Supplement of Weather Clim. Dynam., 7, 633–657, 2026
https://doi.org/10.5194/wcd-7-633-2026-supplement

Supplement of

Understanding biases and changes in European heavy precipitation using dynamical flow precursors

Joshua Oldham-Dorrington et al.

Correspondence to: Joshua Oldham-Dorrington (joshua.dorrington@uib.no)

1 Supplementary Figures

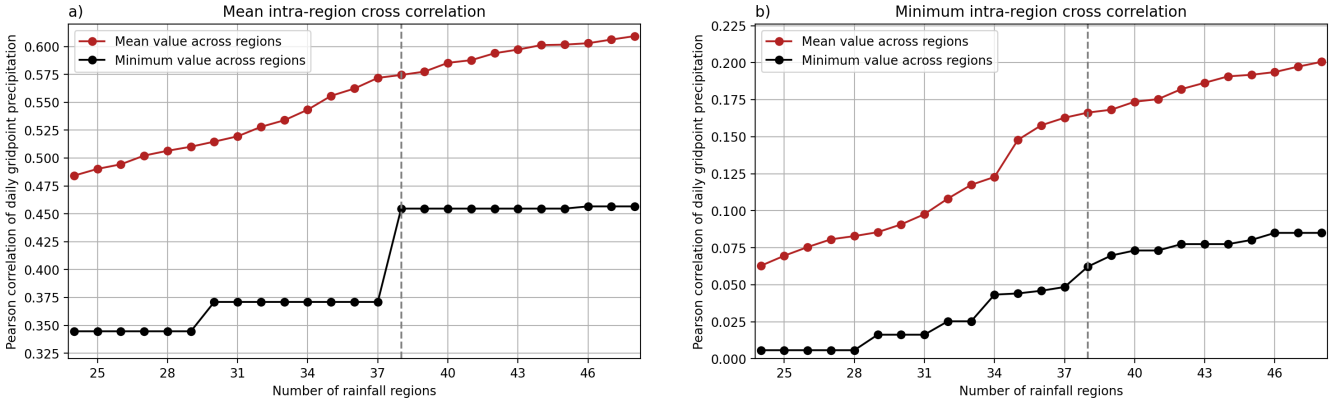

Figure S1. a) Mean (red) and minimum across regions (black) of average pairwise daily precipitation correlation between gridpoints within each region, computed from ERA5 over 1979–2023. b) Mean (red) and minimum across regions (black) of the minimum pairwise daily precipitation correlation within each region.

	95th percentile threshold (mm/day)			
Region	**DJF**	**MAM**	**JJA**	**SON**
Turkey	12.4	8.0	2.6	8.9
Aegean + SW Greece	9.1	5.7	*2.1*	7.4
North Norway + North Finland	13.0	9.3	9.8	12.2
Central Norway	13.1	9.6	11.5	12.5
Finnmark	6.3	6.1	10.2	8.3
West Ukraine	6.2	8.0	11.2	8.4
East Continental Europe	7.2	8.1	9.8	9.4
Central Continental Europe	7.9	8.3	10.7	8.8
North East Europe	5.3	6.0	8.5	6.9
North Adriatic	12.8	11.1	8.6	16.4
Central Alpine	12.5	12.6	12.5	16.4
Northern North Sea	16.3	10.5	10.3	14.8
West Norway	25.8	16.9	15.6	24.0
West Libya	*2.2*	*1.2*	*0.1*	*1.4*
Egypt + East Libya	*2.2*	*1.2*	*0.1*	*1.0*
North Sweden + North Finland	4.8	5.0	8.5	6.7
Central Sweden	6.6	6.5	9.8	8.6
East Baltic	5.9	5.3	8.3	7.3
West Baltic	5.7	5.4	7.7	6.2
Corsica + Sardinia	7.3	7.2	*2.3*	7.9
Tell Atlas	7.4	6.6	*1.4*	6.3
Sicily + Calabria	9.5	7.1	2.7	9.2
South Tunisia + East Algeria	*2.2*	*2.0*	*0.5*	*1.9*
Central Algeria	*1.5*	*1.4*	*0.2*	*1.6*
West Algeria + Balearics	4.1	4.3	*0.9*	4.0
South Morocco + South West Algeria	3.4	2.6	*0.6*	2.9
South East Spain + North Morocco	6.9	6.4	*2.0*	6.9
North East Spain	6.5	7.4	5.6	8.4
Southern France + Coastal North Spain	10.0	10.3	9.1	12.5
NW Iberia	15.7	11.6	5.6	15.0
SW Iberia + Tangier-Tétouan	12.0	8.4	*1.7*	11.4
West France + South England	9.3	7.8	7.9	9.2
Low Countries + Western Germany	9.1	7.5	8.3	8.9
Central UK + Ireland	8.6	7.0	7.5	8.6
Southern North Sea	9.2	7.7	9.2	10.5
Balkan + SE Europe	6.7	8.0	8.4	8.3
South Adriatic	13.6	10.8	6.2	13.8
North Aegean	11.5	8.3	5.0	11.2

Table S1. 95th daily area-averaged precipitation thresholds used to define heavy precipitation for each region. Thresholds below 2.5 mm/day are marked in italics, and excluded from tendency and bias plots.

