https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

Forecast biases of extratropical cyclones classified by their diabatic heating intensity in operational physics-based and machine learning weather prediction models

Qidi Yu[1], Linus Magnusson[2], Clemens Spensberger[1], and Thomas Spengler[1]
[1]Geophysical Institute, University of Bergen, and Bjerknes Centre for Climate Research, Bergen, Norway
[2]European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom
Correspondence: Qidi Yu (qidi.yu@uib.no)

Abstract. Extratropical cyclones (ETCs) are strongly influenced by moist processes, rendering the impact of diabatic heating critical for forecast performance. We systematically evaluate short-range (12-hour) forecast biases of wintertime maritime ETCs over the North Atlantic, North Pacific, and Southern Ocean for the period 2023-2024. Employing a cyclone-centred composite framework, cyclones are categorised into strong and weak diabatic heating groups. We compare forecasts from the ECMWF high resolution operational 9-km Integrated Forecasting System (IFS) and the data-driven Artificial Intelligence Forecasting System (AIFS).

Both the higher resolution of IFS and the data-driven AIFS significantly reduce the ETC propagation biases previously identified in ERA5. However, both models exhibit more pronounced errors in cyclones with strong diabatic heating. In the physics-based IFS, while the previous severe dry and cold biases are largely improved, the model still underestimates cyclone intensity and warm sector wind speeds. Furthermore, IFS displays a distinct spiral-shaped positive bias in the 850-500 hPa temperature difference, suggesting a misrepresentation of the vertical distribution and depth of diabatic heating. AIFS generally yields similar but smaller biases in most fields compared to IFS. Despite these improvements, AIFS features a notable physical inconsistency as it demonstrates a weaker mean sea level pressure (MSLP) bias but a stronger 10 m wind bias compared to IFS. This is related to an underestimation of the near-surface ageostrophic wind speed.

1 Introduction

Extratropical cyclones (ETCs) play a fundamental role in transporting heat and energy from low to high latitudes (Shaw et al., 2016), and are frequently associated with high-impact weather events, such as heavy precipitation and strong surface winds (Catto et al., 2019). While the initial cyclogenesis is primarily driven by dry baroclinic instability, the subsequent rapid deepening is often closely related to moist diabatic processes (Pfahl et al., 2015; Binder et al., 2016). The accurate representation of diabatic heating in models is therefore critical for the forecast performance of ETCs. Previous research Yu et al. (2025) demonstrated that ETCs are the most pronounced contributors to forecast errors among several synoptic weather features. Subsequent, Yu et al. (2026) revealed that cyclones with strong diabatic heating exhibit distinct propagation and intensity biases. Given the significant model updates from the Integrated Forecasting System (IFS) cycle used for ERA5 and

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

the recent introduction of the data-driven model AIFS, we here investigate the forecast performance for ETCs of the currently state-of-the-art IFS and AIFS.

For physics-based models, improving the representation of diabatic heating in ETCs fundamentally depends on two critical factors: parameterisation schemes and resolution. Regarding parameterisations, Joos and Forbes (2016); Binder et al. (2016) found that the specific microphysical processes, as well as the horizontal and vertical distribution of the heating, are equally critical to the forecast of diabatic processes. Martínez-Alvarado et al. (2016b) found that ECMWF forecasts often underestimate the development of upper-level ridges, which is closely related to a southeastward bias in the predicted warm conveyor belt (WCB) outflow location. Similarly, Yu et al. (2026) found that cyclones with strong diabatic heating feature a clear underestimation in both upper level PV and ridge building. Previous studies have shown that higher resolutions can better capture the low-level PV intensification during rapid cyclogenesis caused by diabatic heating (Kuo and Low-Nam, 1990; Willison et al., 2013). Furthermore, Martínez-Alvarado et al. (2018) showed that high resolution is essential for better resolving the filamentary structures of the WCB outflow. Given that Yu et al. (2026) relied on ERA5, i.e., an earlier IFS cycle at coarser resolution, we investigate the impact of different diabatic heating on the forecast biases for ETCs in the updated 9-km IFS.

Unlike physics-based models, data-driven machine learning weather prediction (MLWP) models generate forecasts by learning directly from historical data, rather than numerically integrating explicit physical equations or relying on parameterisation schemes (Lam et al., 2023; Bi et al., 2023; Ben Bouallègue et al., 2024). MLWP has demonstrated comparable or superior performance to traditional NWP in many of the standard forecast scores, such as root-mean-square error (RMSE) and anomaly correlation coefficient (ACC) for geopotential height at 500 hPa (Lam et al., 2023; Lang et al., 2024; Ben Bouallègue et al., 2024). Because of the benefits of high computational efficiency and competitive forecast performance, ECMWF recently took the Artificial Intelligence Forecasting System (AIFS) into operation (Lang et al., 2024). The AIFS outperforms state-of-the-art physics-based models for many measures, including tropical cyclone tracks, with gains of up to 20% (Lang et al., 2024). However, the limited existing evaluations of AI models on ETCs reveal that they systematically underestimate cyclone intensity, specifically manifesting as weaker minimum sea level pressure and peak 10-m winds (Charlton-Perez et al., 2024; Dacre et al., 2026). Hence, we investigate whether forecast errors of AIFS associated to ETCs and distinguish them by the intensity of diabatic heating following Yu et al. (2026).

Yu et al. (2026) identified distinct bias patterns for ETCs in ERA when categorising them into strong and weak diabatic heating. Using a similar methodology, we investigate whether the updated parameterisations and high resolution of the operational 9-km ECMWF IFS can reduce these diabatic heating-related biases and yield better forecast performance and also compare them to the performance of the data-driven AIFS. Following Yu et al. (2026), we employ a cyclone-centred composite framework to quantify short-term (12-hour) forecast biases for groups of strong and weak diabatic heating. As demonstrated by previous studies, making a comparison with the respective analysis at such short lead times is highly effective in isolating 'fast-physics' errors in models (Xie et al., 2012; Klocke and Rodwell, 2014). This approach ensures that the identified biases are mainly dominated by diabatic processes, such as latent heat release (Baumgart et al., 2019). Besides, our study expands the analysis to not only include the North Atlantic but also the North Pacific and the Southern Ocean, enabling a systematic evaluation across different oceanic regions.

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

2 Data and methods

2.1 Data and variables

We use the operational European Centre for Medium-Range Weather Forecasts (ECMWF) 9-km IFS model cycle 49r1 at $0.1°$ horizontal resolution and AIFS version 1 at $0.25°$ resolution (Moldovan et al., 2026). The AIFS model has been trained on the ERA5 reanalysis and fine-tuned on the ECMWF operational analysis (Moldovan et al., 2026). Given that ETCs are generally more frequent and intense during winter (Čampa and Wernli, 2012; Martínez-Alvarado et al., 2016a), our analysis targets the Northern Atlantic and North Pacific during DJF 2023-24, and the Southern Ocean during JJA 2024. The selection of this time period is necessitated by the (pre-)operational availability of the AIFS and to ensure consistency across both models.

We examine forecasts at 00 and 12 UTC, including mean sea level pressure (MSLP), total column water (TCW), 10 m wind, 850 hPa temperature, the 850–500 hPa temperature difference, and 300 hPa geopotential height. Following the methodology of Yu et al. (2026), forecast biases and errors (RMSEs) are calculated by subtracting the analysis fields from the corresponding 12-hour forecasts.

While Yu et al. (2026) employed the ERA5 analysis for verification, we use the IFS analysis as our reference for three reasons. First, both IFS and AIFS exhibit a pronounced, domain-wide positive moisture bias relative to ERA5 (see Supplementary Figs. S4–S6), indicating that ERA5 has systematically drier conditions. Second, using the IFS analysis avoids potential inconsistencies that might arise from the differing data assimilation windows between ERA5 and the operational IFS. Third, as AIFS is initialised from the IFS analysis, the computed biases reflect the analysis increments over this window, i.e., how strongly the assimilated observations adjust the short-range forecast. Evaluating both models against this consistent baseline thus ensures a more robust and direct quantification of systematic biases.

