

Response to Reviewer 1

Yu et al. (2026) examines the short-term forecast errors for ETCs in the updated ECMWF IFS and AIFS for strong-diabatic and weak-diabatic events. Overall, the authors find smaller biases in the IFS compared to the previous version for a number of chosen metrics, and largely smaller biases for AIFS compared to IFS. Generally speaking, strong heating cases show larger biases compared to weak heating cases.

This study is well motivated by the impactful nature of ETCs and does well to build on previous work by the authors by expanding the study regions and using the most current forecast systems, including an ML model. Several aspects of the methods and results, however, need additional justification, clarification, and/or quantification before recommending publication in *Weather and Climate Dynamics*.

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).

General Comments

- 1) Font sizes across all figures could be larger, especially for Fig. 4 and for reporting the biases/RMSEs in the other figures.

We have updated all related figures.

- 2) The differences in biases between strong and weak heating cases are discussed but more could be said throughout the manuscript comparing the patterns themselves between strong and weak heating cases. There are some pretty stark differences in certain fields (e.g., 850–300-hPa temperature difference, 300-hPa heights) that warrant a bit more attention.

We thank the reviewer for this suggestion. We have extended the comparison

between the two heating groups for both fields highlighted by the reviewer. For the 850–500 hPa temperature difference (section 3.5), we now describe the vertical structure of the biases in each group by adding the specific humidity and vertical velocity biases at 850–500 hPa and have rewritten the text accordingly.

For the 300 hPa geopotential height (section 3.6), we have revised the description of the bias patterns in the two groups and added the case-by-case distributions of the biases and RMSEs.

3) For all supplemental figures, should “HRES” be “IFS”?

Updated.

Specific Comments:

- **L8:** Worth specifying here that “ERA5” in this context refers to the previous version of IFS used in ERA5 rather than the ERA5 reanalysis. This made more sense after reading the full paper but readers may initially be confused.

We have added ‘the earlier, coarser-resolution IFS cycle used to produce ERA5’ in the abstract.

- **L19:** Another reference to add here is Büeler and Pfahl (2017): <https://doi.org/10.1175/JAS-D-17-0041.1>

Added as suggested.

- **L24:** Need to define AIFS here.

Added as suggested.

- **L25:** Please include references here for IFS and AIFS.

Added.

- **L43:** Can remove AIFS definition since acronym used already in L24.

Removed.

- **L47–48:** This sentence reads a bit strange. Perhaps “whether” should be “the” instead?

We have replaced "whether" with "the" as suggested, and additionally corrected "associated to" to "associated with".

- **L49:** Should “ERA” be “ERA5”?

Corrected to "ERA5".

- **L61–62:** What is the temporal resolution for forecast output?

We have clarified this in Sect. 2.1: "...for each initialisation at 00 and 12 UTC, we use the single model output at a lead time of 12 hours...".

- **L65:** Need to define DJF and JJA.

Added.

- **L67–68:** Please provide additional context for choices of forecast times and variables. In particular, I’m curious why 300-hPa geopotential height is preferred over 500-hPa height and why jet-level winds aren’t considered.

Previous studies have shown that diabatic heating in extratropical cyclones affects the upper-level flow (e.g. Martínez-Alvarado et al., 2016; Binder et al., 2016), with Yu et al. (2026) identifying corresponding biases in ERA5 at 300 hPa. In this study, we examined geopotential heights at both 300 and 500 hPa (500 hPa now in Fig. S12) and found consistent bias patterns, with neither level providing additional information. As this study is an extension of Yu et al. (2026), we select the 300 hPa geopotential height in the main text for better comparison.

We also examined the 300 hPa wind field. The results only show wind speed biases consistent with the underestimated trough biases, which does not add information beyond the geopotential height, so we did not include it.

We have rephrased section 2.1 accordingly to make this clearer.

- **L69:** What other forecast verification or bias quantification metrics were considered and how do they compare to RMSE?

The composite bias and the RMSE serve complementary purposes in our framework: the composite bias shows the systematic component of the error, whereas the RMSE quantifies the total error magnitude. The comparison between the two allows us to distinguish systematic biases from case-to-case variability, which we further examine through the case-by-case distributions of bias and RMSE. Following the reviewer’s suggestion, we have additionally added the correlation coefficients and linear regressions to Fig. 4. We have also added a corresponding clarification to section 2.1.

