Authors response to reviewer 1

egusphere-2024-3225. Future changes in runoff over western and central Europe: disentangling the hydrological behavior of CMIP6 models.

We thank the reviewer for their time and constructive feedback. We appreciate the points raised and the references provided, which have prompted us to re-examine important aspects of our study and strengthen both the analysis and the presentation of the results.

We address each point below and we will provide a revised version of the manuscript at the end of the interactive discussion. Our responses to the reviewers' comments (in black) are shown below in purple.

Reviewer 1

Summary

In this manuscript, the authors attempt to provide insights into possible future hydrologic changes over Western and Central Europe (WCE). To do so, they make use of historical and future (ssp585) projections from CMIP6 models. They cluster models into 8 groups based on the similarity of their hydrologic responses and show that there is no consistent response across models/clusters in future annual runoff changes over WCE, with half of clusters showing strong decreases and half showing no change or modest increases. They further show that even when clusters of models show similar runoff responses, they may do so for different reasons (e.g., divergent signs/strengths of precipitation and evapotranspiration trends). Additionally, they show that it is challenging to observationally constrain the model projections owing to observational uncertainty in components of the water budget (especially P and ET trends) and a weak linkage between historical and future changes in the models. Finally, they leverage additional CMIP6 experiments from C4MIP and LS3MIP to provide a detailed look at some of the biogeophysical mechanisms driving the models' hydrological responses and find a strong role for plant physiological responses to elevated CO2 and a limited role for soil moisture-atmosphere feedbacks in the spread of model responses.

General comments

Overall, I found the questions the authors posed around future hydrologic changes over WCE both scientifically interesting, owing to the substantial uncertainties, and societally relevant. Their analysis was thorough and technically sound. I particularly appreciated the use of multiple sets of CMIP6 experiments to provide several lines of evidence. I only have two minor concerns that I would like the authors to address before I would consider this manuscript ready for publication.

Specific comments

My first comment pertains to the clustering algorithm used to group models. How was the number of clusters determined? The authors do not describe the dissimilarity measure shown on the x-axis in Figure 1 or how the dissimilarity threshold that determines the final clusters was chosen.

The number of clusters was determined empirically so that there are enough clusters to highlight different behaviors but not so many that there are single model clusters or that the diversity of behaviors is too difficult to analyze. The dissimilarity threshold is chosen accordingly to get eight clusters of models.

The Ward's linkage merges two clusters that result in the smallest increase in intra-cluster inertia. The intra-cluster inertia of a cluster C with centroïd μ_C is computed as follow:

$W(C) = \sum_{x \in C} ||x - \mu_C||^2$. The increase in intra-cluster inertia if clusters A and B were merged is therefore computed as follow: $\Delta W = W(A \cup B) - (W(A) + W(B))$. This increase is the dissimilarity measure that is noted on the x-axis. It can also be expressed as:

$$d(A, B) = \sqrt{|A||B|/(|A| + |B|)}\, ||\mu_A - \mu_B||^2$$

with $|A|$ and $|B|$ the number of points in clusters A and B and $||\mu_A - \mu_B||^2$ the Euclidean distance between the two cluster centroïds.

The method section will be modified to clarify this point.

I feel that these analytical choices do require some justification, and checking the sensitivity of the main results to the threshold/number of clusters would provide a valuable robustness check.

Below is a sensitivity analysis of the results to the number of clusters chosen. We show the clustering and the corresponding classified changes in precipitation, evapotranspiration and runoff with 4, 6 and 10 clusters (figures R1 to R8).

The use of a limited number of clusters (4 in figure R1 and 6 in figure R2) in the classification of models facilitates the identification of predominant behaviours (figures R4 and R5). We clearly see the two types of annual behaviors highlighted in the article: half of the clusters show a decrease in runoff and the other half show an increase. The clusters with a decrease in runoff either project an increase in precipitation and an even greater increase in evapotranspiration or a decrease in precipitation and small changes in evapotranspiration. We also distinguish the outlier behavior of the CCCma models, which are as dissimilar with C3 than C2 is with C1 (figure R1) or with C1/C5 than C2 with C3/C4 (figure R2). We could have stopped the clustering at six clusters but there are still clusters with high intra-variance (C2 and C3 in figure R2). We therefore have added two more clusters without reducing the dissimilarity threshold too much so that TaiESM1 would not be a single model cluster.

