



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

Joshua Oldham-Dorrington^{1,2}, Camille Li^{1,2}, Stefan Sobolowski^{1,2}, and Robin Guillaume-Castel^{1,2}

¹Geophysical Institute, University of Bergen, Norway

²Bjerknes Centre for Climate Research, Norway

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

Abstract. We address the problem of understanding precipitation in climate models. Using a novel decomposition applied to two large ensemble simulations, we disaggregate biases and forced changes in European heavy precipitation occurrence according to different weather conditions and isolate synoptic-scale dynamical contributions from the local-scale conversion of synoptic forcing into precipitation. We categorise weather conditions using multivariate, regionally-specific heavy precipitation precursors that target precipitation-causing flow patterns, revealing a larger role for dynamics in explaining model biases and projected changes than suggested by previous work. We demonstrate that biases in heavy precipitation across models and regions can emerge from errors on very different scales, with compensating biases between scales being common. This has important implications for model selection, for example for downscaling or storyline applications. In terms of forced changes in heavy precipitation, we show that apparent model agreement can arise from markedly different future scenarios with different levels of implied risk.

Our results demonstrate the utility of flow-dependent diagnostics for exposing the origins of climate model biases, which can distort a model's precipitation response in future projections. With an eye to informing researchers in model development and validation, we demonstrate which combinations of dynamical versus conversion biases lead to specific types of distortion, and emphasise that these cannot be corrected for without a flow-dependent perspective. This framework allows us to introduce an intuitive heuristic for guiding model selection and interpretation, and to extract usable climate information from imperfect models.

1 Introduction

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. 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— can produce dynamical precursors that set the potential for precipitation by modulating the availability of moisture and steering low-level winds. Converting this potential for precipitation into precipitation itself involves



25 finer scale processes including cloud microphysics, sub-diurnal heating, boundary-layer and geographical (orographic uplift, coastal gradients) interactions, and mesoscale organisation of convection.

Simulating precipitation presents a major challenge, given the need to represent both synoptic-scale precursors and fine-scale conversion processes. Global climate models have made steady improvements in the spatial distribution and temporal variability of mid-latitude precipitation (Du et al., 2022), but deficiencies remain (Abdelmoaty et al., 2021). Too little heavy
30 precipitation is a widespread problem, largely due to the many parameterised conversion processes that simply cannot be resolved on the $\mathcal{O}(100\text{km})$ grids of climate models. This has motivated the development of higher-resolution models, ultimately moving towards ‘convection-permitting’ (CP) scales below approximately 5km horizontal resolution (Stergiou et al., 2025).

However, it is not currently feasible to run global CP models for more than a few years, much less to produce the multi-scenario, multi-member ensembles required to assess future changes (Stephan et al., 2022). Although regional CP ensembles
35 are becoming available, these suffer from relatively short, time-slice experimental designs, complicating their interpretation, especially for extremes (for example, for Europe (Soares et al., 2024; Pichelli et al., 2021)). Further, while higher model resolution can make certain aspects of precipitation more realistic (e.g., heavy rain intensities), it can degrade other aspects by enhancing some processes disproportionately (Bador et al., 2020; Strandberg and Lind, 2021; Wille et al., 2025). Identifying the scales and conditions responsible for global climate model precipitation biases therefore has benefits both for current and
40 next-generation climate projections. This can not easily be achieved using metrics that aggregate rainfall representation over many different synoptic conditions and scales, motivating a decomposition approach.

A number of decompositions exist to interpret the origins of biases and trends in simulated precipitation. On the global scale, precipitation changes have been decomposed into dynamic and thermodynamic contributions based on Clausius-Clapeyron scaling (Held and Soden, 2006) or changes in vertical velocity and humidity (O’Gorman and Schneider, 2009). Analogue
45 approaches have been used to decompose trends in monthly surface variables into internal and forced contributions (Deser et al., 2016) and into dynamical/residual contributions (Doane-Solomon et al., 2025). However these approaches are less well suited to understanding regional precipitation on daily timescales. To meet this need, Cassano et al. (2007) introduced a decomposition of arctic precipitation trends using self-organising maps to classify regional circulation variability into a small number of states. Under this approach trends were understood as either due to dynamical changes in the frequency of
50 a circulation state or due to ‘thermodynamic’ changes within a state—although we note that all non-synoptic processes are folded into this latter term. Following work has extended this perspective, using weather regimes as a basis for understanding precipitation (Fischer et al., 2025) and temperature (Cattiaux et al., 2013) trends over Europe, and precipitation trends over Morocco (Driouech et al., 2010). Decompositions can also attribute precipitation to single process, e.g. Mediterranean cyclones (Zappa et al., 2015). Attributing precipitation events to intersections of tracked weather objects (Konstali et al., 2024)
55 provides a new and complementary perspective.

Despite the growing body of literature and tools, these daily-timescale diagnostic methods have not reached mainstream or routine application as climate model diagnostics. This may be partly a result of 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). However the choice of which dynamics to focus on strongly impacts the conclusions of a decomposition. Regime



approaches favour dominant modes of variability and variability in rainfall within a regime can be large (Gerighausen et al., 2024). This is especially the case for highly impactful regional-scale heavy precipitation events, which are often driven by uncommon and spatially localised weather patterns.

Here we present a new decomposition approach based on the dynamics that cause heavy precipitation events, and with a focus on decomposing model biases and their interaction with forced changes as well as forced changes themselves. Biases have received comparatively little attention in previous decompositions, yet by accounting for the flow-dependent and scale-dependent nature of model biases, this approach allows us to assess whether models reproduce heavy precipitation realistically, and if so, whether it is for the right reasons. We classify synoptic variability using multi-variate, bottom-up 'flow precursors' (Dorrington et al., 2024a), initially developed for summarising impact-relevant information within ensemble weather forecasts. The scalar precursor indices track the anomalies in geopotential height and low-level wind that are most relevant for causing heavy precipitation in a given region, similar to the 'large-scale meteorological patterns' used in (Grotjahn et al., 2016) to calibrate climate projections of North American heatwaves. We use these precursors to decompose both biases and trends in simulated heavy precipitation into contributions from synoptic-scale dynamics and fine-scale conversion processes. Within this framework, we can then study the interactions that can lead to compensating or compounding effects. Because precursors are captured within a small set of scalar metrics, they can be efficiently diagnosed in large datasets: we consider heavy precipitation in 38 Euro-Mediterranean sub-regions for each season, providing—to our knowledge—the most holistic evaluation to date. This study demonstrates that process- and flow-dependent diagnostics are not only tools for dynamicists and model developers, but also help to interpret and validate projections in a way that enhances their value for stakeholder communities.

Section 2 introduces the datasets used, the clustering method used to define our regions, 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.

2 Data and Methods

2.1 Data

As our observational reference for large-scale dynamics, we use three variables from the ERA5 reanalysis (Hersbach et al., 2020) at 1° resolution: 500 hPa geopotential height (Z500) and 850 hPa zonal (U850) and meridional (V850) wind aggregated to daily means from 6 hourly instantaneous fields. Precursor patterns (see Sect. 2.3) are computed using data over the period 1979–2024, but we compute precursor statistics over 1979–2014 for all comparisons to climate model data. While ERA5 has now been extended back to 1940, we limit ourselves to the satellite era where daily synoptic variability is more reliable. Precipitation data is taken from ERA5-land (Muñoz-Sabater et al., 2021), which has a spatial resolution of 0.25°. While



precipitation in reanalyses is not in general considered as reliable as that from satellite- and gauge-derived products, reanalysis data has the advantage of being temporally and spatially homogeneous and is suitable for our focus on large-scale spatially-
95 aggregated daily precipitation. As a check, we have made a qualitative comparison with 0.1° , 3 hourly precipitation data from MSWEPv3, which blends reanalysis with gauge and satellite data (Beck et al., 2019), and find consistent results.

We analyse two large-ensemble simulations, CESM2 LENS2 (Danabasoglu et al., 2020) and MPI-GE (Olonscheck et al., 2023), using 50 members from each of their historical and SSP3-7.0 scenario runs. While CESM2 has 100 members available, 50 are sufficient to constrain European precipitation statistics (Donat et al., 2023). For the historical runs we use the 1979-2014
100 period shared with ERA5, and for analysis of forced changes we use a ‘future’ period covering 2060-2100. SSP3-7.0 represents the severe but increasingly plausible ‘rocky road’ emissions pathway (O’Neill et al., 2016, 2017), characterised by global end-of-century warming of $2.8\text{--}4.6^\circ\text{C}$. The precipitation response in both models is nearly linear across scenarios (Danabasoglu et al., 2020; Meehl et al., 2020), so we expect our findings for SSP3-7.0 to be generally relevant to any SSP scenario. With only two models, our results will not span the full range of model uncertainty, but CESM2 and MPI-GE come from independent
105 model families (Kuma et al., 2023) and are known to have different bias patterns (Brands, 2022). The models have comparable transient climate response in the middle of the CMIP6 spread – 1.8K in MPI-GE (Scafetta, 2023) and 2.0K in CESM2 (Meehl et al., 2020). CESM2 has a higher horizontal resolution than MPI-GE, 1° vs 1.8° , and a more realistic wintertime jet (Simpson et al., 2020), but both models been found to reproduce observed mid-latitude mean and heavy precipitation with reasonable accuracy (Donat et al., 2023).

110 2.2 Region definition

Rather than aggregating precipitation over regions defined by national boundaries or simple geometry, we define regions based on shared precipitation variability. We take ERA5-land daily precipitation over 1979-2024 at all land gridpoints over the region $[30\text{--}72\text{N}, 12\text{W--}30\text{E}]$, subsample by latitude to maintain a similar spacing between points, and then compute the pearson correlation between each pair of retained gridpoints. This produces a correlation matrix, C , which was converted
115 into a distance matrix $D = \sqrt{1 - C}$, and clustered agglomeratively using Scipy’s linkage function, with method=“complete” (Virtanen et al., 2020). Within this algorithm each gridpoint is initially assigned to its own region, and the number of regions is sequentially reduced at each step by combining the two regions with the lowest maximum distance between the gridpoints they contain. 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. This gave 38 regions in total shown in Fig.
120 1. Supplementary figure 1 shows the dependence of minimum intra-regional correlation on region number. Precipitation from ERA5-land and from CESM2 and MPI-GE (bilinearly interpolated to the ERA5-land grid) were averaged over these regions, using cosine-latitude weighting, to obtain 38 scalar precipitation indices for each data set. We focus on categorical ‘heavy precipitation’ events, which we define separately for each season and regional precipitation index, based on exceedance of the 95th ERA5-land percentile (given explicitly in Supplementary table 1.). We use this same ERA5-land threshold for model
125 precipitation, a choice we justify in Sect. 3.

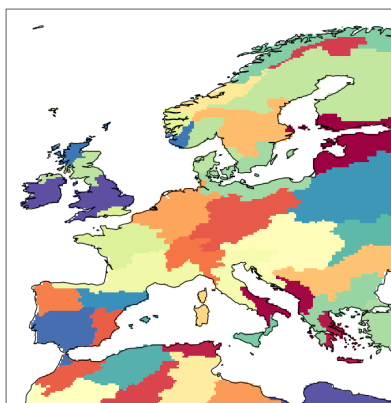


Figure 1. The 38 precipitation regions over which we aggregate precipitation, shown here with arbitrary colouring. Regions were defined through a clustering of 0.25° ERA5-land daily precipitation, as described in the main text.

2.3 Precipitation decomposition and flow precursors

We decompose biases and changes in heavy precipitation occurrence into dynamical contributions and conversion contributions. The dynamical contribution accounts for changes precipitation due to changes in the occurrence of different synoptic flow conditions, based on the observed probability of precipitation under each condition. We categorise different synoptic conditions based on precursor indices, described below. The conversion contribution accounts for changes in precipitation *within* a given synoptic condition, and so is an aggregation of all other factors: differences in microphysics, thermodynamics, boundary layer interactions, the land surface and mesoscale dynamics.

The mathematical formalism behind the decomposition is presented in full and undivided in Appendix A. In the main text, each equation is introduced as it becomes relevant alongside a corresponding concrete example so that each term can be linked directly to a meteorological reference. As a matter of notation, probabilities are always marked with a tilde if they correspond to a potentially biased quantity, and with an asterisk if they correspond to a future climate state.

The flow precursor framework developed in Dorrington et al. (2024a) (described there in full), is the basis of our decomposition. We use this to seek a low-dimensional representation of the synoptic processes that drive heavy precipitation across our 38 regions. On a high level, the approach is to identify the synoptic conditions corresponding to past heavy precipitation events using composite analysis, and define time-evolving ‘precursor activity indices’ based on those composites. Our method for this paper is as follows:

1. For each region and season, composites of deseasonalised ERA5 Z500, U850 and V850 anomalies are computed for heavy precipitation days. Seasonal cycles are computed with a day-of-year climatology, smoothed with a 31-day Gaussian filter.
2. Composites of each variable are partially masked, retaining only gridpoints where anomalies are i) statistically significant ($p < 0.05$), ii) high amplitude (> 0.25 gridpoint standard deviation), and iii) spatially coherent (significant, high amplitude



anomalies form a connected area of $>5e5$ square kilometres). These masked composites are termed *precursor patterns*. This masking is intended to algorithmically imitate the approach a researcher might use to hand-define a variability mode such as the box-based North Atlantic Oscillation (NAO) or El Niño-Southern Oscillation. We maintain the same masking parameters as Dorrington et al. (2024a, b) as they have been shown to produce precursors that can be used to predict daily precipitation in a weather forecast context, which we take as a strong test of their dynamical relevance.

3. Daily ERA5 fields are projected onto these masked precursor patterns and the resulting time series are standardised, producing scalar *precursor activity indices*. By construction, strong positive projection onto these indices is related to increased occurrence of heavy precipitation.

4. Anomaly fields of Z500, U850 and V850 are computed from climate models using the ERA5 seasonal cycle and mean state, thereby preserving all model bias. Model data was not detrended, with the exception of Z500 data for the SSP3-7.0 runs, where a global thickening trend was removed, which is dynamically irrelevant as it has no impact on the geostrophic wind. This was done by computing ensemble- and area-mean Z500 anomaly over the Euro-Atlantic region [80W-60E, 30N-90N], smoothing it with a 21-year time-mean Gaussian filter and subtracting it from each member.

5. Daily anomaly fields for each ensemble member are projected onto the ERA5 precursor patterns, and scaled with ERA5 standardisation parameters.

In combination with the regional precipitation aggregation described in Sect. 2.2, this process reduces the interactions between synoptic dynamics and precipitation in each region to a functional relationship between 4 scalar indices: three precursor indices and one precipitation index. The time evolution of each precursor index is directly interpretable in terms of the evolution of its precursor pattern, just as, for example, the dipole variability of the NAO can be captured neatly within an NAO index. However, in contrast to the NAO which represents the dominant variability mode of Atlantic geopotential height, the flow precursor framework identifies targeted patterns for each region-season combination that explain heavy precipitation variability. The benefit of introducing individual indices for each region is that we obtain far stronger dynamical modulation of precipitation than traditional regime or analogue based approaches can provide. Supplementary figure 2 shows an exemplary comparison for DJF. We discuss only a small number of these precursor patterns explicitly in this paper, but the full set of $3 \times 4 \times 38$ patterns is included in the supplementary material so that specific regional cases can be examined by the interested reader.

2.4 Uncertainty and variability

In our results we quantify the statistical sampling uncertainty in decomposed quantities computed using the full model ensemble, and separately, the internal variability of these quantities in any given ensemble member.

