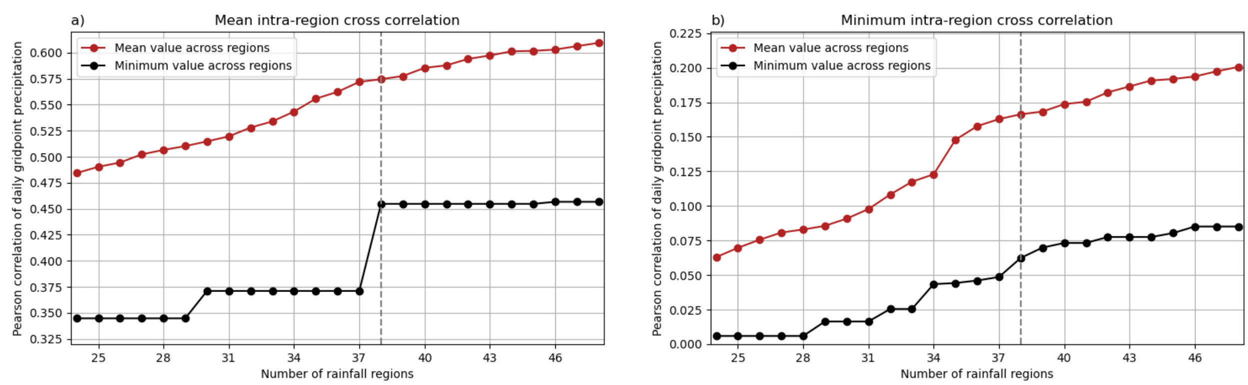
Response to reviewer comments on "Understanding biases and changes in European heavy precipitation using dynamical flow precursors”

We thank the reviewers and the editors for their helpful suggestions and for insightful questions. These have inspired additional supplementary analysis and modifications to the manuscript. We respond to each comment individually below.

Firstly, we need to note an error in our description of the region definition, which we have now rectified. In the original manuscript, starting on L118, we stated: *“Pragmatically, we chose the region number to be as small as possible while still ensuring that precipitation at every gridpoint had a correlation of at least 0.45 with every other gridpoint in its region.”*

We now correctly state: *“Pragmatically, we chose the region number to be as small as possible while still ensuring that the average correlation of precipitation between each pair of gridpoints in a region was at least 0.45, for each region.”*

Associated with this error, Figure S1 was mislabelled and has now been replaced:



(Fig. S1. a) Mean (red) and minimum (black) values across regions of the average correlation of daily precipitation between gridpoints within each region, computed from ERA5 over 1979-2023. b) Mean and minimum values across regions of the minimum correlation between any two gridpoints within each region.)

This error does not impact the validity or interest of our results, as we are already using a more fine-grained spatial discretisation than is typical for such pan-European analyses. However, we apologise for overstating the degree of precipitation co-variability in our regions.

More happily, we can also report the addition of a new visualisation tool for the analyses shown in this paper, available here:

https://uib-precursors-cmip6-interactive.hf.space/app

We plan to expand this tool to include additional models. We now mention this on L529 in our Conclusions as follows:

"To make the details of this analysis as accessible and interpretable as possible, we have developed an interactive interface to visualise our results: https://uib-precursors-cmip6-interactive.hf.space/app"

Response to reviewer 1

This manuscript introduces a flow-dependent decomposition framework for analyzing heavy precipitation biases and forced changes in two major large-ensemble climate simulations (CESM2 LENS2 and MPI-GE). The paper classifies synoptic states using region-specific multivariate precursor patterns (Z500, U850, V850), enabling a novel partition of precipitation errors into dynamical (synoptic forcing occurrence) and conversion (local-scale processes converting forcing to precipitation). The authors apply this to 38 regions across Europe and all seasons.

Overall, this paper is impressively comprehensive, and the results reveal new insights into compensating biases, dynamical controls, and the physical mechanisms behind future changes in heavy precipitation frequency. The paper is clearly written, well structured, and methodologically rigorous. It will be of high interest to the climate dynamics, hydroclimate, and impacts communities. The identification of widespread compensating biases and distortions in forced changes is especially valuable for model evaluation, downscaling, and storyline applications.

I find the manuscript to be a strong and valuable contribution suitable for publication after minor revisions. My comments below aim to enhance clarity, interpretation, and broader applicability.

…

In short, this is a well-designed and insightful manuscript that advances our understanding of flow-dependent heavy-precipitation frequency biases and changes. With clarifications on terminology, broader discussion of intensity considerations, and guidance on ensemble-size requirements, the paper will be even more impactful and accessible to a wide interdisciplinary audience.

We're grateful for the reviewer's positive opinion of our work!

The use of the term "precursor" may inadvertently mislead readers into assuming that these synoptic patterns are derived from lead-time composites—that is, flow anomalies that occur prior to heavy precipitation. As currently implemented, the composites are constructed on the heavy precipitation day itself, meaning they represent heavy precipitation synoptic patterns rather than true temporal precursors. Although I understand the desire for continuity with earlier work (Dorrington et al., 2024), a short clarification of terminology would be helpful. Explicitly *stating that the term "precursor" reflects the synoptic-scale controls on heavy precipitation rather than a temporal lead signal would prevent misinterpretation. You may also wish to reference the terminology adopted in Dorrington et al. (2024) while clarifying its slightly different meaning here.*

We agree this is an important point to clarify. As the reviewer identifies, we do prefer to maintain the terminology to be consistent with related work, but we have made several changes to the manuscript to explain our naming convention and to avoid misunderstanding:

On L124 we now write:

"The flow precursor framework developed in Dorrington et al. (2024a) is the basis of our decomposition. On a high level, the approach identifies the synoptic conditions corresponding to past heavy precipitation events using composite analysis, and defines time-evolving `precursor activity indices' based on those composites. While the flow precursor framework can be used for time-lagged dynamical fields, here we only use fields co-occurring with precipitation to form `lag-0 precursors', but retain the terminology for consistency with Dorrington et al. (2024a,b)"