- **L80–81:** If possible, please include a table of these tracking criteria in supplemental material.

We have added tracking and selection criteria to section 2.2.

- **L85–86:** Please include the classification criteria/threshold for strong vs. weak heating here. Also, what are the sample sizes for each classification?

We have added method details about classification criteria and sample sizes to section 2.3.

- **L87:** I assume the diabatic heating is a direct output of IFS and AIFS? If so, are there any limitations to how it’s computed in the models?

The diabatic heating is derived from ERA5 rather than from IFS or AIFS. We have clarified this, including the computation and our reasons for this choice, in section 2.3.

- **L87:** Is there any sensitivity to this radius selection?

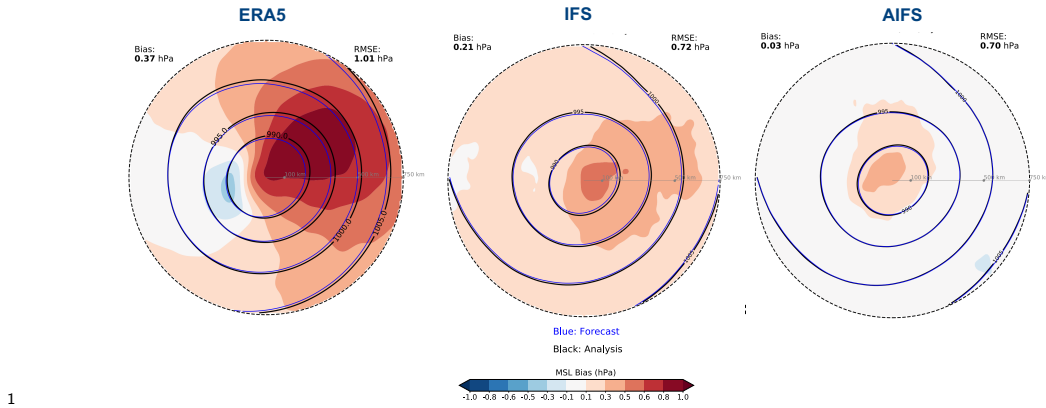
We tested radii of 500, 750, and 1000 km in the early stage of our previous study (Yu et al., 2026), and found that the main patterns were not sensitive to this choice. We retain 750 km for two reasons. First, it is large enough to include most cyclone-related features, such as the frontal structures and the associated regions of latent heat release. Second, as the present study is an extension of Yu et al. (2026), we keep the analysis consistent with that study. We have also added the corresponding explanation to Section 2.3.

- **L98–99:** Can you clarify how large the propagation biases are in ERA5 vs. IFS and AIFS as compared to the reanalysis? Visually, the propagation biases in Fig. S1–S3 arguably aren’t drastically different across the three forecast datasets. The intensity biases, however, are substantially reduced in IFS and AIFS compared to ERA5.

We apologise for the unclear description, which caused this misunderstanding. The comparison we refer to is between Fig. 2a of Yu et al. (2026) and

Fig. 2a, b of this study. For clarity, we show the corresponding panels side by side in Fig.AR1, taking the strong heating group as an example.

In ERA5 (left), the MSLP bias has a clear dipole structure, with negative values west and strongly positive values east of the cyclone centre, which is implying a displacement of the cyclone (intensity bias as well). In the IFS (middle) this dipole is absent, and the bias is a weaker, domain-wide positive bias. In the AIFS (right) the bias is even smaller. We have rewritten the sentence accordingly.



1
2 Figure AR1: Composite MSLP biases for the strong heating group in ERA5 (left;
3 Fig. 2a of Yu et al. (2026)), the IFS (middle; Fig.2a of this study), and the AIFS
4 (right; Fig.2c of this study).

- **L109–110: Can you quantify the magnitude of this positional shift?**
It's very difficult to estimate from the figure.

We apologise for the imprecise wording. In contrast to ERA5 in Fig. 2a,b of Yu et al. (2026), the forecast error in the IFS is dominated by an underestimation of the intensity rather than by a positional error. We have rewritten this part accordingly.