Bias terms are computed using daily model data from each ensemble member concatenated to form a single time series. Change terms are computed similarly as the difference/ratio between terms computed from the full future and historical ensemble datasets. In both cases sampling uncertainty is computed using 400 bootstrap resamples of the daily all-ensemble data

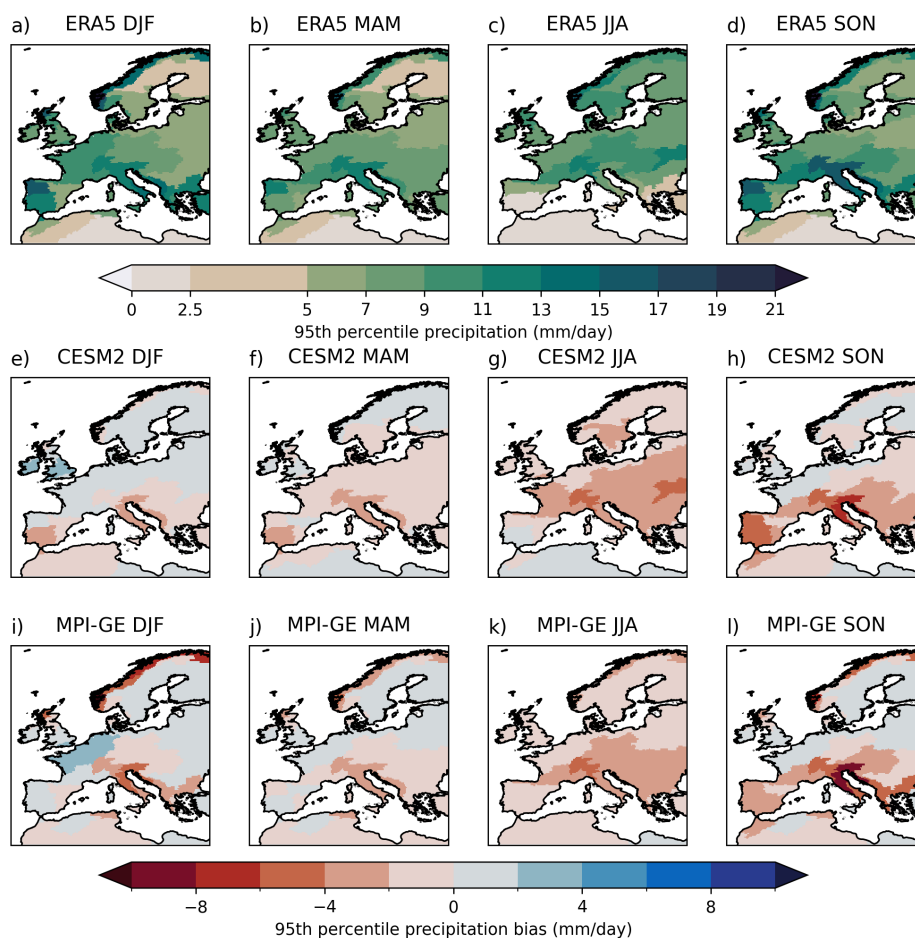


Figure 2. a-d) 95th percentile area-averaged daily precipitation thresholds from ERA5-land used to define heavy precipitation days. e-l) Mean biases in 95th percentile precipitation in the two large ensembles.

with replacement. For internal variability, bias terms are estimated for each ensemble member individually, and change terms are estimated using the difference/ratio between decomposed quantities from each member's future and historical dataset. Historical and future data from different ensemble members are not mixed when estimating internal variability in forced changes.

3 Heavy Precipitation in CESM2 and MPI-GE

Before moving to a flow-dependent analysis we first summarise model performance from the bulk perspective with reference to the ERA5-land climatology of 95th percentile regional precipitation (Fig. 2a-d). We term precipitation over this 95th percentile threshold as *heavy precipitation*, but this should always be understood in local terms. The 95th percentile is highly variable across seasons and regions, ranging from less than 2.5mm/day to as much as 25mm/day (c.f. Supplementary table 1). Both



models generally underestimate the intensity of 95th percentile of precipitation (Fig. 2e-l)), especially in summer. Biases are typically largest in mountainous regions, but this is not universally the case: MPI-GE performs well over the Iberian mountains for example, while CESM2 performs well over the Scandinavian mountains. Pronounced wet biases are rare, but can be seen
190 in DJF in some western European regions in both models.

Going forward we focus on the categorical occurrence of heavy precipitation (Fig. 3a-h)). Even in regions with the largest magnitude dry biases such as the Alps or northern Scandinavia (c.f Fig 2), both models can actually produce real-world heavy precipitation events. Nevertheless occurrence biases lower than -80%—visible for both models in JJA and for MPI-GE in mountainous domains throughout the year—indicate cases where 95th percentile real-world events are akin to 99th percentile
195 or greater events in the model. Positive occurrence biases are also seen in many regions: the Atlantic coast of Europe in SON and DJF receives 30-110% more heavy precipitation events in the two models than in reality. This motivates our choice to define heavy precipitation events with reference to reanalysis precipitation thresholds; for a pan-European analysis using model-derived thresholds would mix events of differing magnitudes.

Throughout the paper we will characterise forced climate signals by analysing changes in heavy precipitation occurrence
200 between the end-of-century period 2060-2100 under the SSP3-7.0 climate scenario and the historical period 1979-2014. Figure 3i-p) shows that these changes are more spatially coherent than the biases, and in general of smaller amplitude. The models show a qualitative consistency in their continental-scale forced response: more heavy precipitation events above 40N in autumn through spring, fewer events in summer (excluding Fennoscandia) and fewer events in southern Iberia and North Africa throughout the year. This is generally consistent with the 'wet gets wetter' heuristic (Held and Soden, 2006). On a region-by-
205 region level however, and with an eye to the actual amplitude of projected changes, the story is richer and more uncertain. We identify a few specific cases that we will discuss in detail later. CESM2 and MPI-GE have pronounced differences in biases and changes in southwestern Iberia during DJF, which notably is a region where climate models have struggled to reproduce recent observational precipitation trends (Donat et al., 2023). The northern Adriatic region features some of the highest precipitation values in Europe (Fig. 2a-d)) and experiences some of the most pronounced societal impacts from flash flooding
210 (e.g. (Dorrington et al., 2024b)), yet is unfortunately one of the regions where the models we consider show consistent heavy precipitation biases of -60 to -90% (Fig. 3a-h). Finally, we note the JJA drying change over the central UK and Ireland, which is seemingly consistent across the two models but is shown in Sect. 5 to be a case of considerable model disagreement.

4 Biases

Heavy precipitation occurs when a) the synoptic environment is favourable and b) the interactions of sub-synoptic processes
215 serve to convert the large-scale forcing into precipitation. Formalising this observation, we can decompose the daily probability of heavy precipitation, P_H , into the probability of a particular synoptic driver occurring, P_S , and the conditional probability of heavy precipitation under that synoptic driver, $P_{H|S}$, summed over a comprehensive set of drivers:

$$P_H = \sum_k P_{H|S_k} \cdot P_{S_k} \quad \sum_k P_{S_k} = 1$$

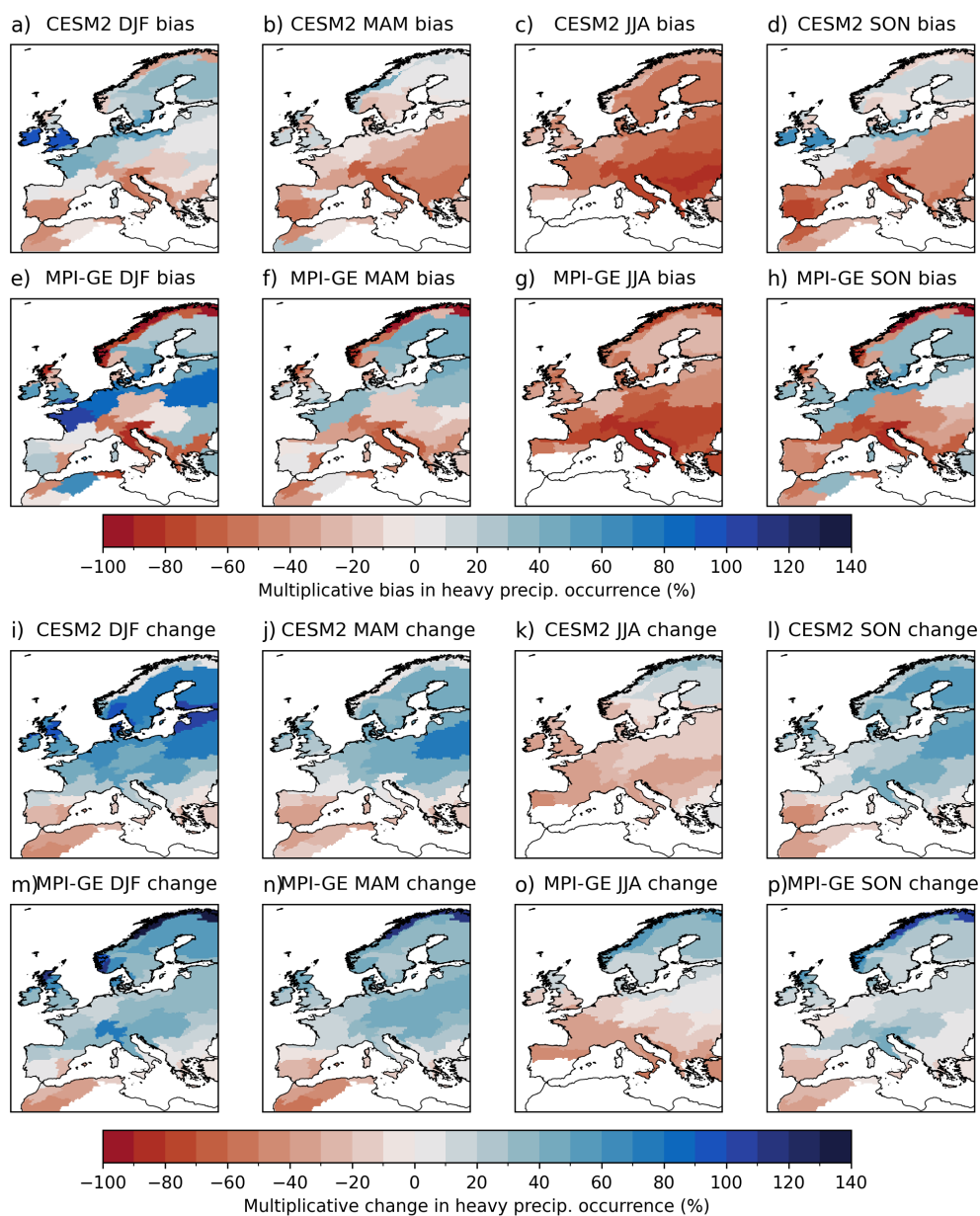


Figure 3. a-h) Multiplicative biases in the occurrence of heavy precipitation days for each season and model over the historical period (1979-2014), with respect to ERA5-land. i-p) Ensemble mean multiplicative changes in heavy precipitation occurrence between the future period (2060-2100, SSP3-7.0 scenario) and the historical period. A bias of -100% would indicate no heavy precipitation occurrence, +100% would indicate a doubling in occurrence probability with respect to ERA5-land.



The stronger the conditioning of heavy precipitation occurrence on this synoptic categorisation, the more useful this decomposition will be. We define $\{P_{S_k}\}$ using the flow precursor indices introduced in Sect. 2.3.

Throughout, *potentially biased quantities are denoted with a tilde, while future quantities are denoted with an asterisk*. For each region-season combination three precursor patterns are defined within Z500, U850, and V850 anomaly fields respectively (examples for three region-seasons are shown in Fig. 4i-iii)). Each pattern is used to define a standardised precursor index, describing the daily variability of projections onto the pattern. We take the first principal component of the three ERA5 precursor indices for each region-season to obtain a single scalar index S that captures the variability of multivariate precipitation-causing dynamics. We then discretise S into 10 bins based on deciles, $\{S_1, \dots, S_K, \dots, S_{10}\}$. In ERA5, each bin is equally likely to occur by construction, and bins labelled with larger K indicate a higher conditional probability of heavy precipitation. Taking the concrete example of southwestern Iberia during DJF, Fig. 4i) shows the identified precursor patterns; a low-level jet shifted to the south, and an omega-type block over the Atlantic. This pattern favours the convergence of moist Atlantic airmasses from the north and west into the region (as also documented in Dorrington et al. (2024a)). Supplementary Fig. 3 shows the precursor patterns for each variable in more detail. The Z500, V850 and U850 precursor indices that summarise the variability of this flow pattern are reduced into the multivariate index S , and Fig. 4i a-d) shows ERA5 circulation composites for different deciles of S . High decile composites ($K=9,10$) closely match the precursor pattern, while those for low deciles ($K=1,2$) show an approximate inverse flow: anticyclonic ridging to the west of Iberia and weakened regional westerlies. Figure 4i e-h) confirms that strong precursors correspond to elevated heavy precipitation probability, validated here with a different precipitation metric than that used to define the patterns—gridpoint-wise heavy precipitation occurrence computed from MSWEP data.

In a historical model simulation with heavy precipitation probability $\tilde{P}_H$, we decompose bias into the occurrence frequency of the synoptic categories as additive, δP_{S_k} , and biases in heavy precipitation conversion, ξ_k , as multiplicative which follows standard precipitation calibration methods (Hewson and Pilloso, 2021):

$$\tilde{P}_H = \sum_k (P_{H|S_k} [1 + \xi_k]) \cdot (P_{S_k} + \delta P_{S_k}) \quad (1)$$

As a note, previous literature has often labelled similar quantities to our conversion term as an ‘intensity’ or ‘thermodynamic’ term. The former is a better descriptor for a continuous precipitation decomposition than for our categorical perspective, while the latter is misleading as a range of small-scale processes are contained within the term. Now, instead of having a single bulk metric of precipitation bias for each region and season, we have 20 metrics (δP_{S_k} and ξ_k for ten values of k), representing biases on different scales and in different flow states.

4.1 Wintertime biases in southwestern Iberia

Fig. 5 visualises these metrics for DJF southwestern Iberia. In this case, both models have large and opposite dynamical biases (Fig. 5a): CESM2 generates too few days with strong precursors (i.e. $\delta P_{S_{\{8,9,10\}}} < 0$) while MPI-GE generates far too many, producing flows like Fig. 4i d) almost a third of the time. The conversion terms from ERA5 clearly show that strong synoptic precursors are almost a prerequisite for heavy precipitation (Fig. 5b). CESM2 simulates the occurrence of heavy precipitation

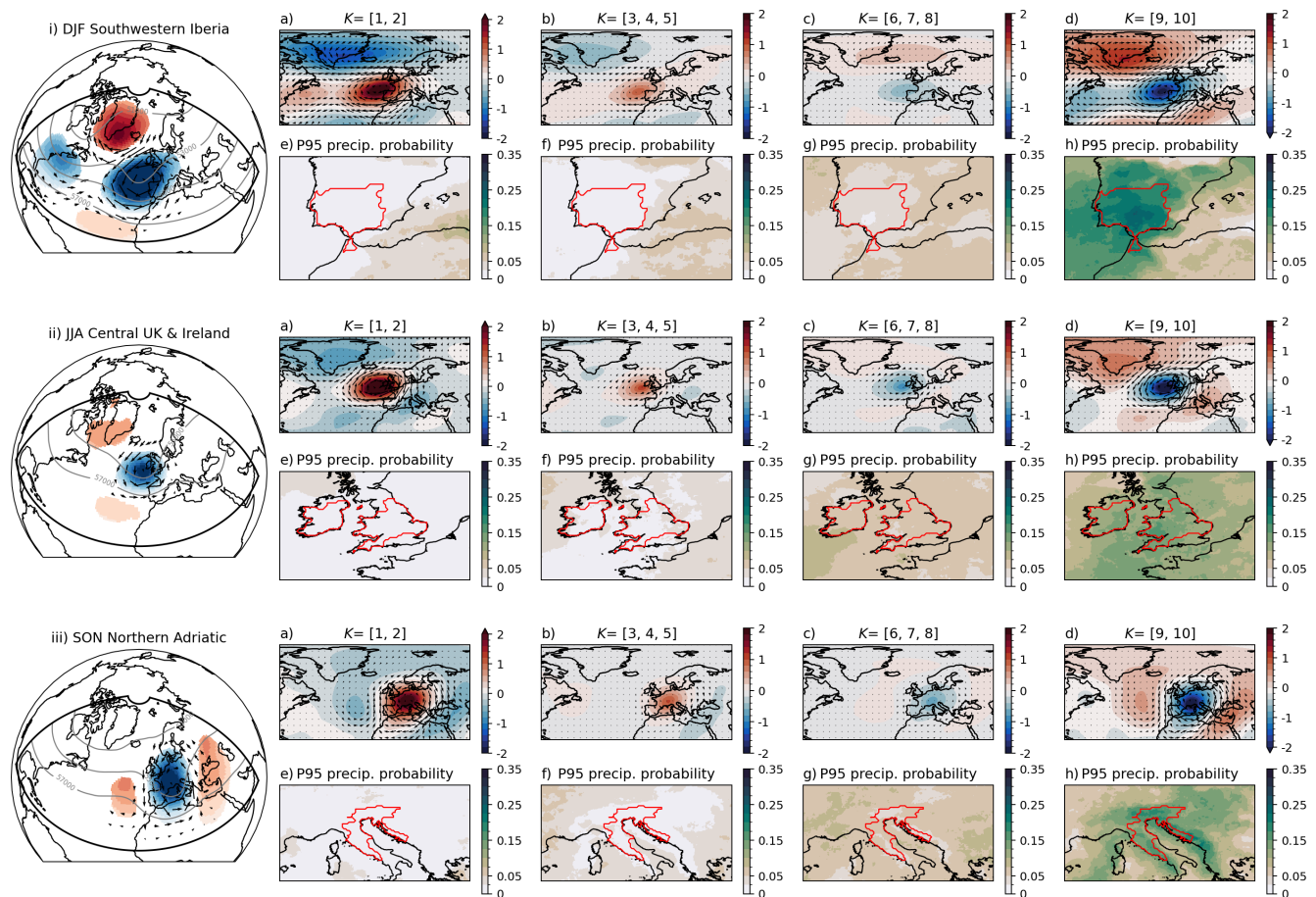


Figure 4. Precursor patterns (i-iii) and circulation composites on bins of the synoptic index S_K (a-d) for three example region-seasons. In all panels, blue-red shading shows standardised anomalies of ERA5 500 hPa geopotential height, and black quivers indicate ERA5 850 hPa wind. Grey contours on the left panels (i-iii) show full geopotential height contours associated with the precursor patterns. Strong precursors lead to increased heavy precipitation probability (e-h), where white-to-green shading shows gridpoint 95th percentile precipitation probability from MSWEP data. Red contours, show the domains over which precipitation is spatially aggregated.

