First reviewer
The response to the reviewer's recommendations is below in blue.

The manuscript applies four complementary detection and attribution diagnostics (SNR/ToE, Knutson & Zeng categorical classification, FAR, and Ribes et al. additive-decomposition optimal fingerprinting) to seasonal temperature and precipitation trends over 1951-2020 in nine sub-regions of the Western Mediterranean, using CMIP6 DAMIP single-forcing experiments and a performance-based model subset. The headline result, a robust GHG-dominated warming signal against a precipitation signal that cannot be distinguished from internal variability, is methodologically careful but is not in itself a new finding for the region.

My main recommendation is that the authors either sharpen what is genuinely new in this analysis or extend it in a direction that adds new information, specifically toward extremes and circulation-conditioned attribution, both of which the manuscript only gestures at. The methodology is sound and the multi-method cross-validation is a strength, but the manuscript currently reads as a careful, more granular confirmation of already-known results. I recommend major revisions, focused on (i) clarifying the novelty claim, (ii) extending the analysis to temperature extremes (and precipitation extremes if feasible) using the same DAMIP framework, and (iii) at minimum discussing, and ideally testing, whether circulation-conditioning changes the precipitation attribution conclusions, with the literature suggested below.

Major comments

1. The result that Mediterranean/European warming is robustly attributable to GHG forcing, partially offset by aerosols, is already established (e.g. Stott, 2003; Feng et al., 2022, both cited). Likewise, the conclusion that Mediterranean precipitation trends stay within the envelope of internal variability is already well documented, including by some of the present authors (Campos et al., 2025, JGR Atmospheres, same lead author, same sub-regional clustering) and by Vicente-Serrano et al. (2025, Nature) and Seager et al. (2025), all cited in this manuscript. The present study's contribution is therefore methodological refinement (sub-regional resolution, DAMIP decomposition, model screening, multi-method cross-checking) rather than a new substantive result. That's a defensible contribution, but it should be argued explicitly: what conclusion follows from this finer-grained analysis that could not already be drawn from Campos et al. (2025) or the broader literature? As written (L86-98), the novelty claim rests on "DAMIP being under-utilised" and "a consistent framework enabling direct comparison," which is a methodological framing, not a scientific one. I'd ask the authors to state directly, e.g. in the Conclusions, one or two concrete things we now know about WM hydroclimate that we didn't before this study.

Two literature pointers the authors may want to add, neither currently cited:

IPCC AR6 WG1, Chapter 11 (Seneviratne et al., 2021), https://doi.org/10.1017/9781009157896.013. This is the relevant AR6 chapter for detection and attribution of regional trends and extremes. The manuscript currently cites "IPCC

(2022)" (L74-76), which in the reference list is the WG2 Impacts/Adaptation/Vulnerability volume, to support a statement about CMIP6 bias being "small enough to support detection and attribution studies." That's a WG1-type claim and should be sourced to WG1 (Ch. 3, Eyring et al., 2021, or Ch. 11).

Tang et al. (2018), ACP, "Dynamical response of Mediterranean precipitation to greenhouse gases and aerosols," https://doi.org/10.5194/acp-18-8439-2018. Decomposes GHG vs. aerosol contributions to Mediterranean precipitation dynamically, directly overlapping with the exercise in Sect. 3.3/Fig. 8.

2. All diagnostics in this paper (Sect. 2.1-2.3) are built on monthly-mean temperature and precipitation, aggregated to seasonal Sen's-slope trends. No extreme indices (TXx, warm-spell duration, Rx1day, R95p, consecutive dry days, etc.) are computed anywhere. This matters for two reasons. First, the attribution literature increasingly finds that the anthropogenic signal in extremes, particularly temperature extremes, emerges more clearly and earlier than in seasonal means, because thermodynamic (Clausius-Clapeyron-type) responses are generally better constrained across models than the dynamical/circulation responses that dominate mean precipitation trends. Second, the authors themselves seem to recognise this: the very last paragraph before the Conclusions (L600-603) states that "future work should also further investigate other aspects of sub-regional climate change in the WM, particularly the behaviour of extreme events under continued global warming and their links to changes in atmospheric dynamics and physical processes," but this is offered as a single deferred sentence rather than pursued. Given that the DAMIP single-forcing runs used here are typically archived at daily resolution, at least a temperature-extremes analogue of the existing Sect. 3.2 (e.g. SNR/FAR for TXx or a heatwave-magnitude index) seems within reach with the data already assembled, and would meaningfully strengthen the paper's claim to novelty (point 1 above). I'd consider this the single highest-value addition to ask for.

