives on the Northern Hemisphere winter storm tracks, Journal of the Atmospheric Sciences, 59, 1041–1061, 2002.

Joos, H. and Forbes, R. M.: Impact of different IFS microphysics on a warm conveyor belt and the downstream flow evolution, Quarterly Journal of the Royal Meteorological Society, 142, 2727–2739, 2016.

Klocke, D. and Rodwell, M.: A comparison of two numerical weather prediction methods for diagnosing fast-physics errors in climate models, Quarterly Journal of the Royal Meteorological Society, 140, 517–524, 2014.

Martínez-Alvarado, O., Joos, H., Chagnon, J., Boettcher, M., Gray, S., Plant, R., Methven, J., and Wernli, H.: The dichotomous structure of the warm conveyor belt, Quarterly Journal of the Royal Meteorological Society, 140, 1809–1824, 2014.

Rodwell, M. and Palmer, T.: Using numerical weather prediction to assess climate models, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 133, 129–146, 2007.

Sawyer, J. S.: The vertical circulation at meteorological fronts and its relation to frontogenesis, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 234, 346–362, 1956.

Xie, S. et al.: On the correspondence between short- and long-timescale systematic errors in CAM4/CAM5, Journal of Climate, 25, 7937–7955, 2012.

Yu, Q., Spensberger, C., Magnusson, L., and Spengler, T.: Forecast Errors Attributed to Synoptic Features, Meteorological Applications, 32, e70 093, 2025.