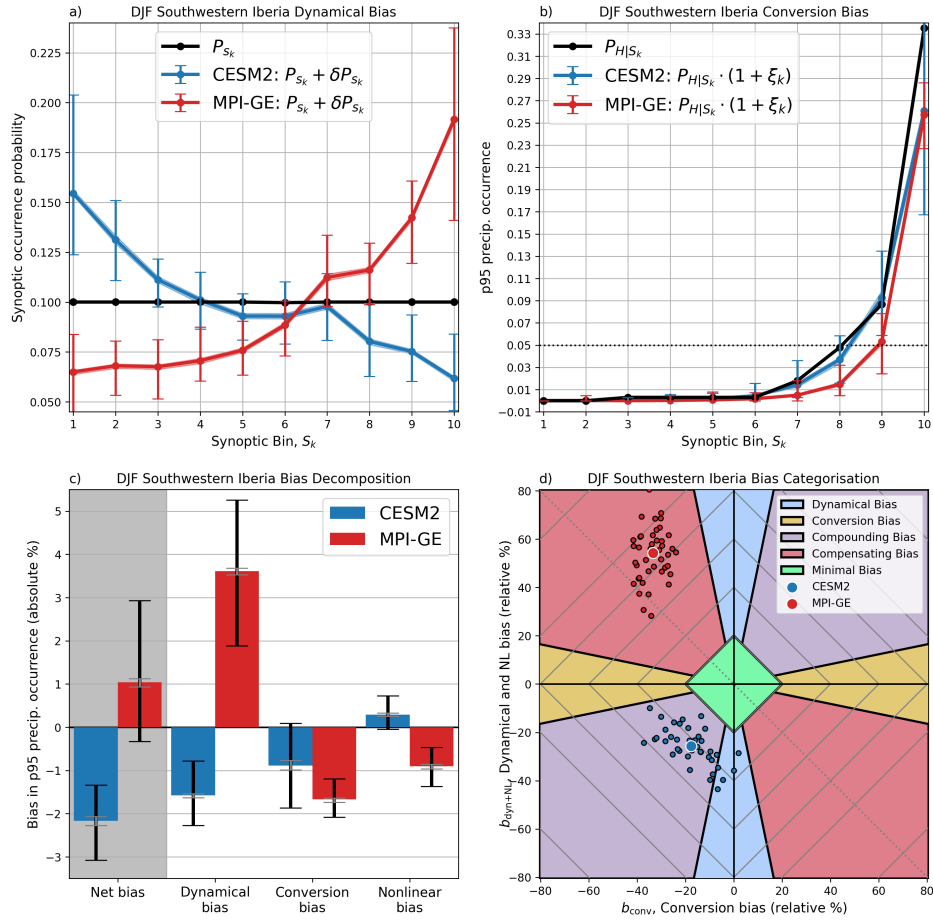


Figure 5. Flow-dependent precipitation bias for southwestern Iberia during DJF. a) Occurrence of different bins of the synoptic index over the historical period. Differences between model (red,blue) and ERA5 reference values (black) determine the **dynamical bias**. b) Heavy precipitation occurrence conditional on synoptic index bin. Dashed black line shows the 5% climatological heavy precipitation rate. Differences between model (red,blue) and ERA5 reference values (black) determine the **conversion bias**. c) Decomposition of net heavy precipitation biases into contributions from dynamical and conversion terms and their non-linear interaction. d) A visualisation of model bias in the two dimensional space of conversion bias and summed dynamical and non-linear bias. Individual ensemble members are shown with black outlines, ensemble mean values with white outlines. The dashed negative diagonal marks the line of zero net bias. Grey diamonds mark contours of constant absolute bias. The significance of the shaded regions is explained in the main text. The 95th percentile sampling uncertainty estimate is shown with shading in a) and b), grey bars in c) and is smaller than the ensemble mean dot in d). The full spread of internal variability is shown in bars in a) and b), black bars in c) and visualised directly in d).



relatively well for moderate precursor strength ($6 \leq K \leq 9$), with some underestimation for the strongest 10% of flows, whereas MPI-GE struggles to sufficiently convert any synoptic precursor into heavy precipitation.

To gain a holistic impression we can aggregate these flow-dependent biases into a three term bias budget:

$$\text{model bias} = \tilde{P}_H - P_H = \sum_k P_{H|S_k} \left(\underbrace{\xi_k P_{S_k}}_{\text{conversion bias}} + \underbrace{\delta P_{S_k}}_{\text{dynamical bias}} + \underbrace{\xi_k \cdot \delta P_{S_k}}_{\text{non-linear bias}} \right) \quad (2)$$

255 These terms are shown in Fig. 5c), with the amplitudes of each term directly interpretable in units of altered heavy precipitation occurrence. CESM2 both undersamples strong precursor circulations and underestimates the conversion of strong precursors into heavy precipitation. The resulting dynamical and conversion biases compound leading to a net bias of -2%; heavy precipitation in the region only occurs 3% of the time in CESM2 versus 5% in ERA5. MPI-GE appears to perform better with a net bias of 1%, but this is a misleading result that occurs due to compensation between dynamical and conversion biases.

260 The positive dynamical bias of MPI-GE would imply a heavy precipitation occurrence of 8.5%, but the model's severe under-conversion reduces this substantially. The two models' biases in this case are then not only of opposing sign, but of opposing type — compensating in MPI-GE and compounding in CESM2.

Different types of bias have different implications for model usability and interpretation. We propose five conceptual categories of bias which we discuss in detail in Sect. 7. To do this, we define a 2-dimensional bias space consisting of the conversion term, b_{conv} , and the sum of the dynamical and non-linear terms¹ $b_{\text{dyn+NL}}$, with an L1-norm $b = |b_{\text{conv}}| + |b_{\text{dyn+NL}}|$ where b is the absolute bias. This can then be used to define and visualise the five categories as follows:

1. Minimal bias: $b \leq b_{\text{max}}$
2. Conversion bias: $\frac{|b_{\text{dyn+NL}}|}{|b_{\text{conv}}|} \leq R$ and $b > b_{\text{max}}$
3. Dynamical bias: $\frac{|b_{\text{dyn+NL}}|}{|b_{\text{conv}}|} \geq \frac{1}{R}$ and $b > b_{\text{max}}$
- 270 4. Compounding bias: $\text{sign}(b_{\text{conv}}) = \text{sign}(b_{\text{therm}})$ and bias not Minimal/Conversion/Dynamical
5. Compensating bias: $\text{sign}(b_{\text{conv}}) \neq \text{sign}(b_{\text{therm}})$ and bias not Minimal/Conversion/Dynamical

We set the two parameters in these definitions to $b_{\text{max}} = 20\%$ (a maximum 'acceptable' bias level) and $R = 0.2$ (a relative importance threshold for different terms) to provide a proof-of-concept, but we stress that these should be tuned for user needs and this categorisation is an interpretive heuristic. Figure 5d) shows the southwestern Iberia biases within this space and

275 visualises the internal variability. For MPI-GE, variability in b_{conv} is small and independent from the bias in $b_{\text{dyn+NL}}$, which ranges from 20-80%. Notably, several ensemble members lie close to the line of zero net bias (marked by the dashed diagonal) while lying on the $b = 80\%$ contour. In contrast, the CESM2 ensemble members are closely clustered around the $b = 40\%$ contour, all with similar net bias and showing strong intra-ensemble correlation between b_{conv} and $b_{\text{dyn+NL}}$. The CESM2 spread

¹ As a note, the non-linear term could be grouped with the conversion term equally naturally. However as the most extreme conversion biases are much larger than the most extreme dynamical biases, a conversion+non-linear grouping can give bias terms with values lower than -100%, which hurts interpretability.



contains both members for which bias is almost entirely dynamical and those for which bias is predominately conversion-
280 related, highlighting the importance of large ensembles to understand model bias in detail.

4.2 Autumn biases in the northern Adriatic

The northern Adriatic in SON provides a second example with very different bias features. Heavy precipitation in this case is not predominately driven by large-scale blocks or jet anomalies but by Rossby wave packets of zonal wavenumber 6-7 (Grazzini et al., 2021), as visible in the precursor patterns and synoptic composites of Fig. 4iii, a-h). Figure 6a) shows
285 moderate dynamical biases: MPI-GE favours strong wave-trains in phases that both favour ($K = 10$) and suppress ($K = 1$) precipitation, while CESM2's bias towards negative precursors indicates a persistent Autumn ridge over central Europe (c.f. Fig. 4iiia)). However these dynamical biases are minor overall in comparison to the conversion biases shown in Fig. 6b). Both models have severe challenges converting dynamical drivers into precipitation, a conversion which in the real world relies on factors these models cannot resolve: convection, diabatic processes over the warm, shallow Adriatic, and uplift and blocking
290 from the Apennines and the Alps. As a result, conversion dominates the bias budget for both models and non-linear terms almost completely cancel the dynamical bias (Fig. 6c-d)). In Sect. 5 we will discuss how usable information about future heavy precipitation can be obtained from models even in such cases of severe historical bias.

4.3 Biases across Europe

Having introduced the bias decomposition and discussed two cases of interest, we now take a holistic view considering every
295 season and region except for those extremely dry cases where the ERA5 heavy precipitation threshold ≤ 2.5 mm. Figure 7 shows the conversion contribution to heavy precipitation bias computed following Eq. 2. The most coherent feature is the systematic under-estimation of conversion by both models during JJA, with occurrence biases of -0.03 or -0.04 in central and southern Europe. Similar, although typically weaker, conversion biases are seen in Southern and Eastern Europe for MAM and SON. Many of these cases correspond to wave-driven flows similar to the SON North Adriatic case (see supplementary material) or
300 cutoff lows over the Mediterranean, and it is plausible the low conversion is due to poorly represented diabatic, convective and orographic processes.

In contrast, conversion biases in Northern Europe during MAM and SON are often positive; cases where large-scale zonal flows with fast eastward-extended jets form the main precursor pattern. Conversion biases are most spatially diverse in DJF, perhaps surprisingly given the larger scale synoptic organisation. MPI-GE in particular struggles to convert dynamical precu-
305 sors to heavy precipitation in the most mountainous domains. Regional downscaling over Norway increases DJF precipitation from convective sources and localises it on the coast (Iversen et al., 2023), suggesting that resolution (possibly compounded by snow parameterisation errors) is indeed the cause of these conversion biases.

The dynamical contributions to heavy precipitation bias are in general more spatially coherent than the conversion biases (Fig. 8), and these features can be partly understood in terms of jet stream biases. CESM2 has positive dynamical biases in
310 wintertime heavy precipitation over northern Europe (Fig. 8a)) due to an overly strong and eastward extended jet (Simpson et al., 2020), with negative dynamical biases over southern Europe due to a corresponding lack of wave-driven and negative

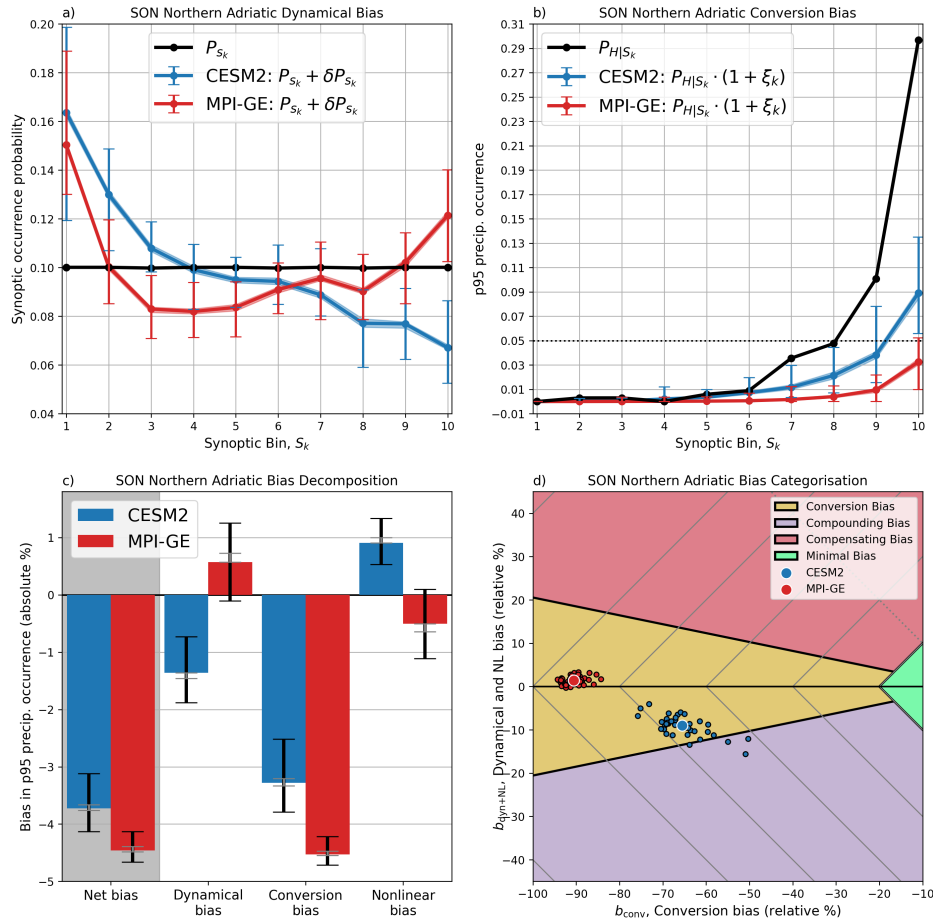


Figure 6. As in Fig. 5 but for the northern Adriatic region in SON.

NAO-like flows. Through the rest of the year CESM2 shows dry biases in western Europe (Fig. 8b-d)) due to a mispositioned jet and a persistent Atlantic ridge anomaly.

