

Reply to Referee #1

Structural warming biases distort extreme rainfall intensification estimates in event attribution

Damián Insua-Costa, Marc Lemus-Cánovas, Martín Senande-Rivera, Victoria M. H. Deman, João L. Geirinhas and Diego G. Miralles

Preprint egusphere-2026-3044 — *Weather and Climate Dynamics*

RC1, 11 June 2026 — <https://doi.org/10.5194/egusphere-2026-3044-RC1>

The referee's comments are reproduced in full and in their original order, followed by our replies.

The manuscript by Insua-Costa et al. focuses on how biases in temperature trends extrapolated from CMIP6 models might lead to the development of incorrect counterfactuals for model-based extreme event attribution studies, taking the Valencia floods of 2024 as a relevant example. The focus is particularly posed on the vertical structure of warming, capable of changing the background stability profile when superimposed to simulations of the event under analysis e.g. in the context of pseudo-global warming approaches. Employing an array of different counterfactual simulations, the authors explain how differences in the dynamical and thermodynamical setup of the counterfactual simulation can substantially affect precipitation totals and, consequently, the attribution outcome.

Reply: We thank the referee for the time spent on the manuscript. We have acted on several specific points, and the sign error identified in comment 4 is a genuine one that we are glad to have had pointed out. On most of the principal criticisms, however, we disagree, and we set out our reasons against each of them below.

While the thermodynamic arguments provided (increased CAPE, increased IWV with global warming) are consistent with the increase in precipitation expected in a warmer climate at a basic level, the dynamical argument based on thermal wind is very likely flawed. Furthermore, with the exception of the changes in stability in top-heavy vs bottom-heavy warming patterns, there is little physical understanding provided to help the reader connect changes in the vertical warming structure and precipitation outcomes. Given the convective nature of the event and the high degree of interaction with orography, small changes in the circulation can result in very large precipitation differences, with maxima of precipitation decaying or shifting away from the region of interest: this implies that the interpretation of precipitation changes between simulations should be done with great care, taking into account the variability within the ensemble.

Reply: On the thermal-wind argument. The passage carried little weight in the submitted paper. It occupies two sentences of Sect. 3.4, was framed as a consistency statement rather than a claim of attribution, and could be deleted without altering a single figure, number or conclusion. The claim of the paper, that CMIP6 models reproduce the integrated magnitude of anthropogenic warming while distorting its vertical structure, and that this propagates into attribution outcomes, rests on Figs. 1–6 and 8, none of which involve thermal-wind reasoning.

Its closing clause (“as a direct consequence of thermal-wind balance”) was nonetheless more assertive than that framing warranted, and a mechanism raised in these terms is better quantified than asserted, so we have done so in comment 2. We also do not accept the physical argument advanced against it, for reasons given under comment 1.

On “little physical understanding”. Section 3.4 of the submitted manuscript is devoted entirely to the causal chain linking the vertical warming structure to the precipitation response, and sets out its steps explicitly. The vertical structure of the warming has two coupled consequences. Under a constant-relative-humidity adjustment it shapes the moistening profile, with bottom-heavy warming concentrating the moisture increase where it matters for convection. It also sets the sign of the lapse-rate change, which is destabilising when the warming is bottom-heavy and stabilising when it is top-heavy. CAPE responds to both, which is why the two perturbation families diverge so strongly in instability, with a median increase of +28 % under ERA5 against +12 % under CMIP6. Each link in that chain is a physical argument supported by diagnostics from the ensemble, not an assertion. We have expanded it in revision following Referee 2’s helpful suggestion, adding the step from instability to enhanced latent heating and stronger updrafts, but we do not think the original was absent.

On precipitation changes and ensemble variability. The study rests on 81 MPAS integrations on a variable-resolution mesh refined to 3 km over the target region, a computational effort well above what is typical in event attribution, where ensembles of a few members at coarser resolution are common. Variability within the ensemble is not merely acknowledged but displayed. Fig. 5 shows all individual members alongside the boxplot and the ensemble mean, and Fig. S8 repeats the analysis over southern Spain rather than the Valencian Community alone. Significance is assessed with a dual criterion, requiring both $p < 0.05$ and $|\text{SNR}| \geq 1$, which is a more conservative standard than is usual in this literature. The concern the referee raises is the one the paper was designed around.