3. SNR, FAR, Knutson-Zeng classification, and Ribes optimal fingerprinting all test for a shift in the marginal distribution of trends. None of them ask whether the (non-)detection of a forced precipitation signal depends on the realised atmospheric circulation. This isn't a minor omission for the WM: several of the manuscript's own citations point directly at circulation as the channel through which forcing affects precipitation. Seager et al. (2025), cited, argue that the observed Mediterranean drying via a positive NAO trend over the last 65 years is "an extreme outlier" relative to the CMIP6 ensemble, i.e. the disagreement between models and observations may be a dynamical (circulation) problem rather than evidence that internal variability alone explains the trend. Olmo et al. (2024) and Olmo et al. (2025), both cited, both by co-authors of this manuscript, already provide circulation-regime classifications and circulation-based model filtering for the Euro-Mediterranean. None of this machinery is brought to bear on the attribution problem here; the model-selection step in Sect. 3.1 filters on spatial-pattern correlation of trends, not on circulation fidelity.

I'd recommend the authors either (a) repeat at least one diagnostic (e.g. FAR, or the additive decomposition) stratified by circulation regime/season-type, using the regime classification from Olmo et al. (2024), to test whether the "non-detection" of a forced precipitation signal is a pooling artefact that disappears once circulation is controlled for; or (b) at minimum, discuss this as a methodological limitation in Sect. 4, citing the conditional/storyline attribution literature that reaches different, sometimes opposite-signed, conclusions for Mediterranean precipitation extremes when conditioning on circulation:

Zappa et al (2015), ERL, "The dependence of wintertime Mediterranean precipitation on the atmospheric circulation response to climate change," https://doi.org/10.1088/1748-9326/10/10/104012

Ginesta et al (2023), Climate Dynamics, "A methodology for attributing severe extratropical cyclones to climate change based on reanalysis data: the case study of storm Alex 2020," https://doi.org/10.1007/s00382-022-06565-x

Faranda et al. (2022), WCD, "A climate-change attribution retrospective of some impactful weather extremes of 2021," https://doi.org/10.5194/wcd-3-1311-2022

Chericoni et al (2025), npj Climate and Atmospheric Science, "Unravelling drivers of the future Mediterranean precipitation paradox during cyclones." Shows mean WM precipitation can decline while cyclone-conditioned precipitation increases, exactly the kind of signal a marginal-trend framework like the one used here can't see.

Points 2 and 3 together suggest a coherent path forward: extend the attribution framework from seasonal means to (a) extreme indices and (b) circulation-conditioned subsets, for temperature in particular. This would both address the novelty concern (point 1) and respond directly to the authors' own "future work" sentence.

4. Sect. 3.1 reports real and substantial limitations: spatial-pattern correlations against ERA5 stay below 0.6 for all models for temperature and below 0.5 for all models for precipitation (L317-326); the multi-model mean smooths out the observed winter precipitation dipole (L264-268); only 7 of 11 models pass even the weak $r > 0$ screening criterion for precipitation (L330-336). These are acknowledged in the Discussion (L522-528, L580-592) but the Abstract (L10-25) reports the precipitation non-detection as a clean physical conclusion, with no indication that CMIP6's skill at reproducing the observed spatial structure of precipitation trends is itself a limiting factor. I'd suggest adding one sentence to the Abstrct, e.g.: "CMIP6 models show limited skill in reproducing the observed spatial structure of precipitation trends, which constrains confidence in the precipitation attribution results."