MPI-GE has a very different dynamical bias pattern, with excessive precipitation-causing dynamics through SON, DJF and MAM. This is a signature of excessive wave activity and wintertime cutoff-lows in the East Atlantic, the latter of which may be related to MPI-ESM-LE's excessively southward and zonal jet stream (Simpson et al., 2020). The wet dynamical bias in MPI-GE is consistent with the 'overshooting' precipitation found in higher resolution MPI-ESM models (Olonscheck et al., 2023), since resolution increases may disproportionately reduce the dry conversion biases relative to the wet dynamical biases.

Fig. 9 synthesises these biases using the 5-category bias categorisation introduced in Sect. 4.1. Results for all seasons and regions are also visualised in bias-space in Supplementary Fig. 4, offering an alternate perspective. Firstly, we note that achieving minimal bias is possible for both models in some regions, even if fairly uncommon — the daily, regional, and near-extreme scales we are interested in are not inherently beyond the capabilities of CMIP6-class models. Secondly, we see that

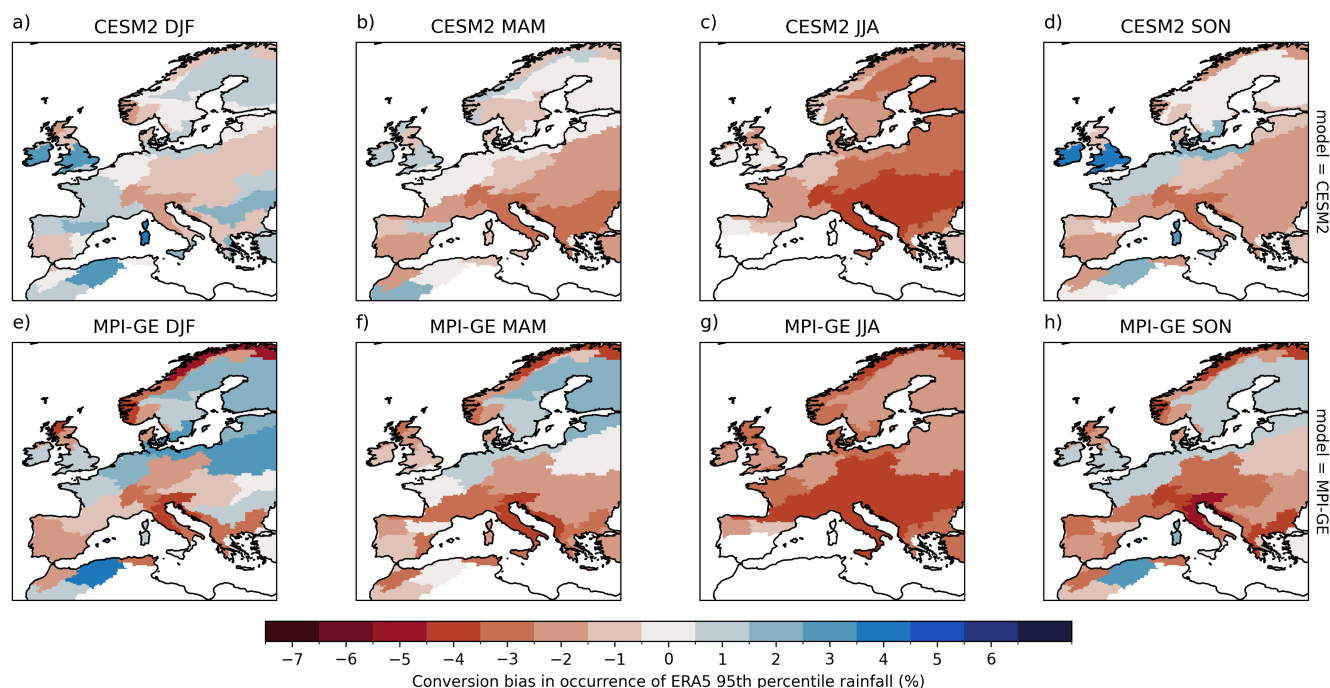


Figure 7. Ensemble mean biases in heavy precipitation occurrence attributable to **conversion biases** (errors in the conversion of synoptic forcing into heavy precipitation). Results are not shown for region-seasons where the heavy precipitation threshold $\leq 2.5\text{mm/day}$.

conversion biases are common, especially in JJA as expected due to the relevance of summertime convection, and yet dominant dynamical biases can occur in any seasons. Perhaps most importantly, we emphasise that compensating biases—as we identified for MPI-GE during DJF over South West Iberia — are common, occurring in both models throughout the year. The two models tend to favour compensating biases of particular sign: CESM2 features excessive conversion for too-rare synoptic conditions, while MPI-GE counterbalances too-common synoptic conditions with weak conversion (c.f. Supplementary Fig. 4).

5 Forced Changes

We now turn to future projections, considering the change in heavy precipitation occurrence between the historical 1979-2015 period and 2060-2100 under the SSP3-7.0 scenario. We focus on the change, ΔP_H , rather than a trend with respect to global temperature to avoid conflating global and local changes. In Section 4 we demonstrated both models contain pronounced heavy precipitation biases for certain regions, coming both from errors in the dynamics of synoptic forcing and small-scale precipitation conversion processes. When considering forced changes can these biases be ignored, or should we expect them to distort the projected forced response or even completely undermine the model's utility?

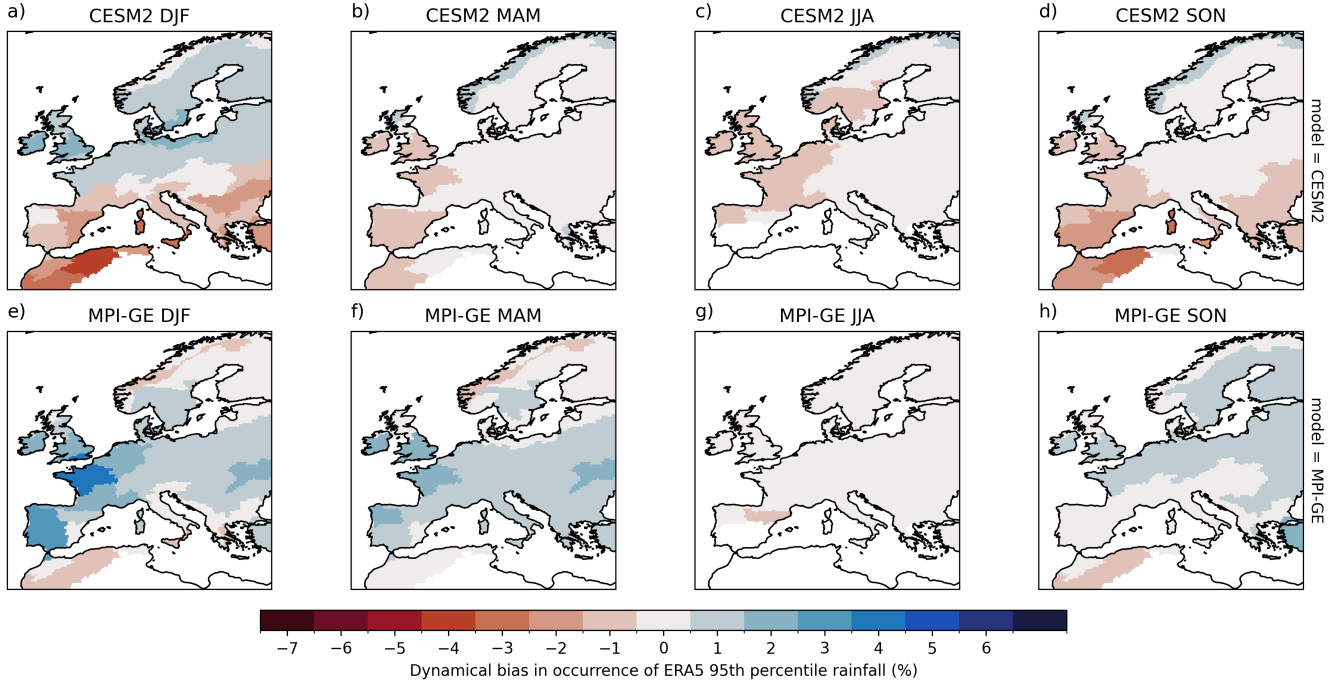


Figure 8. Ensemble mean biases in heavy precipitation occurrence attributable to **dynamical biases** (errors in the distribution of synoptic weather patterns). Results are not shown for region-seasons where the heavy precipitation threshold $\leq 2.5\text{mm/day}$.

335 We first consider an estimate where we assumed mean-state biases have no impact on forced changes. This is the perspective used, often implicitly, when directly computing the bulk *relative* change in model precipitation, which we denote $\tilde{\beta}$. As before, tildes denote biased model quantities, and now asterisks denote future quantities:

$$P_{H,\text{bulk estimate}}^* = P_H \cdot (1 + \tilde{\beta}) \quad (3)$$

$$\tilde{\beta} = \tilde{P}_H^* / \tilde{P}_H - 1 \quad (4)$$

340 Extending Eq. 1 to include forced changes reveals that $\tilde{\beta}$ can be distorted by model bias, and that this can be corrected for while also allowing us to decompose and interpret precipitation changes into synoptic and conversion-related contributions, just as we did for model biases. A model's future precipitation occurrence, P_H^* , can be decomposed into an additive dynamical change with respect to its historical state, ΔP_{S_k} , which captures changes in the occurrence of precipitation precursors, and a multiplicative change in precipitation conversion probability with respect to its historical state, α_k , which captures changes in the precipitation associated with a given dynamical precursor:

$$\tilde{P}_H^* = \sum_k (P_{H|S_k} [1 + \xi_k] [1 + \alpha_k]) \cdot (P_{S_k} + \delta P_{S_k} + \Delta P_{S_k}) \quad (5)$$

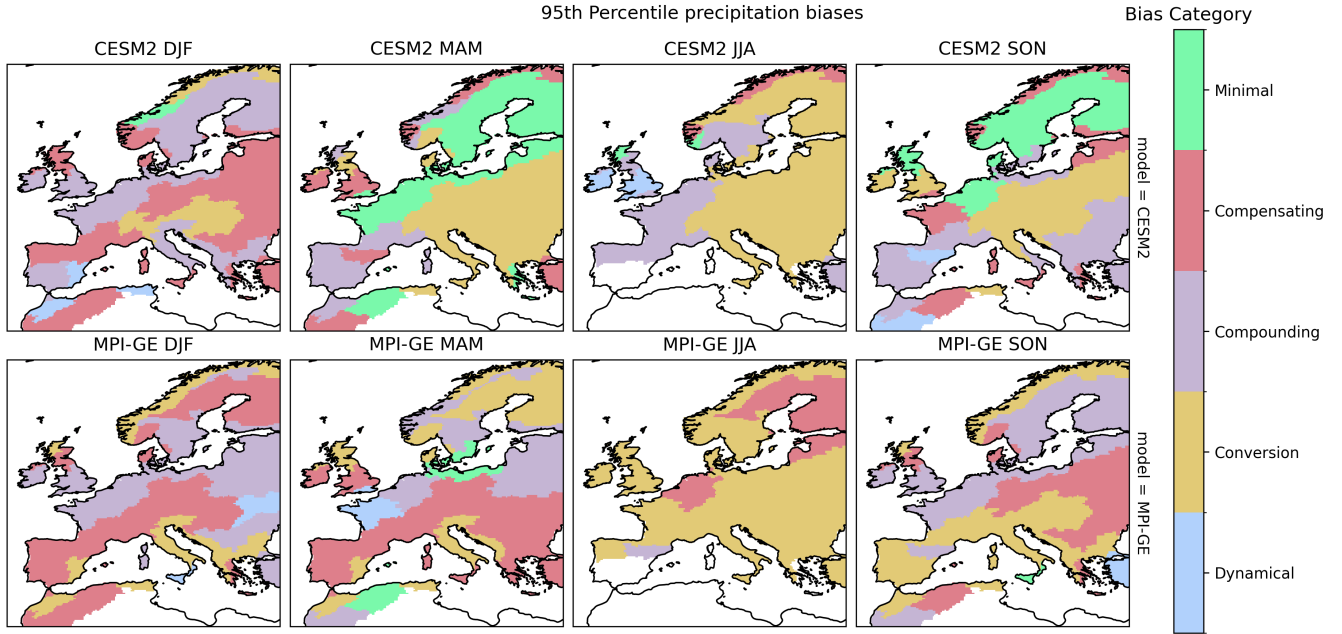


Figure 9. Categorisation of model biases based on the relative contribution of their conversion and dynamical biases. See main text for details.

As a note, a direct computation of $\alpha_k = \frac{\tilde{P}_{H|S_k}^*}{\tilde{P}_{H|S_k}} - 1$ can provide unphysical changes if the historical flow-dependent precipitation probability, $\tilde{P}_{H|S_k}$ is close to zero (see Appendix A.3). In these cases, we can reformulate the conversion change as additive by redefining $\alpha_k = \frac{\tilde{P}_{H|S_k}^* - \tilde{P}_{H|S_k}}{\tilde{P}_{H|S_k}}$, but such strongly biased cases will mostly be excluded from discussion unless explicitly indicated.

Equation 5 contains products between biases and forced changes, revealing that precipitation statistics in future simulations are distorted by historical biases. As Eq. 5 is the numerator of $\tilde{\beta}$ from Eq. 3 these distortions also impact the bulk estimate of forced precipitation changes. We quantify this distortion in Sect. 6. From a flow-dependent perspective we can drop these bias-change cross terms, and obtain a corrected estimate of the overall heavy precipitation change, β , which decomposes into dynamical, conversion, and non-linear contributions:

$$P_{H,\text{flow dependent estimate}}^* = P_H \cdot (1 + \beta) \quad (6)$$

$$\beta = \frac{1}{P_H} \sum_k P_{H|S_k} [\Delta P_{S_k} + \alpha_k \cdot P_{S_k} + \alpha_k \cdot \Delta P_{S_k}] \quad (7)$$

The power of this correction should not be overemphasised. The models' estimates of the flow-dependent forced changes α_k and ΔP_{S_k} may still ultimately be incorrect—and indeed given disagreements between models, most must be. However these corrected forced changes are, at a minimum, physically consistent with the current observed climate.



5.1 Summertime forced changes in the central UK and Ireland

Decomposed precipitation changes can offer new insights into future scenarios, reveal sources of model disagreement, and uncover cases where model disagreement is greater than it initially appears. A good example of hidden model disagreement is given by JJA forced changes over Central UK and Ireland (Fig. 10). Current large-scale summertime heavy precipitation in this region is driven by strong, localised cyclones (c.f. Fig 4ii), and is dynamically suppressed by anticyclonic ridges over the east Atlantic (Fig 4ii a,b,e,f). Both climate models project a net decrease in heavy precipitation, a shift from 5% to 4% occurrence probability in CESM2, and to 4.5% in MPI-GE. However Fig. 10c) shows that the decomposed budget of these forced changes is very different. MPI-GE's forced change is dominated by a negative dynamical contribution (Fig. 10a). CESM2's forced change by contrast is dominated by a negative conversion contribution (Fig. 10b). These are very different scenarios with differing implications: CESM2 describes a world with a similar number of strong summertime cyclones to today, but each less likely to cause heavy precipitation. MPI-GE describes a world with far fewer cyclonic anomalies but each more likely to produce heavy precipitation than those of today. Arguably the smaller net change in MPI-GE corresponds to the more volatile and impactful case.

Just as we did for biases in Sect. 4.1, we can construct a 2-dimensional space of conversion changes and summed dynamical and non-linear changes to define heuristic categories of forced changes. Figure 10d) shows this change space, and highlights how ensemble mean signals from both models lie close to the same line of -20% relative net change (dashed black), therefore appearing in agreement from a bulk perspective. Figure 10c-d) also demonstrates high internal variability in regional heavy precipitation changes, although mean signals are larger than the sampling uncertainty. Internal variability in forced changes is necessarily higher than for biases, because of compounding variability from historical and future periods. For any given 40-year period, it is therefore not a given that the ensemble-mean change we have described will actually be seen. The actual real-world implications of this are unclear as climate models have both insufficient low-frequency variability (Mann et al., 2020) including in European circulation features (Dorrington et al., 2022) and also suffer from a signal-to-noise problem which may lead to an overly-weak forced response (Scaife and Smith, 2018). In any case, this example demonstrates how decomposition of changes can also help identify signals on which models *are* confident: while the MPI-GE absolute net change in individual members spans from -1.7% to 1.5 %, the dynamical change is confidently negative for each member, -1.5% to -0.3%.