The null hypothesis that what has been observed in the Valencia floods was specific to the event, to the simulations, or to the region, and to reasons other than the vertical structure of the warming, is not excluded convincingly. The generalization attempt is, therefore, not supported enough by the result shown. Given that I feel my comments challenge some very central points of the paper, I have to recommend very major revisions or rejection in its current form.

Reply: The concern that the result might be specific to the event, the simulations or the region is addressed by the design, which comprises six independent CMIP6 models, nine initialisation times, two independent counterfactual constructions (CMIP6-derived and ERA5-derived), two integration domains, and a global assessment of the warming-structure bias in Fig. 8. We also now provide, at the referee’s request, difference maps for both wind (comment 3) and precipitation (minor comment 5) showing that the response is a systematic change rather than a spatial displacement.

In our view, the principal criticisms in this report are not supported by what the manuscript actually contains, and the assessment is unduly severe. The recommendation of rejection rests substantially on the thermal-wind interpretation, which carries no result and whose removal would leave every figure and conclusion of the paper intact. The evidence presented here comprises an 81-member ensemble at 3 km mesh refinement, two independent counterfactual constructions, member-level display of ensemble spread, dual-criterion significance testing, and two integration domains. If that is held to be insufficient to exclude the null hypothesis, we would note that the resulting standard is one that few event-attribution studies currently meet. We are of course willing to be shown otherwise on any specific point, and we address each in turn below.

Major comments

Major comment 1

I believe that the explanation of dynamical changes provided by the authors is not consistent with the basic synoptic meteorology concepts behind thermal wind balance. The explanation provided by the authors for the difference in upper-level flow strength involves changes in the land-sea thermal contrast that would propagate upward and affect free tropospheric winds at 500 hPa (Fig. 5f). Thermal wind balance descends directly from the assumptions of geostrophic and hydrostatic balance, and those assumptions are not satisfied at the typical scales of land-sea thermal contrasts. Such temperature gradients originate other types of mesoscale circulation, such as sea breezes, that are non-geostrophic and limited to the height of the boundary layer with no effect on the upper troposphere. Otherwise, we would have a NE/SW oriented jet stream along the western coast of the Mediterranean Sea forming virtually every afternoon during summer as solar heating warms the surface and the ocean stays relatively cool... which is not the case.

Reply: We respectfully disagree with this argument as a general statement, and we would set out why before describing what we have changed.

Circulations driven by surface thermal contrasts are not in general confined to the boundary layer, and several attain approximate geostrophic balance. Monsoonal low-level jets, such as the Somali and West African jets, are driven by land–sea and land–land thermal contrasts and have cores at 1–1.5 km and above, at or beyond the top of the boundary layer. The California coastal jet, forced by a land–sea contrast sharpened by coastal upwelling, is approximately geostrophic in its “expansion fan” structure. A criterion that rules out an upper-level response to such contrasts on the grounds of horizontal scale alone would exclude these as well.

The governing criteria, in our reading, are two, and neither is horizontal scale. The thermal wind integrates vertically, so what matters is the depth over which the horizontal temperature gradient extends. Geostrophic adjustment proceeds on a timescale of order of the inertial period $2\pi/f$, so what also matters is whether the forcing persists long enough for balance to be established. The two criteria do different work. Persistence alone determines whether a thermally driven circulation can reach approximate balance at all. Depth determines how far up that balanced flow extends.

The referee’s counter-example is consistent with these criteria. The summer Mediterranean land–sea gradient reverses overnight, giving an effective forcing time of 6–8 h against an inertial period of ≈ 19 h at 40°N , and it reverses close to the half-period, so the flow executes an inertial hodograph rather than settling into a balanced jet. It is also confined to a shallow layer, so even the circulation it does drive has little vertical reach. Where a land–sea gradient is persistent, as on the northwestern Iberian and Californian coasts or in the monsoon regions, balanced coastal jets of just this kind do form. The absence of a Mediterranean summer jet is thus evidence about that particular gradient, not about thermal-wind reasoning in general.