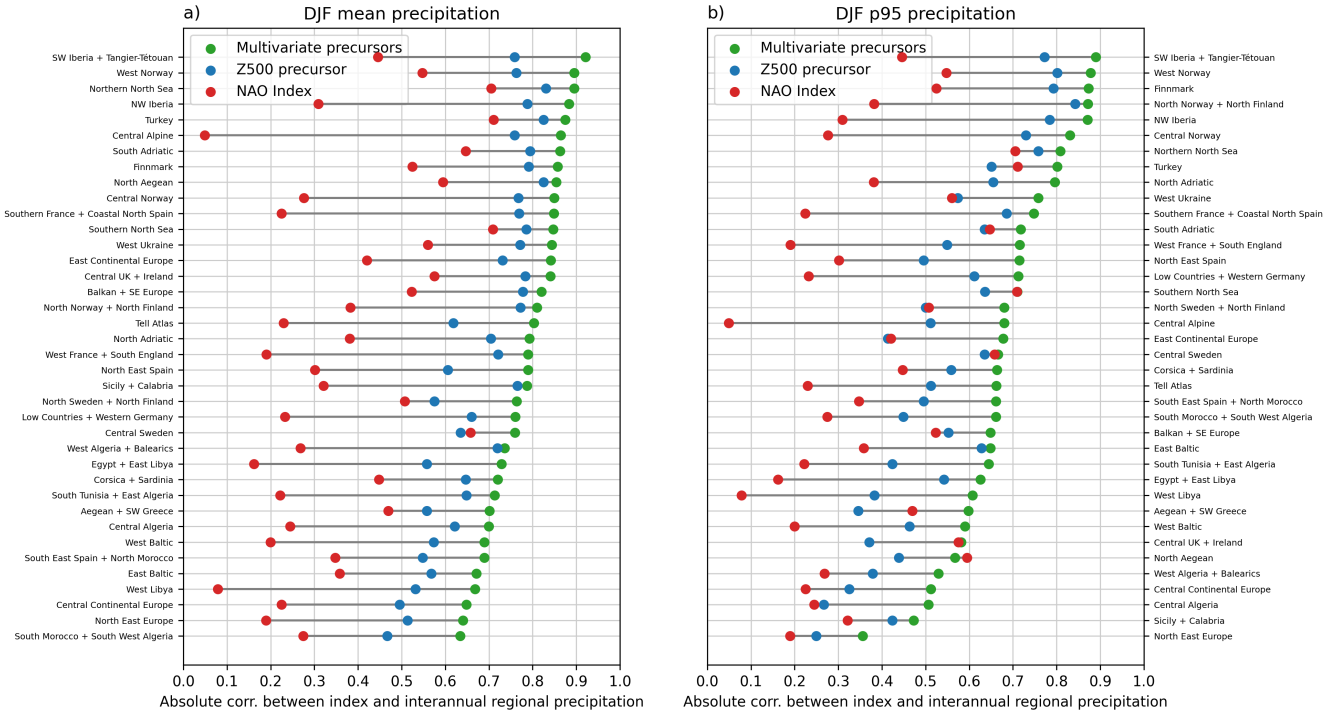

Figure S2. Positively oriented pearson correlations are shown between DJF mean values of different dynamical indices and DJF mean (a) / DJF 95th percentile (b) seasonal precipitation totals for each region over the period 1979-2023. The NAO index is included for reference, defined as the first cosine-latitude weighted principal component of Z500 over [80W-40E,30-85N], and is the same index for every region. The Z500 precursor is defined as in the main text, and is distinct for each region. 'Multivariate precursors' are the first principal component of U850, V850 and Z500 precursor indices, as used to define S in the main text. Regions have been ordered by multivariate precursor correlation, and are ordered differently in each panel. Z500 precursors show stronger and more consistent correlations with both mean and heavy precipitation than the NAO index in almost every region. Multivariate precursors perform better still, even though the U850, V850 and Z500 precursors are often closely correlated.

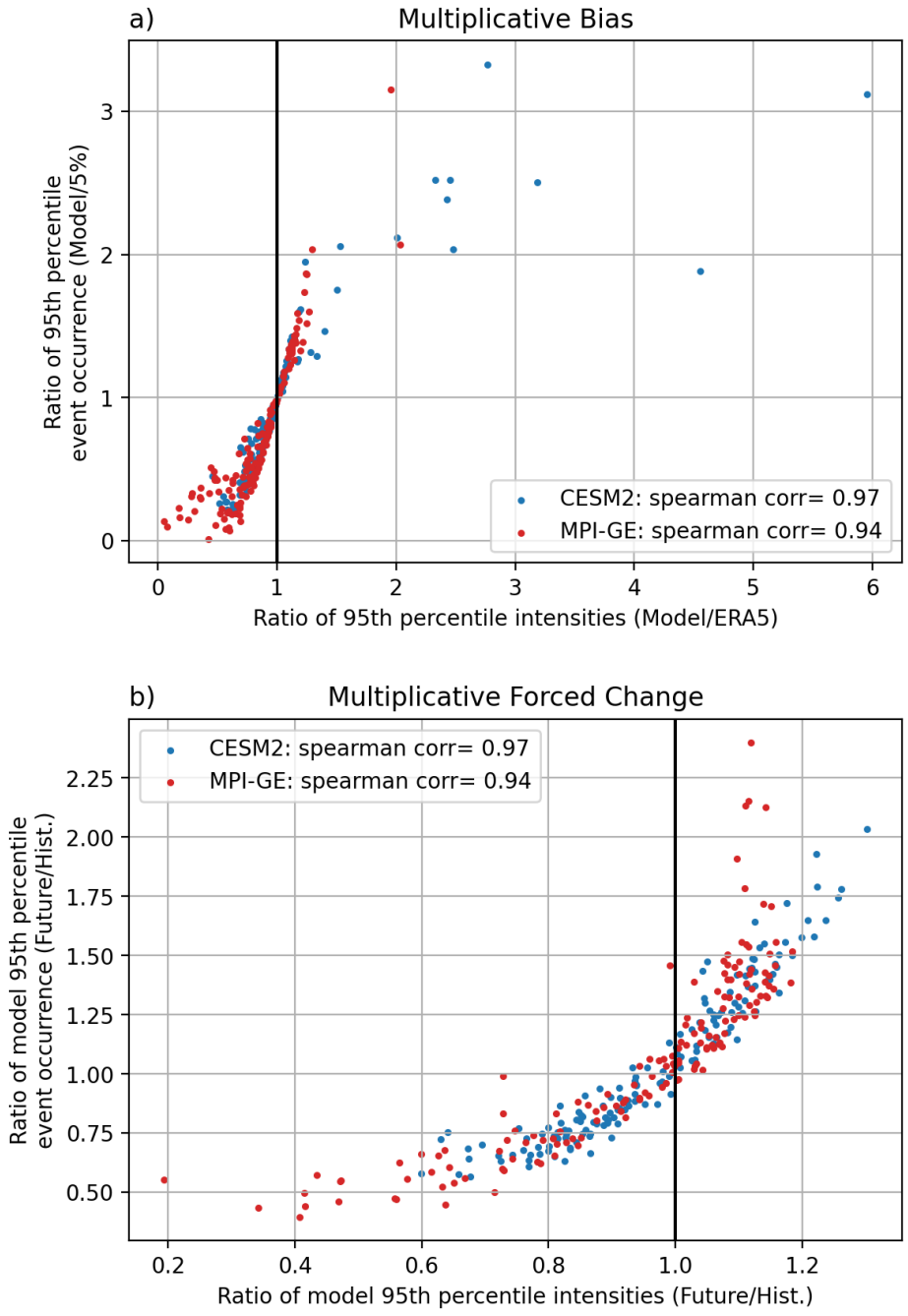

Figure S3. a) Comparison of multiplicative biases in the intensity of 95th percentile precipitation and exceedance of the ERA5 95th percentile precipitation threshold, for each region and season. b) Comparison of multiplicative forced changes between future and historical scenarios in the intensity of 95th percentile precipitation and in exceedance the ERA5 95th percentile precipitation threshold.