2.2 Cyclone detection and tracking

We identify cyclone tracks using the tracking algorithm implemented and refined by Spensberger and Marcheggiani (2024) and Marcheggiani et al. (2025) (algorithm is publicly available as part of dynlib, Spensberger, 2024). Using the same selection criteria as Yu et al. (2026), we excluded non-developing and spurious systems and focus on the timestep of maximum intensification. Our analysis focuses on three ocean basins: the North Atlantic (30°–68°N, 80°W–12°E), the North Pacific (30°–68°N, 120°E-120°W), and the Southern Ocean (30°–70°S, 180°W–180°E).

2.3 Compositing based on vertically integrated diabatic heating

Following the classification framework of Yu et al. (2026), we categorise the selected ETCs into strong and weak diabatic heating groups (red and blue circles in Fig. 1, respectively). This classification is performed independently for each ocean basin using the vertically integrated non-radiative diabatic heating, averaged within a 750 km radius of the cyclone centre.

As illustrated in Fig. 1, the spatial distribution of the two cyclone groups exhibits regional characteristics. In general, strong-heating cyclones are predominantly found at relatively lower latitudes compared to the weak-heating cyclones, consistent

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

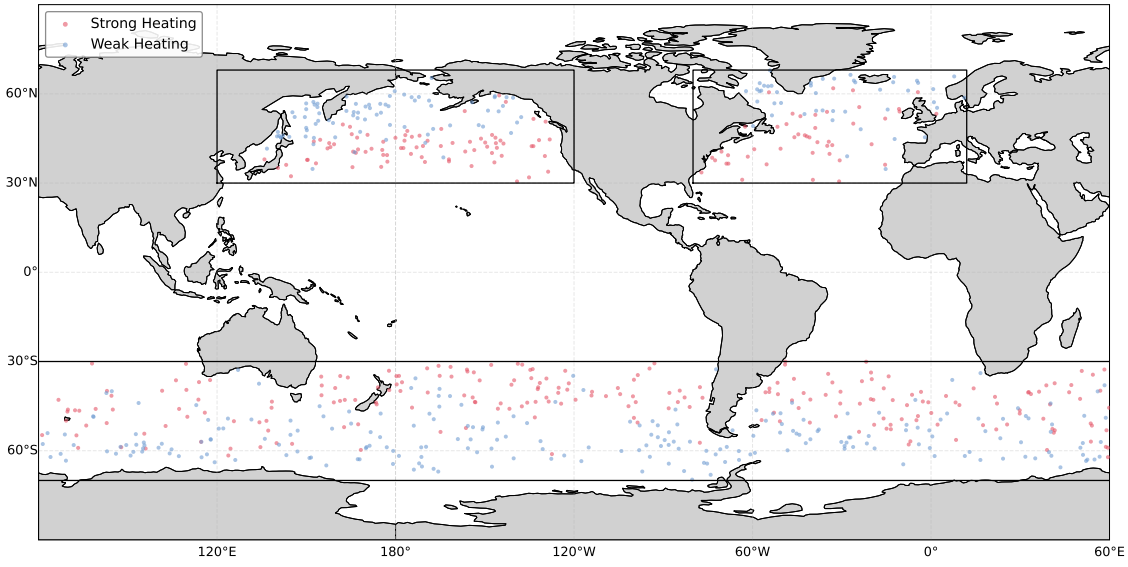

Figure 1. Geographical distribution of cyclone centres for the strong heating (SH) and weak heating (WH) groups at the time of maximum intensification. Study regions are outlined by black boxes (North Atlantic, North Pacific, and Southern Ocean).

with higher background moisture availability. However, this distribution is not zonally uniform. For example, strong heating cyclones in the North Atlantic gather closely along the warm currents, resulting in a slanted, southwest-to-northeast distribution.

To confirm that our classification is not merely reflecting latitudinal biases, we also evaluated the forecast performance for cyclones in the two different heating categories restricted to the same latitudes (i.e., $45°$–$55°$). Given that the forecast bias patterns for the strong and weak heating remained consistent (not shown), we use diabatic heating, rather than latitude, as the key factor for our cyclone classification.

While Yu et al. (2026) employed both the track centres as well as the realigned centres to isolate position biases, our study only applies the track centre for our composite analysis. This methodological choice is supported by our finding that the pronounced propagation biases previously identified in ERA5 are markedly reduced in both IFS and AIFS (Supplementary Figs. S1–S3).

Furthermore, given that the forecast bias patterns exhibit similarities across the three respective oceanic basins (Supplementary Figs. S1–S3), we aggregate cyclones from all three basins into combined strong and weak heating composites to yield statistically more robust results. As cyclonic systems in the Southern Hemisphere rotate in the opposite direction compared to their Northern Hemisphere counterparts, the Southern Ocean cyclones are mirrored meridionally (i.e., flipped along the north-south axis). This flip ensures spatial alignment across all samples in the presented composite.

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

3 Results

3.1 MSLP

Both groups of IFS forecasts exhibit a systematic, domain-wide underestimation of intensity in terms of MSLP, i.e a positive bias (Fig. 2a,b). For the strong heating group (Fig. 2a), the positive MSLP bias is relatively pronounced near the cyclone centre and within the right-hand quadrant (the warm sector). Furthermore, a comparison of the composite contours shows a relative westward position shift of the forecast (blue) compared to the analysis (black). Given that all cyclones within this composite framework are aligned to propagate eastward, this westward shift indicates a slow propagation bias in the forecasts. However, compared with the propagation bias ERA5 (IFS Cy41r2, ∼55-km resolution) (Figs. S1–S3, (Yu et al., 2026, Fig. 2a, b)), the 9-km IFS (IFS Cy49r1) forecasts represent a substantial improvement in both intensity and propagation speed.

Regarding the weak heating group of IFS, ETCs exhibit a domain-wide, relatively weaker intensity underestimation (Fig. 2b). However, the basin decomposition reveals a hemispheric asymmetry (Fig. S1-S3): the bias is remarkably small in the Northern Hemisphere, indicating the averaged intensity underestimation in Fig. 2b is primarily attributable to cyclones over the Southern Ocean.

In contrast to IFS, the AIFS forecasts show broadly similar bias patterns for both heating groups, characterised primarily by an intensity underestimation confined near the cyclone centre (Figs. 2c,d). The spatial extent and magnitude of this central positive MSLP bias are slightly larger in the strong heating group than in the weak heating group. This systematic underestimation of the cyclone core aligns with previous studies (Dacre et al., 2026), which note that while AI models generally perform well in track prediction, they suffer from a persistent intensity bias.

Compared with IFS, the composite AIFS MSLP bias is substantially weaker in both spatial extent and magnitude. However, the area-averaged RMSE of AIFS is only slightly smaller than that of IFS. To further understand this difference, we examined the case-by-case MSLP biases (Supplementary Figs. S7) and found that the IFS biases tend to have the same sign across all cases, whereas the AIFS biases are less sign-coherent and therefore partly cancel when averaged.

3.2 10 m wind

The 10 m wind speed biases in the IFS forecasts are generally modest in both heating groups. For the strong heating group (Fig. 3a), a slight underestimation of the wind speed is observed in the right quadrant of the cyclone, mainly in the upper right quadrant. The wind speed underestimation aligns with the weakened horizontal pressure gradients associated with the positive MSLP bias in Figs. 2a. The weak heating group (Fig. 3b) exhibits no obvious spatial bias pattern, featuring only some slight near-centre wind speed underestimation. Compared to the previous evaluation using ERA5 (Yu et al., 2026, Fig. 2c, d), the wind speed biases in the IFS forecasts are substantially reduced in both heating groups. However, for the strong heating group, the underestimation in the right quadrant of the cyclone is still present.