Our imposed perturbation differs from a sea breeze on both counts. It is a stationary shift of the background state with no diurnal cycle, so its forcing timescale is effectively unbounded relative to the inertial period. On depth, the warming contrast it imposes is not confined to a shallow surface layer. A shallower one would in any case leave a signature aloft unless the gradient reversed higher up, since the thermal wind integrates the gradient over the whole layer.

We should also be clear about what we are not claiming. The upper-level circulation of this event does not originate in the land–sea contrast. The cut-off low comes from the breakdown of the jet stream, which is itself maintained by the latitudinal temperature gradient. Our argument concerns the modulation of that pre-existing flow: a perturbation that strengthens the land–sea contrast adds a geostrophic increment on top of it, which is what Fig. R1 quantifies.

We have quantified the mechanism (comment 2) and rewritten the two sentences accordingly. The revised text retains the enhanced land–sea warming contrast as one candidate contribution among others, rather than as the explanation.

Major comment 2

Knowing the temperature gradient at one layer should not be used to predict the wind several thousand of meters above it, especially in a region with complex orography (rather than in the simplified set up of idealized jet streams over oceans, in which thermal wind is usually introduced during introductory meteorology classes). If the authors would like to continue with the thermal wind argument, they should at least compute the geostrophic and thermal wind and produce actual evidence that specific differences in low-level thermal wind are the dominant contributors to the 10-15% difference in 500 hPa wind speed depicted in Fig. 5f, which then might impact vertical shear and convection organisation. This would result in a more mechanistic understanding of why changes in vertical warming structure should affect attribution results.

Reply: We have carried out this calculation. From the imposed δT we integrate the thermal-wind relation to obtain the geostrophic increment between 1000 and 500 hPa,

$$\delta \mathbf{v}_g(p) = \frac{R}{f} \mathbf{k} \times \int_p^{p_0} \nabla_h \delta T \, d \ln p',$$

evaluated on the ERA5 grid with f varying with latitude. We note that the horizontal gradient is taken of δT at every level and then integrated, so no single-layer gradient is used to predict the wind aloft. The vertical integral is the whole content of the estimate.

The result is shown in Fig. R1. The implied increment is organised as a coherent south-west to north-east band running from northern Morocco across the western Mediterranean and over the Valencian Community, reaching +6.9 % of the factual 500 hPa wind speed where it is strongest and averaging +3.3 % over the Valencian Community itself. Under the CMIP6 perturbation the same calculation gives +0.1 % there, with no comparable organised structure. The imposed temperature perturbation therefore does, by itself, imply stronger upper-level flow over the region in the ERA5 experiment and essentially none in the CMIP6 one, which is the qualitative contrast the manuscript reported, now quantified. This increment is smaller than the $\sim 13\%$ difference simulated at 500 hPa, so it cannot account for that difference on its own, but it is of the right sign and order and is clearly capable of contributing to it. The revised text presents it as one candidate contribution among others, alongside the latent-heating pathway raised by Referee 2 and a possible change in the position or shape of the cut-off low.