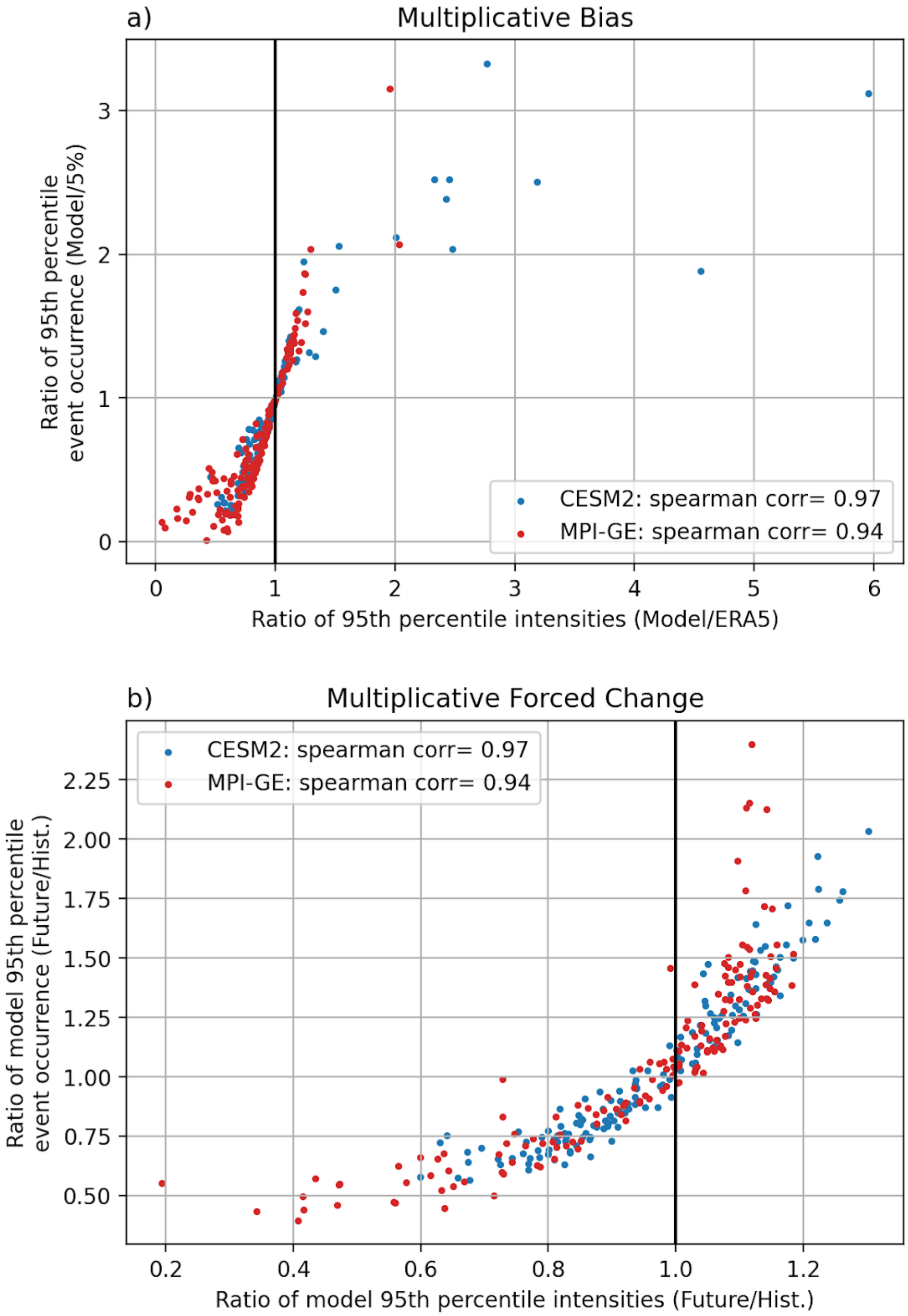
And on L541:

" In subtropical regions, and to some extent the midlatitude summer, heavy precipitation is less organised by synoptic dynamics and has stronger upscale feedbacks. In these contexts the use of time-lagged precursors as in Dorrington et. al. 2024a may prove important to clearly establish causality."

"A minor weakness of the manuscript is that it focuses almost exclusively on the occurrence (frequency) of heavy-precipitation days. Beyond Figure 2, there is relatively little discussion of the associated intensity (magnitude) biases or projected changes. While the emphasis on categorical heavy precipitation is scientifically justified, real-world hazards depend jointly on both frequency and magnitude. I recommend adding a short contextual discussion. Alternatively, the title and narrative could be adjusted slightly to emphasize "heavy precipitation frequency" to better reflect the paper's scope.

We agree that precipitation is too complex a variable/hazard to sum up completely in a single metric. To maintain a manageable scope for the paper, we limited ourselves to p95 occurrence. This variable choice is conceptually well aligned to the inherently probabilistic link between synoptic dynamics and surface hazards.

However, given the shape of rainfall distributions, the intensity and occurrence of extreme events scale near-monotonically with each other: so we would not actually gain much additional information from an intensity analysis. This is evidenced in the new Supplementary 3, which shows that occurrence biases/forced changes in 95th percentile ERA5 events are highly correlated with intensity biases/forced changes in the model's own 95th percentile events:



Reviewer Fig 1: Comparison of multiplicative biases in the intensity of 95th percentile precipitation and exceedance of the ERA5 95th percentile precipitation threshold, for each region and season. b) Comparison of multiplicative forced changes between future and historical scenarios in the intensity of 95th percentile precipitation and in exceedance the ERA5 95th percentile precipitation threshold.

For the most extreme intensity biases this close link is sometimes violated, generally if the probability of an extreme event is no longer small in the model, but in general the two variables are closely linked.

On L111 we now write:

“For the two models we consider, biases and changes in heavy precipitation occurrence are closely related to intensity (c.f. Supplementary Fig 3). “

On L174:

“We focus on the occurrence probability of heavy precipitation (Fig. 3a-h)), but there is a strong monotonic relation between deviations in extreme precipitation intensity and occurrence (c.f. Supplementary Fig. 3)”

And on L539:

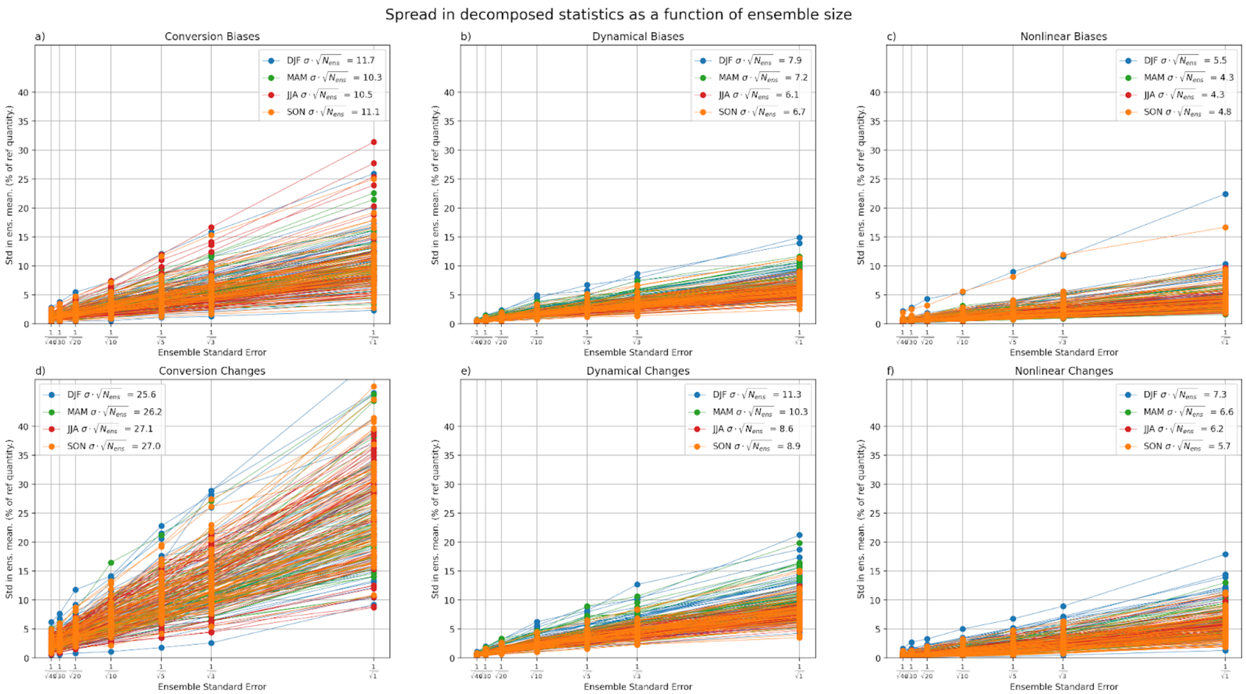
“While we consider only the occurrence of heavy rainfall rather than rainfall intensity, the two quantities are closely related in models. Our results therefore also expose where the quantity of heavy precipitation may change.”