Turning to the AIFS forecasts, the 10 m wind field exhibits markedly stronger biases compared to IFS, which is consistent with the wind speed underestimation reported in recent studies (Charlton-Perez et al., 2024; Dacre et al., 2026). AIFS displays a similar spatial pattern in both heating groups, characterised by a pronounced underestimation of wind speed near the cyclone

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

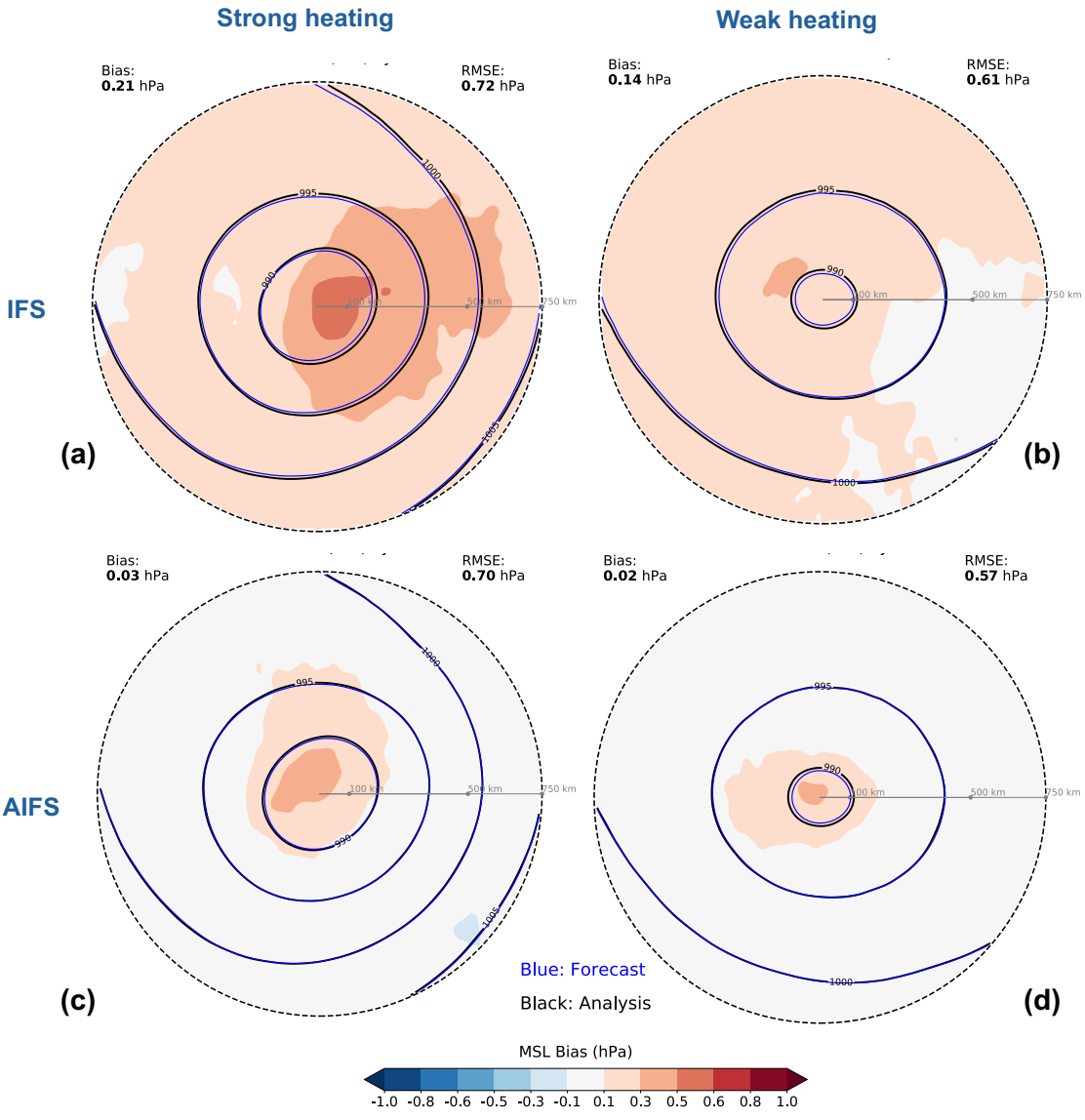

Figure 2. Composite mean sea level pressure (MSLP) biases (shading) over the North Atlantic, North Pacific, and Southern Ocean, verified against IFS analysis for the (a, c) strong heating and (b, d) weak heating groups. Black (blue) contours indicate the analysis (forecast) composite. Panels (a, b) show results for IFS and (c, d) for AIFS.

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

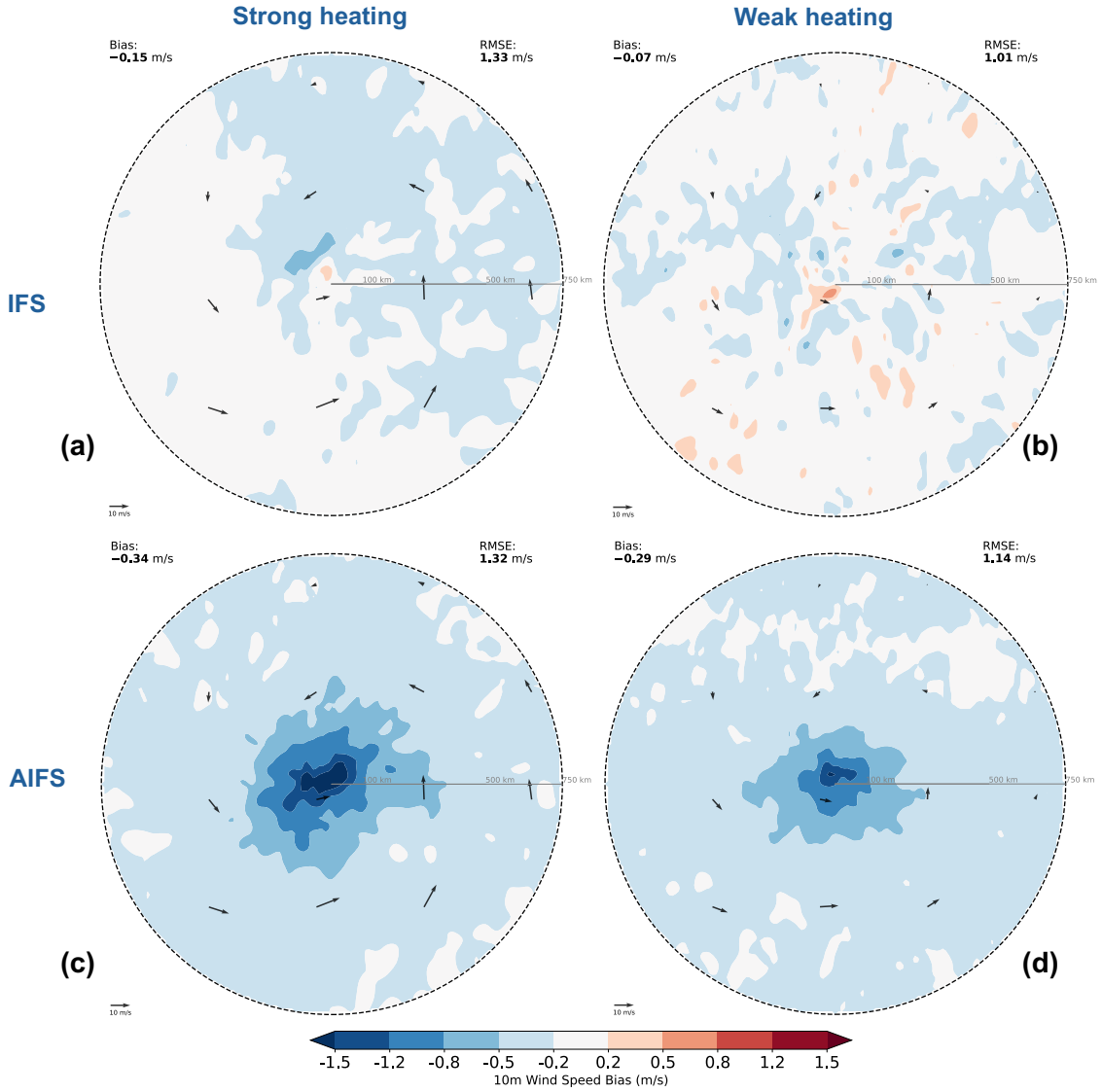

Figure 3. As in Figure 2, but for 10 m wind speed biases (shading). Black quivers show the analysis wind composite.

