

## Response to Reviewer 1

Overall, I find the study interesting and the results valuable. In particular, the successful implementation of trajectory calculations within a climate model framework is a notable strength of the paper, as such analyses are often challenging due to the limited temporal and spatial resolution of typical climate model output.

That said, I believe the presentation of the material could be substantially improved. My comments below are therefore mainly concerned with the clarity and structure of the presentation, rather than with the underlying scientific approach or results.

We would like to thank the reviewer for their helpful comments. Our responses are shown in blue, while the reviewer's questions are in black. We also appreciate the suggestion to improve the clarity and structure of the presentation. In our revised manuscript, we will carefully go through the manuscript again and incorporate the reviewers' comments to enhance readability and make the narrative as clear as possible for the reader.

### Major Comments

1. The descriptions in the manuscript are often quite lengthy and at times somewhat repetitive. For example, results are first presented for DJF warming events in the present and future, then DJF cooling events, followed by JJA warming and cooling events, and subsequently for the tropics. To enhance reader engagement, it might be helpful to introduce the figures once and then focus more on highlighting the key differences, while keeping the descriptions concise and emphasising the essential points.

**Response:** We would not change the entire structure of the manuscript. We think the details are valuable and can be interesting for readers interested in specific types of events. Furthermore, the common structure makes it easy for readers not interested in specific details to skip some parts of the manuscript, and the highlights are summarised in section 4. We will go through the manuscript again and try to shorten it as much as possible based on the reviewers' comments.

2. Playing the devil's advocate: How physically meaningful are the computed trajectories within the atmospheric boundary layer and especially in the tropics, where turbulent mixing is intense? I think a brief discussion on the limitations or uncertainties associated with the trajectories in these regions would strengthen the study.

**Response:** Thank you for this insightful comment. We agree that discussing the limitations of trajectories in the (tropical) boundary layer will strengthen the study.

We will add a brief discussion acknowledging that:

Trajectory calculations rely on resolved wind fields and therefore cannot capture the full spectrum of convective and turbulent motions. This limitation is especially pronounced in the tropics, where intense moist convection introduces significant uncertainty into the calculated transport pathways (Bao & Stevens, 2021; Bergman & Sardeshmukh, 2004; Stohl, 1998).

## Minor Comments

L15/16: The mention of a “clear dipole pattern” is somewhat confusing to me. You mention a “clear dipole pattern”, but then for JJA, the pattern does not clearly take a dipole form. Consider rephrasing.

**Response:** We will rephrase that sentence as: "The projected changes in (extreme) DTDT variations display a seasonal contrast: during December–February (DJF): weakening in mid- to high latitudes and intensification in the tropics, while during June–August (JJA), tropical intensification is more widespread, and only some extratropical locations experience reductions in DTDT variations”.