- **L112: Should “propagation bias ERA5” be “propagation bias of ERA5”?**

Corrected to "the propagation bias of ERA5".

- **L113: Please quantify (for propagation speed) and clarify these improvements. For example, the average bias and RMSE for IFS in Fig. 2a is larger than for ERA5 in Fig. S1a (though the color shading indicates the intensity bias is noticeably weaker).**

We apologise for the unclear description. The comparison refers to Fig. 2a,b of Yu et al. (2026), not to Fig. S1, which shows the IFS bias verified against ERA5 and is included only to justify our choice of verification dataset.

The improvement in intensity is evident in Fig. AR1, where both the bias and the RMSE decrease from ERA5 to the IFS and further to the AIFS. We have also removed the reference to propagation speed, as discussed above, and rewritten the sentence accordingly.

- **L115: I'm finding it difficult to directly compare Fig. S1–S3 to Fig. 2 given the difference in baseline analysis (i.e., ERA5 reanalysis vs. IFS analysis). How do the basin-specific forecast differences compare when using the IFS analysis as reference?**

We have added the basin-specific composites verified against the IFS analysis for MSLP(Figs. S16,S17), and updated the corresponding text and citations in the manuscript.

- **Fig. 2 (and others): What are the contour values for the blue contours?**

We have clarified this in the caption of Fig. 2, which applies to the subsequent figures as well.

- **Fig. 3: Can you increase the quiver density a bit and add quivers for the forecast composite?**

Thank you for this comment.

The main purpose of Figure 3 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.

- **L143–145/Fig. 4:** Can you report the correlation coefficient and plot the linear regression lines (or some other relevant metric) to quantify (1) the match between forecast and analysis and (2) the difference in match strength between IFS and AIFS?

Thank you for the suggestion.

We have added the correlation coefficients and the linear regression lines to Fig. 4, and rewritten the corresponding part of Section 3.2.

- **L152:** Can you quantify? Narrower by how much? The difference in contour placement is difficult to estimate on the figure.

Thank you for pointing out that "narrower" was not an appropriate description here. What we intended to convey is that the forecast overestimates TCW within the high-moisture tongue while underestimating it elsewhere, most notably in the lower left quadrant. We have rephrased the sentence accordingly.

- **L153:** TCW acronym already defined.

Corrected.

- **L177:** Is this the primary explanation? How do the vertical differences in humidity compare to support this hypothesis? Could differences in trough depth also explain an upper-level cold bias?

We thank the reviewer for these questions. We have added the specific humidity and vertical velocity biases at 850, 700 and 500 hPa, and have revised the interpretation in Sect. 3.5 accordingly. The upper-level cold bias is now related to the corresponding band of underestimated ascent at 500 hPa rather than to the trough depth.

- **L184–185: Could you show this case-to-case variability in the supplemental figures like MSLP in Fig. S7?**

Thank you for this suggestion. We have added the case-by-case distributions for $\Delta T_{850-500}$ Fig. S8 and additionally for the 300 hPa geopotential height (Fig. S9) and revised section 3.5 and section 3.6 accordingly.

- **L223: Being in the gray zone, does the IFS use a convective parameterization scheme?**

Yes, the operational IFS Cycle 49r1 retains the mass-flux convection scheme of Tiedtke (1989), with subsequent ECMWF modifications (ECMWF, 2024). We have rephrased corresponding content in the manuscript.

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, <https://doi.org/10.1175/jas-d-15-0302.1>, 2016.
- ECMWF: IFS Documentation CY49R1 – Part IV: Physical Processes, IFS documentation, European Centre for Medium-Range Weather Forecasts, <https://www.ecmwf.int/en/elibrary/81626-ifs-documentation-cy49r1-part-iv-physical-processes>, 2024.
- Martínez-Alvarado, O., Madonna, E., Gray, S. L., and Joos, H.: A route to systematic error in forecasts of Rossby waves, *Quarterly Journal of the Royal Meteorological Society*, 142, 196–210, <https://doi.org/10.1002/qj.2645>, 2016.
- Tiedtke, M.: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models, *Monthly Weather Review*, 117, 1779–1800, [https://doi.org/10.1175/1520-0493\(1989\)117<1779:ACMFSF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2), 1989.
- 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.