

Response to Reviewer 2

This manuscript provides a concise overview of errors in the 12-hour forecast from the IFS and AIFS for a subset of cyclones in the extratropical storm tracks of the Northern and Southern hemisphere. They do a good job of identifying a series of variables linked to cyclone intensity and generally to diabatic heating. The results are concise and clear—though perhaps to the point of missing opportunities to expand their explanations or depth of work (which would broaden the impact of the paper). I’ve summarized the comments below as a combination of major and minor comments.

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).

Major comments:

- 1) **Depth of referenced literature:** In general, I felt like the authors narrowed their literature review/connection to prior work more than I would like. For example, there’s a large body of work that has explored the impacts of diabatic heating on forecast errors – I would suggest expanding the citations here beyond those of the author(s). In addition, I some of the issues in, for example, boundary layer representation between a physics model and AI model have been explored, even if with other models.

We thank the reviewer for this suggestion. We have expanded the discussion of diabatic heating and forecast errors in the introduction with references to Pickl et al. (2023) and Berman and Torn (2022), and we have added a reference to Charlton-Perez et al. (2024) regarding the underestimation of surface wind speeds by machine-learning models.

- 2) **Attribution to diabatic heating:** Correct me if I’m wrong, but can’t you get the explicit diabatic heating terms from the IFS?

I know this isn't available for the AIFS, but some of your leaps (attributing variables to physical processes) could be shored up by showing some of these terms at least for the IFS.

The operational IFS output only provides the total temperature tendency from parametrised processes, without separating the radiative contribution, which is why we base our classification on ERA5 (added in section 2.3). To examine the process from additional perspectives, we have added analyses of the biases in vertical velocity and specific humidity, which are closely linked to diabatic heating (section 3.5). Based on these results, we have revised the interpretation of the temperature biases in section 3.5.

- 3) Extent of results: I understand (and appreciate) the narrow focus of this work (on 12-hour biases). However, the role of diabatic heating can be an integrated process (to an extent) on the atmosphere, and plays a distinct role in the temporal evolution of systems (eg. feedbacks on cyclone growth via baroclinic processes). It may be beyond what the authors can do, but given the brevity of the manuscript it feels like there's an opportunity to look at what these errors look like either at longer forecast lead times, or at different steps in the cyclone evolution.**

We thank the reviewer for this suggestion. We focus on the 12-hour lead time because, at longer lead times, the forecast errors are increasingly affected by the limited predictability of the atmosphere and a mismatch between the detected cyclones. This makes it more difficult to attribute the biases to the specific processes and characteristic structures within the cyclone.

Regarding the different time steps of the cyclone evolution, Yu et al. (2026) examined the forecast errors at multiple stages along the cyclone tracks and found them to be consistently largest at the time of maximum intensification. We therefore focus on this stage, which also allows a direct comparison between the two studies. We have added a statement of this in section 2.2.

Minor comments:

- 1) L20-21: Please add a 'by' between research and Yu.**

Corrected as suggested.

- 2) L23-25: This last sentence is worded a bit awkwardly (particularly at the end ‘...investigate the forecast performance for ETCs of the ...’) – please consider reworking.**

We have reworded this sentence.

- 3) L27: Please rework as ‘Joos and Forbes (2016) and Binder et al. (2016)’**

Corrected as suggested.

- 4) L41: Please remove the ‘Ben’ from your citation.**

Thank you for pointing this out. We have double-checked the reference. "Ben" is part of the author's surname ("Ben Bouallègue") as it appears in the published article.

- 5) L47-48: The sentence is unclear/has some grammar issues as written – please aim to rework.**

We have reworded this sentence for clarity: "whether" has been replaced by "the", and "associated to" corrected to "associated with".

- 6) L49: ERA5?**

Corrected.

- 7) L75 and elsewhere: You often are missing using a definite article (eg. ‘the’) in front of IFS or AIFS (eg. it should be ‘the AIFS’ or ‘the IFS’).

We have gone through the manuscript and now use ‘the’ consistently when referring to the IFS and the AIFS.

- 8) L82-83: You may also want to note that you appear to have employed a land mask within these domains (based on Fig. 1).

We have clarified this in Sect. 2.2.

- 9) Section 2.3: Given how new the Yu et al. 2026 methodology is, I would aim to add more detail on how the method works here. Though this may have been addressed in the other manuscript (if so, you may want to say so), it would be good to elaborate on why you focused on just the amount of diabatic heating compared to, say, anomalous diabatic heating relative to the background climatology (which would likely remove the latitudinal bifurcation of the majority of your cases). I know you get at this a bit with L92-95, but I think a bit more justification/reasoning could be applied here.