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

centre (Figs. 3c,d). Consistent with the MSLP bias patterns for the different heating groups, the spatial extent and magnitude of the near-centre wind bias are notably larger in the strong heating group compared to the weak heating group.

Comparing MSLP and 10 m wind biases between the two models, AIFS exhibits a weaker MSLP bias than IFS, yet a stronger 10 m wind bias. This indicates a physical inconsistency. Using the ratio between ageostrophic and geostrophic wind speed, previous studies found that some data-driven models show physical inconsistency over extratropical regions (Bonavita, 2024; Kochkov et al., 2024; Hu et al., 2026). Examining the geostrophic and ageostrophic components of the 10 m wind, we find that the geostrophic wind speed is well captured by both models (Fig. 4). The ratio of ageostrophic to geostrophic wind speed is reduced for many cases in AIFS, whereas the IFS forecasts remain closer to the analysis. Furthermore, the ratio differs little between IFS and AIFS at other pressure levels (not shown). These results suggest that the pronounced underestimation of 10 m wind speed in AIFS is associated with the underestimated near-surface ageostrophic component, rather than a general weakening of the ageostrophic flow throughout the troposphere. This points to a limitation in representing boundary-layer and friction-related wind components in AIFS.

3.3 Total column water

The IFS weak heating group mainly shows a slight overall underestimation in total column water (Figs. 5b). For the strong heating group, the forecast moist tongue (blue) is narrower than the analysis (black); and the bias shading shows an overestimation of total column water (TCW) within this high-moisture tongue (Figs. 5a). Compared to the pronounced moisture underestimation in the ERA5 (Yu et al., 2026, Fig. 4a), the IFS no longer exhibits such a substantial dry bias.

The weak heating group of AIFS exhibits almost no obvious bias in TCW (Figs. 5d). For the strong heating group, the AIFS presents a bias pattern partially similar to that of the IFS, featuring an overestimation around the high-moisture area. However, unlike the IFS, the composite contours do not feature a narrowing of the moist tongue. Instead, the AIFS exhibits underestimations in regions where the TCW contours are denser. This shows that the data-driven model struggles to represent sharp atmospheric gradients, which is consistent with previous studies (Selz and Craig, 2023; Bi et al., 2023; Lang et al., 2024; Charlton-Perez et al., 2024).

3.4 Temperature at 850 hPa

For 850 hPa temperature (T_{850}), the strong heating group of IFS exhibits a marked positive (warm) bias near the cyclone centre and along the warm tongue, particularly pronounced within regions of tight temperature gradients (Figs. 6a). As for the weak heating group, T_{850} shows only a slight warm bias in the warm sector and a slight cold bias in the cold sector with negligible biases across the regions of denser temperature gradients (Figs. 6b). Consistent with the moisture field, IFS no longer exhibits a pronounced cold bias identified in ERA5 (Yu et al., 2026). Although the IFS strong heating group no longer shows the pronounced cold and dry biases evident in ERA5 (Yu et al., 2026, Fig. 6e), similar underestimations in cyclone intensity (MSLP) and warm sector wind speed prevail.

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

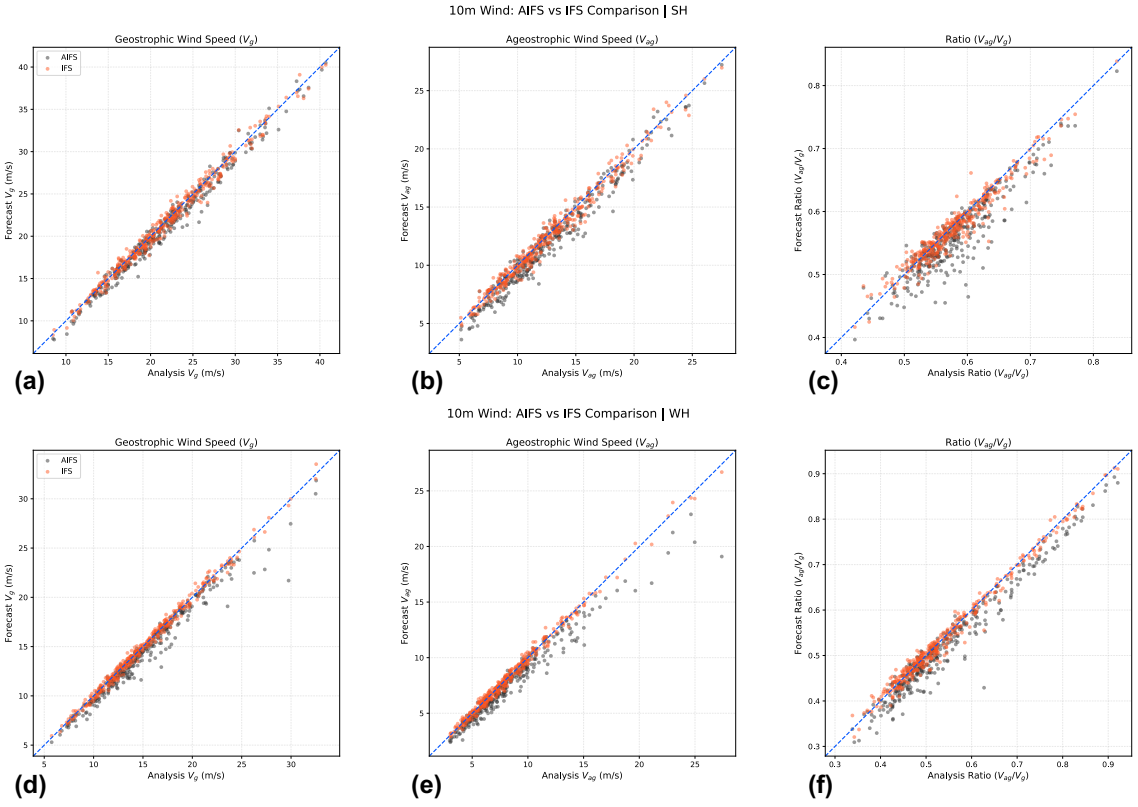

Figure 4. Scatter plots comparing 10 m area-averaged wind properties (within a 750-km radius) from AIFS (black dots) and IFS forecasts (coral dots) with the corresponding IFS analysis for the strong heating (upper row) and weak heating (lower row) group. (a,d) geostrophic wind speed (V_g), (b,e) ageostrophic wind speed (V_{ag}), and (c,f) ratio of ageostrophic wind speed over geostrophic wind speed (V_{ag}/V_g). The blue dashed line indicates the 1:1 reference line.