The proposed dynamical and conversion decomposition relies on well-populated synoptic bins and robust estimates of the conditional probability as well as dynamical and conversion biases. The authors use 50-member ensembles from CESM2 and MPI-GE, but many modeling centers and CMIP6/CMIP7 simulations provide only a handful of ensemble members. To help readers assess the broader applicability of the framework, I encourage the authors to estimate the minimal ensemble size required to obtain stable decomposition results. Even a qualitative guideline or a discussion of how the required size depends on the choice of the number of synoptic bins (e.g., K = 5 K = 10) would greatly improve the usability of the method for other studies.

Thanks for this nice suggestion. Applying the approach used here to smaller ensembles/single members is an area of active investigation for us, and so we have taken this as an opportunity to analyse this question systematically; we now include 3 additional Supplementary Figures within a Supplementary Discussion on Internal Variability and Parameter Dependence. Specifically, we quantify:

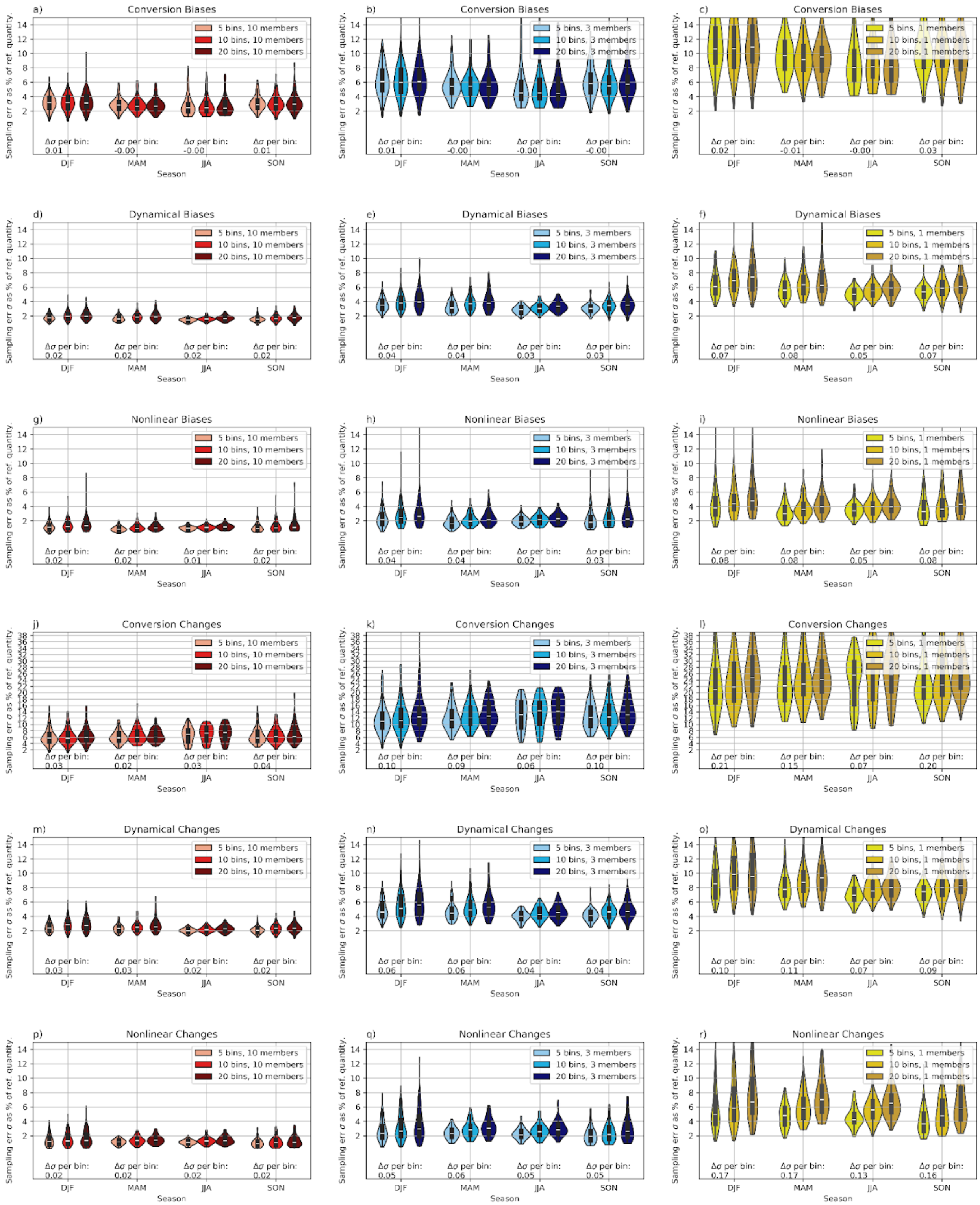
- **How does the sampling error in decomposed statistics scale with the size of the model ensemble?**
- **Does using a different number of synoptic bins change the estimates of decomposed statistics? (i.e. does the estimate of the conditional pdfs converge for K=5,10,20?)**
- **How does using a smaller/greater number of synoptic bins impact sampling variability?**

These questions are answered using data for all seasons, regions and models.



Reviewer Figure 2: Sampling error (1 standard deviation) in ensemble mean decomposed quantities as a function of ensemble size. Results are shown for all seasons, regions and models. A near-linear relationship between theoretical standard error, $\frac{1}{\sqrt{N_{ens}}}$, is apparent, with the estimated slope computed across both models and all regions indicated for each season and quantity.