We have added more detail in the methods in sections 2.2 and 2.3, including the tracking and selection criteria, the computation of the diabatic heating, as well as the percentile thresholds and sample sizes. We also note that Yu et al. (2026) has now been published and is available to readers.

Regarding the heating: our interest is in how the diabatic heating affects the cyclone structure and the associated forecast errors, which depends on the actual amount of heating rather than on its anomaly relative to the local climatology. The percentiles are also computed independently for each basin, which already accounts for part of the regional differences in background moisture. To verify that the classification does not simply reflect a latitudinal separation, we restricted both groups to 45°–55° and found the distinct bias patterns between the two heating groups still exist (see Fig. AR1).

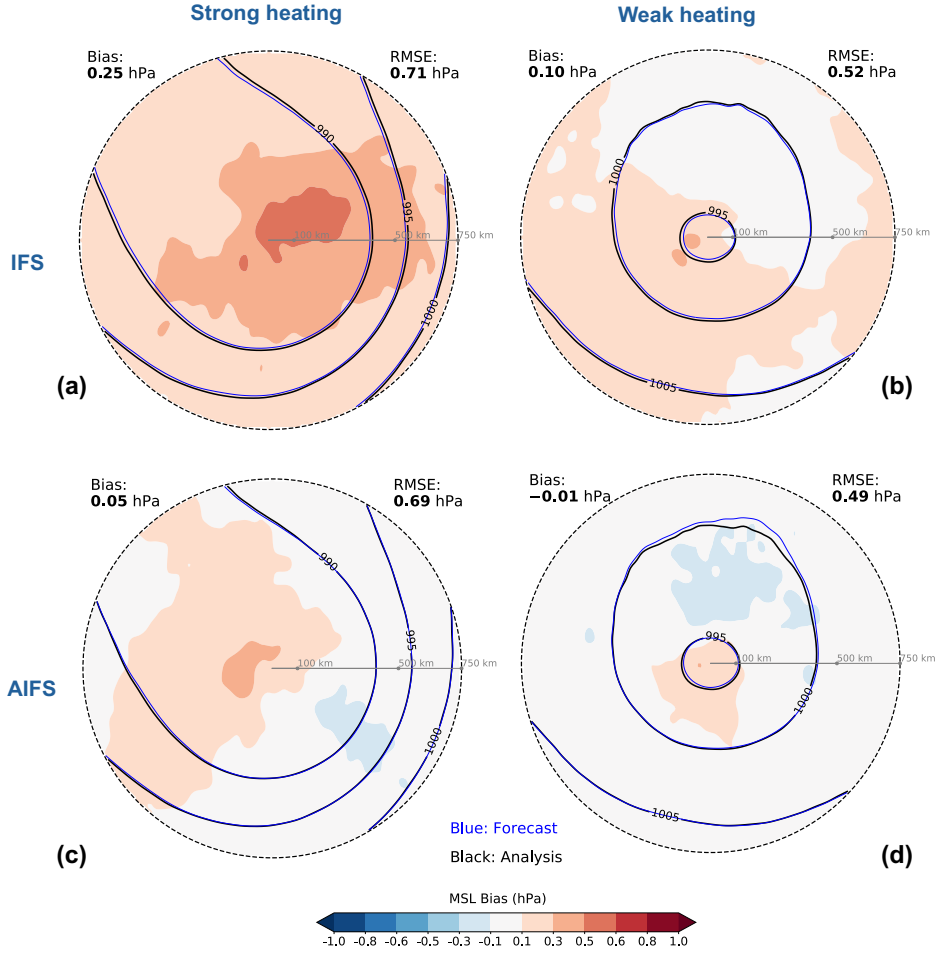


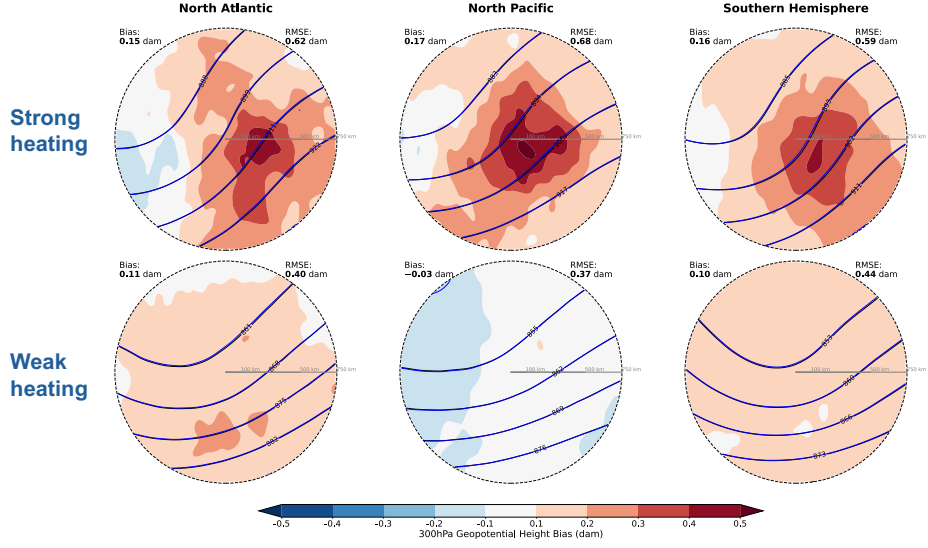
Figure AR1: As in Figure 2, but only for latitudes 45° – 55° N and 45° – 55° S.