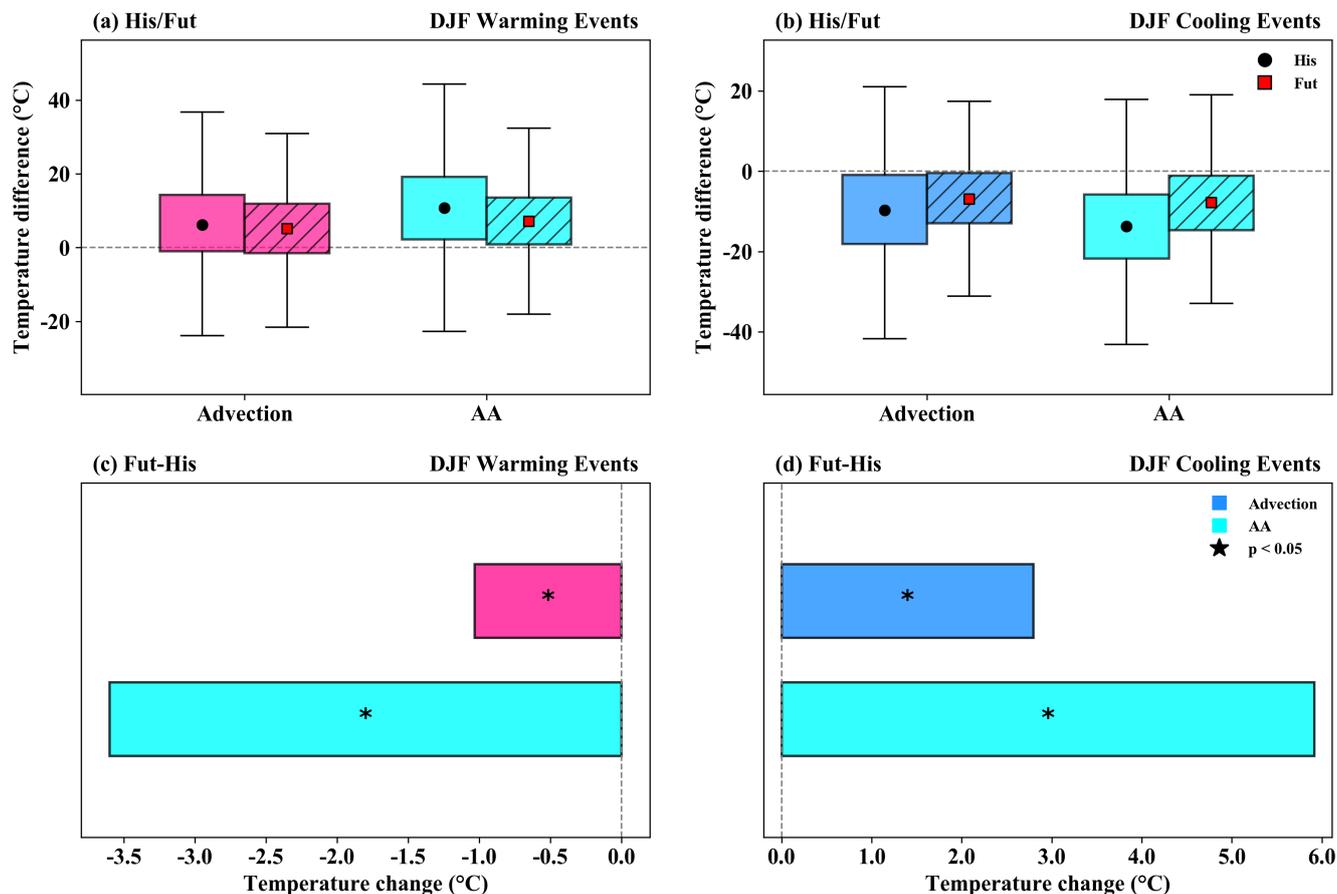
L18: “only” instead of “also”?

**Response:** We will change this.

L19: I think you should be more careful here when writing “due to Arctic Amplification”.

**Response:** We have added some results based on this as per the suggestion of Reviewer 2.

Our trajectory analysis reveals a clear fingerprint of Arctic amplification: air originating from the Arctic experiences consistently larger future warming than air from lower latitudes (Figure 5, main paper). This reflects a weakening of the meridional temperature gradient. Furthermore, as shown in Figure R6 (response to Reviewer 2), air masses on both day  $t-1$  and day  $t$  are substantially modified by the warming climate. The ratio of local temperature change to global warming exceeds 1 on all days ( $-72$  h to 0 h), and is above 1.4 for both day  $t-1$  for warming and day  $t$  for cooling events, satisfying the quantitative definition of Arctic amplification. This confirms that air masses participating in these extreme DTDT events carry the signature of Arctic amplification, as they originate from regions warming at a rate significantly faster than the global average.



**Figure R1.** (a,b) Box plots comparing the contribution of advection and the estimated effect of Arctic Amplification (AA) for (a) warming events and (b) cooling events in North America, on a 3-day time scale. For each process, the left box (solid edge) represents the historical climate, and the right box (hatched) represents the future climate. The box spans the 25th to 75th percentiles; the black dot marks the historical mean; the red square marks the future mean; whiskers extend to 1.5 times the interquartile range. The effect of AA is estimated by comparing the seasonal-mean projected temperature change at the origins (at -72 h) of the historical backward trajectories between day  $t$  and day  $t-1$ . (c,d) Stacked horizontal bars showing the mean change (future minus historical) for advection and AA for (c) warming and (d) cooling events. The \* indicates that the difference is statistically significant at the 95 % confidence level based on a t-test.

We isolated the Arctic Amplification (AA) contribution to extreme temperature events by comparing the seasonal-mean projected temperature change at the origins (at -72 h) of the historical backward trajectories between day  $t$  and day  $t-1$  in Figure R1. In this way, only the warming at the origins (which differs between  $t$  and  $t-1$  due to the typically different latitudes of air parcels at -72h) is considered under the assumption that the circulation is unchanged. The magnitudes of the AA effects are more pronounced than those from the changes in advection determined from equation 2. Specifically, for warming events, the change in advection (-1.1 °C) is smaller in magnitude than the change in AA (-3.5 °C), while for cooling events, the change in advection (+2.8 °C) is also smaller than the change in AA (+5.9 °C). The effect of AA explains the weakening of the advective contribution to both warming and cooling events, but the magnitude of the AA effect is mitigated by circulation changes. Also, the event-to-event variability (boxes and whiskers in Figure R1) is very similar between advection and estimated AA contributions.