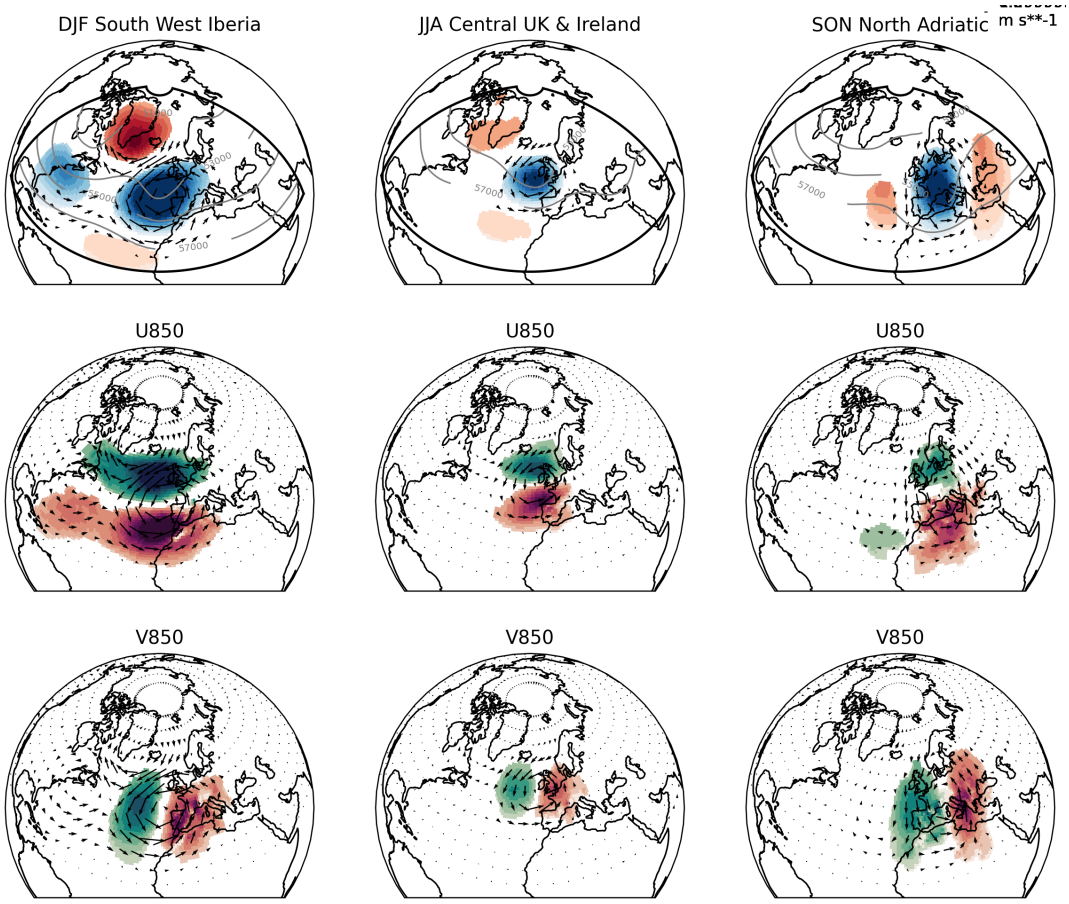

Figure S4. ERA5 precursor patterns for the three region-seasons discussed individually in the main text, and as in Figure 4 but with more detail. Row 1 shows 500 hPa geopotential height (Z500) precursor patterns in shading, with grey lines marking selected full field contours of geopotential. Row 2 shows 850 hPa zonal wind (U850) precursors in shading, and row 3 shows 850hPa meridional wind (V850) precursors in shading. In each panel shaded anomalies are on an arbitrary standardised scale, with red colors indicating positive anomalies. Arrows in all panels show 850hPa wind anomalies, in Row 1 only for gridpoints where a U850 and/or V850 precursor is present.

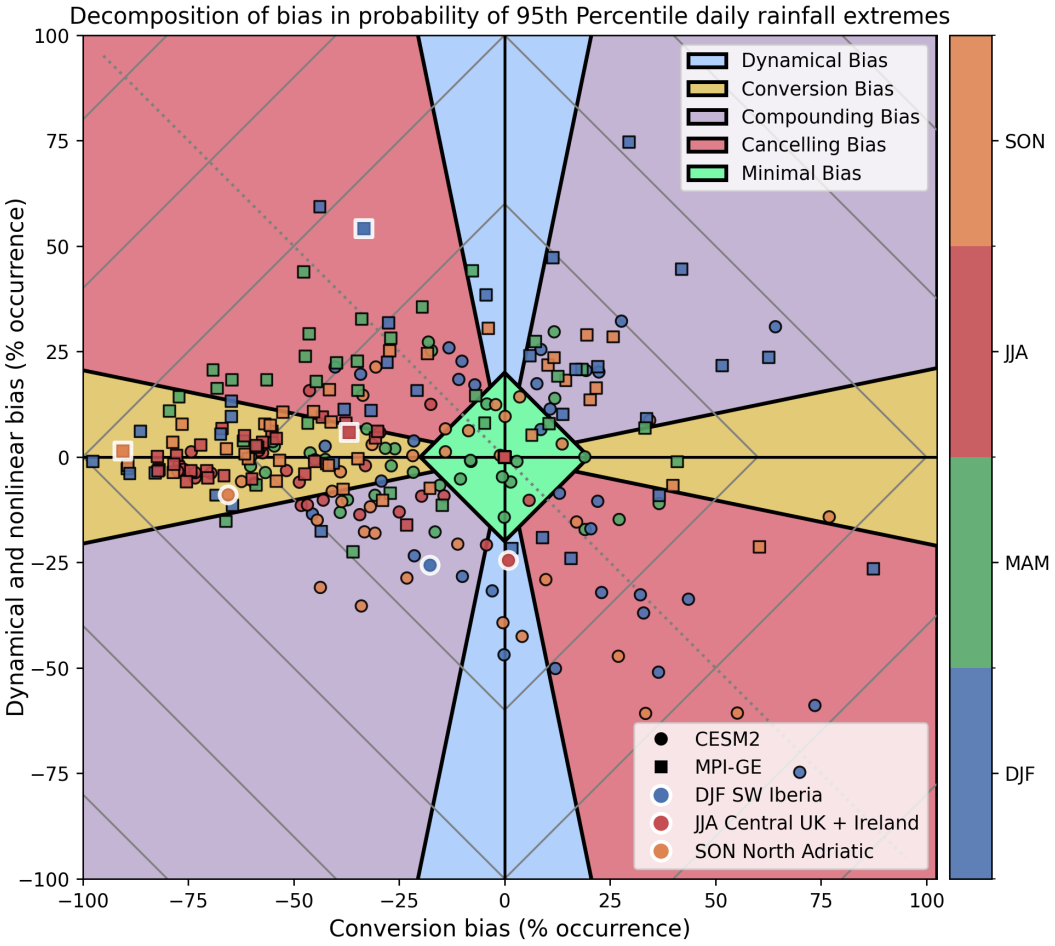

Figure S5. Heavy precipitation biases visualised in a bias space as described in the main text. Ensemble mean results are shown for every region-season excluding cases where the heavy precipitation threshold is below 2.5mm/day. The three cases discussed in the main text are distinguished with a white border.