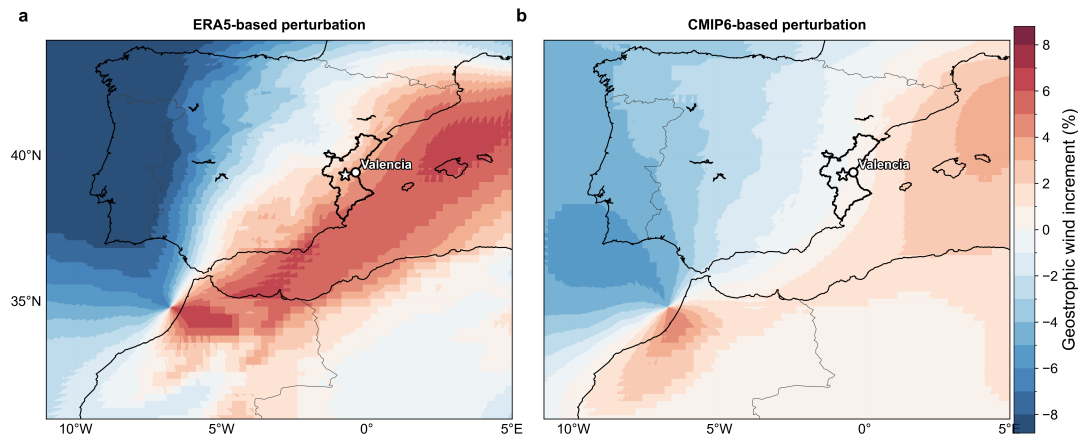


Figure R1. Geostrophic wind increment between 1000 and 500 hPa implied by the imposed temperature perturbation, for (a) the ERA5-based and (b) the CMIP6-based perturbation. The increment is a vector field determined solely by $\nabla_h \delta T$, and is shown projected onto the factual 500 hPa flow, as a percentage of the factual wind speed, so that positive values strengthen that flow. The projection is applied identically to both panels. The outline marks the Valencian Community, the star Turís and the circle the city of Valencia. This figure has been added to the revised Supplement.

Two technical notes. The integration is truncated at the local surface pressure at each grid point, since in parts of the domain the 1000 hPa level lies below ground. And the ERA5 trend, computed on the native 0.25° grid, retains orographic fine structure that the gradient operator amplifies, whereas the CMIP6 signal is interpolated from coarser grids and is already smooth. We therefore smooth δT to 150 km before differentiating, identically for both families.

We should be equally clear about the limits of this calculation. It shows what the imposed thermodynamic state implies geostrophically at initialisation, not that the simulated response follows from it. The perturbation is applied 24–48 h before the event and evolves with the developing cut-off low.

Major comment 3

Lacking a clear explanation of the connection between the vertical warming pattern and the wind, one might come to the simpler hypothesis that weaker upper-level winds over the Valencia region are simply due to a lateral shift of the region with the maximum winds. In Fig. 4 the ERA5 counterfactual features a stronger sea-level pressure gradient with a deeper minimum east of Gibraltar, that enhances easterly winds over the southern portion of Spain (Fig. 4c): a similar shift might be present also at 500 hPa, and would be consistent with the enhanced precipitation in the Murcia/Almería region visible in Fig. 4f. Please provide 500 hPa wind difference maps between simulations to actually show the spatial pattern of wind difference.

Reply: This is a fair and testable objection, and we have tested it. The two hypotheses make different predictions. Under a pure displacement, what is gained on one flank is lost on the other, so the area-integrated change over a region enclosing the shift should be close to zero. We therefore computed the 500 hPa wind difference fields (Fig. R2) and integrated them over the southern Spain domain already used in the manuscript (Figs. S7 and S8).

The result does not support a displacement. Under the ERA5 perturbation the mean change over that southern Spain domain is $+0.63 \text{ ms}^{-1}$, the wind is strengthened over 61 % of its

area, the normalised asymmetry is $+0.49$ where zero would indicate a pure dipole, and the pattern is a coherent band of intensification extending from the western Mediterranean across eastern Iberia rather than a pair of opposing lobes. The CMIP6 ensemble gives -0.17 m s^{-1} and an asymmetry of -0.40 over the same domain, a weak signal of the opposite sign, consistent with its much weaker response. Both the non-zero integral and the band structure hold throughout the event, not only in the daily mean.

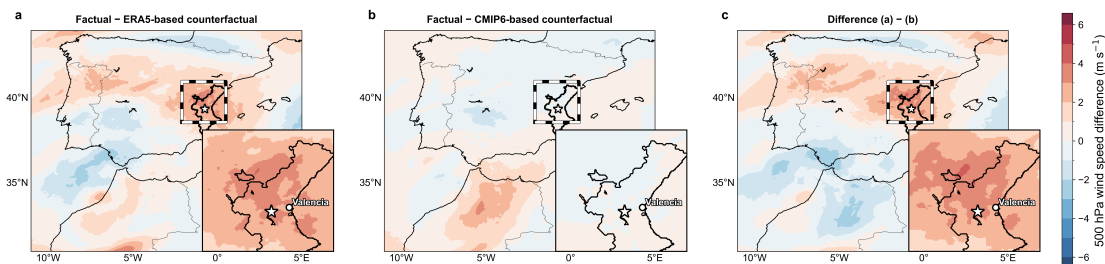


Figure R2. Difference in 500 hPa wind speed on 29 October 2024 (daily mean), factual minus counterfactual. (a) ERA5-based counterfactual, (b) CMIP6-based counterfactual, (c) their difference. Positive values indicate stronger winds in the factual simulation. The outline marks the Valencian Community, the star the location of the observed precipitation maximum at Turís. This figure has been added to the revised Supplement.