We will add the Figure to the supplementary material and discuss this result in the main paper.

L37: “imperative” is a very strong word. I would be a bit more moderate here.

**Response:** We will change this. Now the sentence reads “Therefore, studying DTDT changes and their extremes in a warming climate is important”.

L56: “for the past” instead of “in the past”?

**Response:** We will change this.

L63: I would avoid citing Mayer (2025) when discussing the importance of diabatic heating, as Mayer (2025) emphasises the role of advection in temperature extremes rather than adiabatic or diabatic processes.

**Response:** We will change this.

L69: “process understanding” instead of “processes understanding”?

**Response:** We will change this.

L95: “use” instead of “utilise”

**Response:** We will change this.

L106: In its current position, the formula does not appear to be well integrated into the text. The same applies to Eq. (2).

**Response:** In the revised manuscript, we will restructure the surrounding text to properly introduce and discuss each equation before it appears.

L110: Why are these seasons “key seasons”? Rather explain or omit the “key”.

**Response:** We will remove this and improve the sentence.

L124-127: I would omit the lengthy description of all the individual grid points in the supplement.

**Response:** We agree and will omit the lengthy description of individual grid points and add to supplement.

Eq. (2): It might be helpful to write out the integrals explicitly to clarify exactly which terms are being computed.

**Response:** As the explicit integral forms are the same as those shown in Part I, we will add a cross-reference in the text to guide readers to the complete expressions.

We will add this: "Here, the DTDT change ( $\delta_T^0$ ) is decomposed into three contributing factors in Eq. (2), with full integral expressions calculated as in Part I (equations A4-6 in (Hamal & Pfahl, 2025))."

Eq. (2): Why do you accumulate over 3 days? Is there a physical reason? Have you tested the sensitivity of your results to other accumulation periods?

**Response:** We use LAGRANTO to calculate 10-day backward trajectories initialised at 18 UTC on both the preceding day ( $t-1$ ) and the event day ( $t$ ), at 10, 30, 50, and 100 hPa above the surface, for the corresponding grid boxes. The first part of our study, as well as previous studies, suggests that extremes typically develop on a timescale of 2–3 days (Bieli et al., 2015; Röthlisberger & Papritz, 2023). Therefore, we focus on 3-day backward trajectories for our analysis. For tropical locations, the time scale is further reduced to 1-day, since the relevant processes appear to occur on shorter time scales (see section 3.2.2 and Fig. 10). This has been discussed in Part I of our study.

L134-136: It is not clear what is meant by “mean temperature difference” or “mean adiabatic compression.” Consider clarifying the meaning of “mean” here.

**Response:** We will clarify that "mean" here refers to the trajectory-average value of each process over the 3-day back trajectory.

L147/148: There seems to be a contradiction: first, the results are said to fit ERA5 “in many regions,” then to deviate “in large parts.” Consider clarifying this.

**Response:** In the original manuscript, we used a different method to compare ERA5 and CESM. Following Reviewer 2's recommendation, we now employ bootstrapping with the FDR test for this analysis. The text will be updated accordingly in the revised manuscript.

Figure 1, 2, 3, etc.: I think it would be helpful to use white color for small deviations around 0.

**Response:** The newly introduced significance test based on bootstrapping indicates that relatively small changes can also be statistically significant (see our responses to Reviewer 2 for more details). To prevent these signals from being masked, we did not adjust the color scale.

L162: The pattern does not appear as “distinct” to me.

**Response:** We will remove this.

L166/167/177: Phrases like “changes are driven by”, “increases due to”, or “influence” imply causality to me. Since the decomposition (into  $\sigma_T$  and autocorrelation) is “just” descriptive, I think you avoid implying causality.

**Response:** We thank the reviewer for this important distinction. We will revise the text to avoid causal language (e.g., "driven by," "due to") and instead use descriptive terms (e.g., "associated with," "contributed to by," or "coincides with") that align with the decomposition framework.

L179/180: The statements about Chile are contradictory: first mentioning it as an exception, then saying “(apart from Chile).” Consider clarifying.