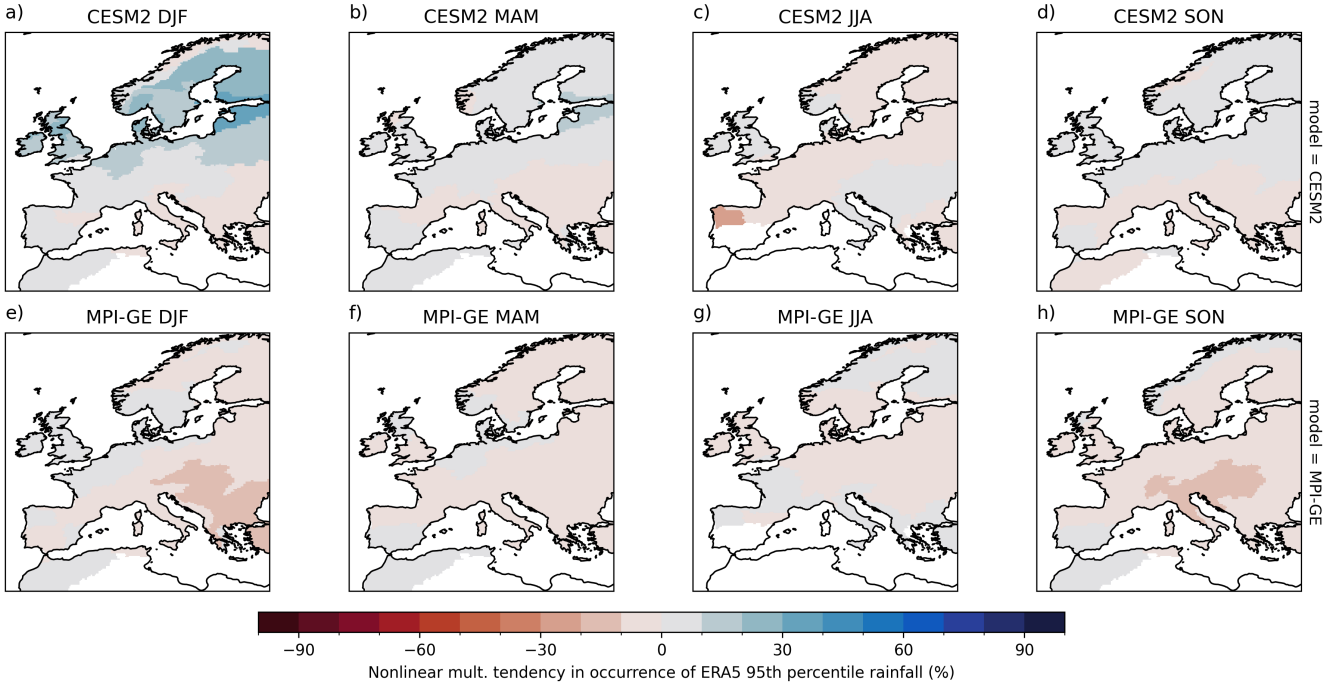

Figure S6. Ensemble mean relative tendencies (2060-2100, SSP3-7.0) in heavy precipitation occurrence attributable to nonlinear interactions of dynamical and conversion tendencies. Results are not shown for region-seasons where the historical heavy precipitation threshold is less than 2.5mm/day

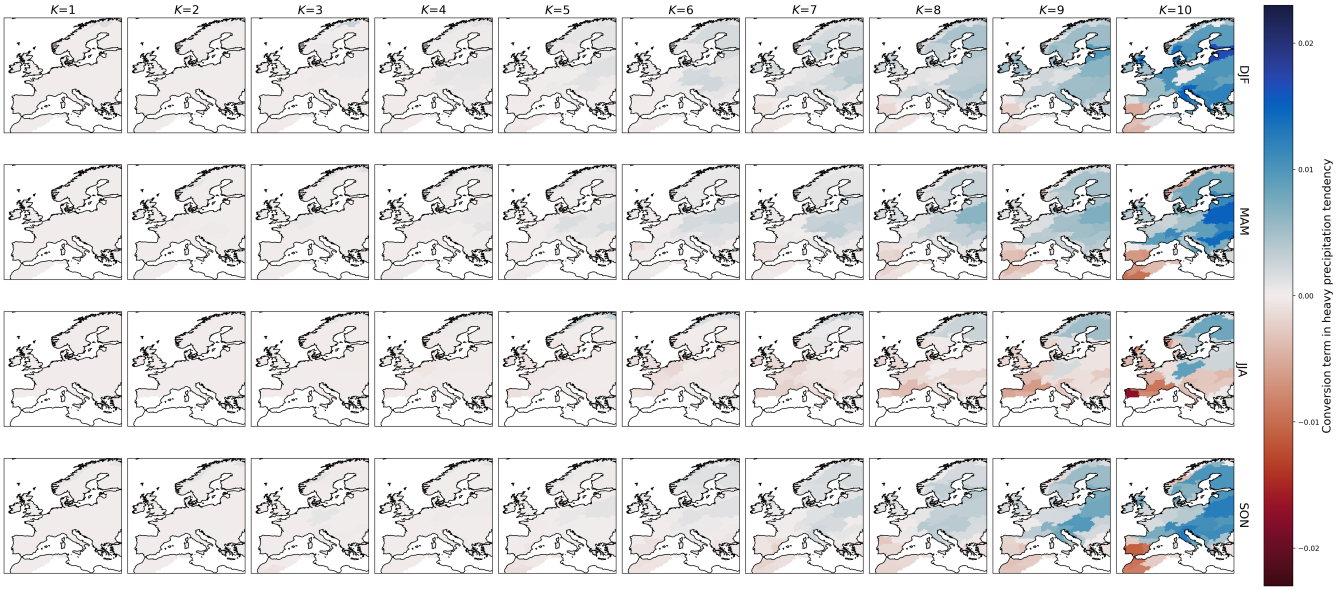

Figure S7. Ensemble mean conversion tendencies decomposed by synoptic bin for CESM2 for each season. If new dynamical processes were causing heavy precipitation in the future simulations this would be visible as increased conversion rates for weak or negative precursors. Instead the majority of conversion signal is combined to bins K=[8,9,10].

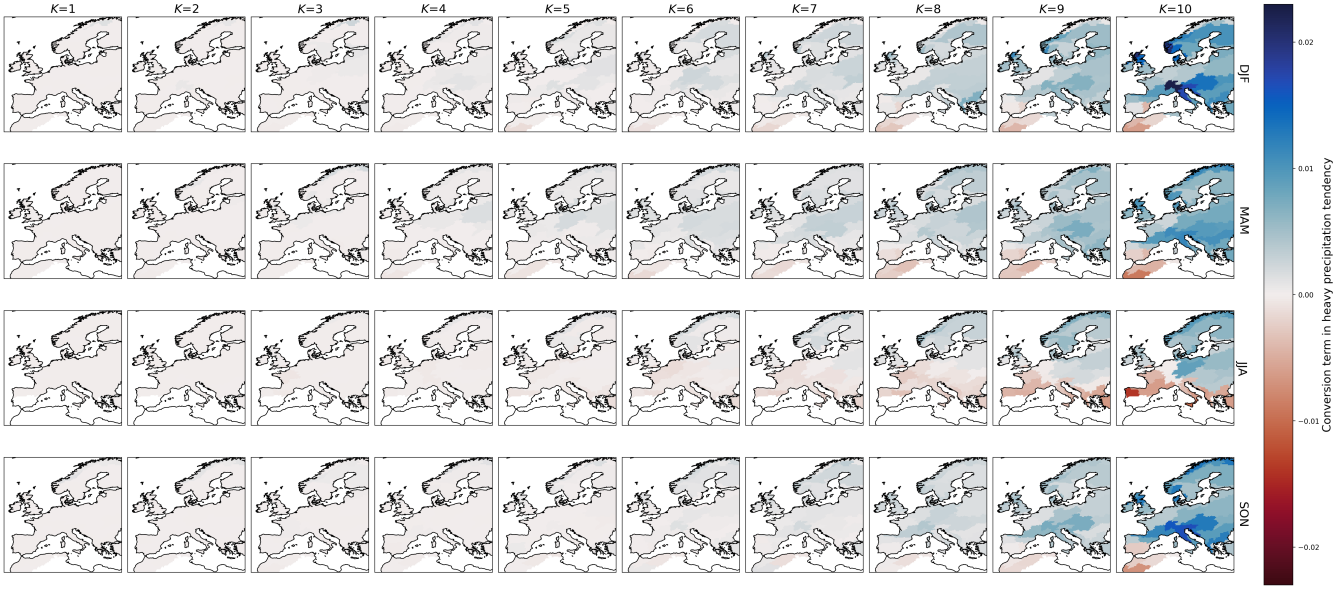

Figure S8. Ensemble mean conversion tendencies decomposed by synoptic bin for MPI-GE for each season. If new dynamical processes were causing heavy precipitation in the future simulations this would be visible as increased conversion rates for weak or negative precursors. Instead the majority of conversion signal is combined to bins K=[8,9,10].