- 10) L100-104: Understood on the biases being similar amongst the three basins. It may not be relevant, but are the impacts of these heating biases on the jet (and thus the baroclinic evolution of the system) the same? The North Atlantic jet tends to operate more as an eddy-driven jet, while the North Pacific and Southern Ocean jets tend to be more a merged eddy-driven and subtropical jet. Does this influence the interpretation of the results given the time evolution aspect of your study?

We agree that the North Atlantic jet is predominantly eddy-driven, whereas

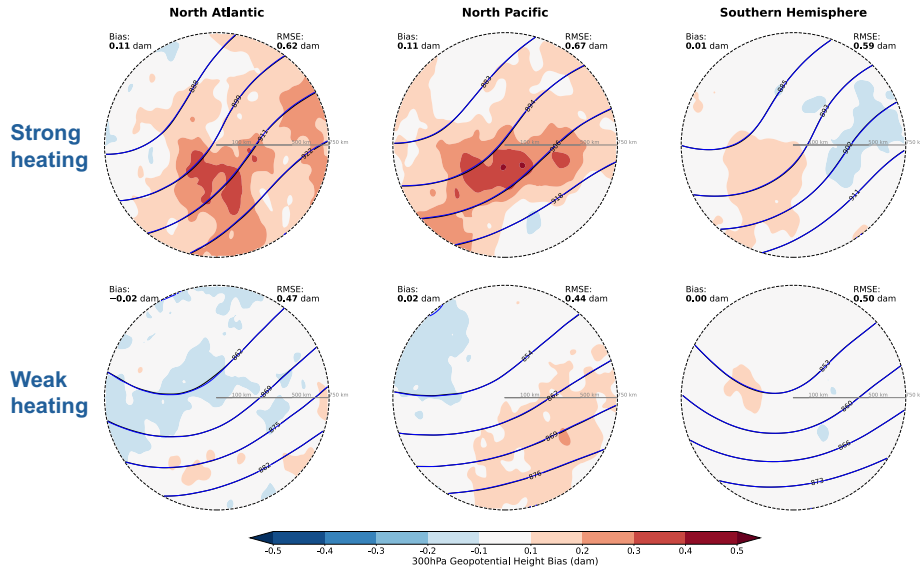
the North Pacific and Southern Ocean jets are more often a merged eddy-driven and subtropical jet. However, the modulation of the baroclinic evolution by the jet type operates through eddy-jet feedbacks on timescales of one to several days, which is beyond the 12-hour window analysed here. Our study is restricted to a single time step and does not address the temporal evolution of the cyclones, so this aspect does not enter the interpretation of our results.

Figs. AR2 and AR3 show similar bias patterns across the three basins, which means that the differences in jet structure between the basins are not reflected in the upper-level biases at this lead time. We agree that the influence of the jet type would be an interesting question at longer lead times, and note it as an avenue for future work.



2

3 Figure AR2: As in Figure 8, but shown separately for individual ocean basins for
4 IFS.



5

6 Figure AR3: As in Figure 8, but shown separately for individual ocean basins for
7 AIFS

- 11) L108-109: While I agree the warm sector is likely in this vicinity, many oceanic storms at peak intensity have also started to undergo occlusion or seclusion processes. Is it possible that this could also be a residual of the diabatic heating process from an earlier warm sector that has since been displaced? Or is this analysis based off your figures 5 and 6?

We have removed the inaccurate label "warm sector" and replaced it with a more accurate description, based on Figs. 5 and 6. This region corresponds to the tight temperature and moisture gradients. The wording in the manuscript has been updated accordingly.

- 12) In general, when referencing your quadrant diagrams – please keep in mind that a quadrant can only be $\frac{1}{4}$ of your domain. For example, in section 3.2 you discuss the ‘right quadrant’ – this should either be written as ‘right quadrants’ or ‘right hemisphere’.

Thank you for pointing this out. We have gone through the manuscript and corrected all references to quadrants, using either the specific quadrant where the whole half of the domain is meant.