The AIFS T_{850} exhibits a similar but notably weaker bias pattern compared to the IFS (Fig. 6c. For the strong heating group, AIFS displays a positive (warm) bias mainly in the warm tongue (Fig. 6c). For the weak heating group, AIFS only features a marginal weak cold bias, located across the upper quadrants of the composite domain (Fig. 6d).

3.5 850–500 hPa temperature difference

We examine the 850–500 hPa temperature difference ($\Delta T_{850-500}$), which serves as a direct indicator of the vertical temperature lapse rate in the lower-to-mid troposphere. In the strong heating group, IFS shows a clear spiral-shaped overestimation around the centre and the warm sector, particularly where the contours are dense (Fig. 7a). This steeper lapse rate consists of a warm bias at 850 hPa (Fig. 6a), together with a slight warm bias near the warm tongue at 700 hPa (Fig. S9a), and a cold bias at 500 hPa (Fig. S10a). This implies an underestimation of the vertical extent of diabatic heating. In contrast, the biases in the IFS

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

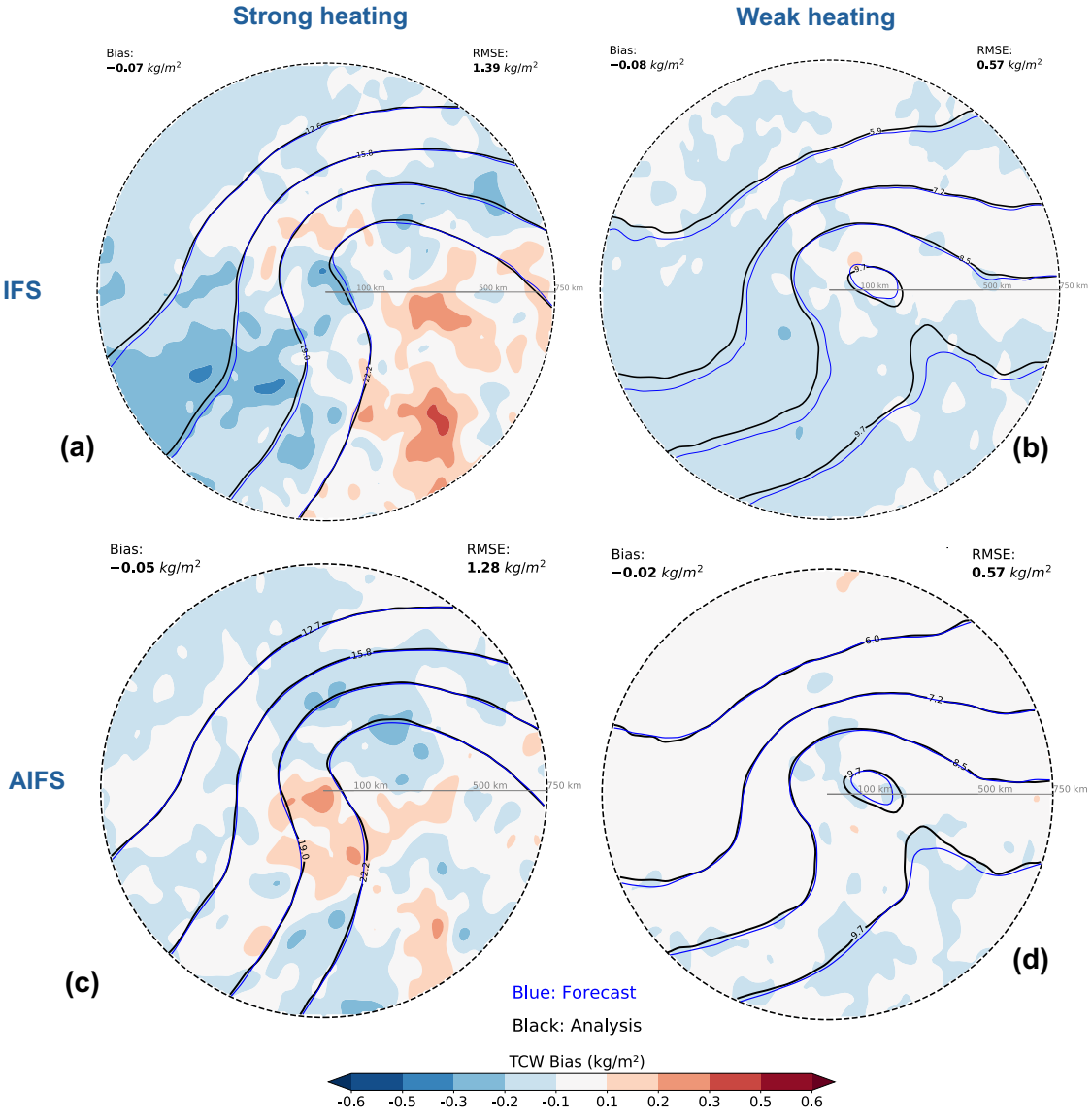

Figure 5. As in Figure 2, but for the total column water (TCW) biases (shading). Black (blue) contours indicate the analysis (forecast) composite.

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

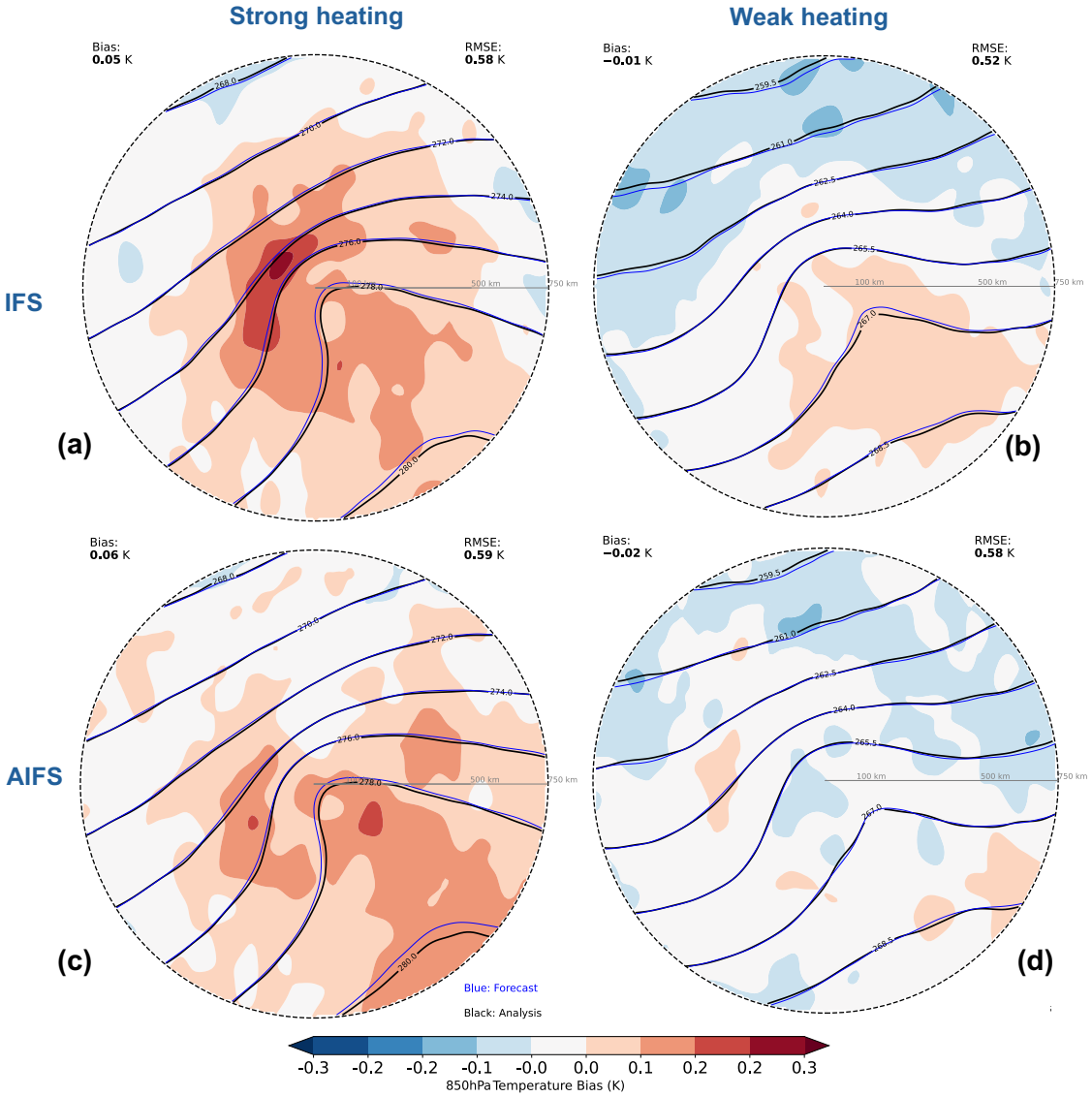

Figure 6. As in Figure 2, but for temperature at 850 hPa biases (shading), black (blue) contours indicate the analysis (forecast) composite.