Minor comments

L74-76: see point 1, citation mismatch (WG2 report cited for a WG1-type claim).
L86-98: please state explicitly how this study's conclusions differ from or add to Campos et al. (2025), which uses the same sub-regional clustering and the same observational datasets.
L330-336: the pragmatic $r > 0$ selection criterion is very permissive (it retains models with essentially no skill beyond the sign of the spatial pattern). Please justify why this threshold, rather than e.g. a minimum correlation, was chosen, and report how sensitive the attribution results (Sect. 3.3) are to it. The manuscript states a "sensitivity analysis without model selection" was performed (L590-592) but results are "not shown"; given how central the

filtering step is to the paper's claimed methodological contribution, this should be shown, at least in the Supplement.
L544-547: the "exploratory analysis of the period 1981-2020" suggesting a shift in aerosol response is mentioned but "not shown". If this is a meaningful result it should be included, even briefly, since it speaks directly to the brightening/dimming literature already discussed.
L600-603: as above, this sentence anticipates exactly the kind of analysis (extremes plus circulation) that would strengthen the present submission. Consider whether at least a partial version belongs in this paper rather th an being deferred entirely.
Conclusions (L605-619): consistent with point 1, the conclusions largely restate the abstract. Consider adding an explicit statement of what is new relative to prior basin-scale work and the authors' own sub-regional work.

Response:

We thank the reviewer for this careful and constructive revision of our manuscript. We appreciate the recognition of the methodological robustness of the study and agree that the current version might not sufficiently distinguish between methodological novelty and the new scientific understanding provided by the analysis. We will work further to clarify this distinction and define more explicitly the contribution of the study relative to previous work, focusing on the importance of multi-method attribution.

Regarding the first major comment, we agree that the manuscript builds directly upon Campos et al. (2025), particularly through the use of the same climatically homogeneous sub-regions and observational datasets. However, while Campos et al. (2025) primarily focused on the detection and characterisation of observed seasonal trends, the present study goes a step further by formally assessing their attribution using CMIP6 DAMIP single-forcing experiments and several complementary detection and attribution methods (FAR, Optimal Fingerprints). We intend to make this distinction much clearer throughout the Introduction and Discussion in the revised manuscript. The regional framework applied in this manuscript allows us to examine whether attribution conclusions remain consistent across areas with substantially different climatological conditions, seasonal variability, and observed trend characteristics. We agree that the current manuscript might not sufficiently articulate what can be learned from this finer spatial perspective, and we intend to state more explicitly the concrete conclusions that emerge from the sub-regional and multi-method analysis.
We will also incorporate the recommended WG1 references and Tang et al. (2018), and correct the current mismatch involving the citation of the IPCC WG2 report.

Regarding the second major comment, we agree that extending the analysis to climate extremes would strengthen the manuscript and provide additional information beyond the seasonal-mean trends. We therefore intend to make every reasonable effort to include ETCCDI indices for both temperature and precipitation, subject to the availability and consistency of daily data across the required DAMIP experiments. The intention is to select a limited number of representative indices and apply a subset of the detection and attribution methods already used in the manuscript. To preserve the existing seasonal perspective of the study, we plan to calculate and analyse seasonal rather than annual extreme indices.

This extension will focus on adding insightful information without substantially increasing the manuscript's length or altering its overall structure.

Regarding the third major comment, we agree that circulation-conditioned attribution is a highly relevant and promising direction. This approach could provide a substantially different perspective on Mediterranean precipitation trends. Nevertheless, we believe that implementing a robust circulation-conditioned attribution analysis would exceed the intended scope of the present manuscript. Such an extension would require introducing and evaluating an additional methodological framework, including selecting and validating circulation classifications, assessing model skill in representing circulation regimes and their frequencies, and treating the uncertainties associated with regime-conditioned attribution. We are concerned that adding this full analysis would substantially change the scope and balance of the paper rather than constituting a limited extension of the current framework. We will therefore not propose a full circulation-conditioned attribution analysis in the revised manuscript. It will be considered as a prioritised follow-up study within our research group. However, we fully accept the reviewer's recommendation that this issue should receive considerably more attention. We intend to expand the Discussion to explicitly distinguish marginal attribution from circulation-conditioned and storyline approaches, and to acknowledge that the non-detection of a forced precipitation trend does not exclude the existence of forced thermodynamic responses within particular circulation states or event types. We will discuss the possibility that the pooling of dynamically different situations may obscure such responses and will incorporate the literature recommended by the reviewer.