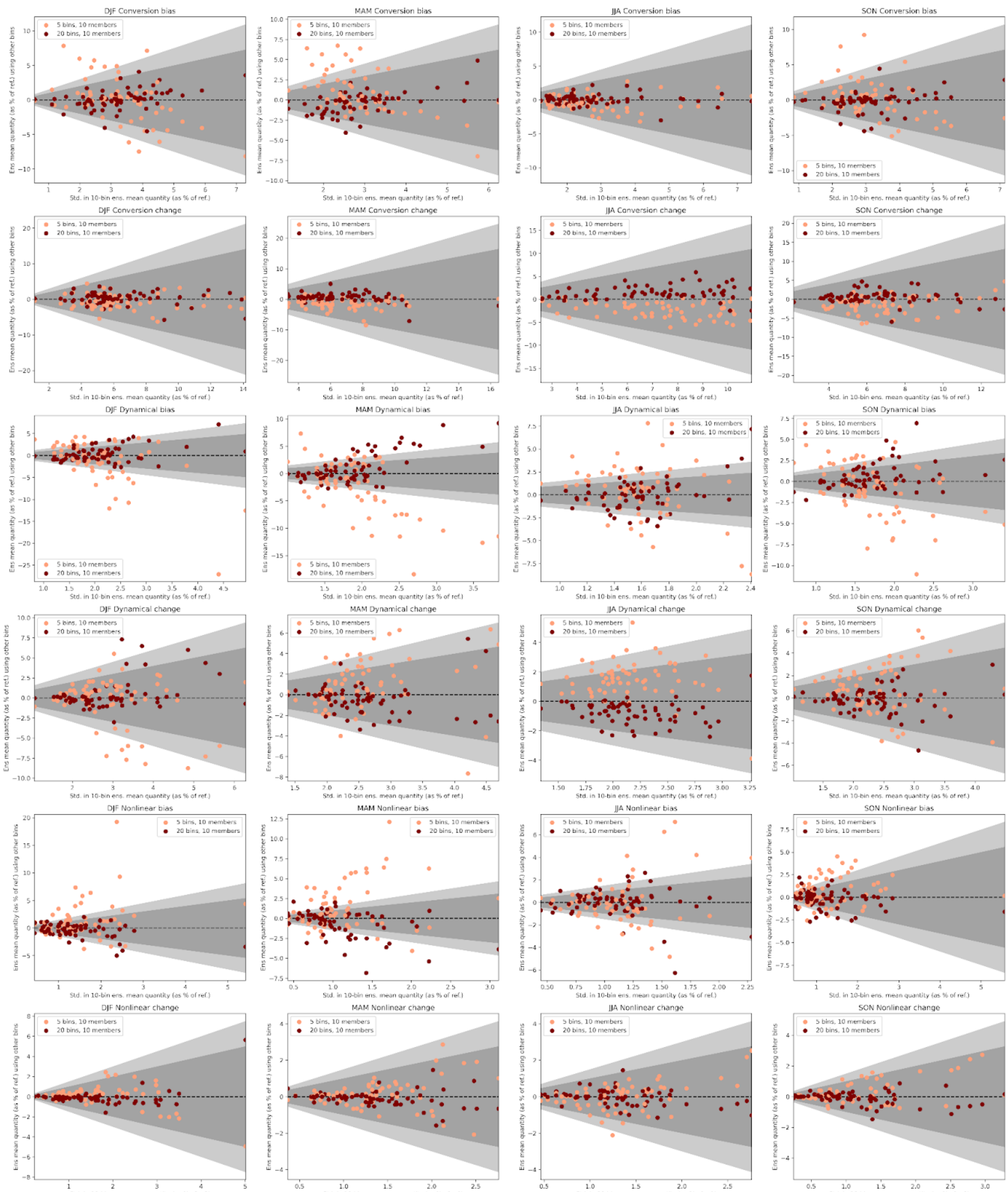
This shows that the sampling error scales as expected with the size of the ensemble, and gives some quantitative guides as to how large a signal you should be able to robustly detect with ensembles of different size, or a single member simulation.



Reviewer Figure 3: Sampling error in ensemble mean decomposed quantities as a function of the number of ensemble members used, and the number of synoptic bins used in the precursor decomposition. Sampling errors are disaggregated by season and are shown as distributions across regions. Columns show the dependence on number of ensemble members: 10 (left), 3 (centre) or 1 (right) member/s. Colours within a panel indicate the number of synoptic bins used. Beneath each set of distributions, annotations show the estimated linear relationship between sampling error and number of synoptic bins used.

This shows that sampling error increases with number of bins, and that this effect is proportionately greater for small ensembles, but that this is a small effect relative to the impact of ensemble size: a 5-bin decomposition of a single member simulation has larger sampling error than a 20-bin decomposition of a 3-member simulation, and the same is true for a 3-member and 10-member simulation. This is perhaps unsurprising as even bulk precipitation metrics can be subject to substantial internal variability in single member simulations.

This figure also provides some quantitative guides: for example, to robustly detect conversion-driven changes in a single-member simulation (panel L), they would have to be greater than approx 25%: a change in event probability from 0.05 to <0.0325 or >0.0625.



Reviewer Fig 4: Difference between decomposed quantities computed using 5 (pale dots) or 20 (dark dots) synoptic bins and those computed using 10 synoptic bins, shown as a function of sampling error in the 10-bin ensemble mean estimate. Dark/light grey shading shows the +/-1 and +/-1.5 std error intervals, with differences outside these intervals indicating a meaningful impact of bin number on the decomposition.

This shows that using 5 synoptic bins can systematically misestimate the magnitudes of e.g. MAM dynamical and nonlinear biases, and DJF conversion changes. While the difference in estimated quantities between 10 and 20 synoptic bins is smaller than between 5 and 10 bins, we do see alterations in the assignment of e.g. JJA forced changes between dynamical and conversion contributions.

Synthesising our interpretation of these results:

- **We advocate the use of 10-member and 3-member ensembles for analyses like ours, and emphasise that only large (> approx 25%) differences in event probability in single-member simulations should be considered meaningful.**
- **We would recommend caution in using fewer than 10 synoptic bins: the marginally decreased sampling error can be associated with a biased decomposition as a result of the coarser estimate of the conditional probability distribution P(H|S).**
- **For ensembles of more than 10 members it may be useful to consider even more synoptic bins (e.g. K=20) as we find the decomposition has not fully converged for K=10. The main implication for our own results is that the forced drying change in JJA may be more dynamically driven than our decomposition suggests.**

We now mention these results (in brief form) in the main manuscript in the following places:

On L213:

"ERA5, each bin is equally likely to occur by construction, and bins labelled with larger K indicate a higher conditional probability of heavy precipitation. Using a coarser discretisation can slightly reduce sampling uncertainty (see supplementary discussion) while a finer discretisation will more accurately estimate dynamical impacts. We use 10 bins, discretising S into deciles."

On L532:

"Our results showed high variability in future changes across ensemble members, emphasising the importance of ensemble projections. For single member or small-ensemble simulations, limiting the flow decomposition to a smaller number of synoptic bins (e.g. $K=5$) can decrease sampling error by $\approx20\%$ but at the cost of decreasing the accuracy of the decomposition. Indeed, for large ensembles a larger number of synoptic bins may be justifiable, as supplementary analysis has revealed small but systematic refinements in the decomposition when using $K=20$, for JJA in particular. "

<u>Response to reviewer 2</u>

The authors present an interesting framework to attribute climate model heavy precipitation biases over European regions to biases in different drivers and residual (nonlinear) terms, as *well as to attribute changes in heavy precipitation to changes in drivers respectively. The framework is comprehensive, useful and applicable to other regions and variables. It also allows for the identification of compensating biases.*

We're glad the reviewer had a generally positive opinion of the paper.