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

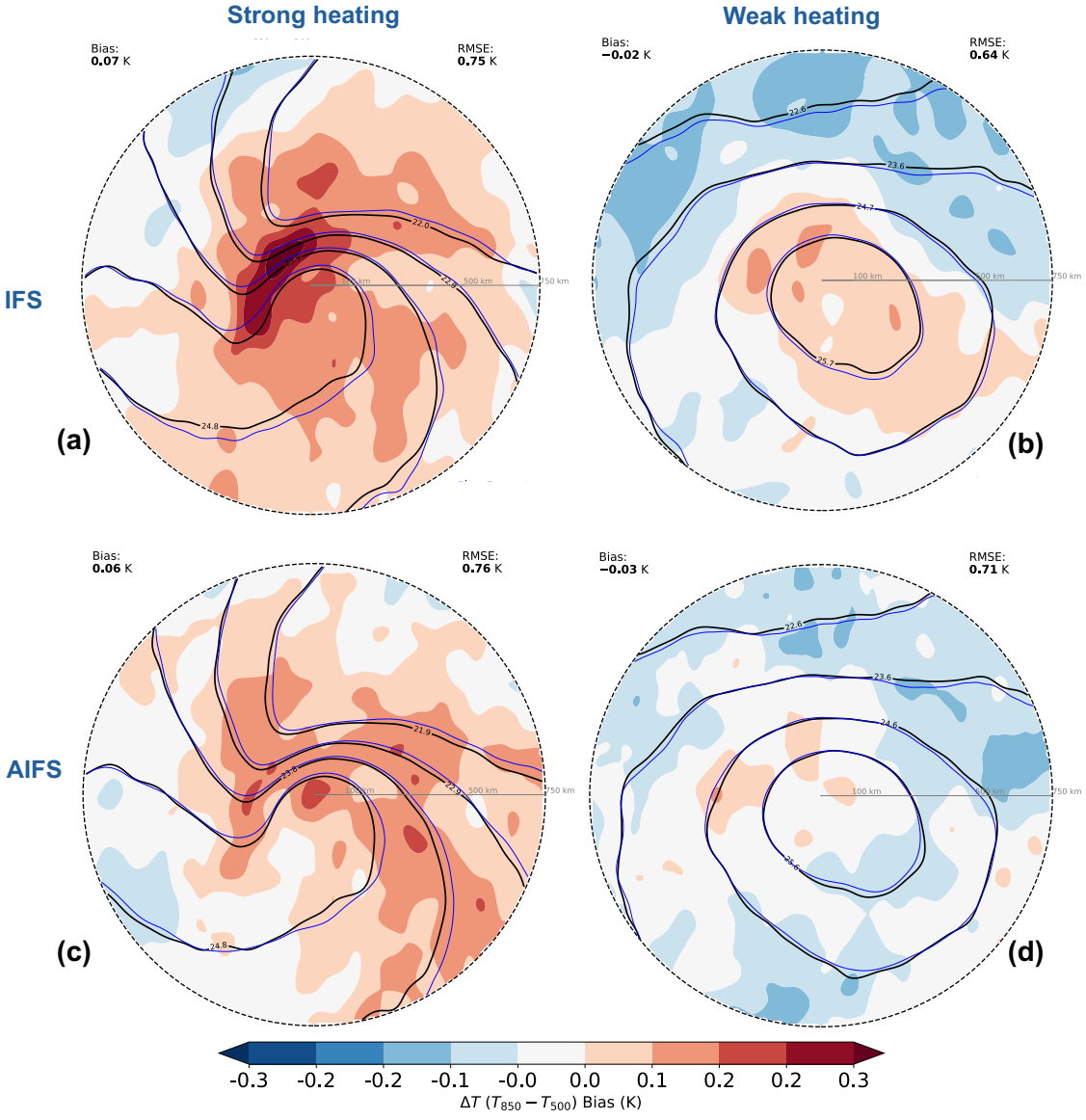

Figure 7. As in Figure 2, but for biases in the 850–500 hPa temperature difference (shading), black (blue) contours indicate the analysis (forecast) composite.

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

weak heating group are substantially smaller, featuring a slight overestimation near the centre and an underestimation in the upper quadrant (Fig. 7b).

AIFS exhibits broadly similar bias patterns compared to IFS. In the strong heating group, while the intensity of the positive bias is slightly weaker than in IFS, it still shows a distinct overestimation along the inward-spiraling warm sector (Fig. 7c). The AIFS weak heating group displays only a minor negative bias across the domain (Fig. 7d). However, the domain-mean RMSE increases from 0.64 in IFS to 0.71 in AIFS, indicating that the reduced composite biases do not correspond to a smaller error. Instead, this discrepancy suggests a stronger cancellation of AIFS biases in the composite mean due to case-to-case variability in sign and/or space.

3.6 Geopotential height at 300 hPa

In the IFS weak heating group, the forecast exhibits a weak underestimation of the trough (Fig. 8b). This pattern is similar to the biases found in the ERA5 weak heating group (Yu et al., 2026), though with a slightly reduced magnitude. In the IFS strong heating group, there is a distinct positive bias within the trough as well as the extended downstream region, indicating an underestimation of the trough (Fig. 8a). For ERA5, the strong heating case displayed an underestimation further downstream along the ridge side (Yu et al., 2026, Fig. 7c), suggesting a different impact of diabatic heating on the upper-level flow for the two model versions. Overall, for both strong and weak heating, the underestimation of the trough aligns with the reduced surface cyclone intensity (MSLP), which is consistent with previous findings for upper-level trough development and surface cyclone intensification (Čampa and Wernli, 2012; Binder et al., 2016).

Regarding AIFS, both heating groups show a much weaker positive bias within the trough (Figs. 8c,d). However, similar to the other fields, the weaker composite bias is not accompanied by a comparable reduction in RMSE, indicating that case-to-case errors remain.

4 Conclusions

We systematically evaluated the short-range (12-hour) forecast performance of the operational 9-km ECMWF IFS and the data-driven AIFS for wintertime maritime extratropical cyclones (ETCs) across the North Atlantic, North Pacific, and Southern Ocean. Compared to a previous study using ERA5 (Yu et al., 2026), our results demonstrate that higher resolution and updated physics as well as data-driven approaches significantly reduce propagation biases. Overall, both IFS and AIFS exhibit more pronounced forecast biases for cyclones with strong diabatic heating compared to cyclones with weak diabatic heating. While the bias patterns are distinctly different between the two heating groups for the physics-based IFS, AIFS displays less pronounced differences between the two groups with some bias patterns being similar to those of IFS.