We would not go so far as to rule out any contribution from a displacement of the system. The referee’s hypothesis is a reasonable one, and the revised manuscript now lists a change in the position or shape of the cut-off low among the candidate contributions to the upper-level wind response. What the difference maps show is that it does not appear to be the leading one. A displacement acting alone would leave the integrated change near zero and would affect both perturbation families similarly, and neither is what we find.

Major comment 4

The explanation of the results is very difficult to understand and at times contradictory. Let us take the sentences at lines 271-274 and the 500 hPa wind speed as example, but the lack of clarity is widespread across Section 3.3 and 3.4 and in the Conclusions. Here, we know that the ERA5 counterfactual features an increased land-sea contrast and this would correspond (by the authors’ hypothesis) to a weakening of upper-level winds in the counterfactual simulation with respect to the factual (I guess because the thermal wind vector would be roughly aligned with the coastline and point south-westward if land is warmer than the sea, opposing the southerly flow driven by the DANA south of Gibraltar we see in Fig. 4). The $C < F$ is consistent with $(F - C)/F > 0$, as shown in Fig. 5f. The problem is in the wording used: when the authors write “the ERA5-based perturbations yield a significant strengthening of upper-level wind” (line 272) the reader might be confused, because in the ERA5-based counterfactual the upper-level winds are actually weaker than in the factual. This confusion becomes a contradiction in the conclusions, when it is said “The ERA5-based experiment yields enhanced low-level moistening, stronger convective instability and a more vigorous upper-level jet, producing a super-Clausius–Clapeyron precipitation response”. Please revise the explanations to ensure consistency and clarity.

Reply: We agree, and we are grateful for the careful reading. This is an error in our wording, not in the analysis.

We have verified the sign against the source data. The $+13\%$ in Fig. 5f corresponds to

winds that are weaker, not stronger, in the ERA5 counterfactual, exactly as the referee infers. The sentence at line 272 takes the counterfactual as its subject and therefore states the opposite of what the figure shows.

We have rewritten that sentence, corrected the Conclusions so that the strengthening is unambiguously attributed to the factual simulation, and swept Sects. 3.3 and 3.4 for the same ambiguity. The sign convention is now stated explicitly on first use (all relative changes are $(F - C)/F$, so positive values denote a stronger factual response) and repeated in the caption of Fig. 5.

Major comment 5

The generalization beyond Valencia presented in Section 4.3 is also not convincing, in my opinion. What is shown in Fig. 8 is simply that CMIP6 models, in most cities, tend to feature a upper-level warming bias with respect to the estimate derived by ERA5: such a result is consistent with the bias described by Keil et al. (2021) and mentioned in the introduction. However, this does not automatically imply that the precipitation response in London or Los Angeles will follow the same one as Valencia, when a similar approach is applied to study extreme events over those cities. The Valencia event was convectively dominated in the context of a DANA, while extreme precipitation in other parts of the world might be more stratiform and associated with other mechanisms (e.g., atmospheric rivers, cyclones). More physics-based understanding would support better the generalization attempt, also allow to hypothesized for which climatic region the proposed mechanism is expected to be more (or less) relevant.

Reply: We accept part of this and have moderated Sect. 4.3, but we disagree with the premise that the generalisation lacks a physical basis.

We would first clarify what Sect. 4.3 claims. The section does not state that the precipitation response elsewhere would follow the Valencia case. What it says is that in many mid-latitude settings, CMIP6-derived warming is likely to produce weaker precipitation responses than an observationally constrained warming structure, and it already carries an explicit geographical caveat: where the structural bias reverses sign, the same mechanism would act in the opposite direction. The claim concerns the direction of the bias, not the size of the response.