The authors are surprisingly ignorant of existing literature on trend and bias attribution. While the latter is indeed still limited, a large body regarding the first exists. A key methodology in this respect is dynamical adjustment, which aims at identifying (and removing) trends caused by large-scale circulation changes (Deser et al., 2016; Smoliak et al., 2015, see also Sippel et al., 2019, Vautard et al., 2023, and IPCC AR6 WG1 Chapter 10 for applications; it could be that some of the papers cited in this manuscript are also examples).

Line 25: I am missing a discussion of large-scale circulation errors (and projection uncertainties) along the line of the discussions in Shepherd, 2014, in particular given the scope of the paper.

A paper of particular importance here is the one by Pfahl et al (2017), decomposing changes in extreme precipitation into dynamic and thermodynamic components. While these studies have a different purpose than this one here, the approaches still deserve to be made explicit. The body on attributing local surface climate model biases to large-scale biases is much less developed, but there also initial studies exist (e.g., Respati et al., 2024 on drivers of tropical rainfall biases;

We have completely rewritten our introduction, which now reads as follows:

"*Precipitation is one of the most important processes in the Earth system. Dynamically, it is a major source of diabatic heating;societally, it shapes global ecology and agriculture and, in its extreme form, is responsible for some of the most deadly and damaging weather events (EEA, 2022). Understanding how precipitation will change in a warmer world is therefore a question of key importance. Yet simulating precipitation remains a major challenge. While the spatial distribution and temporal variability of mid-latitude precipitation in global climate models has steadily improved (Du et al., 2022), deficiencies remain (Abdelmoaty et al., 2021), contributing to large uncertainties in future projections.*

These persistent model deficiencies highlight a need for methods that can identify where errors arise within the diversity and complexity of precipitation processes. There are many kinds of precipitation (broadly categorisable as stratiform, convective or orographic), each the result of non-linear interactions between processes across scales. In the mid-latitudes, synoptic-scale variability—the passage of weather systems—sets the potential for precipitation by modulating the availability of moisture and steering low-level winds. Converting this potential for precipitation into precipitation itself involves finer scale processes: mesoscale organisation of convection, boundary layer, coastal and orographic interactions, sub-diurnal heating and cloud microphysics. Many of these conversion processes must be parameterised as they simply cannot be resolved on the O(100km) grids of climate models. Purely statistical correction of the *resulting net rainfall bias is insufficient to credibly inform downstream climate applications (Addor et al., 2016; Maraun et al., 2017). Rather, the sources of biases should be first understood before calibration is considered (Maraun, 2012). However, process understanding in climate simulations can be challenging to obtain especially when, as for rainfall, the processes vary significantly across time and space.*

One approach to building better understanding of model behaviour has been to decompose climate signals into dynamical and thermodynamic contributions. The various decomposition frameworks, which target different goals, have produced seemingly different answers about the importance of dynamical drivers, largely because the choice of which dynamics to focus on strongly shapes the conclusions. For example, the dynamical adjustment approach uses circulation analogues to isolate (circulation-induced) internal variability from forced thermodynamic signals (Deser et al., 2016; Sippel et al., 2019), thereby reducing uncertainty in the latter Shepherd (2014). It reveals substantial dynamical contributions to trends in regional mean and extreme temperature (Deser et al., 2016; Terray, 2021; Vautard et al., 2023) and in monthly precipitation (Guo et al., 2019; Doane-Solomon et al., 2025), but may be unreliable for understanding daily extreme rainfall (Thompson, 2025).

Regime-based decomposition approaches target these shorter timescales explicitly, defining dynamics in terms of dominant modes of synoptic variability in a region, e.g. (Cassano et al., 2007; Cattiaux et al., 2013). However, several regime-based decompositions found that forced dynamical changes in regime frequency are far less important for understanding climate model precipitation trends than the changes in precipitation intensity within regimes (Driouech et al., 2010; Fischer et al., 2025) possibly due to the strong internal variability of daily rainfall within regimes (Gerighausen et al., 2024). Finally, decompositions based on local precipitation dynamics have been applied on the global (Held and Soden, 2006; O'Gorman and Schneider, 2009), monthly (Respati et al., 2024) and daily (Pfahl et al., 2017) scale, typically defining dynamics at the gridpoint level in terms of vertical velocity. These approaches connect model behaviour directly to moist dynamical theory, but can obscure the synoptically-modulated relationship between humidity and ascent, and the role of dynamics in setting local thermodynamic properties through e.g. airmass advection.

Our work extends this growing literature on decomposition approaches by characterising dynamics in terms of flow precursors: multivariate characterisations of the weather patterns most relevant for heavy precipitation, specific to a particular region and season (Dorrington et al., 2024a). As high-impact heavy precipitation is often driven by relatively uncommon and spatially localised weather patterns, this 'bottom up' approach can identify a stronger dynamical conditioning of the events than traditional regime or area-analogue methods. Precursors are used to decompose both historical biases and future changes into contributions from synoptic-scale dynamics and from the fine-scale conversion of dynamical forcing into precipitation. This decomposition framework allows us to explicitly account for the flow- and scale-dependent nature of model biases and their interaction with forced changes, and therefore assess not only whether a model reproduces heavy precipitation realistically, but whether it does so for the right reasons. The approach scales efficiently to large datasets such as large, multi-model ensembles, distilling the rainfall-relevant dynamics to a small number of *scalar indices. The resulting diagnostics are thus not only tools for model development, but also offer a practical framework for visualising and interpreting projections and for extracting usable climate information from imperfect models — a contribution we aim explicitly at the downscaling and climate services communities.*

Section 2 introduces the datasets used, our region definitions, and a summary of the flow-precursor approach. Section 3 summarises known results on the bulk representation of precipitation in our two large ensembles and on the ERA5 climatology of heavy precipitation, providing context for later sections. The formalism for the precipitation decomposition is introduced gradually, alongside demonstrative examples: the decomposition of biases is introduced in Sect. 4, the decomposition of forced changes in Sect. 5, and the interactions between biases and forced changes in Sect. 6. A self-contained theoretical discussion of the precipitation decomposition is given in Appendix A. Section 7 synthesises and discusses our key results while Sect. 8 provides a summary and forward perspective."

More general, statements about "previous literature" (line 241) should be backed up by actual references to the literature.

All such statements have been removed or had relevant citations added.

2. I am not quite sure whether equation A3 (line 579) really is useful to understand the influences of model biases on the representation of trends. In many situations, model biases may be time invariant, but, e.g, feedback processes can induce time-varying climate model biases (e.g. Maraun, 2012). Does equation A3 give useful results in such a situation? As far as I understand, biases are implicitly assumed to be time invariant.

We thank the reviewer for the question.