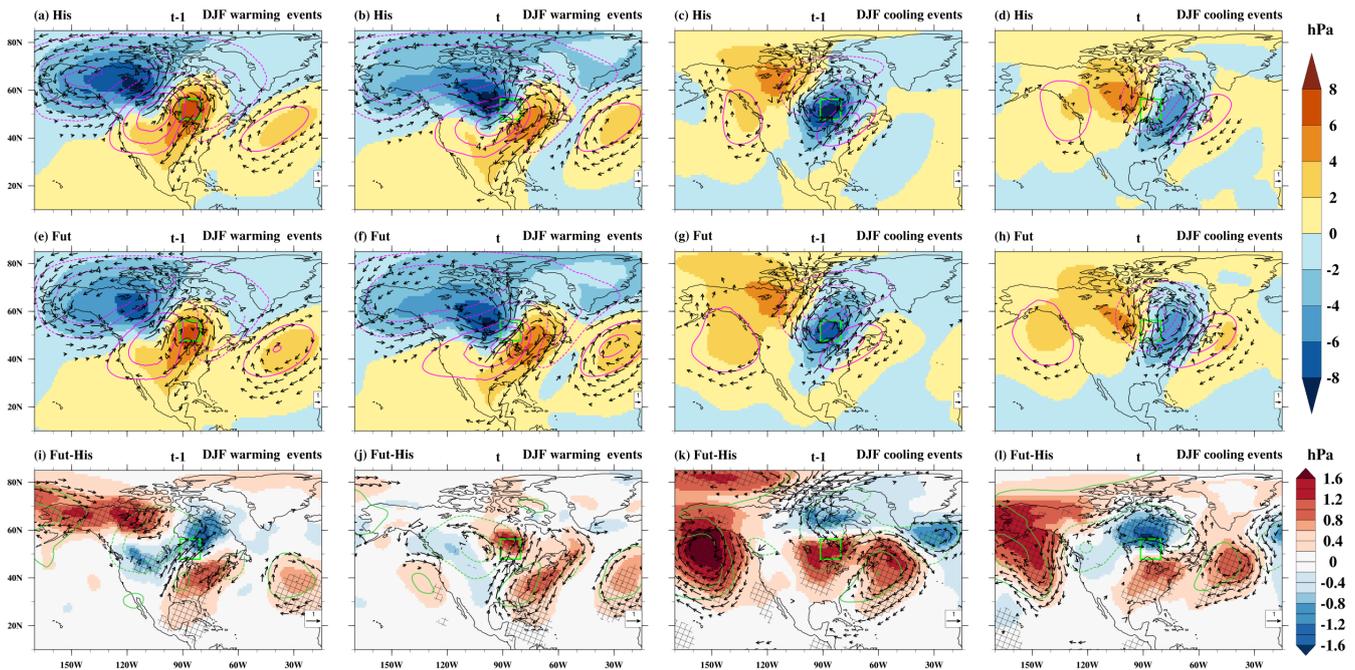
**Response:** We will make this clear based on the revised Figure.

Figure S3: The description is very detailed, but the figure is in the supplement. Consider either shortening the description or moving the figure to the main text.

**Response:** The description largely follows that of Figures 1 and 2 (in the main paper), with the only difference being the statistical analysis. We will therefore shorten the paragraph in the revised manuscript.

Figure 3/4/7/12/13: The figures are very small, which makes it difficult for the reader to fully appreciate their content.

**Response:** Thank you for the suggestion. We have now slightly enlarged the composite Figures, changed the stippling to more visible cross-hatching, and reduced the spacing between the geopotential height and wind panels, as shown in the revised Figure R2 below.



**Figure R2.** Composite of sea level pressure anomalies (hPa, color shading), wind anomalies at 850 hPa ( $m s^{-1}$ , vectors), and geopotential height anomalies at 500 hPa (gpm, magenta and darkgreen contours) relative to the seasonal mean on the (a, e, i, c, g, k) previous day ( $t-1$ ) and (b, f, j, d, h, l) the event day ( $t$ ) of the warming (a-b, e-f and i-j) and cooling (c-d, g-h and k-l) events during December-February (DJF) in (a-d) historical climate (His), (e-h) future climate (Fut), and (i-l) projected future changes (Fut-His) at a selected grid box in North America (green box). Note that, in (a-h), wind vector anomalies  $\geq 2 m s^{-1}$  and in (i-l), wind vector difference anomalies  $\geq 0.5 m s^{-1}$  are plotted. The dotted and bold contours indicate negative and positive geopotential height anomalies, respectively. Additionally, the cross-hatched area indicates where the ensemble mean of sea level pressure differences exceeds the 95% confidence threshold based on a t-test.

L208: omit the “future” as projected already implies future?