- 13) L130-131: Is the PGF really weakened? Yes there are pressure differences, but it's unclear that the pressure gradient itself is reduced (for example, the blue lines are all biased 'inside' of the black lines, but the horizontal distance between your blue isobars and black isobars doesn't appear to be different).**

Thanks for pointing this out. A positive MSLP bias does not by itself imply a reduced pressure gradient, and our decomposition in Fig. 4 shows that the geostrophic wind speed, which is determined by the pressure gradient, is well captured by the IFS (regression slope 0.99, correlation 0.994). We have removed the attribution to a weakened pressure gradient and now only note that the wind speed underestimation occurs in the same region as the positive MSLP bias. The interpretation of the 10 m wind speed biases has been modified according to the updated decomposition results.

- 14) L140-145: This is a nice summary of the issue at hand (r.e. boundary layer representation). Maybe it's beyond the scope of the project, but can you elaborate more on the differences between the two models for the boundary layer? Are the other studies that have looked at boundary layer processes between physics models and AI models?**

We thank the reviewer for this suggestion. We have added a description of how the two models differ in obtaining the near-surface wind at the paragraph, with the corresponding references.

Regarding existing work, few studies have specifically compared boundary-layer processes between physics-based and data-driven models for extratropical cyclones. We have added references in the introduction to studies reporting an underestimation of near-surface wind speeds in machine-learning models. However, this underestimation has not previously been examined by decomposing the near-surface wind into its geostrophic and ageostrophic components, in particular for the AIFS.

- 15) **L177:** Have you also explored biases in vertical velocity for these event? While you are generally diagnosing diabatic heating, you need vertical motion (and condensation) to unlock the heating process. Your diagnosis here captures the moisture availability, but not the vertical velocity.

We thank the reviewer for this comment. We have now examined the vertical velocity biases at 850, 700 and 500 hPa, together with the specific humidity biases at the same levels, and have added these to the supplement. We have rewritten this part of the section accordingly.

In the IFS strong heating group, the ascent in the warm sector is underestimated at 850 and 700 hPa, co-located with the warm bias, while the specific humidity shows no clear bias there. Since ascent drives the latent heat release, but also cools the air adiabatically, a weaker ascent reduces both, and the sign of the temperature bias depends on which reduction is larger. The warm bias at 850 hPa indicates that the reduced adiabatic cooling dominates, whereas at 500 hPa a band of cold bias coincides with a band of underestimated ascent, where the reduced latent heating appears to dominate.

- 16) **Section 3.6:** The looks more like a position/tilt bias rather than an intensity of trough bias, in particular for the IFS. Also, wouldn't positive (negative) biases in diabatic heating result in positive (negative) biases in geopotential heights (through a variety of arguments – thickness, thermal wind, QG, PV, etc.) – maybe link into this here as well?

A displaced or tilted trough would produce a dipolar bias across the trough axis. In Fig. 8a the bias is positive throughout the domain, with the largest values confined to a closed maximum around the trough, with the same at 500 hPa (Fig. S12). We therefore interpret it as an underestimation of the trough amplitude and have made the basis for this explicit in the revised text.

Regarding the diabatic heating, we do not evaluate its bias: it is taken from ERA5 and used only to group the cyclones, for the reasons now given in Sect. 2.3. Our revised analysis in Sect. 3.5 also shows that the temperature bias in the ascending region is the residual of the adiabatic cooling and the diabatic heating, so its sign does not indicate the sign of a heating bias. What

our results do show is that the underestimation of the upper-level trough is consistent with the underestimation of the surface cyclone intensity.

References

- Berman, J. D. and Torn, R. D.: The Sensitivity of Downstream Ridge Building Forecasts to Upstream Warm Conveyor Belt Forecast Uncertainty Using MPAS, *Monthly Weather Review*, 150, 2573–2592, <https://doi.org/10.1175/MWR-D-21-0048.1>, 2022.
- Charlton-Perez, A. J., Dacre, H. F., Driscoll, S., Gray, S. L., Harvey, B., Harvey, N. J., Hunt, K. M., Lee, R. W., Swaminathan, R., Vandaele, R., et al.: Do AI models produce better weather forecasts than physics-based models? A quantitative evaluation case study of Storm Ciarán, *npj Climate and Atmospheric Science*, 7, 93, <https://doi.org/10.1038/s41612-024-00638-w>, 2024.
- Pickl, M., Quinting, J. F., and Grams, C. M.: Warm conveyor belts as amplifiers of forecast uncertainty, *Quarterly Journal of the Royal Meteorological Society*, 149, 3064–3085, <https://doi.org/10.1002/qj.4546>, 2023.
- Yu, Q., Spensberger, C., Magnusson, L., and Spengler, T.: Distinct bias structures for extratropical cyclones with strong or weak diabatic heating, *Weather and Climate Dynamics*, 7, 1117–1131, <https://doi.org/10.5194/wcd-7-1117-2026>, 2026.