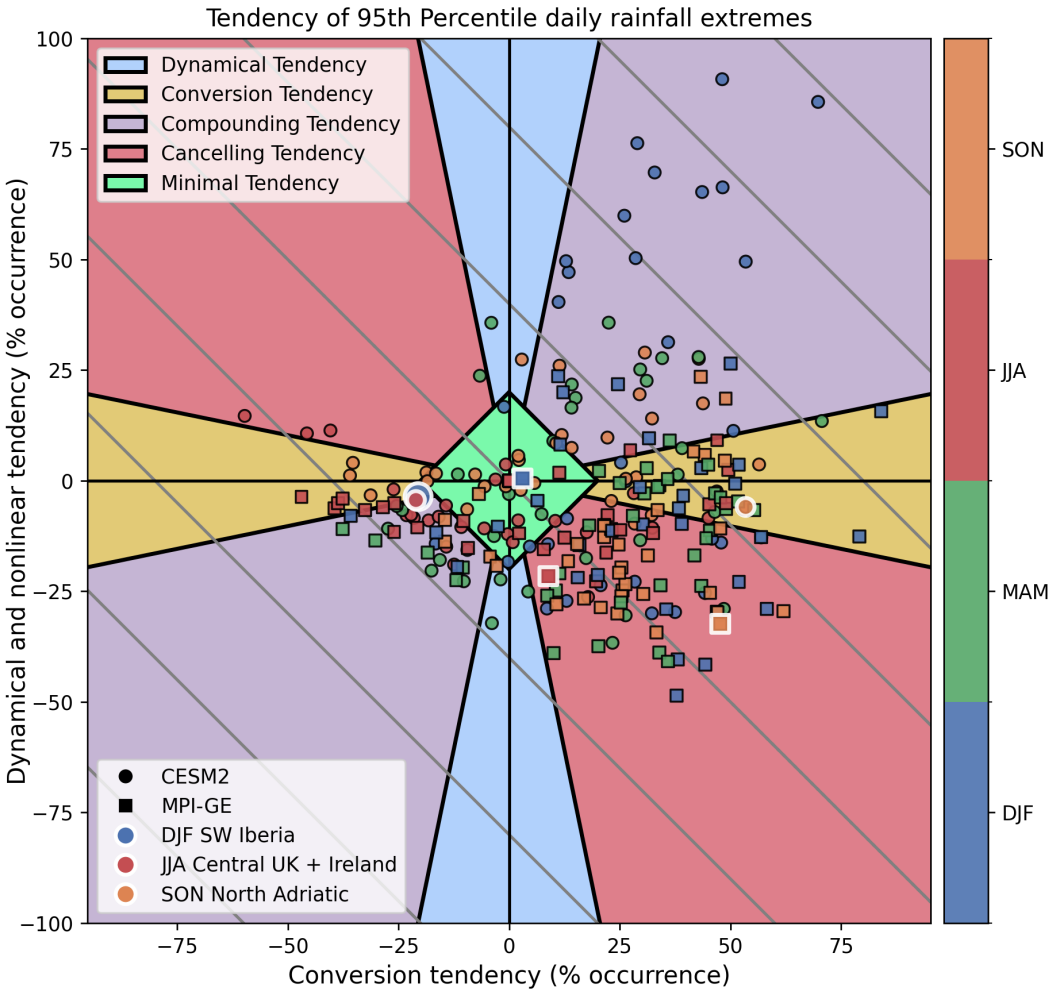

Figure S9. Heavy precipitation tendencies visualised in a tendency space as described in the main text. Ensemble mean results are shown for every region-season excluding cases where the heavy precipitation threshold is below 2.5mm/day. The three cases discussed in the main text are distinguished with a white border

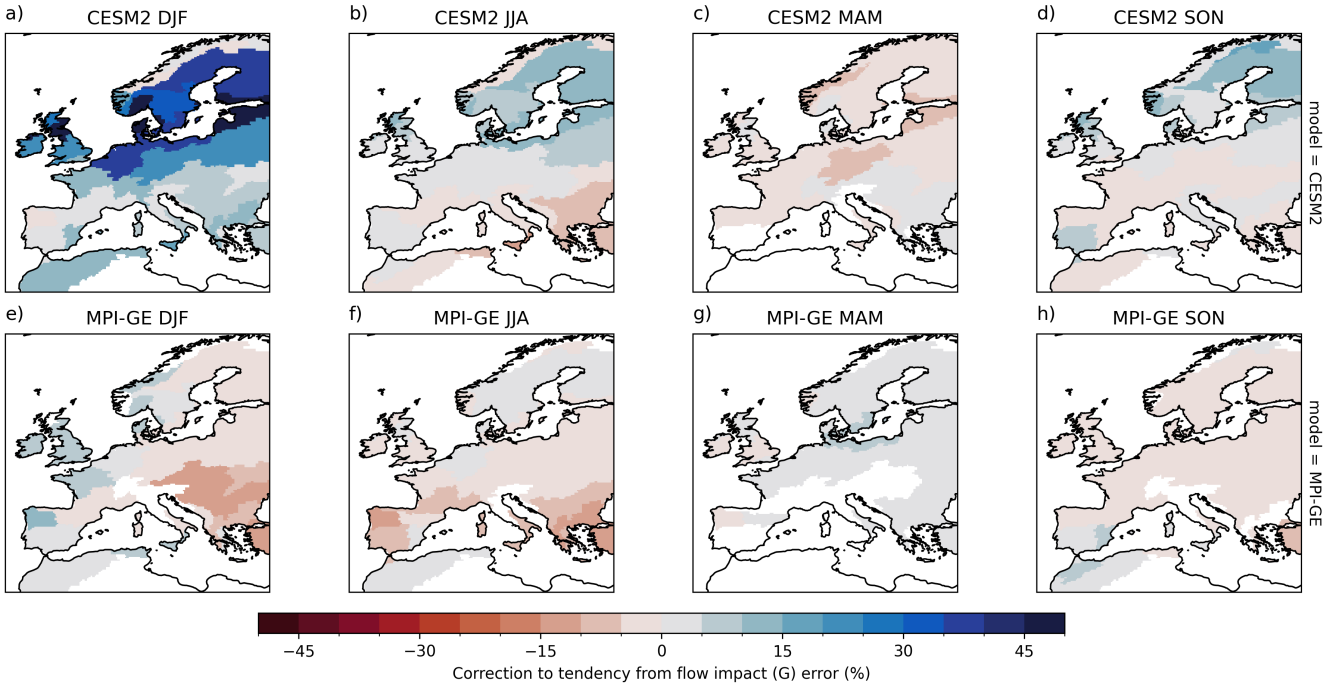

Figure S10. The flow-dependent correction to the relative precipitation tendency coming from model flow impact errors (term 2 in Equation 8 of the main text).