The updated IFS also no longer exhibits the dry and cold biases present for the ERA5 strong heating cases (Yu et al., 2026). Instead, the IFS produces a narrower but overestimated high-moisture tongue, along with a pronounced warm bias at 850 hPa. Despite this warmer and moister environment, MSLP and lower wind speeds still show similar patterns as ERA5 (Yu et al., 2026). Specifically, cyclone intensity and warm sector wind speeds remain underestimated, though the bias magnitudes have

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

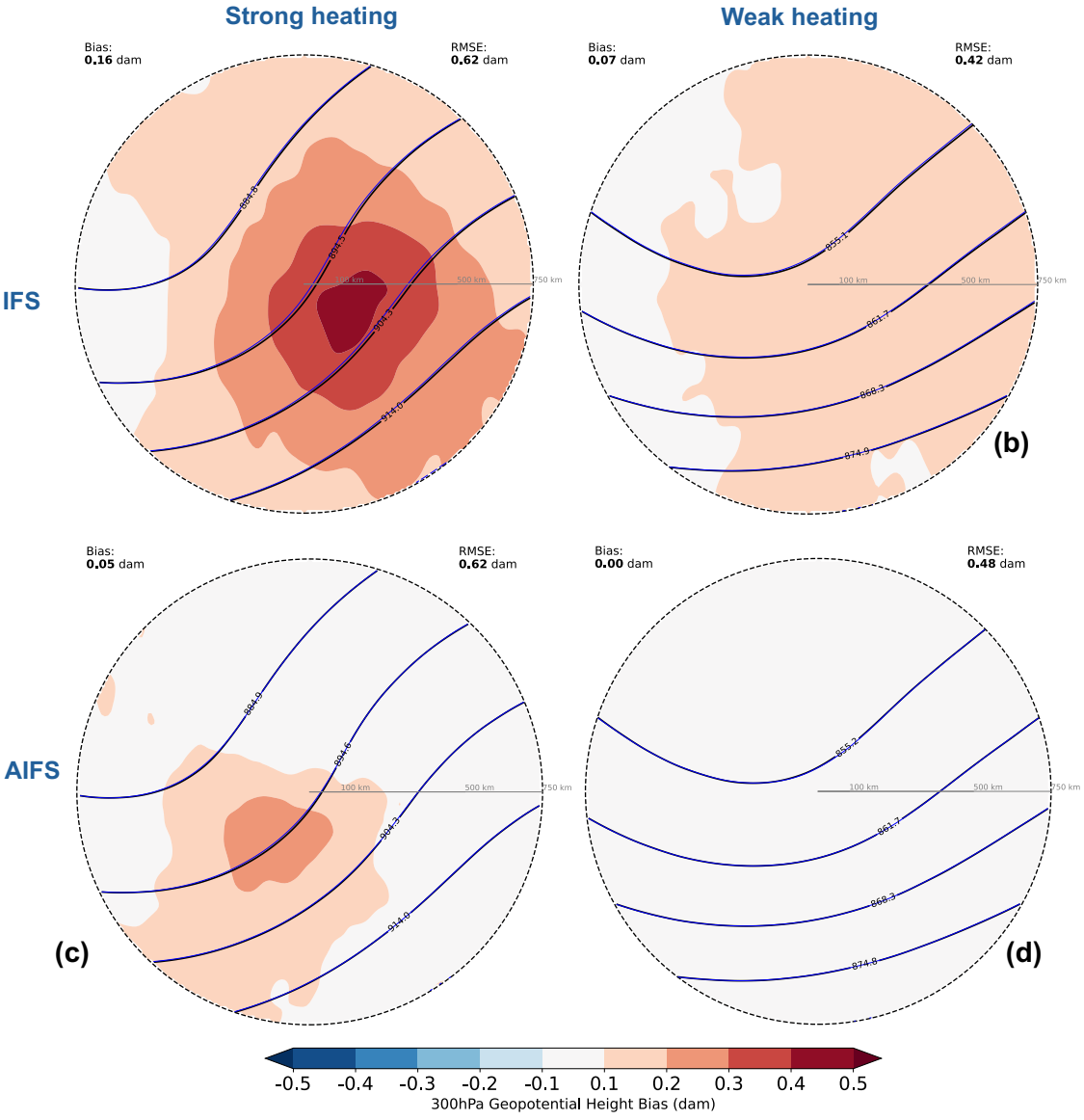

Figure 8. As in Figure 2, but for geopotential height at 300 hPa biases (shading), black (blue) contours indicate the analysis (forecast) composite.

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

decreased. The 850–500 hPa temperature difference displays a distinct spiral-shaped positive bias. Together with a warm bias at 850 hPa and a cold bias at 500 hPa, this hints at a misrepresentation of the vertical distribution of diabatic heating processes in the forecasts.

Furthermore, at 300 hPa, the geopotential height exhibits an overestimation within the trough that extends downstream. This differs from ERA5, where an underestimation of geopotential height was located further downstream along the ridge. While the underestimated ridge development in ERA5 is likely associated with underestimated heating intensity, the underestimated depth of the trough in IFS is more likely related to the underestimation of the vertical extent of diabatic heating, weakening its impact on the upper-level circulation. As for cyclones with weak diabatic heating, the biases remain relatively small and consistent with the ERA5 results, featuring only slight underestimations in cyclone intensity, low-level wind, moisture, and depth of the upper-level trough.

The misrepresentation of the vertical distribution of diabatic heating for strong heating cyclones in IFS is likely associated with the interplay between its microphysical parameterisations and resolution. As demonstrated by Joos and Forbes (2016), changes in microphysical parameterisations can directly alter the vertical heating distribution within extratropical cyclones. Furthermore, despite the increase in resolution to 9 km for IFS, it is still within the convective 'grey zone' (Arakawa and Wu, 2013; Gerard, 2007), which could yield further discrepencies.

Biases for AIFS are generally smaller than those of IFS, whereas the area-averaged RMSEs do not show a similar reduction. This is associated with the AIFS biases being more variable in sign across different cases, with greater cancellation in the composite mean. Furthermore, the bias patterns in AIFS are similar to those in IFS for most fields, which might be related to the way AIFS is adapted to the operational analysis system. Given that AIFS is fine-tuned to the operational IFS analyses, the model becomes more sensitive to the initial analysis (Ben Bouallègue et al., 2026). The IFS analysis is itself produced by assimilating observations into a short-range forecast background (ECMWF, 2026), which may affect AIFS through both its initial conditions and fine-tuned data. However, AIFS exhibits a more pronounced underestimation of the 10 m wind field, despite having a smaller MSLP bias compared to IFS, revealing a potential physical inconsistency in AIFS. A decomposition of the surface wind shows that this is related to an underestimation of the near-surface ageostrophic wind speed, which suggests a limitation in AIFS's representation of boundary-layer and friction-related wind components.

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

Code and data availability. Data from IFS and AIFS are available under ECMWF's Open Data Creative Commons licence (https://www.ecmwf.int/en/forecasts/datasets/open-data) and DOI https://doi.org/10.21957/open-data. Cyclone tracks are openly available at https://doi.org/10.11582/2024.00023. The Python library dynlib (Spensberger, 2024) is available https://doi.org/10.11582/2024.00023.

Competing interests. The authors declare that they have no conflict of interest.

Author contributions. QY performed data analyses and prepared the paper. LM, CS and TS contributed to the interpretation of the results and to the writing of the paper

Acknowledgements. We thank the ECMWF for making the data openly available. This study was supported by the Research Council of Norway (Norges Forskningsråd, NFR) through the BALMCAST project (NFR grant number 324081).

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

References

Arakawa, A. and Wu, C.-M.: A unified representation of deep moist convection in numerical modeling of the atmosphere. Part I, Journal of the Atmospheric Sciences, 70, 1977–1992, https://doi.org/10.1175/jas-d-12-0330.1, 2013.

