

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.

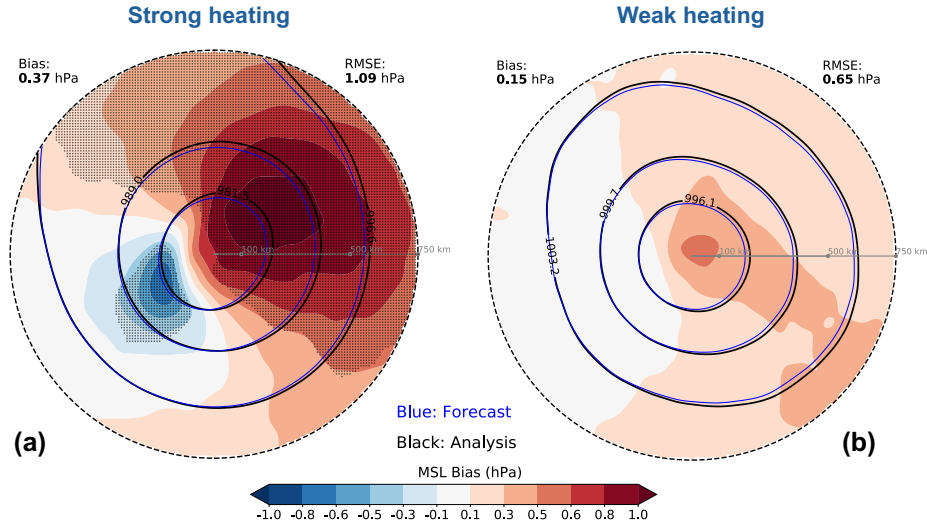


Figure AR1: As in Figure 2, but only for latitude 45° – 55° N

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 $0.5^{\circ} \times 0.5^{\circ}$ 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/e-library/16648-part-ii-data-assimilation>, 2016.
- Eliassen, A.: On the vertical circulation in frontal zones, *Geophys. 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.