5.2 Forced changes across Europe

Figs. 11 and 12 show the model conversion ($\sum_k P_{H|S_k} \cdot \alpha_k \cdot P_{S_k}$) and dynamical changes ($\sum_k P_{H|S_k} \cdot \Delta P_{S_k}$) respectively. Non-linear changes ($\sum_k P_{H|S_k} \cdot \alpha_k \cdot \Delta P_{S_k}$) are small, as shown in supplementary figure 5. As our decomposition is based on the dominant precipitation driver in observations, it is conceivable that truly dynamical changes could be misattributed to the conversion term if new dynamics become important drivers of heavy precipitation in the future. Supplementary figures 6 and 7 indicate no evidence for this, with future conversion changes driven almost exclusively by changed conversion during strong precursor situations ($K=[8,9,10]$).

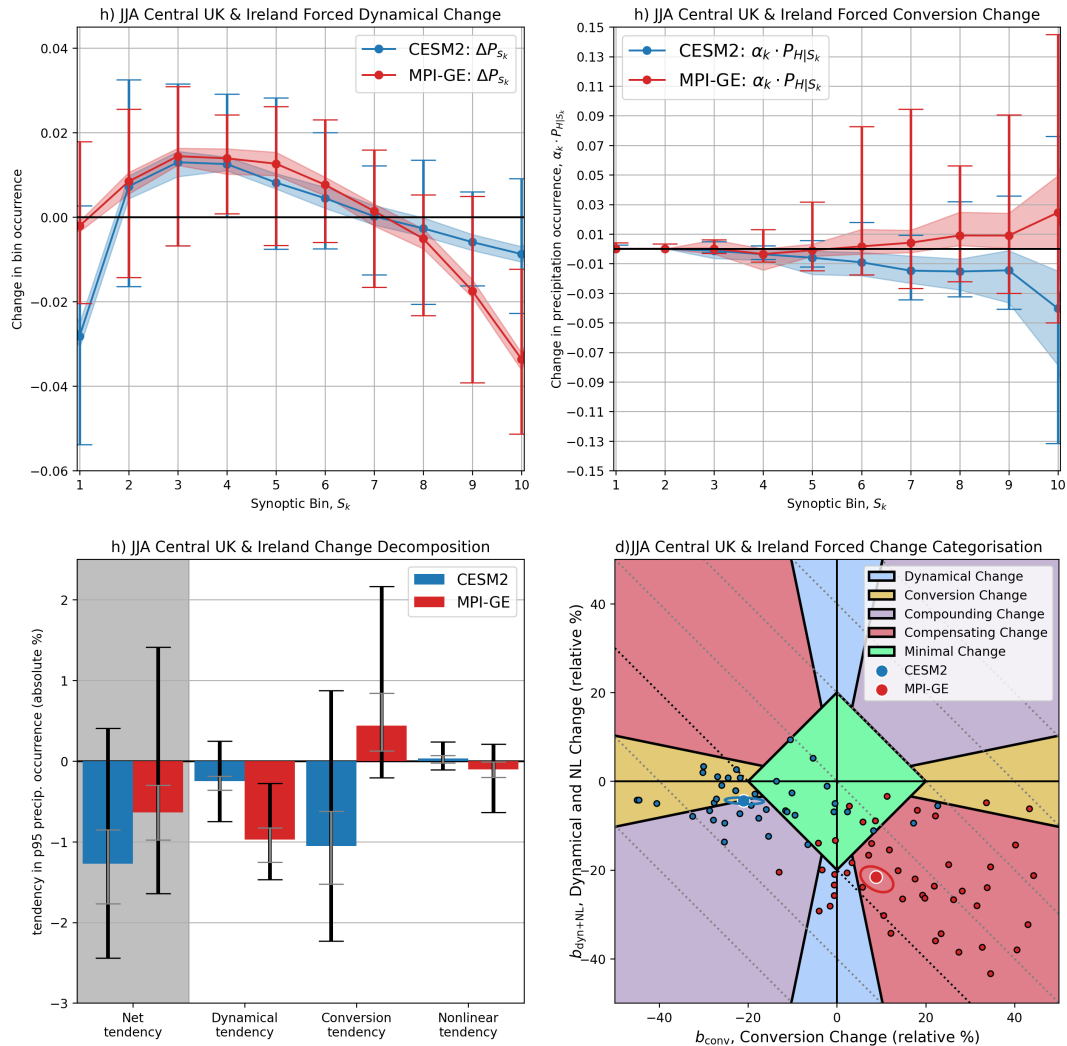


Figure 10. Flow-dependent forced changes in precipitation for the central UK and Ireland during JJA. a) Changes in occurrence of different bins of the synoptic index, determining the **dynamical change**. b) Changes in heavy precipitation occurrence conditional on synoptic index bin, determining the **conversion change**. c) Decomposition of net heavy precipitation changes into contributions from dynamical and conversion terms and their non-linear interaction. d) A visualisation of model forced change in the two dimensional space of conversion changes and summed dynamical and non-linear changes. Individual ensemble members are shown with black outlines, ensemble mean values with white outlines. Ellipses show the 2-standard deviation confidence interval around ensemble mean changes. Dashed lines mark contours of constant net change, with the -20% net change shown in bold. The 95th percentile sampling uncertainty estimate is shown with shading in a) and b) and with grey bars in c). The full spread of internal variability is shown with bars in a) and b), black bars in c) and visualised directly in d).



The bulk forced changes shown in Fig. 3 showed qualitative agreement between the two models. As a reminder, both projected less wintertime heavy precipitation in southern Europe, less heavy precipitation in southwestern Europe during spring and autumn, and less summertime heavy precipitation almost everywhere. In broad terms, both project less (more) wintertime heavy precipitation in southern (northern) Europe, less spring and autumn heavy precipitation in southwestern Europe and more elsewhere, and less summertime heavy precipitation, excepting Scandinavia.

In wintertime, both models show positive conversion changes of approx. 40%, except in southwestern regions where conversion declines (Fig. 11a,e)). Over southeastern Europe (Fig. 12a,e)) it is dynamical changes that drive decreased heavy precipitation. This implies a decline of Mediterranean winter cyclones, each individually more likely to cause heavy precipitation but leading to a net negative change overall. This is consistent with CMIP6 results from a cyclone-focused analysis (Chericoni et al., 2025) although we warn that they found the forced signal changed dramatic dramatically in higher resolution simulations. The models disagree on dynamical changes in northern Europe, with little signal in MPI-GE and 50-90% increases in heavy precipitation in CESM2, caused by more frequent and stronger eastward jet extensions. A careful reader may notice that CESM2's DJF conversion and dynamical changes Fig. 11a) and Fig. 12a)) sum to far more than the bulk change shown in Fig. 3i); this is a consequence of our correction of distorted trends and we revisit this in Sect. 6. CESM2's negative dynamical changes along the north Norwegian coast may be linked to the model's pronounced North Atlantic warming hole (Iversen et al., 2023) and the decline of cold-air outbreaks into the Greenland Sea (Konstali et al., 2024).

In MAM and SON, MPI-GE predicts a decline in heavy precipitation over southwestern regions due to compounding conversion and dynamical changes (Figs. 12 and 11 f and h). In JJA MPI-GE shows negative conversion changes in much of west and south-east Europe (Fig. 11e)). CESM2 has similar conversion changes to MPI-GE (Fig. 11b-d)), but different dynamical changes. CESM2 shows positive dynamical changes in MAM and SON in Northern Europe (Fig. 12b-d)) again related to a projected eastward jet extension, while negative dynamical changes in JJA relate to a projected less wavy summertime-flow (the reliability of which remains contentious (Stendel et al., 2021)). It has recently been suggested that JJA extreme precipitation may be 'shifting' to MAM and SON (Zhu et al., 2025), as summer rainfall decreases and shoulder season rainfall increases, consistent with Fig. 3. We see here that this is not a coherent phenomena, but rather is dynamically driven in some regions and seasons, and driven by local-scale conversion changes in others.

The categorisation of ensemble mean heavy precipitation changes is shown for all regions in Fig. 13 (and also in a 2D-space in Supplementary Fig. 8), and further highlights the differences between the two models, which initially seemed mainly in agreement from bulk metrics (Fig. 3). Predominately dynamical changes are rare; future changes in the relationship between synoptic dynamics and the surface (that is, a change in conversion) are nearly ubiquitous. However this does not mean dynamical changes can be neglected: the majority of cases feature compensating and compounding changes, where dynamical and conversion changes are of similar size.

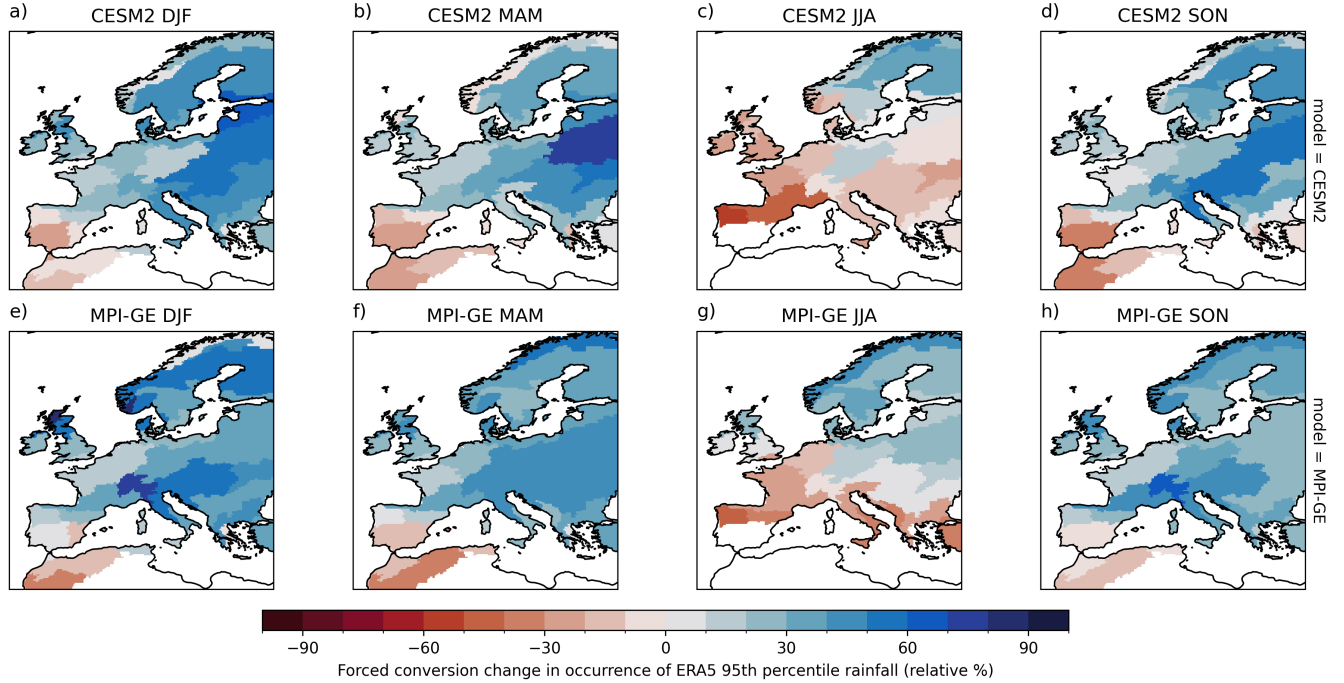


Figure 11. Ensemble mean relative changes (2060-2100, SSP3-7.0) in heavy precipitation occurrence attributable to changes in precipitation conversion. Results are not shown for region-seasons where the historical heavy precipitation threshold $\leq 2.5\text{mm/day}$.

6 Bias-Change Interactions

425 In the section we showed that the flow-corrected heavy precipitation changes β computed from Eq. 5 are not equivalent to the bulk changes $\tilde{\beta}$ computed from Eq. 3 as a result of dropping terms that mix biases and forced changes. How important is this difference in practice? Because Eqs. 5 and 3 have different denominators, the cross-terms alone do not explain the difference between bulk and flow-corrected changes. Instead the difference between bulk ($\tilde{\beta}$) and flow-corrected (β) multiplicative changes is given by (see Appendix A for derivation):

$$430 \quad \tilde{\beta} - \beta = \sum_k \left(\alpha_k \cdot [\tilde{F}_k - F_k] + (1 + \alpha_k) \cdot \Delta P_{S_k} \cdot [\tilde{G}_k - G_k] \right), \quad \text{where:} \quad (8)$$

$$F_k = \frac{P_{H|S_k} \cdot P_{S_k}}{P_H} = P_{S_k|H} \quad \tilde{F}_k = \frac{\tilde{P}_{H|S_k} \cdot \tilde{P}_{S_k}}{\tilde{P}_H} = \tilde{P}_{S_k|H} \quad G_k = \frac{P_{H|S_k}}{P_H} = \frac{P_{S_k|H}}{P_{S_k}} \quad \tilde{G}_k = \frac{\tilde{P}_{H|S_k}}{\tilde{P}_H} = \frac{\tilde{P}_{S_k|H}}{\tilde{P}_{S_k}}$$

F_k is the fraction of a dataset's heavy precipitation events (either ERA5 or historical model) that occur during a given synoptic condition, which we term the *flow relevance*. G_k is the relative odds of heavy precipitation under a given synoptic condition

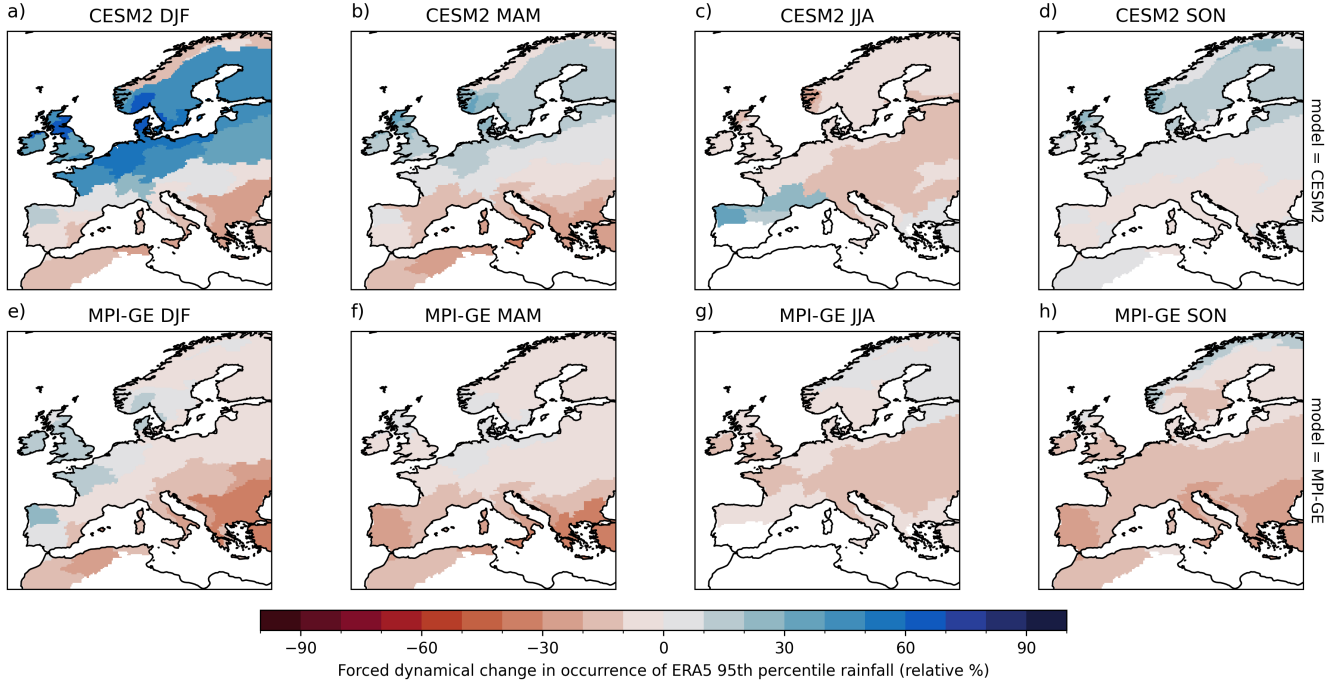


Figure 12. Ensemble mean relative changes (2060-2100, SSP3-7.0) in p95 heavy precipitation occurrence attributable to changes in synoptic dynamics. Results are not shown for region-seasons where the historical heavy precipitation threshold $\leq 2.5\text{mm/day}$.

within a dataset, or equivalently, the relative odds of a synoptic condition when heavy precipitation occurs. We term G_k the
 435 *flow impact*.