Baumgart, M., Ghinassi, P., Wirth, V., Selz, T., Craig, G. C., and Riemer, M.: Quantitative view on the processes governing the up-scale error growth up to the planetary scale using a stochastic convection scheme, Monthly Weather Review, 147, 1713–1731, https://doi.org/10.1175/mwr-d-18-0292.1, 2019.

Ben Bouallègue, Z., Clare, M. C. A., Magnusson, L., Gascón, E., Maier-Gerber, M., Janoušek, M., Rodwell, M., Pinault, F., Dramsch, J. S., Lang, S. T. K., Raoult, B., Rabier, F., Chevallier, M., Sandu, I., Dueben, P., Chantry, M., and Pappenberger, F.: The Rise of Data-Driven Weather Forecasting: A New Era in Numerical Weather Prediction?, Bulletin of the American Meteorological Society, 105, E1045–E1064, https://doi.org/10.1175/BAMS-D-23-0162.1, 2024.

Ben Bouallègue, Z., Raoult, B., and Chantry, M.: Farewell to the external AI models, ECMWF AIFS Blog, https://www.ecmwf.int/en/about/media-centre/aifs-blog/2026/farewell-external-ai-models, accessed: 2026-05-26, 2026.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q.: Accurate medium-range global weather forecasting with 3D neural networks, Nature, 619, 533–538, https://doi.org/10.1038/s41586-023-06185-3, 2023.

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.

Bonavita, M.: On Some Limitations of Current Machine Learning Weather Prediction Models, Geophysical Research Letters, 51, e2023GL107 377, https://doi.org/10.1029/2023GL107377, 2024.

Čampa, J. and Wernli, H.: A PV perspective on the vertical structure of mature midlatitude cyclones in the Northern Hemisphere, Journal of the atmospheric sciences, 69, 725–740, https://doi.org/10.1175/JAS-D-11-050.1, 2012.

Catto, J. L., Ackerley, D., Booth, J. F., Champion, A. J., Colle, B. A., Pfahl, S., Pinto, J. G., Quinting, J. F., and Seiler, C.: The future of midlatitude cyclones, Current Climate Change Reports, 5, 407–420, https://doi.org/10.1007/s40641-019-00149-4, 2019.

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.

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, https://doi.org/10.1175/bams-d-25-0129.1, 2026.

ECMWF: Data assimilation, https://www.ecmwf.int/en/research/data-assimilation, 2026.

Gerard, L.: An integrated package for subgrid convection, clouds and precipitation compatible with meso-gamma scales, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 133, 711–730, https://doi.org/10.1002/qj.58, 2007.

Hu, Y., Yin, F., and Deng, K.: A Hybrid Framework for Global Weather Forecasting via Low-Resolution Dynamical Core and Multigrid Neural Operator, npj Climate and Atmospheric Science, https://doi.org/10.1038/s41612-026-01374-z, 2026.

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, https://doi.org/10.1002/qj.2863, 2016.

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

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, https://doi.org/10.1002/qj.2172, 2014.

Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., Klöwer, M., Lottes, J., Rasp, S., Düben, P., Hatfield, S., Battaglia, P., Sanchez-Gonzalez, A., Willson, M., Brenner, M. P., and Hoyer, S.: Neural General Circulation Models for Weather and Climate, Nature, 632, 1060–1066, https://doi.org/10.1038/s41586-024-07744-y, 2024.

Kuo, Y.-H. and Low-Nam, S.: Prediction of nine explosive cyclones over the western Atlantic Ocean with a regional model, Monthly Weather Review, 118, 3–25, https://doi.org/10.1175/1520-0493(1990)118<0003:poneco>2.0.co;2, 1990.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., et al.: Learning skillful medium-range global weather forecasting, Science, 382, 1416–1421, https://doi.org/10.1126/science.adi2336, 2023.

Lang, S., Alexe, M., Chantry, M., Dramsch, J., Pinault, F., Raoult, B., Clare, M. C., Lessig, C., Maier-Gerber, M., Magnusson, L., et al.: AIFS–ECMWF's data-driven forecasting system, arXiv preprint arXiv:2406.01465, https://doi.org/10.48550/arXiv.2406.01465, 2024.

Marcheggiani, A., Dacre, H., Spensberger, C., and Spengler, T.: Weather features drive free-tropospheric baroclinicity variability in the North Atlantic storm track, Quarterly Journal of the Royal Meteorological Society, p. e5061, https://doi.org/10.1002/qj.5061, 2025.

Martínez-Alvarado, O., Gray, S. L., and Methven, J.: Diabatic processes and the evolution of two contrasting summer extratropical cyclones, Monthly Weather Review, 144, 3251–3276, https://doi.org/10.1175/mwr-d-15-0395.1, 2016a.

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, 2016b.

Martínez-Alvarado, O., Gray, S. L., Hart, N. C., Clark, P. A., Hodges, K., and Roberts, M. J.: Increased wind risk from sting-jet windstorms with climate change, Environmental research letters, 13, 044 002, https://doi.org/10.1088/1748-9326/aaae3a, 2018.

Moldovan, G., Nemesio, A. P., Pinnington, E., Lang, S., Polster, J., O'Brien, C., Santa Cruz, M., Alexe, M., Cook, H., Forbes, R., et al.: Improving AIFS Forecast Skill through Fine-Tuning across Spatial Resolutions and Datasets, Tech. rep., Copernicus Meetings, https://doi.org/10.5194/egusphere-egu26-6950, 2026.

Pfahl, S., Schwierz, C., Croci-Maspoli, M., Grams, C. M., and Wernli, H.: Importance of latent heat release in ascending air streams for atmospheric blocking, Nature Geoscience, 8, 610–614, https://doi.org/10.1038/ngeo2487, 2015.

Selz, T. and Craig, G. C.: Can artificial intelligence-based weather prediction models simulate the butterfly effect?, Geophysical Research Letters, 50, e2023GL105 747, https://doi.org/10.1029/2023gl105747, 2023.

Shaw, T., Baldwin, M., Barnes, E. A., Caballero, R., Garfinkel, C., Hwang, Y.-T., Li, C., O'gorman, P., Rivière, G., Simpson, I., et al.: Storm track processes and the opposing influences of climate change, Nature Geoscience, 9, 656–664, https://doi.org/10.1038/ngeo2783, 2016.

Spensberger, C.: Dynlib: a library of diagnostics, feature detection algorithms, plotting and convenience functions for dynamic meteorology, https://doi.org/10.5281/zenodo.10471187, 2024.

Spensberger, C. and Marcheggiani, A.: ERA5 cyclone tracks, https://doi.org/10.11582/2024.00023, 2024.

Willison, J., Robinson, W. A., and Lackmann, G. M.: The importance of resolving mesoscale latent heating in the North Atlantic storm track, Journal of the Atmospheric Sciences, 70, 2234–2250, https://doi.org/10.1175/jas-d-12-0226.1, 2013.

Xie, S. et al.: On the correspondence between short- and long-timescale systematic errors in CAM4/CAM5, Journal of Climate, 25, 7937–7955, https://doi.org/10.1175/jcli-d-12-00134.1, 2012.

Yu, Q., Spensberger, C., Magnusson, L., and Spengler, T.: Forecast Errors Attributed to Synoptic Features, Meteorological Applications, 32, e70 093, https://doi.org/10.1002/met.70093, 2025.

https://doi.org/10.5194/egusphere-2026-3727
Preprint. Discussion started: 2 July 2026

Yu, Q., Spensberger, C., Magnusson, L., and Spengler, T.: Distinct bias structures for extratropical cyclones with strong or weak diabatic heating, EGUsphere, 2026, 1–22, https://doi.org/10.5194/egusphere-2026-257, 2026.