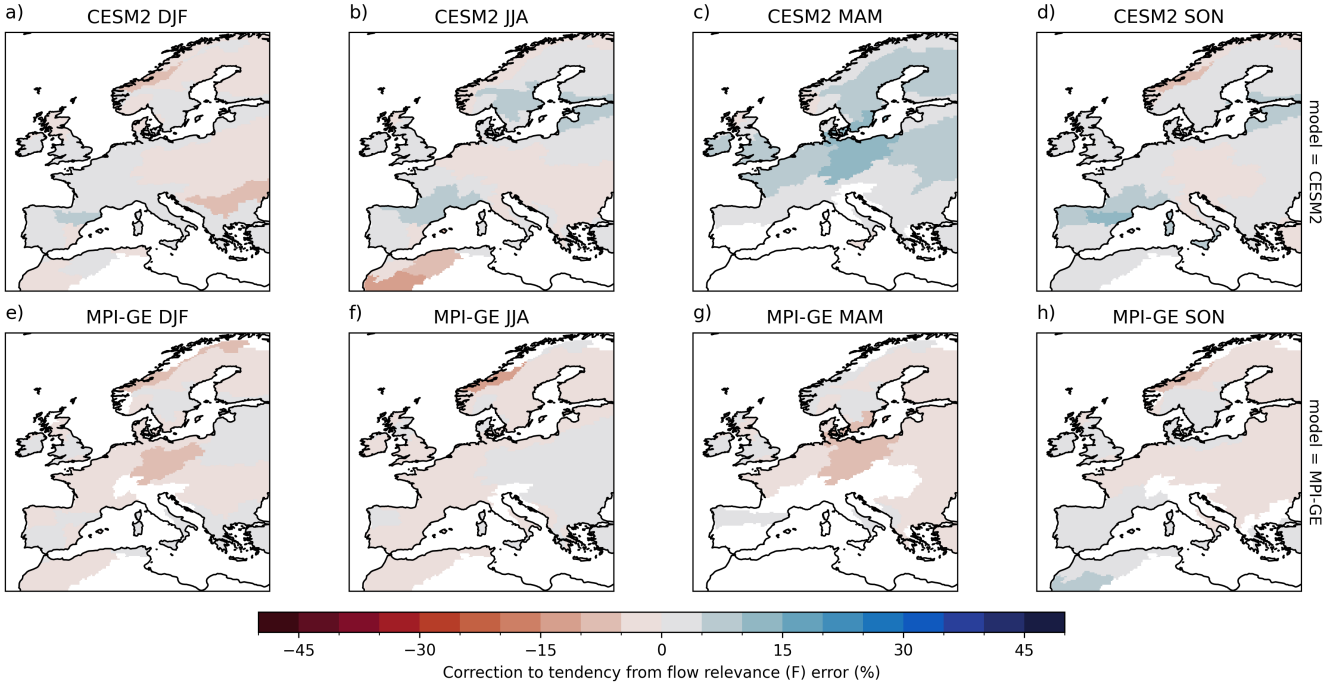

Figure S11. The flow-dependent correction to the relative precipitation tendency coming from model flow relevance errors (term 1 in Equation 8 of the main text).

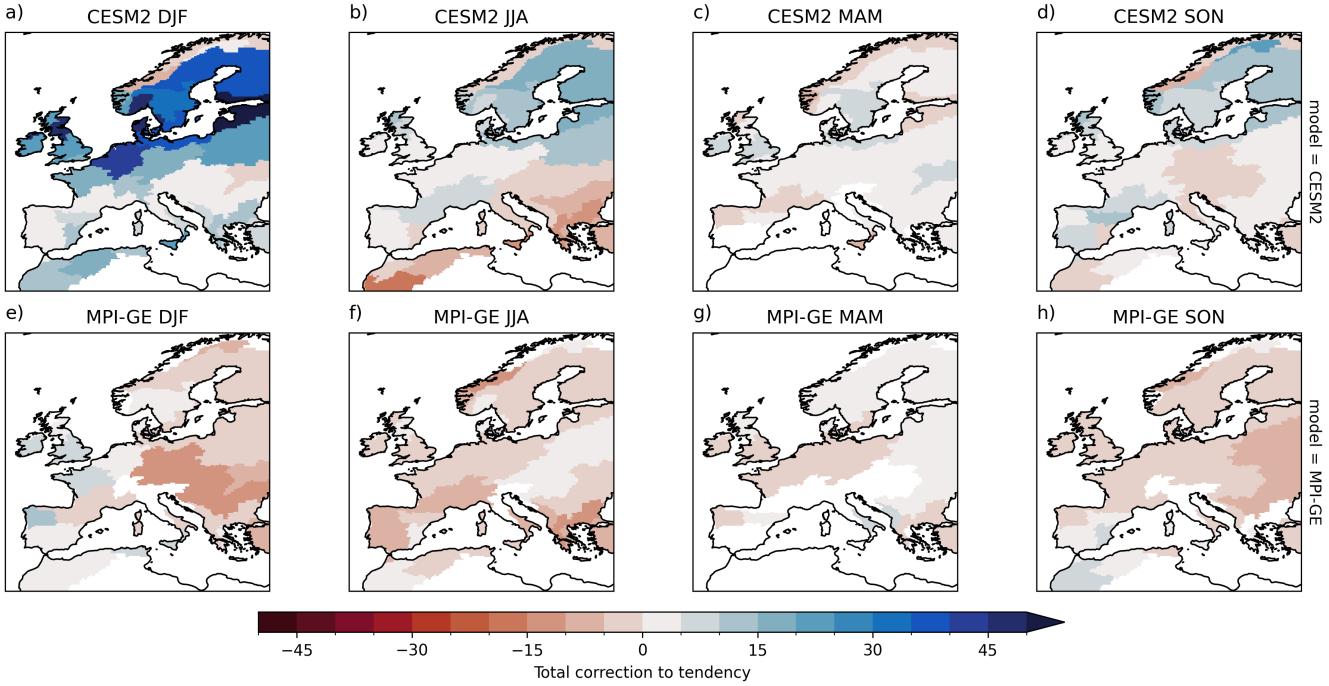

Figure S12. The total flow-dependent correction to the relative precipitation tendency ($\tilde{\beta} - \beta$).

2 Supplementary Discussion: Internal Variability and Parameter Dependence

In this paper we have analysed two large ensemble simulations, allowing us to limit the impact of internal variability on our results. However such large ensembles are not the norm. Some climate modelling centres have contributed 10-member simulations to CMIP6, with the majority of others providing 3-member simulations for core scenarios. We now consider how the analyses we develop in this paper might be impacted when applied to single-member or small ensemble simulations.

To assess sampling error in decomposed biases and forced changes, we used a bootstrapping approach for both the CESM2 and MPI-GE grand ensembles. All days for the historical/future periods were shuffled across all members and 400 random samples were drawn, each of length equivalent to using 1, 3, 10, 20, 30 or 40 ensemble members. The bias and change decomposition with respect to ERA5 was applied to each of these resamples, giving a distribution for each decomposed quantity.

Supplementary Figure. S13 shows the sampling uncertainty in each of the decomposed quantities as a function of ensemble size, given by the standard deviation of the bootstrapped distributions. Data are shown for every season and region and for both models. As expected from statistical theory, the sampling uncertainty decreases with the inverse square root of ensemble size. As such, a 10 member ensemble can detect signals that are 3x as small as those that can be detected in a single member simulation. Conversion quantities are more uncertain than dynamical or nonlinear quantities, and changes are more uncertain than biases, with conversion changes having the highest sampling variability. In a single member simulation a forced conversion change of less than 40% lies within 1 standard deviation of sampling errors in some cases, which would obscure most signals shown in main Figure 12.