Equation 8 shows that only model biases in F_k and G_k will distort forced changes. Biases in F_k will distort conversion
 changes whereas biases in G_k will distort dynamical and non-linear changes. As a concrete example: for the central UK
 and Ireland in JJA, heavy precipitation probability during strong precursors ($K=10$) is 0.25 in ERA5 and 0.12 in CESM2,
 while the unconditional probability of heavy precipitation is 0.05 and 0.034 respectively. Therefore $G_{10} = 0.25/0.05 = 5$,
 440 while for CESM2 $\tilde{G}_{10} = 0.12/0.034 = 3.5$. This means that the flow impact of strong precursors in CESM2 is too small
 ($\tilde{G}_{10} - G_{10} < 0$) and so dynamical changes in heavy precipitation due to changes in the occurrence of strong precursors,
 $\Delta P_{S_{10}}$, will be underestimated.

For the two models considered here, the relative amplitude of errors in F and G is mostly $< 20\%$ (that is; a ± 0.01 change in
 heavy precipitation probability) and so provide only a minor refinement to the bulk change, as shown in Supplementary Fig.s
 9-11. The standout exception is for CESM2 DJF in northern Europe, where the flow-dependent decomposition reveals that
 445 changes are underestimated by as much as 50% due to errors in the flow impact, G . Figure. 14a-c) shows $\delta G = \sum_k [\tilde{G}_k - G_k]$
 summed over the lowest 4 synoptic bins (a), the middle 4 bins (b) and the top 2 bins (c). In northern Europe, CESM2 system-
 atically underestimates the flow impact of the strongest precursors, which capture strong eastward jet anomalies, while mildly

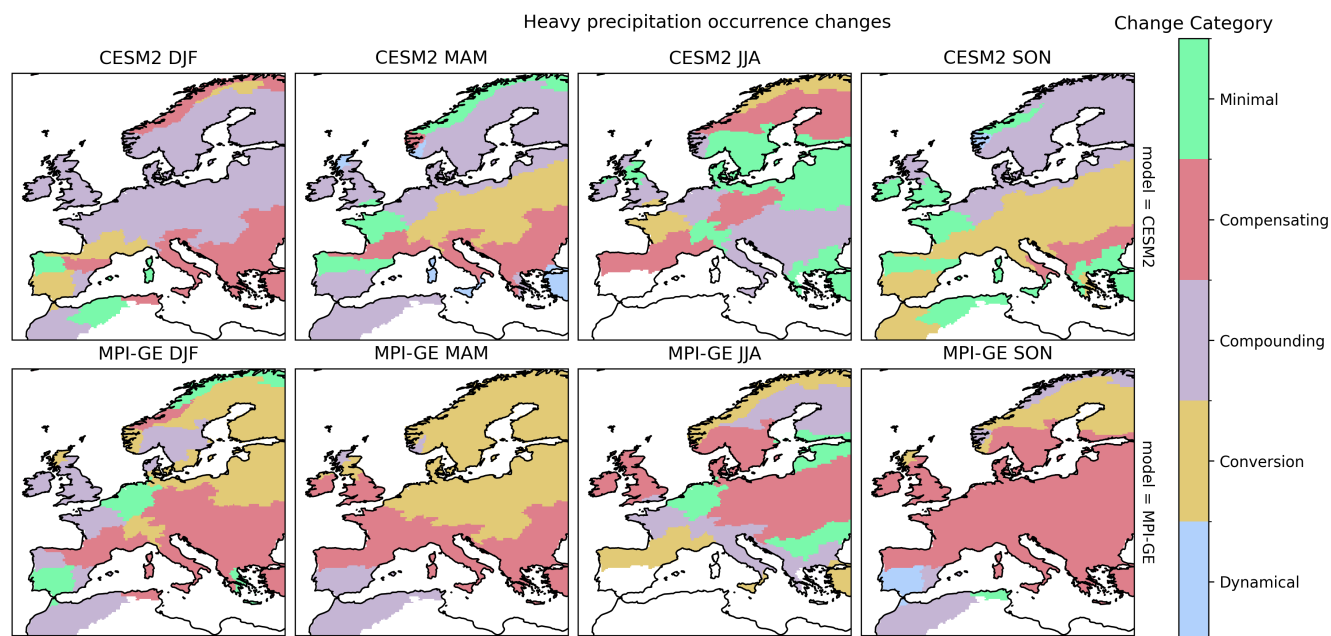


Figure 13. Categorisation of ensemble mean model changes based on the relative amplitude of their conversion and dynamical changes. See main text for details.

overestimating the impact of weak-to-moderate jet anomalies. Figure 14d) shows the spread of G_k over these regions and highlights a systematically low sensitivity in the heavy precipitation response: changes in synoptic conditions have too little impact on heavy precipitation probability. This lack of sensitivity means that CESM2's projected increase in strong, eastward jet days in a warmer climate (c.f. Fig. 12a) does not increase the model's precipitation as much as that same dynamical change would in the real world. As we live in the real world, it is the corrected precipitation change we are interested in.

6.1 Autumn forced changes in the northern Adriatic

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. However, flow decomposition can allow usable information to be extracted even in such cases. Considering again the northern Adriatic in SON, where both models showed severe conversion biases (c.f. Sect. 4.2) we find that both models project future increases in heavy precipitation probability (positive net change in Fig. 15c). We can use the additive reformulation of α_k introduced in Sect. 5 to compute conversion changes of $\approx +2\%$, but it is debatable whether this information should be used: the magnitude of the conversion biases indicates that the underlying real-world processes that will drive future conversion changes are simply not accounted for. The dynamical changes however can be considered quite independently. CESM2 projects negligible changes in the wave-driven dynamics that drive North Adriatic precipitation (Fig. 15a). MPI-GE shows signs of fewer mid-Mediterranean troughs and more ridges, resulting in a negative (-1%) dynamical

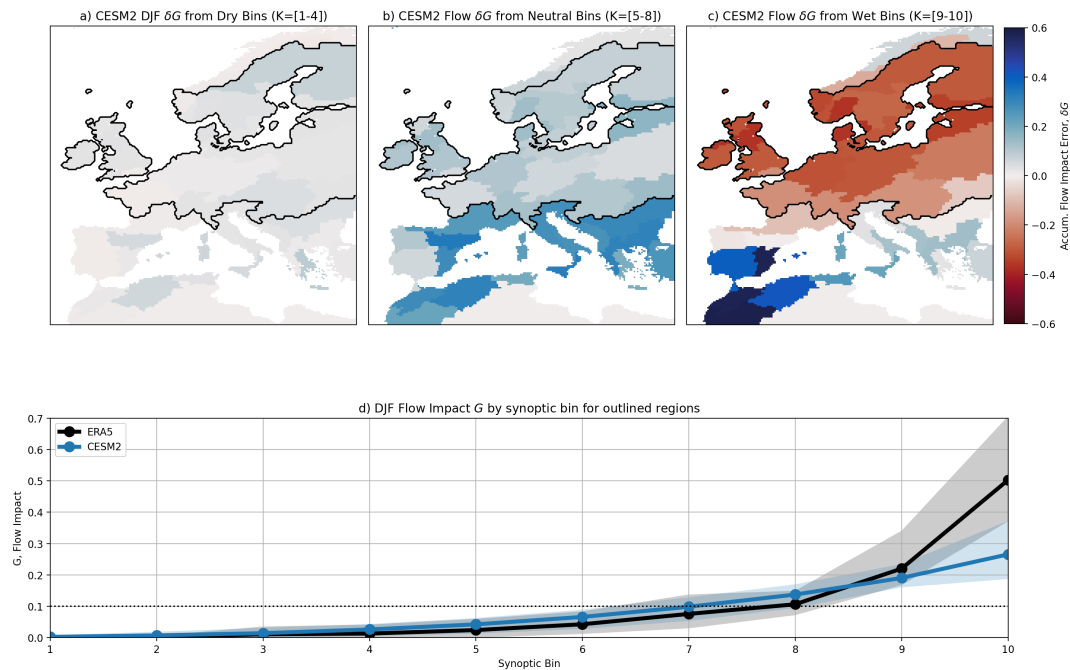


Figure 14. DJF biases in the flow impact, G_k , aggregated for the 40% of synoptic conditions least favouring heavy precipitation (a), the middle 40% of conditions (b) and for the 20% of conditions that most favour heavy precipitation (c). Negative/positive biases indicate a given synoptic condition is less/more likely to cause heavy precipitation in a model than in observations, after accounting for the model's bulk bias in heavy precipitation occurrence. The black contour outlines 15 regions with large under-sensitivity to strong synoptic precursors. d) G_k averaged over the 15 outlined regions for ERA5 and for CESM2, with the spread in values between regions indicated by shading.

contribution to changes in heavy precipitation, and a negative dynamical change in every ensemble member (Fig. 15c)). These dynamical changes are based on the model's dynamics and real-world precipitation conversion, and so can be considered a valid projection despite the model's conversion errors. Decomposing the overall change into these contributions allows each to be assessed and acted upon on its own merits, increasing the usability of the simulations overall.

7 Discussion

We have not discussed details of heavy precipitation biases and changes for all 152 region-season combinations computed for each model. Nevertheless, through a thorough investigation of European heavy precipitation in CESM2 and MPI-GE, we have identified a number of generalisable points of interest.

Model biases cannot be simply subtracted or divided out: As we have shown in Sect. 5 and 6, flow-dependent biases—especially compensating biases—can distort changes. In the worst case, these distortions can be of the same magnitude as the

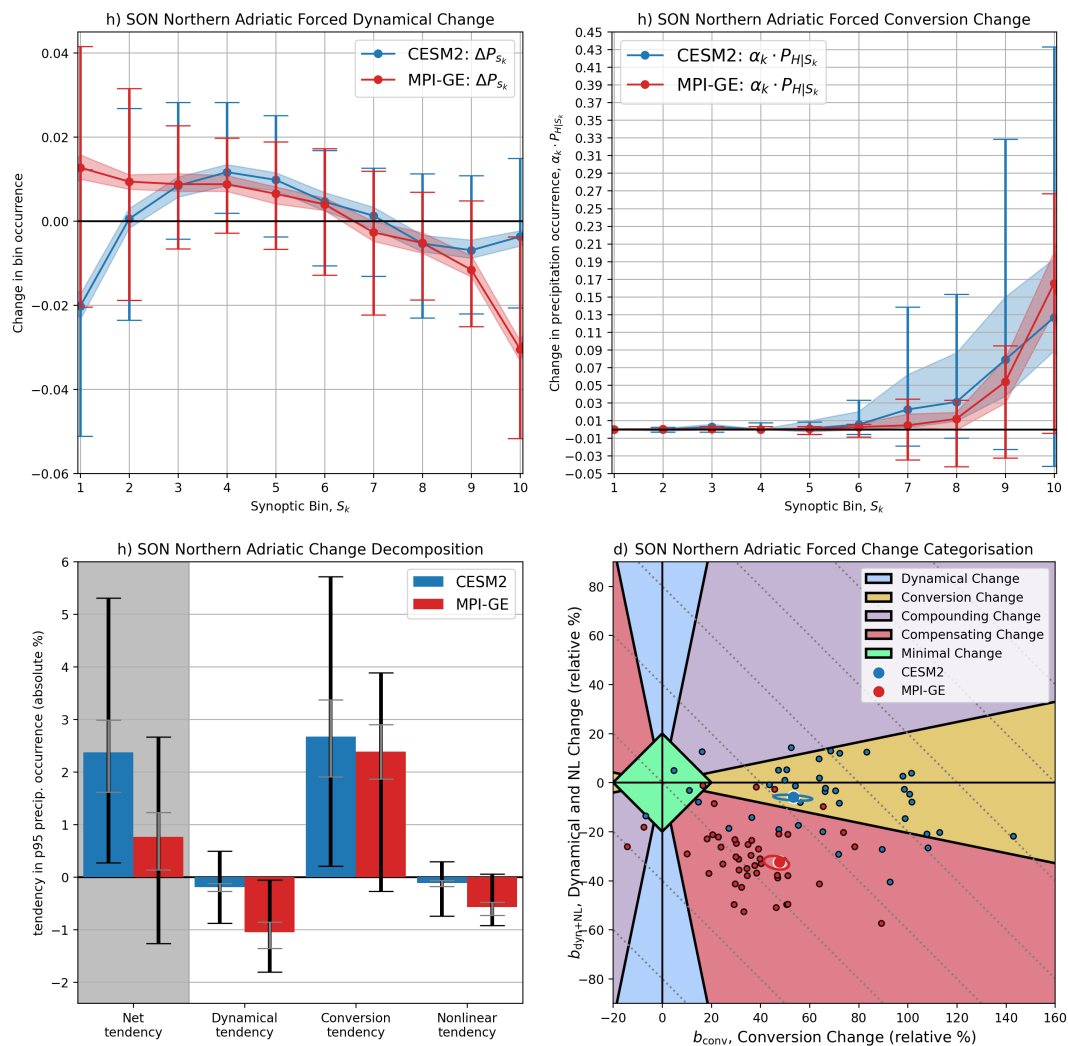


Figure 15. As for Fig. 10 but for the northern Adriatic region in SON.



475 true change. This was the case for CESM2's projections of wintertime heavy precipitation: the model's projected strengthening of the wintertime jet would imply an increase in northern European heavy precipitation probability 30-60% larger than direct model output indicates, given the observed synoptic-to-precipitation relationship.

Compensating heavy precipitation biases are common: Errors in a model's synoptic circulation often partially cancel out aggregated deficiencies in the conversion of dynamical forcing into heavy precipitation by mesoscale, boundary layer, land surface, thermodynamic and microphysical processes (c.f. Fig. 9). In these cases models are less trustworthy than they initially appear, and improvements of model processes may worsen their bulk performance. This helps to explain the sometimes surprisingly poor results of increasing model resolution without retuning and warns that expensive regional downscaling efforts may have disappointing results when driven by models with compensating errors. While regional modellers select global models with an eye to dynamical biases (Sobolowski et al., 2025) this focus is mostly limited to dominant variability modes, e.g. the Atlantic jet position. However we have seen that conversion biases are often compensated by dynamical biases in precursor flows rarely discussed in the literature, and so would be missed. With a flow-dependent decomposition of precipitation, we can identify, explain and quantify these and other cases, visualising them either in a 2D space (e.g. Fig. 5d)) or on a map (e.g. Fig. 9).

490

Flow-dependent model biases can be categorised to inform the downscaling and climate services communities: While the exact thresholds used to define categories should be tuned to different use cases, the practical interpretation of these conceptual categories remains the same:

1. *Minimal bias:* With both realistic dynamics and conversion processes, regional precipitation from the model can be used at face-value.
2. *Conversion bias:* Deficiencies in conversion processes but accurate large-scale dynamics make this model an excellent candidate for regional downscaling.
3. *Dynamical bias:* Accurate conversion of dynamical forcing into precipitation but a skewed distribution of synoptic circulations mean the model could be useful for storyline analysis, but should not be used to derive bulk probabilities or drive regional simulations unless nudging or bias correction techniques are employed.
4. *Compounding bias:* The physical realism of this model for the region is better than it appears, and could be somewhat improved by downscaling or improving model processes.
5. *compensating bias:* The model should be used with caution in this region and can be expected to actively worsen locally when improvements are made to either the large-scale circulation or the conversion processes alone.