That said, we agree that the strength of the response is event-dependent, and the submitted text did not say so clearly enough. In a predominantly stratiform event, or one strongly forced by the large-scale dynamics, sensitivity to low-level stability will be weaker than in a convectively dominated case such as this one. We have added that qualification.

On the physical basis, however, we would push back. The chain connecting the warming structure to the precipitation response is not event-specific. A bottom-heavy warming profile steepens the lapse rate and, under a constant-relative-humidity adjustment, concentrates the moisture increase in the lower troposphere. Both act to increase convective available potential energy, and both follow from thermodynamics rather than from the particular synoptic setting of this case. A counterfactual built with a top-heavy warming profile will therefore understate instability and low-level moisture wherever it is applied. What varies between event types is how strongly the precipitation responds to that bias, not the sign of the bias itself.

The revised section states this distinction explicitly, and presents the size of the effect as requiring testing on other event types. But the direction of the error introduced by a misrepresented warming profile is a thermodynamic consequence, and we do not think it is

specific to Valencia.

Minor comments

Minor comment 1

Section 3.1: The global-scale discussion presented in this section is not specifically relevant to the western Mediterranean case study and to the conclusions, and is in this sense a deviation with respect to the point of the paper. Consider removing or drastically shortening, by placing the changes in the western Mediterranean in the context of the North Atlantic/European/Mediterranean region only.

Reply: We would keep this section, and we think the choice is defensible.

The structure of the paper is deliberately global-to-regional. We chose to establish the planetary-scale structure of the CMIP6–ERA5 discrepancy first, so that the reader comes to the Mediterranean case already knowing that what follows is an instance of a systematic bias rather than a regional curiosity. The opposite order, from the case study outwards, would have been equally legitimate.

We would add that the global analysis has value in its own right. We are not aware of a comparable level-by-level comparison of CMIP6 against ERA5 warming structure across the globe, and some readers may find it useful independently of the attribution case. On length, Sect. 3.1 is $\sim 11\%$ of the body text, in line with the other substantive sections rather than disproportionate.

We would prefer to retain the global framing, but this is a matter of presentation rather than of substance, and if the referee is not persuaded we are willing to shorten the section.

Minor comment 2

Line 144-145: the role of the lapse rate feedback in inducing Arctic Amplification is being currently reconsidered critically, consider rephrasing this sentence (see Caballero, R., and T. M. Merlis, 2025: Polar Feedbacks in Clear-Sky Radiative–Advection Equilibrium from an Airmass Transformation Perspective. J. Climate, 38, 3399–3416).

Reply: Rephrased. The sentence no longer presents the lapse-rate feedback as an established primary driver of Arctic amplification, but as one of several contributing mechanisms whose relative importance remains under debate, citing Caballero and Merlis (2025).

Minor comment 3

Line 171: the 700 hPa level is not usually considered “mid-tropospheric”, as it still bears substantial influences from the surface (e.g., see the North Atlantic “warming hole”). What about 600 or 500 hPa?

Reply: We take the point, and Referee 3 raises it as well. We would note that 700 hPa is conventionally the boundary between the lower and middle troposphere, so it sits at the edge of both labels and neither is strictly wrong. We agree, nonetheless, that “mid-tropospheric” invites the reading that the level is free of surface influence, which at 700 hPa it is not.

We have therefore avoided the ambiguity rather than swapping one imprecise label for another. Where a genuinely mid-tropospheric reference is needed we now use 500 hPa,

which is also the level at which the wind response is diagnosed, and elsewhere we refer to 700 hPa by its pressure rather than by a layer name.

Minor comment 4

Line 193: it looks to me more vertically homogeneous with a relative maximum around 850 hPa, while the magnitude of the warming trend near the surface does not really stand out. The CMIP6 trend instead is clearly top-heavy.

Reply: Corrected. The ERA5 trend is better described as vertically homogeneous with a relative maximum near 850–925 hPa than as surface-intensified.

The contrast on which our argument depends is unaffected. What matters is the ERA5 profile relative to the clearly top-heavy CMIP6 one, and hence the sign of the lapse-rate change, rather than the presence of a near-surface maximum in ERA5.