Regarding the fourth major comment, we agree that the limited ability of CMIP6 models to reproduce the observed spatial structure of precipitation trends should be more clearly reflected in the Abstract and in the interpretation of the attribution results. We intend to add an explicit statement to the Abstract noting that limited model skill constrains confidence in the precipitation attribution conclusions. We will also revise the wording throughout the manuscript to clarify that the forced precipitation signal cannot be robustly distinguished from internal variability within the evaluated CMIP6 framework, rather than implying that the analysis demonstrates the absence of a forced response. Following a recommendation from the second reviewer, we will divide the Discussion section into subsections. The first will focus on the strengths and limitations of the CMIP6 DAMIP-based detection and attribution framework. The second will focus on the attribution of temperature and precipitation changes across the nine Western Mediterranean subregions, emphasising regional and seasonal differences. And, depending on the results, we will include a third subsection on the attribution of extreme indices.

We also agree with the minor comments and intend to address them in the revised manuscript.
We thank the reviewer again for the constructive feedback and for identifying these opportunities to sharpen our manuscript's contribution.

Second reviewer
The response to the reviewer's recommendations is below in blue.

First of all, I would like to thank the Editor and the Authors for the opportunity to review this interesting and well-motivated paper. I learnt a lot about how the models were selected to create the subset and DAMIP experiments.

The paper addresses an interesting topic, namely the Contrasting Detection and Attribution of Temperature and Precipitation Changes in the Western Mediterranean from CMIP6 DAMIP Experiments. This subject has not been extensively investigated. Therefore, this study represents a valuable contribution to existing literature. However, the manuscript requires major revisions.

The motivation for this study is clear, and the conclusions are both interesting and supported by a novel methodology and set of metrics. However, I have some concerns regarding the presentation of the methodology and results. In particular, it is challenging to follow the methodological workflow and the progression of the results throughout the manuscript. Furthermore, precipitation analysis requires careful assessment, particularly in subregions 2, 3, and 5, where observed and historical trends exhibit opposing signals, specifically in summer.

Major comments:

1. The models exhibit larger biases in precipitation estimates, particularly over subregions 2, 3, and 5 (Fig. 1h and Fig. 8b). Morin 2011, based on GPCC observations (https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2010WR009798), identified statistically significant precipitation trends over the Mediterranean basin (see Fig. 1b). These areas also show large relative precipitation changes when normalized by mean annual precipitation (Figs. 4a and 4b in Morin, 2011). Therefore, it would be valuable to further discuss the reasons why the models exhibit larger biases in these regions and whether their inability to reproduce these observed trends is related to model limitations, internal variability, or the representation of regional precipitation processes.
2. The signal-to-noise ratios in Figs. 7a and 7b remain between −0.5 and 0.5 for all subregions, suggesting weak attribution signals. This may be influenced by the length of the record or by the magnitude of the detected trends. Please discuss this issue in relation to Morin (2011) (see Figs. 5 and 6). Considering that the analysis uses a 69-year record, the minimum detectable absolute trend should be around ~10 mm/decade (see Fig. 7 in Morin, 2011). The authors should clarify whether the low signal-to-noise ratios arise from record-length limitations or other factors affecting trend detection.
3. I suggest including a flowchart that provides a simple overview of the methodology, as well as the associated metrics, tests, and the processes used for the detection and attribution of temperature and precipitation changes. Such a figure would help readers navigate the manuscript more easily and follow the progression of the analysis without having to repeatedly move back and forth between sections.
4. The Discussion section could be divided into two subsections: (1) the advantages and limitations of climate change detection using CMIP6 DAMIP models, and (2) the

attribution of the detected changes across the nine subregions of the Western Mediterranean. This structure would allow for a clearer assessment of both the strengths and limitations of the modeling framework, while also highlighting regional differences in the attribution of climate change signals.