505 **Apparent agreement on bulk statistics between models is not always actually agreement:** As demonstrated (c.f. Fig 10), models may produce the same net change through very different future scenarios. This can give a misleading sense of con-



confidence in changes when actually the models have inconsistent narratives. Different future scenarios can be efficiently distinguished with flow-dependent metrics and quickly visualised with a categorisation of forced changes, just as for biases:

1. *Minimal change*: Future heavy precipitation will resemble the present.
- 510 2. *Conversion change*: Individual synoptic conditions can be expected to become more/less reliably extreme, altering precipitation volatility.
3. *Dynamical change*: Heavy precipitation will be caused by a change in the frequency of synoptic conditions, each behaving the same as today.
4. *Compounding change*: Heavy precipitation will increase due to a greater number of more extreme synoptic conditions,
 515 or decrease due to fewer, less extreme conditions.
5. *Compensating change*: Fewer, stronger synoptic conditions or (rarely) more frequent but weaker conditions. The former case poses the greater risk for historically unseen precipitation extremes.

Flow-dependent approaches can make biased models more useful: Decomposing model behaviour is not simply a way to critique existing climate simulations, but offers opportunities to improve their value. For moderate biases, we have shown
 520 how flow-dependent decomposition provides a flow-aware bias correction and recalibration of forced changes. Even when models arguably cannot represent regional conversion processes well enough to model their future change, as we suggest for the North Adriatic region (c.f. Figs. 6 and 15), the dynamical component of heavy precipitation change can still be isolated and understood.

8 Conclusions and outlook

525 How can we understand the enormous complexity and spatiotemporal heterogeneity of precipitation in climate models? This study has tackled this question with an eye to both informing model development and ensuring existing projections can be well used, focusing on biases and changes in European daily heavy precipitation in large-ensemble experiments from two leading climate models.

We introduced a multi-scale, flow-dependent decomposition, analysing separately biases and forced changes in the synoptic
 530 dynamics that cause precipitation and in the of conversion of those dynamics into heavy precipitation. The dynamics are represented by flow-precursors that efficiently encode the most important local precipitation-causing processes, enabling us to provide a comprehensive 4-season analysis covering 38 regions. We show that compensating biases and changes between scales are common, that these have serious implications for model interpretation, and that our flow-dependent approach is well suited to understanding these. We find a more central role for dynamics in explaining biases and changes than previous work. We
 535 suggest this is due to the use of flow-precursors rather than regimes to compute our decomposition but several methodological differences may also be relevant. Namely, our choice of climate models, our focus on categorical regionally-aggregated heavy precipitation events, and our particular decomposition formula may explain some differences in our findings.



A natural next step would be to apply this analysis to the full range of CMIP6 and soon-to-come CMIP7 simulations, providing a breakdown of structural uncertainties in the precipitation projections that will support the next IPCC assessment.

540 Our results showed high variability in future changes across ensemble members, pointing to the need for large ensemble simulations where possible. Limiting the flow decomposition to a smaller number of synoptic bins (e.g. $K = 5$) may aid extension of the approach to single-member model runs.

An exciting prospect is to apply our analysis to the growing wealth of convection permitting regional simulations over Europe, which inherit much of their precipitation-causing dynamics from the driving global climate models, but promise to
545 resolve the conversion processes more faithfully. The forced change budget we derived can combine information from a variety of sources, allowing one to compute dynamical precipitation changes from large ensemble output and conversion changes from convection-permitting regional models, for example. Such a quantitative synthesis of the increasingly diverse climate modelling landscape is so-far lacking, and pursuing this is an area of active work for the authors.

The general framework of our approach could be easily extended beyond Europe to other mid-latitude regions, to other
550 precipitation metrics, and beyond precipitation to any surface hazard strongly conditioned on synoptic dynamics, such as wind gusts or fire risk. In subtropical regions, and to some extent the midlatitude summer, heavy precipitation is less organised by synoptic dynamics, and extending the precursor approach down to include 3-dimensional mesoscale dynamics is another area of planned work for the authors.



Appendix A: Extended derivation of precipitation decomposition

555 In this appendix we expand upon and consolidate the derivation of the bias and change decomposition introduced in the main text, and discuss a small number of special cases which ultimately had a negligible role in our analysis of CESM2 and MPI-GE, but may have more relevance in other contexts.

A1 Decomposition

560 We represent the observed probability of a particular hazard H , as conditional on some parameterisation of the large-scale flow S :

$$P_H = \int_{-\infty}^{\infty} P(H|S) \cdot P(S) ds \quad (\text{A1})$$

To allow empirical estimation of both $P(H|S)$ and $P(S)$ we discretise into k bins, where the equality with P_H is still exact but where the decomposition into P_S and $P_{H|S}$ is now approximate:

$$P_H = \sum_k P_{H|S_k} \cdot P_{S_k}$$

565 $1 = \sum_k P_{S_k}$

This formulation is now mathematically equivalent to e.g. a decomposition into k categorical flow regimes, although of course the exact choice of S will significantly impact the results. We define S by discretising the first principal component of Z500, U850 and V850 precursor indices into k evenly spaced quantiles. This decomposition is then specific to a given season and region.

570 Throughout we will denote biased model quantities using tildes, and future quantities (whether model or best real-world estimate) with asterisks. We denote P_{S_k} as the dynamical term and $P_{H|S_k}$ as the conversion term. For an imperfect model with hazard probability $\tilde{P}_H$ we represent errors in the occurrence frequency of the synoptic states $\{S_k\}$ as additive errors δP_{S_k} . We represent errors in the conditional conversion term as multiplicative errors ξ_k , which is a choice informed by the known dynamics of precipitation. As such:

$$575 \quad \tilde{P}_H = \sum_k (P_{H|S_k} [1 + \xi_k]) \cdot (P_{S_k} + \delta P_{S_k}) \quad (\text{A2})$$

To consider forced changes, we denote the future hazard probability of the model as P_H^* , and again formulate dynamical changes, ΔP_{S_k} , as additive, and conversion changes, α_k , as multiplicative. Rewriting P_H^* in terms of deviations from the observed P_H caused by both biases and forced changes we obtain:

$$P_H^* = \sum_k (P_{H|S_k} [1 + \xi_k] [1 + \alpha_k]) \cdot (P_{S_k} + \delta P_{S_k} + \Delta P_{S_k}) \quad (\text{A3})$$

580 Why replace the concise left hand side of this equation with a lengthy decomposition? Ultimately, we wish to know the future real-world hazard probability P_H^* , but this is inaccessible to us—we only have estimates from imperfect models. While



there are surely errors in the model changes ΔP_{S_k} and α_k , we have no way to identify these a priori. Ultimately we must trust some aspect of the model output or else we cannot proceed. Our contention is that **errors in the decomposed, synoptically-conditioned changes $\{\alpha_k\}$ and $\{\Delta P_k\}$ will be smaller, in aggregate, than the change error coming from a direct bulk computation $\tilde{P}_H^*/\tilde{P}_H$** . 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. If we then accept the hypothesis that flow dependent changes provide a more trustworthy assessment of future change, it is useful to rewrite Eq. A3 as:

$$P_H^* = \sum_k \left[\underbrace{P_{H|S_k} \cdot P_{S_k}}_{\text{Observed reference}} + \underbrace{P_{H|S_k} [\delta P_{S_k} + \xi_k (P_{S_k} + \delta P_{S_k})]}_{\text{Historical model bias}} + \underbrace{P_{H|S_k} [\alpha_k P_{S_k} + \Delta P_{S_k} (1 + \alpha_k)]}_{\text{Adjusted model change}} + \underbrace{P_{H|S_k} \left[\underbrace{\alpha_k \cdot \delta P_{S_k}}_{\text{Dynamical error}} + \underbrace{\xi_k [\alpha_k \cdot P_{S_k} + (1 + \alpha_k) \cdot \Delta P_{S_k}]}_{\text{Thermodynamic error}} + \underbrace{\xi_k \alpha_k \cdot \delta P_{S_k}}_{\text{non-linear error}} \right]}_{\text{Spurious model change}} \right] \quad (\text{A4})$$

We see that the observed conversion rate for a given synoptic state, $P_{H|S_k}$, factors out completely from all terms, meaning that changes and biases in states where $P_{H|S_k}$ is large will have greater importance. We now have 4 groupings of terms. The first is simply the observational reference. The three terms in the second grouping contain the dynamical, conversion and non-linear contributions to model bias respectively. The third grouping contains all terms which contain a change in hazard, but with no contribution from model bias. They therefore tell us what the model implies about real-world hazard changes. Finally, the fourth grouping of terms contains products of changes and model bias. We consider these last terms to be spurious changes as they represent changes in future hazard probability that are not consistent with our current observed reality. If hazard changes are considered as additive then the differences between the bulk change, $\tilde{P}_H^* - \tilde{P}_H$, and the flow-dependent change is simply this fourth group of terms. However, precipitation changes are typically considered as multiplicative:

$$\tilde{\beta} = \frac{\tilde{P}_H^*}{\tilde{P}_H} - 1 \quad (\text{A5})$$

where $\tilde{\beta}$ is the bulk multiplicative change. The flow-dependent multiplicative change is simply:

$$\beta = \frac{\sum_k P_{H|S_k} [\alpha_k P_{S_k} + \Delta P_{S_k} (1 + \alpha_k)]}{P_H} \quad (\text{A6})$$

but now computing $\tilde{\beta} - \beta$ is complicated by the different denominators in Eqs. A5 and A6, which mean the spurious terms cannot be directly disentangled into conversion and dynamical contributions. However we can proceed as follows:



$$\tilde{\beta} - \beta = \frac{\sum_k \tilde{P}_{H|S_k} \cdot (\tilde{P}_{S_k} + \Delta P_{S_k}) \cdot (1 + \alpha_k)}{\tilde{P}_H} - \frac{\sum_k P_{H|S_k} \cdot (P_{S_k} + \Delta P_{S_k}) \cdot (1 + \alpha_k)}{P_H} \quad (\text{A7})$$

$$= \sum_k \left[\Delta P_{S_k} \cdot (1 + \alpha_k) \left[\frac{\tilde{P}_{H|S_k}}{\tilde{P}_H} - \frac{P_{H|S_k}}{P_H} \right] + \alpha_k \left[\frac{\tilde{P}_{H|S_k} \cdot \tilde{P}_{S_k}}{\tilde{P}_H} - \frac{P_{H|S_k} \cdot P_{S_k}}{P_H} \right] \right] \quad (\text{A8})$$

$$\equiv \sum_k \left[\Delta P_{S_k} \cdot (1 + \alpha_k) [\tilde{G}_k - G_k] + \alpha_k [\tilde{F}_k - F_k] \right] \quad (\text{A9})$$

Where in Equation 17 we have newly defined F_k and G_k , which we term the flow relevance and flow impact respectively. Both new terms can be written in alternate forms, which aids with their interpretation:

$$F_k = \frac{P_{H|S_k} \cdot P_{S_k}}{P_H} = P_{S_k|H} \quad G_k = \frac{P_{H|S_k}}{P_H} = \frac{P_{S_k|H}}{P_{S_k}}$$

The same alternate forms hold for $\tilde{F}_k$ and $\tilde{G}_k$, with tildes on all quantities. The flow relevance is therefore the fraction of hazard events that occur within a particular synoptic bin, while the flow impact is the conditional odds ratio of a hazard.

Additive errors in G_k and F_k *weighted by the flow-dependent forced changes* explain the differences between the net bulk and flow dependent multiplicative changes. If a model has constant conversion bias (that is $\xi_k = \xi \ \forall \ k$) then $\tilde{G}_k = G_k$ and dynamical changes are not distorted. However even under these conditions, nonzero dynamical bias will still result in distortions in conversion changes.

A2 Novel synoptic conditions

In principle it is possible for a model to produce a novel synoptic state entirely outside the observed distribution, which we can assign to a new bin n . The frequency of occurrence of these conditions is $\delta P_{S,n}$ and where for lack of more information we assume $\xi_n = 0$:

$$\tilde{P}_{H,n}^* = \tilde{P}_{H|S_n} \delta P_{S_n} + \tilde{P}_{H|S_n} \Delta P_{S_n} [1 + \alpha_n] + \alpha_n \tilde{P}_{H|S_n} \delta P_{S_n} \quad (\text{A10})$$

which gives us the bias, the adjusted trend, and the spurious trend respectively. If novel synoptic states occur in the future simulations only, then we can estimate none of the biases or the trends, and so we can only include the contribution of the bin as a whole: $P_H^* = \sum_k P_{H_k}^* + P_{H_n}^*$. In fact, novel states have no impact in the two models we consider in this paper, and so they are not included in our main analysis. However we cannot rule out their relevance for more severely biased models or in more extreme climate scenarios.

A3 Conversion changes under severe bias

α_k is formulated as multiplicative, but this can cause issues if $P_{H|S_k}, P_{H|S_k}^* \gg \tilde{P}_{H|S_k}$, that is, if the historical simulation shows near-zero conversion for a particular weather pattern but does exhibit conversion in a future climate. As a concrete example,



consider $P_{H|S_k} = 0.04$, $\tilde{P}_{H|S_k} = 0.002$ and $P_{H|S_k}^* = 0.02$. A direct reading of this would give $\alpha = 10$ and an estimate of real-
630 world future hazard probability rising from 4% to 40%. This is clearly unrealistic, and in most cases conversion changes under
such conditions should be disregarded as uninformative. However for completeness and to handle border-line cases, we model
conversion changes as additive when $1 + \xi_k$ is close to zero:

$$P_{S_k} P_{H|S_k}^* = P_{S_k} P_{H|S_k} \cdot (1 + \alpha_k) = \left. \begin{array}{l} P_{S_k} P_{H|S_k} \left(\frac{\tilde{P}_{H|S_k}^*}{\tilde{P}_{H|S_k}} \right), \quad \text{for } \xi \gg -1 \\ P_{S_k} P_{H|S_k} \left(1 + \frac{\tilde{P}_{H|S_k}^* - \tilde{P}_{H|S_k}}{P_{H|S_k}} \right) \approx P_{S_k} P_{H|S_k} \left(1 + \frac{\tilde{P}_{H|S_k}^*}{P_{H|S_k}} \right), \quad \text{for } \xi \approx -1 \end{array} \right\}$$

To avoid a sharp switch in the algorithm which might be sensitive to resampling or parameter values and can distort ensemble
635 spread, we compute α with a blending function:

$$\alpha_k = (1 - w(\xi_k)) \cdot \frac{\tilde{P}_{H|S_k}^*}{\tilde{P}_{H|S_k}} + w(\xi_k) \cdot \left(\frac{\tilde{P}_{H|S_k}^*}{\tilde{P}_{H|S_k}} - 1 \right) \quad (\text{A11})$$

where

$$w(\xi) = \frac{(1 + \xi)^4}{(1 + \xi)^4 + 0.1^4} \quad (\text{A12})$$

For small or positive conversion biases $w = 1$ with a deviation of only $\left(\frac{0.1}{1 + \xi} \right)^4$, while for strong negative bias, $w = 0$ with
640 an deviation of only $\left(\frac{1 + \xi}{0.1} \right)^4$. Testing has showed this reduces sensitivity of the decomposition results to the number of bins,
 k , chosen and gives more ‘sensible’ results for these special cases.

Author contributions. JOD was responsible for conceptualisation, methodology, software, visualisation, original draft preparation and editing.
CL and SS were involved in conceptualisation, review and editing. RGC was involved with methodology and review and editing.

Competing interests. CL is a member of the editorial board of Weather and Climate Dynamics.

645 *Acknowledgements.* This work was funded by European Union’s Horizon 2020 research and innovation programme under the Marie
Skłodowska-Curie grant agreement No 101151904.