We found it enlightening to consider this from the perspective of Maraun 2012, and we appreciate the reviewer pointing us to that paper. There, the sources of time varying biases are divided into sensitivity-related changes (e.g. due to cloud or land surface feedbacks), apparent changes arising from sampling variability, and mixture-related/ flow-dependent changes from shifts in weather patterns. This third category of change is what we explicitly account for through our use of flow-dependent bias and trend terms δ, ξ, Δ and α. The second category of change is minimised in our case as we use large ensembles and we quantify the impact through bootstrapping of all our statistics. Only sensitivity-related changes are a possible source of time-varying biases which eq. A3 cannot account for.

Our assumption of constant bias for a given weather pattern is less stringent than a constant bias assumption, and perhaps is the best 'one-size fits all' assumption we can make, but we have made sure to acknowledge the relevant caveats.

We now introduce the initial discussion of bias-trend interactions in the main text as follows, on L339:

“Within Eq. 4 are products between flow-dependent forced changes and flow-dependent biases, which can be thought of as either non-stationary biases or spurious forced signals. As Eq. 4 is the numerator of $\widetilde{\beta}$, these distortions also impact the bulk estimate of forced precipitation changes. We quantify the impact of this distortion in Sect. 6. However, if we accept the flow-dependent forced changes as credible estimates of true climate response and if we have no prior reason to expect a model's flow dependent biases to be non-stationary, then as part of our decomposition we can drop these bias-change cross terms. We then obtain a calibrated estimate of the overall heavy precipitation change, β.”

When analysing forced changes below, we use this calibrated estimate, but the effect of this calibration should not be overstated. The models' estimates of the flow-dependent forced changes may still ultimately be incorrect---and indeed given disagreements between models, most must be. However these corrected forced changes are at least physically consistent with synoptic dynamics and their link to precipitation in the current observed climate.

And on L445, we begin the discussion of bias-trend interactions by saying:

“If a model is very severely biased in a particular region we may not trust even a corrected estimate of the model's future forced changes as discussed above.”

In the Appendix on L582, we introduce Eq. A3 by saying:

“As the bulk change is simply the aggregate of changes under different conditions, then errors in the bulk change error can only be smaller if the flow-dependent errors cancel each other out. This would be a shaky scenario to rely upon. In Section A2 we consider the possibility of time-varying biases model bias”

And include a new section in the Appendix Section A2, entitled "Time evolving biases":

Eq. A3 is based on the assumption that model biases in the occurrence of synoptic weather patterns, δP_{S_k}, and in the occurrence of heavy precipitation under a weather pattern, ξ_k, are stationary in time. However as pointed out in e.g. Maraun (2012), a new climate state might result in models being biased in new ways if, for example, climate change forces a shift in the ratio of different cloud types and a model has biases in those cloud processes. To address this, we can extend Eq. A3 to include a further set of changes in conversion and dynamics, $\Delta\xi_k$ and $\Delta\delta P_{S_k}$ respectively:

$$\widetilde{P}_H^* = \sum_k \left(P_{H|S_k}\left[1+\xi_k\right]\left[1+\Delta\xi_k\right]\left[1+\alpha_k\right]\right)\cdot\left(P_{S_k}+\delta P_{S_k}+\Delta\delta P_{S_k}+\Delta P_{S_k}\right) \tag{A10}$$

These terms can be interpreted in two ways: we could consider $\delta P_{S_k} + \Delta\delta P_{S_k}$ as representing a time evolving bias superimposed on the true climate response ΔP_{S_k}, or we could consider a modelled climate response $\Delta P_{S_k} + \Delta\delta P_{S_k}$ with true and spurious components, superimposed on a stationary bias δP_{S_k}. The same consideration can be given to the conversion terms. As such, non-stationary bias and spurious climate trends are mathematically equivalent and can be viewed as interchangeable ideas in the absence of any concrete physical hypothesis for a particular model and region/season.

If we view $\Delta\xi_k$ and $\Delta\delta P_{S_k}$ as spurious changes, but acknowledge that in general we have no way to disentangle them from the true forced changes, α_k and ΔP_{S_k}, then we can set them to 0, and return to Eq. A3. However, if through prior analysis we have obtained some emergent constraint or isolated a known process-dependent bias then this can be used to specify any or all of the terms $\Delta\xi_k$ and $\Delta\delta P_{S_k}$ as a known quantity. These known non-stationary biases/ spurious trends can then be grouped with the historical bias terms, and Eq. A4 can be applied as written.

3. Is it useful to apply PCA to create the scalar index defining the extremeness of the precursors? PCA is linear and based on the correlation matrix and may not capture the asymmetries and tail behaviour in the "precursor" time series. This choice should at least be discussed.

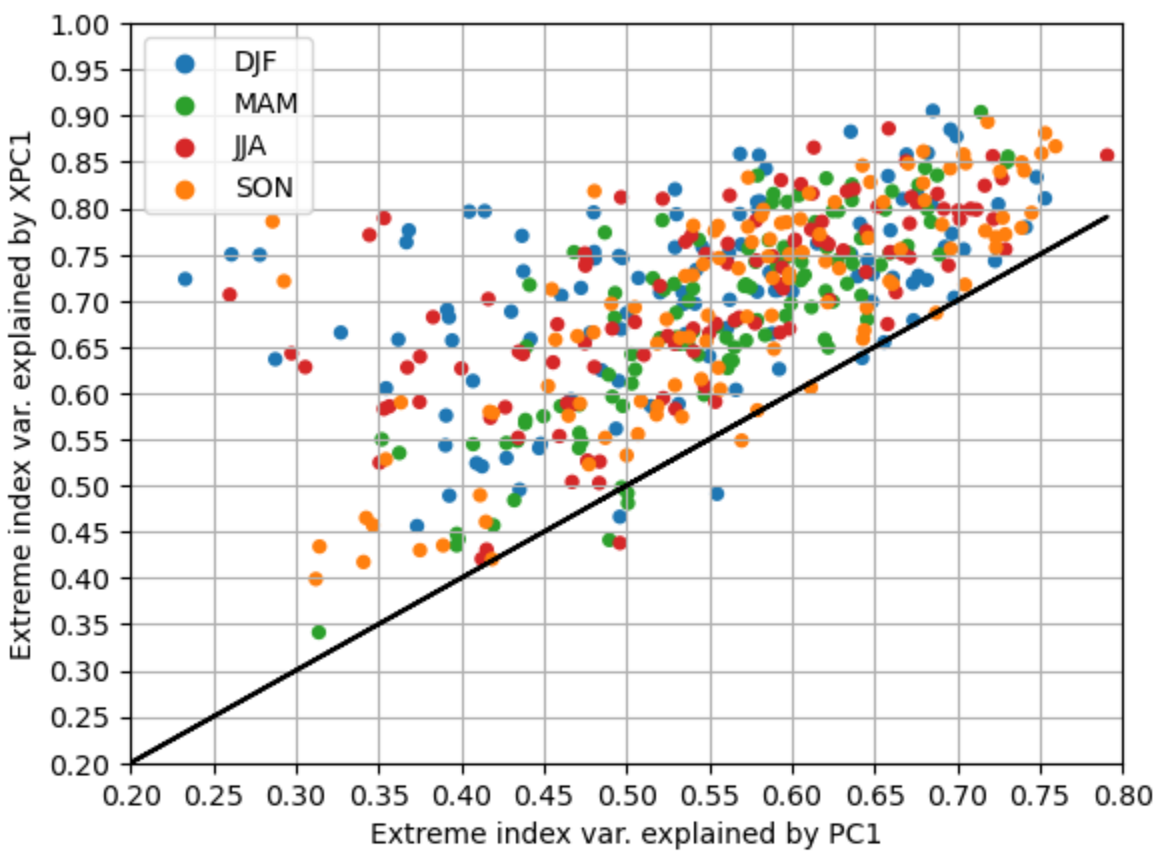
Thanks for another thought provoking question. Applying PCA is useful because it allows us to obtain a single scalar index on which we can condition rainfall; the existing SI Figure 2 demonstrates the benefits of this approach over a single univariate index.