One natural approach to counteracting the increased sampling variability in a small ensemble is to consider decreasing the number of synoptic bins used within the decomposition. By doing this, each bin contains more data samples and so should allow a more accurate statistical estimation. From a meteorological perspective, decreasing the number of bins means grouping together synoptic conditions that are less similar and so this has the effect of decreasing the accuracy of the decomposition.

To assess this, we repeat the approach used above, for 1, 3 and 10 member ensembles and now for a decomposition using 5, 10 or 20 synoptic bins. In Supplementary Figure S14 we show the distribution across regions of sampling errors in each decomposed quantity, as a function of both the number of synoptic bins and ensemble members. As in Supplementary Fig. S13, sampling error is higher for smaller ensembles. In contrast, the impact of increasing or decreasing bin number on sampling error is comparatively small. Proportionately, the decrease in sampling error from decreasing bin number is greatest for single member ensembles, and becomes less impactful as ensemble size increases. Decreasing bin number alone however does not manage to decrease sampling error as much as adding additional members. This is perhaps not surprising, as even bulk heavy precipitation statistics (i.e. a bin size of 1) can be difficult to estimate from single realisations.

Finally, Supplementary Fig. S15 indicates the 'cost' of decreasing bin number, showing how the mean of the bootstrapped distributions for each decomposed quantity shift when using 5 or 20 bins, vs using 10 bins as in the main manuscript. These are plotted against the sampling variability for each season and region so that the impactfulness of these mean changes can be evaluated. Estimates of dynamical and nonlinear biases are often markedly different when using 5 synoptic bins, clearly

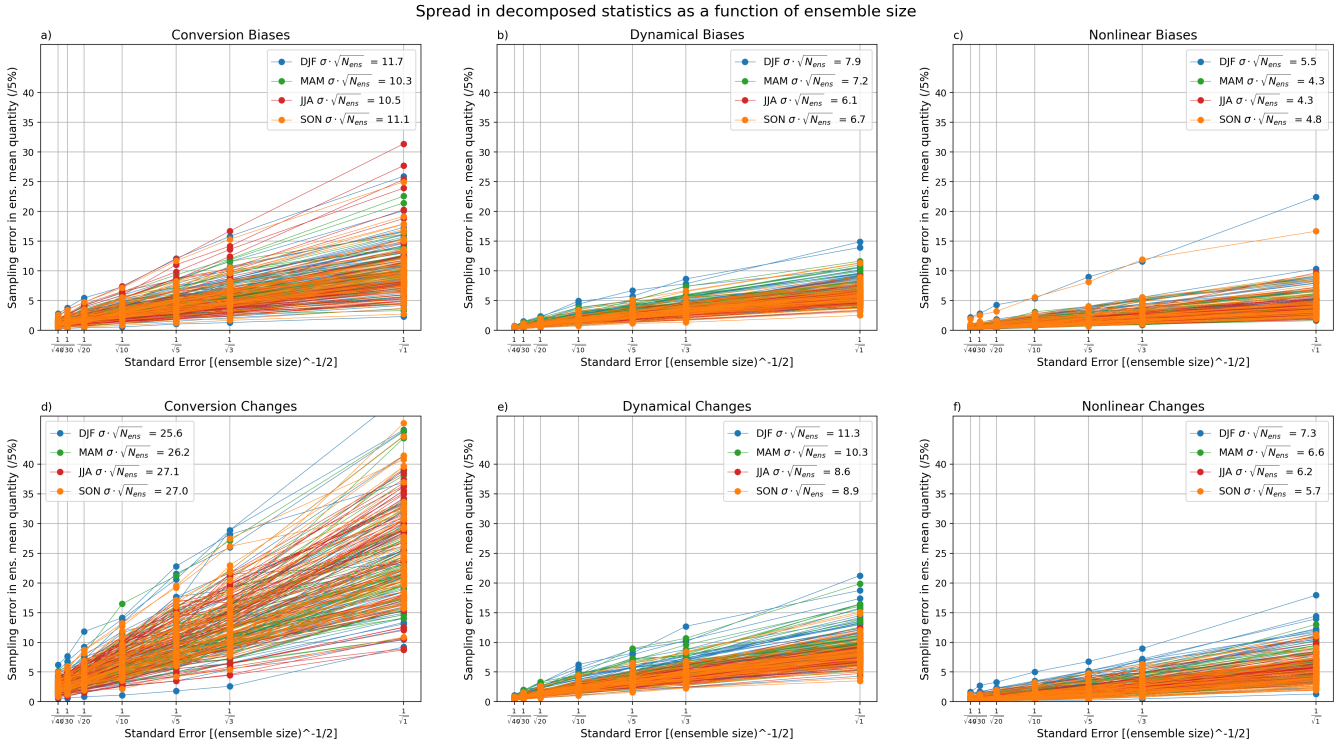

Figure S13. Sampling error (1 standard deviation) in ensemble mean decomposed quantities as a function of ensemble size. Results are shown for all seasons, regions and models. A near-linear relationship between theoretical standard error, ($\frac{1}{\sqrt{Ens}}$), is apparent, with the estimated slope computed across both models and all regions indicated for each season and quantity.

lying outside the variation expected from sampling. Dynamical and conversion changes indicate relatively small but systematic misestimates in JJA. Specifically a decrease in conversion change and an increase in dynamical change is seen. As the prevaling forced changes are negative (towards less heavy precipitation), this indicates the 5-bin approach placing a greater emphasis on conversion changes. As a result, we do not advocate for using a reduced number of synoptic bins for the cases we have considered in this paper: the decreases in sampling error are mostly small, and the distortion to the decomposition is not negligible.

In fact, when considering the use of 20 synoptic bins, there are suggestive signs the number of bins should be increased when analysing model data from a sufficiently large ensemble. While the changes in mean quantities between 10 and 20 bins are less than those between 5 and 10 bins (indicating a convergence toward the true, continuous conditional distribution), there is a systematic shift in the decomposition in JJA. This is consistent with the shift mentioned above, pointing towards an even higher role for dynamical drivers in the general decrease in JJA heavy precip than noted in the main manuscript.