**Response:** We agree and will remove "future" since "projected" inherently refers to future conditions.

L239: Out of curiosity: Do you have an idea why the contribution of the diabatic heating is larger in the CESM-LE compared to ERA5? Could this relate to vertical resolution?

**Response:** Thank you for your curiosity. We will add the new description and analysis in the revised manuscript.

Approximately 40–70% of global land regions exhibit systematic biases in daily temperature and related metrics simulated by CESM-LE compared with ERA5 (Figures R2–R3 in our response to Reviewer 2). While the model captures large-scale spatial patterns (cf. Part I), there are notable discrepancies in the magnitude of anomalies (Figure R3) and in the contributions of underlying physical processes. These biases include an overestimation of DTD extremes, most pronounced in the mid-to-high latitudes (e.g., over North America and large parts of Asia), and an underestimation at lower latitudes during DJF and across most regions during JJA. The overestimation of extremes is linked to advection (primarily in the extratropics) and diabatic processes, with the relative influence of these processes varying across events and regions. This aligns with recent studies indicating an overestimation of daily temperature extremes due to advection and amplified diabatic heating from sensible heat fluxes (Röthlisberger et al., 2025). In contrast, the underestimation in the tropics is primarily associated with biases in the adiabatic contribution and, consequently, in vertical motion, likely stemming from inadequate representation of convective processes and turbulent fluxes (Bao & Stevens, 2021; Bergman & Sardeshmukh, 2004; Stohl, 1998). Similarly, during JJA, the widespread underestimation across most regions results mainly from a combination of underestimated advective and adiabatic processes and also overestimated diabatic processes. These seasonal biases have important implications for targeted model development and improvement.

L245: When mentioning “Rossby wave propagation,” consider providing supporting evidence or omitting the comment.

**Response:** We will revise the wording to avoid any reference to Rossby waves.

L253: I was stumbling across the formulation “This reduction is because ...”

**Response:** We will improve this as “This advective reduction is related to the fact that the future warming at the origin of the traced air masses (at -3d) is 9.4°C at  $t-1$  but only 8.3°C at  $t$  (Figure 5n)”.

L585: “driven by” and “due to” as before. I think you should refrain from implying causality here. Several transition words (e.g., “conversely” in L627, “however” in L583, and “in contrast” in L660) do not seem appropriate in their current context and may be misleading. Revisiting these connectors could help improve clarity and flow.

**Response:** We will revise causal phrasing throughout and reconsider transition words to improve clarity and accuracy.

## References

- Bao, J., & Stevens, B. (2021). The Elements of the Thermodynamic Structure of the Tropical Atmosphere. *Journal of the Meteorological Society of Japan. Ser. II*, 99(6), 1483-1499. <https://doi.org/10.2151/jmsj.2021-072>
- Bergman, J. W., & Sardeshmukh, P. D. (2004). Dynamic Stabilization of Atmospheric Single Column Models. *Journal of Climate*, 17(5), 1004-1021. [https://doi.org/https://doi.org/10.1175/1520-0442\(2004\)017](https://doi.org/https://doi.org/10.1175/1520-0442(2004)017)
- Bieli, M., Pfahl, S., & Wernli, H. (2015). A Lagrangian investigation of hot and cold temperature extremes in Europe. *Quarterly Journal of the Royal Meteorological Society*, 141(686), 98-108. <https://doi.org/https://doi.org/10.1002/qj.2339>
- Hamal, K., & Pfahl, S. (2025). Physical processes leading to extreme day-to-day temperature change – Part 1: Present-day climate. *Weather Clim. Dynam.*, 6(3), 879-899. <https://doi.org/10.5194/wcd-6-879-2025>
- Röthlisberger, M., & Papritz, L. (2023). Quantifying the physical processes leading to atmospheric hot extremes at a global scale. *Nature Geoscience*, 16(3), 210-216. <https://doi.org/10.1038/s41561-023-01126-1>
- Röthlisberger, M., Sprenger, M., Beyerle, U., Fischer, E. M., & Wernli, H. (2025). Advective, adiabatic and diabatic contributions to heat extremes simulated with the Community Earth System Model version 2. *EGUsphere*, 2025, 1-32. <https://doi.org/10.5194/egusphere-2025-5146>
- Stohl, A. (1998). Computation, accuracy and applications of trajectories—A review and bibliography. *Atmospheric Environment*, 32(6), 947-966. [https://doi.org/https://doi.org/10.1016/S1352-2310\(97\)00457-3](https://doi.org/https://doi.org/10.1016/S1352-2310(97)00457-3)