Minor comment 5

Fig. 4: Accumulating precipitation in presence of very extreme peaks might lead to interpretation problems, as changes in single precipitation maxima should not be seen as representative of the behaviour of large area averages. The color scale of precipitation does not increase monotonically, further complicating the interpretation. Please provide spatial precipitation difference maps between simulations to actually show that there is a systematic reduction of precipitation and not a shift or just a loss individual maxima.

Reply: We have produced these maps (Fig. R3) and they do show a systematic reduction rather than a displacement.

On the colour scale, we would prefer to keep the one used in Fig. 4. Its class boundaries follow the discrete intervals widely used in radar and precipitation products, and the non-monotonic steps are what make threshold-crossing visible: in the zoom panels of Fig. 4d–f, for instance, the contraction of the area exceeding 70 mm between the factual and the ERA5 counterfactual is immediately apparent, which a continuous brightness ramp would render as a gradual fade. The new difference maps use a perceptually uniform diverging scale precisely because they serve the complementary purpose of showing the sign and magnitude of the change rather than where thresholds are crossed. We think the two are better kept distinct, but if the referee still prefers a perceptually uniform scale in Fig. 4 we are happy to make the change.

Applying the same displacement test as in comment 3 to accumulated precipitation, over the Valencian Community the factual ensemble exceeds the ERA5 counterfactual by +20.5 mm on average, or +39 % expressed as $(F - C)/F$ in the convention of Fig. 5, and the counterfactual is drier over 96 % of the area (and over 98.8 % of the points where rain actually falls), the normalised asymmetry is +0.99, and the centroid of the precipitation field shifts by only 6 km, of the order of the grid spacing. Over the wider southern Spain domain of Figs. S7 and S8 the contrast is smaller but of the same sign, with a mean increase of +14.5 % under ERA5 against +4.7 % under CMIP6.

On the concern that single maxima may not represent area averages, our diagnostics point the other way. Over the Valencian Community the domain mean falls by 39 %, the 99th percentile by 32 % and the single maximum by 30 %, so the reduction is not carried by a few grid points.

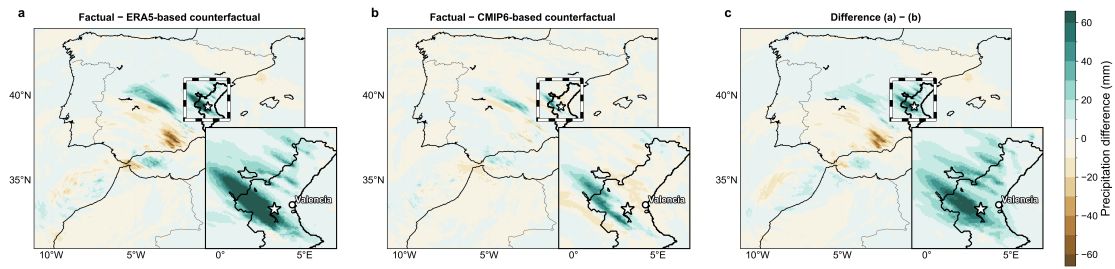


Figure R3. Accumulated precipitation on 29 October 2024, factual minus counterfactual (mm). (a) ERA5-based counterfactual, (b) CMIP6-based counterfactual, (c) their difference. Positive values indicate more rainfall in the factual simulation. The outline marks the Valencian Community, the star the location of the observed maximum at Turís. Insets zoom on the affected region. This figure has been added to the revised Supplement.

Finally, we would note that the manuscript already reports both of the analyses the referee asks for. Fig. 5 shows the two metrics side by side over the Valencian Community, the spatial mean (P_{mean}) in panel a and the spatial maximum (P_{max}) in panel b, so the behaviour of the area average and that of the storm peak can be compared directly. Figs. S7 and S8 give the corresponding analysis integrated over southern Spain, referred to in Sect. 3.3. In the revision we have made these cross-references more prominent.

References

- Caballero, R. and Merlis, T. M.: Polar feedbacks in clear-sky radiative–advective equilibrium from an airmass transformation perspective, *Journal of Climate*, 38, 3399–3416, <https://doi.org/10.1175/JCLI-D-24-0031.1>, 2025.