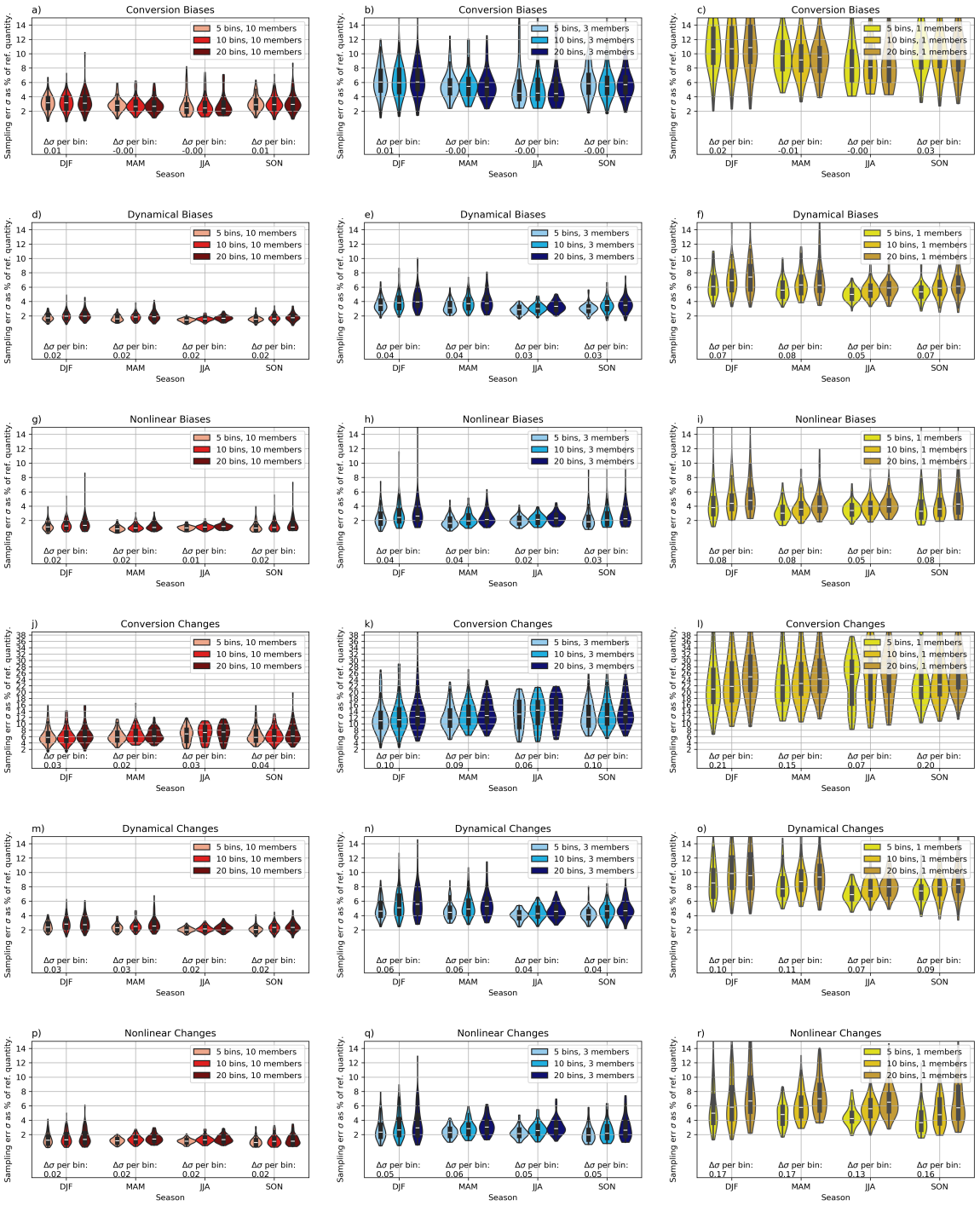

Figure S14. Sampling error in ensemble mean decomposed quantities as a function of the number of ensemble members used, and the number of synoptic bins used in the precursor decomposition. Sampling errors are disaggregated by season and are shown as distributions across regions. Columns show the dependence 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.

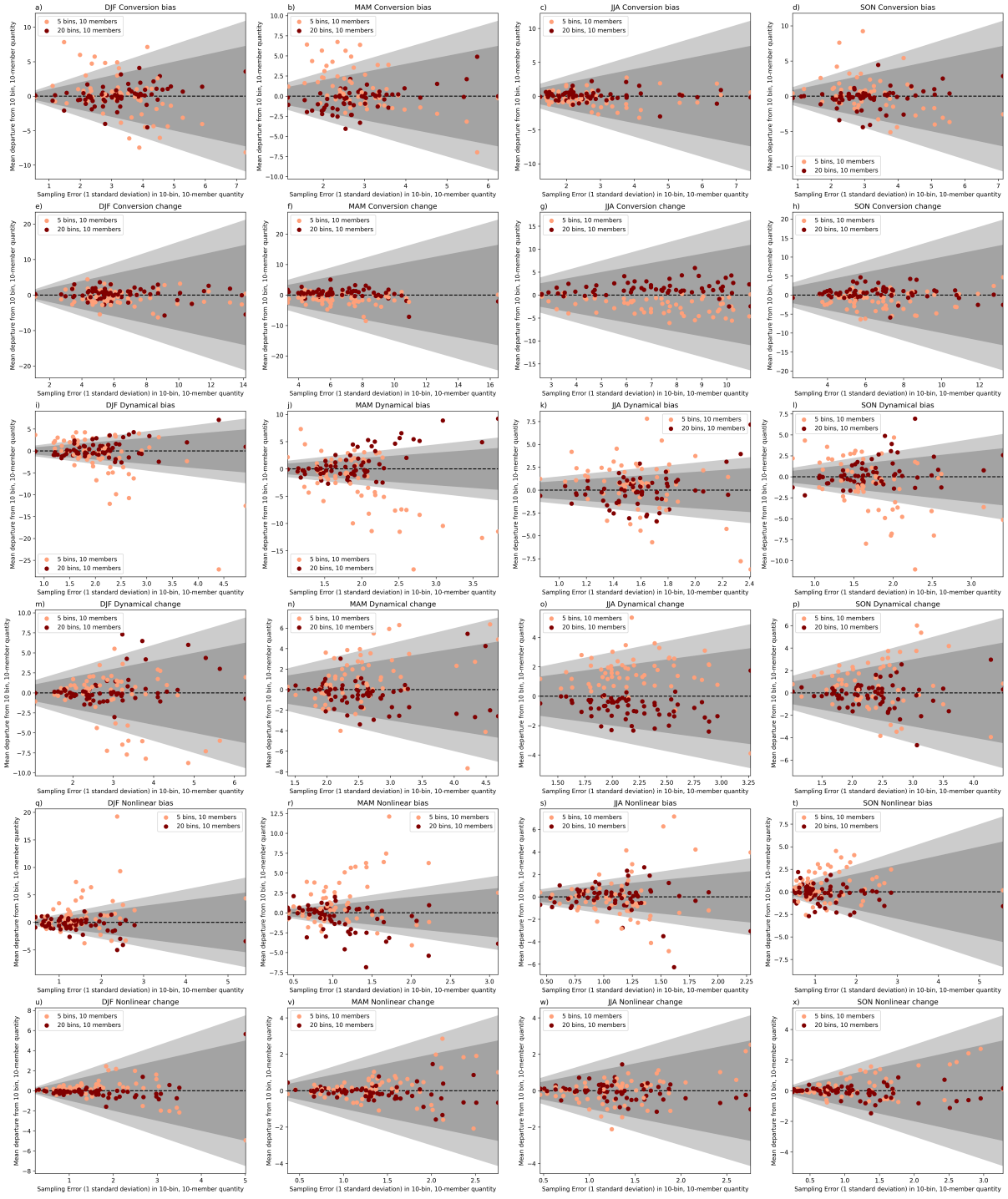

Figure S15. Difference between decomposed quantities computed using 5 (pale dots) or 20 (dark dots) synoptic bins and those computed using 10 synoptic bins, shown as a function of sampling error in the 10-bin ensemble mean estimate. Dark/light grey shading shows the +/-1 and +/-1.5 std error intervals, with differences outside these intervals indicating a meaningful impact of bin number on the decomposition.