5. I suggest including a map as the final figure that summarizes the key metrics, statistical tests, and drivers responsible for detecting and attributing changes in precipitation and temperature. The map should highlight the nine subregions of interest and emphasize that, although all these areas are part of the Western Mediterranean region, they are influenced by different climate-change drivers and therefore exhibit distinct patterns of change.

Minor comments:

1. I suggest using different color palettes in Fig. 5 and Fig. 7: one palette to represent the subregions (a,b) and another palette to represent the attribution categories (c,d).
2. Figures 5a,5b, 7a,7b: The trends are difficult to distinguish clearly. Consider adding a zoomed-in inset for the 2005–2015 period to better illustrate the differences and the divergence among the subregional trends.
3. Figures 5c and 5b are predominantly characterized by attribution category 2 (detectable and attributable increase). Please consider using a classification scheme ranging from −4 to 4, including the corresponding category names or acronyms. Furthermore, if subregional differences are identified, clarify why all subregions exhibit dominance of the same attribution category.
4. Figures 7a and 7b: Please consider restricting the y-axis range to −2 to 2 in order to better highlight the differences among the subregions. Additionally, please discuss the results presented in Figs. 7c and 7d, particularly regarding the occurrence of attribution categories −3, −2, 0, 1, and 4. Explain the possible mechanisms or factors that lead to the dominance of these categories.

Response:

We thank the reviewer for the positive and constructive assessment of our manuscript. We appreciate the recognition of the relevance of the topic, the value of the DAMIP-based framework, and the novelty of the methodological approach and set of diagnostics. We also agree that the methodological workflow and the progression of the results may not be sufficiently clear in the current version. We will work to improve the structure and presentation of the methodology, strengthen the interpretation of the precipitation results, and make the regional differences more accessible to the reader.

Regarding the first major comment, we agree that the discrepancies between observed and simulated precipitation trends in some subregions deserve a more detailed discussion. We will expand the relevant sections to discuss the possible origins of these differences, based on the current literature. We will also discuss more explicitly whether the observed-model disagreement should be interpreted as a model limitation, a manifestation of internal variability, or a combination of both.

We also thank the reviewer for suggesting Morin (2011). We will incorporate this study into the Discussion and compare its results with ours.

Regarding the second major comment, we will discuss more explicitly the extent to which these low SNR values may reflect the length of the period studied, or the relatively small magnitude of the forced precipitation response, and the large internal variability. We agree that the 1951-2020 period may limit the emergence of weak forced precipitation signals. We will therefore strengthen the discussion of record-length dependence and clarify that low precipitation SNR values should not be interpreted solely as evidence of an absent forced response. We will also clarify the conceptual relationship between the minimum detectable trend discussed by Morin (2011) and the SNR diagnostic applied in our study.

We fully agree with the suggestion to include a methodological flowchart. We will add a schematic figure or table providing a concise overview of the complete workflow. The flowchart will also indicate the main purpose of each metric and how the different analyses are connected. We expect this addition to make the methodological progression substantially easier to follow.

We also agree that the Discussion would benefit from a clearer internal structure. We will reorganise it into the two subsections suggested. The first will focus on the strengths and limitations of the CMIP6 DAMIP-based detection and attribution framework, including model performance and model selection. The second will focus on the attribution of temperature and precipitation changes across the nine Western Mediterranean subregions, emphasising regional and seasonal differences and the contribution of the different external forcings. Depending on the results, we will include a third subsection on the attribution of extreme indices, as suggested by the first reviewer.

We appreciate the suggestion to include a final synthesis map. We will explore including a summary figure showing the dominant detection and attribution outcomes for temperature and precipitation across the nine subregions.

Regarding the minor comments, we will review the colour palettes, legends, axis ranges, category labels, and captions to improve the readability of our figures.

We thank the reviewer again for these helpful suggestions. We believe that the proposed revisions, particularly the addition of a methodological flowchart and the reorganisation of the Discussion, will make the manuscript considerably clearer and strengthen the interpretation of its main findings.