A multidimensional conditioning, or a non-linear dimensionality reduction focused on extremes could potentially help us gain more insight. However, as PCA is a well-known method for dimensionality reduction, and as we introduce several methodological novelties in this paper, we wanted to avoid further complexity in either methods or data visualisation.

While on the one hand, the strong relationship between dynamical indices and precip we observe (e.g. in Fig 4h, 5b and 6b) shows that this approximation has proven 'useful enough', we have taken this opportunity to quantify the role of non-linear relationships between the Z500, U850 and V850 precursor indices.

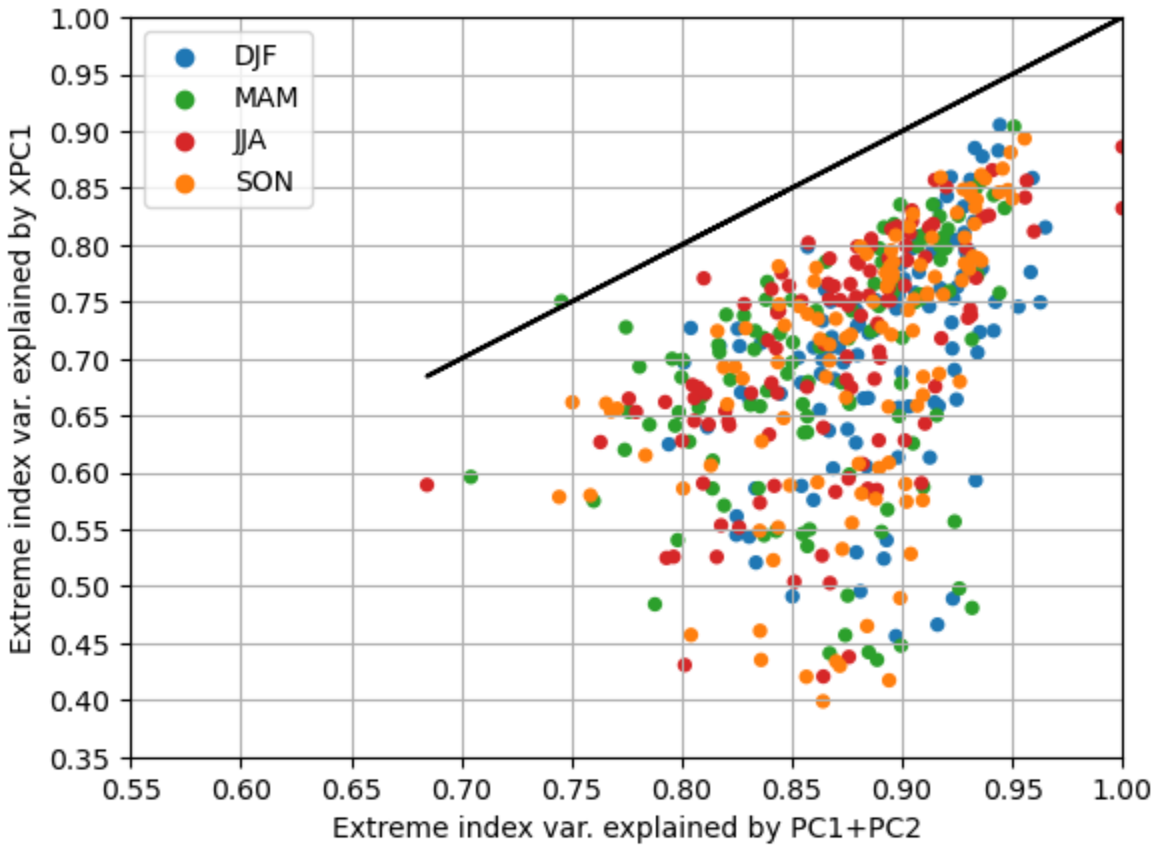
Our method is as follows:

- **For each season and region, we separate ERA5 precursor indices into an ~3000 day training dataset and ~1000 day test dataset.**
- **We compute the first principal component (denoted PC1) of the 3 precursor indices using the training dataset.**
- **For comparison, we again compute the first PC in the training dataset now using only days when the index for a specified variable >1.5 (denoted XPC1). That is, we apply PCA only on days when the specified dynamical precursor was strong, using approximately ~7% of the days in the training data for each case. Conditioning on each of the 3 variables (z500, u850 and v8500) means we compute 3 versions of XPC1 for each region and season.**
- **PC1 aims to capture the dominant climatological covariance structure between the dynamical indices, whereas each XPC1 is focused on the covariance structure only under extreme, rain-favouring circulations. If the covariance structure is flow invariant, then PC1 and XPC1 will be equally effective at explaining variance in extreme circulations in the test dataset.**
- **To test whether this is the case, we compute the variance explained by PC1 and XPC1 for days in the test dataset when index i >1.5. We plot the two explained variances against each other for all regions, and seasons:**



Reviewer Fig. 5, showing variance explained in strong precursor days for a PC computed on all days (PC1) or only on strong days (XPC1).

As the data lie above the diagonal in almost all cases, we clearly see that there is nonlinearity in the inter-variable correlations under strong dynamical conditions which simple PCA is not capturing. This could motivate the use of a nonlinear dimensionality reduction technique in future work, although we observe that using two PCs is superior to using 1 XPC, which may provide a simpler way forward:



Reviewer Fig. 6, showing variance explained in strong precursor days for two PCs computed on all days (PC1) or for only 1 precursor computed on strong-index days (XPC1).

We don't include this level of detail in the manuscript for reasons of focus and brevity, but we now identify this more clearly as a methodological simplification that could be addressed in future work, writing on L211 in our methods:

" Reducing three precursors to a single index removes information, and principal component analysis cannot account for any nonlinearity in the covariance between precursors. Nevertheless, we use this simplification here to maintain a manageable analytic scope."