References

- Abdelmoaty, H. M., Papalexiou, S. M., Rajulapati, C. R., and AghaKouchak, A.: Biases Beyond the Mean in CMIP6 Extreme Precipitation: A Global Investigation, *Earth's Future*, 9, e2021EF002 196, <https://doi.org/10.1029/2021EF002196>, 2021.
- 650 Bador, M., Boé, J., Terray, L., Alexander, L. V., Baker, A., Bellucci, A., Haarsma, R., Koenigk, T., Moine, M.-P., Lohmann, K., Putrasahan, D. A., Roberts, C., Roberts, M., Scoccimarro, E., Schiemann, R., Seddon, J., Senan, R., Valcke, S., and Vanniere, B.: Impact of Higher Spatial Atmospheric Resolution on Precipitation Extremes Over Land in Global Climate Models, *Journal of Geophysical Research: Atmospheres*, 125, e2019JD032 184, <https://doi.org/10.1029/2019JD032184>, 2020.
- Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I. J. M., McVicar, T. R., and Adler, R. F.: MSWEP V2 Global
655 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment, *Bulletin of the American Meteorological Society*, 100, 473–500, <https://doi.org/10.1175/BAMS-D-17-0138.1>, 2019.
- Brands, S.: Common Error Patterns in the Regional Atmospheric Circulation Simulated by the CMIP Multi-Model Ensemble, *Geophysical Research Letters*, 49, e2022GL101 446, <https://doi.org/10.1029/2022GL101446>, 2022.
- Cassano, J. J., Uotila, P., Lynch, A. H., and Cassano, E. N.: Predicted Changes in Synoptic Forcing of Net Precipitation in Large Arctic River
660 Basins during the 21st Century, *Journal of Geophysical Research: Biogeosciences*, 112, 0, <https://doi.org/10.1029/2006JG000332>, 2007.
- Cattiaux, J., Douville, H., and Peings, Y.: European Temperatures in CMIP5: Origins of Present-Day Biases and Future Uncertainties, *Climate Dynamics*, 41, 2889–2907, <https://doi.org/10.1007/s00382-013-1731-y>, 2013.
- Chericoni, M., Fossier, G., Flaounas, E., Gaetani, M., and Anav, A.: Unravelling Drivers of the Future Mediterranean Precipitation Paradox during Cyclones, *npj Climate and Atmospheric Science*, 8, 260, <https://doi.org/10.1038/s41612-025-01121-w>, 2025.
- 665 Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2
670 (CESM2), *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001 916, <https://doi.org/10.1029/2019MS001916>, 2020.
- Deser, C., Terray, L., and Phillips, A. S.: Forced and Internal Components of Winter Air Temperature Trends over North America during the Past 50 Years: Mechanisms and Implications, *Journal of Climate*, 29, 2237–2258, <https://doi.org/10.1175/JCLI-D-15-0304.1>, 2016.
- Doane-Solomon, R., Woollings, T., and Simpson, I. R.: Dynamic Contributions to Recent Observed Wintertime Precipitation Trends in Mediterranean-Type Climate Regions, *Geophysical Research Letters*, 52, e2024GL114 258, <https://doi.org/10.1029/2024GL114258>,
675 2025.
- Donat, M. G., Delgado-Torres, C., De Luca, P., Mahmood, R., Ortega, P., and Doblas-Reyes, F. J.: How Credibly Do CMIP6 Simulations Capture Historical Mean and Extreme Precipitation Changes?, *Geophysical Research Letters*, 50, e2022GL102 466, <https://doi.org/10.1029/2022GL102466>, 2023.
- Dorrington, J., Strommen, K., Fabiano, F., and Molteni, F.: CMIP6 Models Trend Toward Less Persistent European Blocking Regimes in a
680 Warming Climate, *Geophysical Research Letters*, 49, e2022GL100 811, <https://doi.org/10.1029/2022GL100811>, 2022.
- Dorrington, J., Grams, C., Grazzini, F., Magnusson, L., and Vitart, F.: Domino: A New Framework for the Automated Identification of Weather Event Precursors, Demonstrated for European Extreme Rainfall, *Quarterly Journal of the Royal Meteorological Society*, 150, 776–795, <https://doi.org/10.1002/qj.4622>, 2024a.



- Dorrington, J., Wenta, M., Grazzini, F., Magnusson, L., Vitart, F., and Grams, C. M.: Precursors and Pathways: Dynamically Informed
685 Extreme Event Forecasting Demonstrated on the Historic Emilia-Romagna 2023 Flood, *Natural Hazards and Earth System Sciences*, 24,
2995–3012, <https://doi.org/10.5194/nhess-24-2995-2024>, 2024b.
- Driouech, F., Déqué, M., and Sánchez-Gómez, E.: Weather Regimes—Moroccan Precipitation Link in a Regional Climate Change Simula-
tion, *Global and Planetary Change*, 72, 1–10, <https://doi.org/10.1016/j.gloplacha.2010.03.004>, 2010.
- Du, Y., Wang, D., Zhu, J., Wang, D., Qi, X., and Cai, J.: Comprehensive Assessment of CMIP5 and CMIP6 Models in Simulating and
690 Projecting Precipitation over the Global Land, *International Journal of Climatology*, 42, 6859–6875, <https://doi.org/10.1002/joc.7616>,
2022.
- EEA: Economic Losses and Fatalities from Weather- and Climate-Related Events in Europe — European Environment Agency, <https://www.eea.europa.eu/publications/economic-losses-and-fatalities-from>, 2022.
- Fischer, L. J., Bresch, D. N., Büeler, D., Grams, C. M., Noyelle, R., Röthlisberger, M., and Wernli, H.: How Relevant Are Frequency Changes
695 of Weather Regimes for Understanding Climate Change Signals in Surface Precipitation in the North Atlantic–European Sector? A Con-
ceptual Analysis with CESM1 Large Ensemble Simulations, *Weather and Climate Dynamics*, 6, 1027–1043, [https://doi.org/10.5194/wcd-](https://doi.org/10.5194/wcd-6-1027-2025)
6-1027-2025, 2025.
- Gerighausen, J., Dorrington, J., Osman, M., and Grams, C. M.: Quantifying Intra-Regime Weather Variability for Energy Applications,
<https://doi.org/10.48550/arXiv.2408.04302>, 2024.
- 700 Grazzini, F., Fragkoulidis, G., Teubler, F., Wirth, V., and Craig, G. C.: Extreme Precipitation Events over Northern Italy. Part II: Dynamical
Precursors, *Quarterly Journal of the Royal Meteorological Society*, 147, 1237–1257, <https://doi.org/10.1002/qj.3969>, 2021.
- Grotjahn, R., Black, R., Leung, R., Wehner, M. F., Barlow, M., Bosilovich, M., Gershunov, A., Gutowski, W. J., Gyakum, J. R., Katz, R. W.,
Lee, Y.-Y., Lim, Y.-K., and Prabhat: North American Extreme Temperature Events and Related Large Scale Meteorological Patterns:
A Review of Statistical Methods, Dynamics, Modeling, and Trends, *Climate Dynamics*, 46, 1151–1184, [https://doi.org/10.1007/s00382-](https://doi.org/10.1007/s00382-015-2638-6)
705 015-2638-6, 2016.
- Held, I. M. and Soden, B. J.: Robust Responses of the Hydrological Cycle to Global Warming, *Journal of Climate*, 19, 5686–5699,
<https://doi.org/10.1175/JCLI3990.1>, 2006.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Sim-
mons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren,
710 P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J.,
Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Vil-
laume, S., and Thépaut, J.-N.: The ERA5 Global Reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049,
<https://doi.org/10.1002/qj.3803>, 2020.
- Hewson, T. D. and Pillosu, F. M.: A Low-Cost Post-Processing Technique Improves Weather Forecasts around the World, *Communications*
715 *Earth & Environment*, 2, 132, <https://doi.org/10.1038/s43247-021-00185-9>, 2021.
- Iversen, E. C., Hodnebrog, Ø., Seland Graff, L., Nygaard, B. E., and Iversen, T.: Future Winter Precipitation Decreases Associated With
the North Atlantic Warming Hole and Reduced Convection, *Journal of Geophysical Research: Atmospheres*, 128, e2022JD038374,
<https://doi.org/10.1029/2022JD038374>, 2023.
- Konstali, K., Spengler, T., Spensberger, C., and Sorteberg, A.: Linking Future Precipitation Changes to Weather Features in CESM2-LE,
720 *Journal of Geophysical Research: Atmospheres*, 129, e2024JD041190, <https://doi.org/10.1029/2024JD041190>, 2024.



- Kuma, P., Bender, F. A.-M., and Jönsson, A. R.: Climate Model Code Genealogy and Its Relation to Climate Feedbacks and Sensitivity, *Journal of Advances in Modeling Earth Systems*, 15, e2022MS003 588, <https://doi.org/10.1029/2022MS003588>, 2023.
- Mann, M. E., Steinman, B. A., and Miller, S. K.: Absence of Internal Multidecadal and Interdecadal Oscillations in Climate Model Simulations, *Nature Communications*, 11, 49, <https://doi.org/10.1038/s41467-019-13823-w>, 2020.
- 725 Meehl, G. A., Arblaster, J. M., Bates, S., Richter, J. H., Tebaldi, C., Gettelman, A., Medeiros, B., Bacmeister, J., DeRepentigny, P., Rosenbloom, N., Shields, C., Hu, A., Teng, H., Mills, M. J., and Strand, G.: Characteristics of Future Warmer Base States in CESM2, *Earth and Space Science*, 7, e2020EA001 296, <https://doi.org/10.1029/2020EA001296>, 2020.
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-
- 730 N.: ERA5-Land: A State-of-the-Art Global Reanalysis Dataset for Land Applications, *Earth System Science Data*, 13, 4349–4383, <https://doi.org/10.5194/essd-13-4349-2021>, 2021.
- O’Gorman, P. A. and Schneider, T.: The Physical Basis for Increases in Precipitation Extremes in Simulations of 21st-Century Climate Change, *Proceedings of the National Academy of Sciences*, 106, 14 773–14 777, <https://doi.org/10.1073/pnas.0907610106>, 2009.
- Olonscheck, D., Suarez-Gutierrez, L., Milinski, S., Beobide-Arsuaga, G., Baehr, J., Fröb, F., Ilyina, T., Kadow, C., Krieger, D., Li, H.,
- 735 Marotzke, J., Pléziat, É., Schupfner, M., Wachsmann, F., Wallberg, L., Wieners, K.-H., and Brune, S.: The New Max Planck Institute Grand Ensemble With CMIP6 Forcing and High-Frequency Model Output, *Journal of Advances in Modeling Earth Systems*, 15, e2023MS003 790, <https://doi.org/10.1029/2023MS003790>, 2023.
- O’Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6,
- 740 *Geoscientific Model Development*, 9, 3461–3482, <https://doi.org/10.5194/gmd-9-3461-2016>, 2016.
- O’Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., van Ruijven, B. J., van Vuuren, D. P., Birkmann, J., Kok, K., Levy, M., and Solecki, W.: The Roads Ahead: Narratives for Shared Socioeconomic Pathways Describing World Futures in the 21st Century, *Global Environmental Change*, 42, 169–180, <https://doi.org/10.1016/j.gloenvcha.2015.01.004>, 2017.
- Pichelli, E., Coppola, E., Sobolowski, S., Ban, N., Giorgi, F., Stocchi, P., Alias, A., Belušić, D., Berthou, S., Caillaud, C., Cardoso, R. M.,
- 745 Chan, S., Christensen, O. B., Dobler, A., de Vries, H., Goergen, K., Kendon, E. J., Keuler, K., Lenderink, G., Lorenz, T., Mishra, A. N., Panitz, H.-J., Schär, C., Soares, P. M. M., Truhetz, H., and Vergara-Temprado, J.: The First Multi-Model Ensemble of Regional Climate Simulations at Kilometer-Scale Resolution Part 2: Historical and Future Simulations of Precipitation, *Climate Dynamics*, 56, 3581–3602, <https://doi.org/10.1007/s00382-021-05657-4>, 2021.
- Scafetta, N.: CMIP6 GCM Validation Based on ECS and TCR Ranking for 21st Century Temperature Projections and Risk Assessment,
- 750 *Atmosphere*, 14, 345, <https://doi.org/10.3390/atmos14020345>, 2023.
- Scaife, A. A. and Smith, D.: A Signal-to-Noise Paradox in Climate Science, *npj Climate and Atmospheric Science*, 1, 28, <https://doi.org/10.1038/s41612-018-0038-4>, 2018.
- Simpson, I. R., Bacmeister, J., Neale, R. B., Hannay, C., Gettelman, A., Garcia, R. R., Lauritzen, P. H., Marsh, D. R., Mills, M. J., Medeiros, B., and Richter, J. H.: An Evaluation of the Large-Scale Atmospheric Circulation and Its Variability in CESM2 and Other CMIP Models,
- 755 *Journal of Geophysical Research: Atmospheres*, 125, e2020JD032 835, <https://doi.org/10.1029/2020JD032835>, 2020.
- Soares, P. M. M., Johannsen, F., Lima, D. C. A., Lemos, G., Bento, V. A., and Bushenkova, A.: High-Resolution Downscaling of CMIP6 Earth System and Global Climate Models Using Deep Learning for Iberia, *Geoscientific Model Development*, 17, 229–259, <https://doi.org/10.5194/gmd-17-229-2024>, 2024.



- 760 Sobolowski, S., Somot, S., Fernandez, J., Evin, G., Brands, S., Maraun, D., Kotlarski, S., Jury, M., Benestad, R. E., Teichmann, C., Christensen, O. B., Bülow, K., Buonomo, E., Katragkou, E., Steger, C., Sørland, S., Nikulin, G., McSweeney, C., Dobler, A., Palmer, T., Wilcke, R., Boé, J., Brunner, L., Ribes, A., Qasmi, S., Nabat, P., Sevault, F., and Oudar, T.: GCM Selection & Ensemble Design: Best Practices and Recommendations from the EURO-CORDEX Community, *Bulletin of the American Meteorological Society*, -1, <https://doi.org/10.1175/BAMS-D-23-0189.1>, 2025.
- 765 Stendel, M., Francis, J., White, R., Williams, P. D., and Woollings, T.: Chapter 15 - The Jet Stream and Climate Change, in: *Climate Change (Third Edition)*, edited by Letcher, T. M., pp. 327–357, Elsevier, ISBN 978-0-12-821575-3, <https://doi.org/10.1016/B978-0-12-821575-3.00015-3>, 2021.
- Stephan, C. C., Duras, J., Harris, L., Klocke, D., Putman, W. M., Taylor, M., Wedi, N. P., Žagar, N., and Ziemann, F.: Atmospheric Energy Spectra in Global Kilometre-Scale Models, *Tellus A: Dynamic Meteorology and Oceanography*, 74, <https://doi.org/10.16993/tellusa.26>, 2022.
- 770 Stergiou, I., Traka, N., Kaskaoutis, D. G., Tagaris, E., and Sotiropoulou, R.-E. P.: Do the Use of a Convection Scheme in the Convective “Gray Zone” and the Increase in Spatial Resolution Enhance the WRF’s Precipitation Predictive Capability?, *Theoretical and Applied Climatology*, 156, 196, <https://doi.org/10.1007/s00704-025-05415-0>, 2025.
- Strandberg, G. and Lind, P.: The Importance of Horizontal Model Resolution on Simulated Precipitation in Europe – from Global to Regional Models, *Weather and Climate Dynamics*, 2, 181–204, <https://doi.org/10.5194/wcd-2-181-2021>, 2021.
- 775 Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., and van Mulbregt, P.: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nature Methods*, 17, 261–272, <https://doi.org/10.1038/s41592-019-0686-2>, 2020.
- 780 Wille, J. D., Koch, R., Becker, T., and Fischer, E.: Extreme Precipitation Depiction in Convection-Permitting Earth System Models Within the nextGEMS Project, *Journal of Advances in Modeling Earth Systems*, 17, e2024MS004 840, <https://doi.org/10.1029/2024MS004840>, 2025.
- Zappa, G., Hawcroft, M. K., Shaffrey, L., Black, E., and Brayshaw, D. J.: Extratropical Cyclones and the Projected Decline of Winter Mediterranean Precipitation in the CMIP5 Models, *Climate Dynamics*, 45, 1727–1738, <https://doi.org/10.1007/s00382-014-2426-8>, 2015.
- 785 Zhu, D., Pfahl, S., Knutti, R., and Fischer, E. M.: Future Extreme Precipitation May Shift to Colder Seasons in Northern Mid- and High Latitudes, *Communications Earth & Environment*, 6, 657, <https://doi.org/10.1038/s43247-025-02651-0>, 2025.