4. I am not convinced by the term "precursor". Precursor definitely has a temporal prediction aspect to it. E.g., there is a community working on identifying precursors to natural hazards such as earth quakes or health issues such as epileptic seizures. If I understand correctly though, the precursors are evaluated on the same day as the precipitation event, i.e., there is *no prediction in time. I would suggest to replace precursor by predictor, which is commonly used in regression analysis and does not necessarily imply predictions in time. Another, even more neutral term would be covariate.*

We agree that this terminology needs clarification, as suggested by reviewer 1 as well. We do prefer to keep the terminology of 'precursors', which reviewer 1 points out would maintain consistency with the work of Dorrington et al. 2024a,b and Dorrington & Messori 2026. In general, the methodology is oriented to time-lagged dynamical forcing, and our use only of 'lag 0' patterns here is a matter of convenience. Even at lag 0, the synoptic control of mid-latitude precipitation is fairly causally direct on the scales we consider here. To avoid misunderstanding and to make this explicit we now make the following comments in our manuscript:

On L124 we now write:

"The flow precursor framework developed in Dorrington et al. (2024a) is the basis of our decomposition. On a high level, the approach identifies the synoptic conditions corresponding to past heavy precipitation events using composite analysis, and defines time-evolving `precursor activity indices' based on those composites. While the flow precursor framework can be used for time-lagged dynamical fields, here we only use fields co-occurring with precipitation to form `lag-0 precursors', but retain the terminology for consistency with Dorrington et al. (2024a,b)"

And on L541:

" In subtropical regions, and to some extent the midlatitude summer, heavy precipitation is less organised by synoptic dynamics and has stronger upscale feedbacks. In these contexts the use of time-lagged precursors as in Dorrington et. al. 2024a may prove important to clearly establish causality."

Line 38: "enhancing some processes disproportionately" sounds odd.

Line removed as part of rewrites.

Line 77: evaluate, not validate. A climate model cannot be validated, but only evaluated.

Changed to 'evaluate'.

Line 137: "described there in full" is not necessary.

Removed as suggested.

Line 191: "categorical occurrence". This is tautological and sounds odd or pretentious.

Changed to "occurrence probability".

Line 241: then cite the literature!

Now done.

Line 252: "whereas MPI-GE struggles to sufficiently convert any synoptic precursor into heavy precipitation". This sounds overly negative and subjective. The difference between the two models is quantitative, not qualitative.

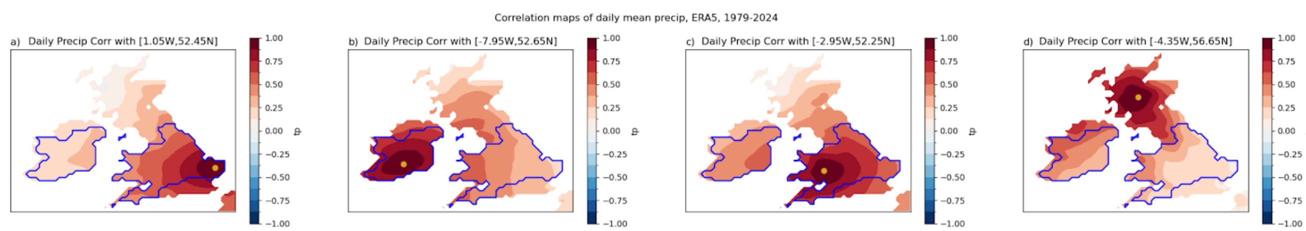
Reworded to “whereas MPI-GE exhibits markedly low heavy precipitation conversion for the strongest 40% of precursor flows.”

Line 253: "bias budget" sounds strange.

We favour maintaining this terminology, as it is usual for discussing decomposed quantities, and emphasises that we don’t exclude any residual contributions.

Figure 3: I am surprised by the inferred regions: e.g. heavy precipitation in East Anglia (highest for north-easterly flow) behaves very different to precipitation over the West of England, Wales or Ireland (highest for south-westerly flow). Why are they combined in the same region?

We show some results for the UK and Ireland to help explain the regional grouping in more detail. The quantitative basis for our region definition is to divide the whole Euro-Mediterranean domain into a fixed number of regions, while minimising the grouping together of uncorrelated points. Given the scale of the domain and the number of regions, Ireland and East-anglia actually are not so different in their precipitation variability, which we evidence with the below correlation maps:



Precipitation in East Anglia (Panel a)) shows correlations of 0.5-0.6 with precipitation in central England and >0.25 in eastern Ireland. Precipitation in Ireland (Panel b)) meanwhile shows reasonably high correlations with Wales and central England. Further, precipitation in the midlands (Panel c)) shows correlations above 0.5 with much of England, Wales and east Ireland, but not Scotland, which is only weakly correlated with the rest of the UK and all but the North West of Ireland (Panel d)).

In the context of balancing variability within the other European regions, the assignment of East Anglia and much of Ireland to a region shared with England and Wales serves to avoid Scotland being grouped with much more weakly covarying southern regions.

For further context, increasing to 39 regions results in a split of the joint Finland-Sweden region which is several times larger than the UK. Even if we go to 49 regions (the highest number we trialled), the central England domain remains undivided:

the algorithm prioritises further subdividing precip over the Alps and along the Dalmatian coast.

Figure 4: the labeling is odd. The i) (meaning 1) can easily be misinterpreted as the letter i. In the caption, replace quiver by arrow. You are talking to climate scientists, not mathematicians.

We have changed quiver to arrow, and we have modified the row labelling in the Figure to avoid confusion with the panel labelling.

Figure 5a,b: add ERA5 to label Psk

We have made this label change, also in Figure 6.

Response to editor comment

I may add a specific comment about your quote from the study by Fischer et al. (2025) [which I co-authored]. In L57 you wrote "… previous work finding little role for changes in dynamics, concluding that they "are of secondary importance for explaining climate change signals in [precipitation]"(Fischer et al., 2025). In the next sentence you then write "However the choice of which dynamics to focus on strongly impacts the conclusions of a decomposition." I could not agree more with this statement, and I therefore think that your way of quoting Fischer et al. is slightly misleading. Fischer et al., when writing about the "secondary importance" wrote explicitly "… frequency changes of weather regimes are of secondary importance for explaining climate change signals in precipitation". The mention of "weather regimes" here is important, the statement is not about "dynamics in general" as your quote in L57 might imply. The understanding of Fischer et al. is that dynamics is clearly more than weather regimes (in particular variations on smaller scales not captured by the regimes), which is in line with your statement about the "choice of which dynamics to focus on …". I would be glad if you modified this paragraph in your paper to avoid the impression that Fischer et al. regarded dynamics in general as of little importance.

We apologise for the unintentional implication given by this quote.We agree that the formulation was misleading. In our newly structured introduction we avoid a direct quote, and say on L44:

"However, several regime-based decompositions found that forced dynamical changes in regime frequency are far less important for understanding climate model precipitation trends than the changes in precipitation intensity within regimes (Driouech et al., 2010; Fischer et al., 2025), possibly due to the strong internal variability of daily rainfall within regimes (Gerighausen et al., 2024)."