With ten clusters, the C5 and C6 clusters are split (figure 1) to give the new C5, C9, C6 and C10 clusters (figure R3) with C10 being a single model cluster. In this classification, C5 and C9 have very similar behavior (their seasonal changes in R, P and E are close) except in summer. C6 and C10 also have similar behavior except for runoff in DJF and MAM with C10 showing a stronger than usual increase (figure R6). Additionally, it is harder to identify the mechanisms that might be responsible for a particular behavior with a single model cluster (especially as there are not enough models available in the MIPs we analyzed to have at least one model per cluster). The issues are the same with 10 clusters or more.

The choice of eight clusters is therefore a good compromise between a diversity of hydrological behavior without having single model clusters or clusters with similar behaviors. Furthermore, as discussed above, having fewer or more clusters would not change the main physical interpretations of our study based on 8 clusters.

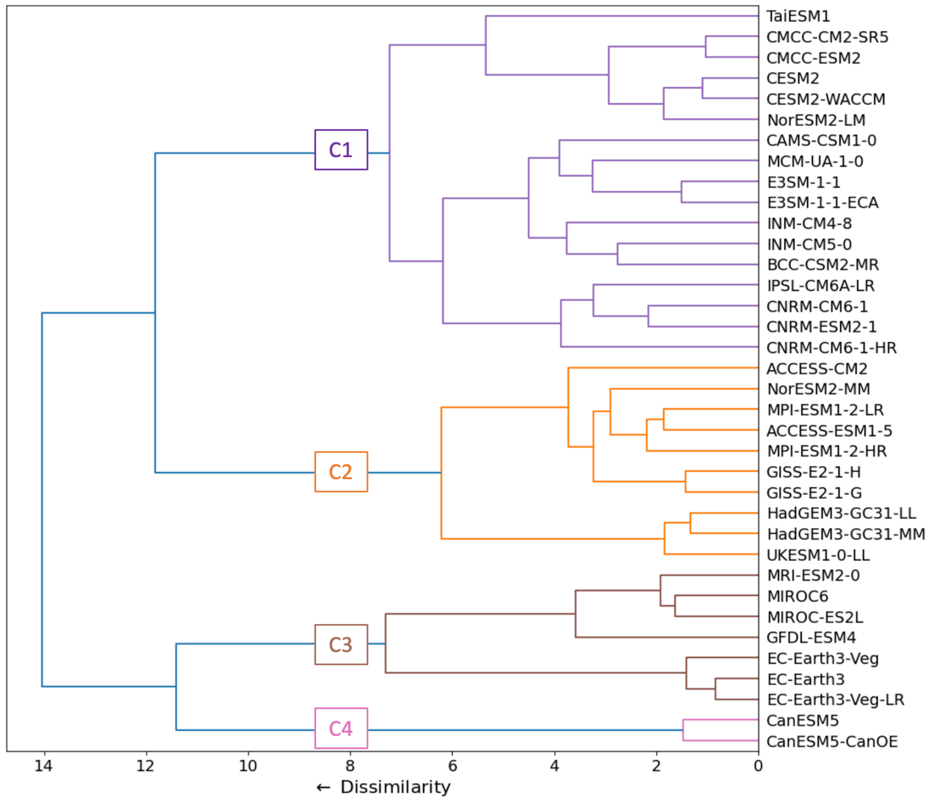

Figure R1: Same as Figure 1 but with four clusters of models

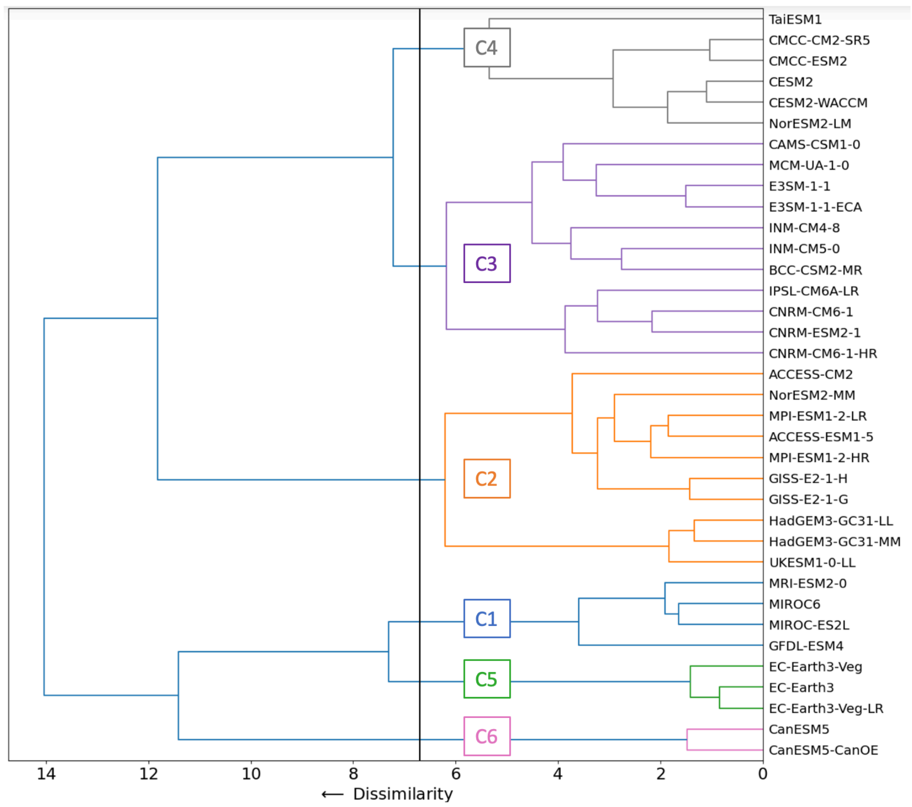

Figure R2: Same as Figure 1 but with six clusters of models

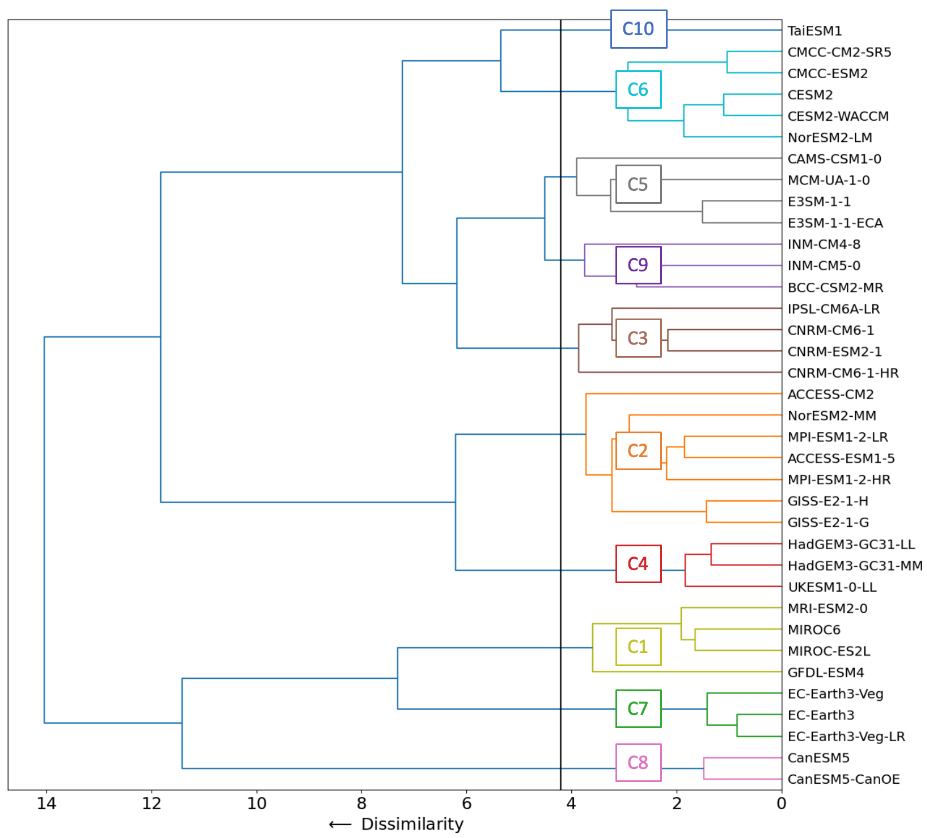

Figure R3: Same as Figure 1 but with ten clusters of models

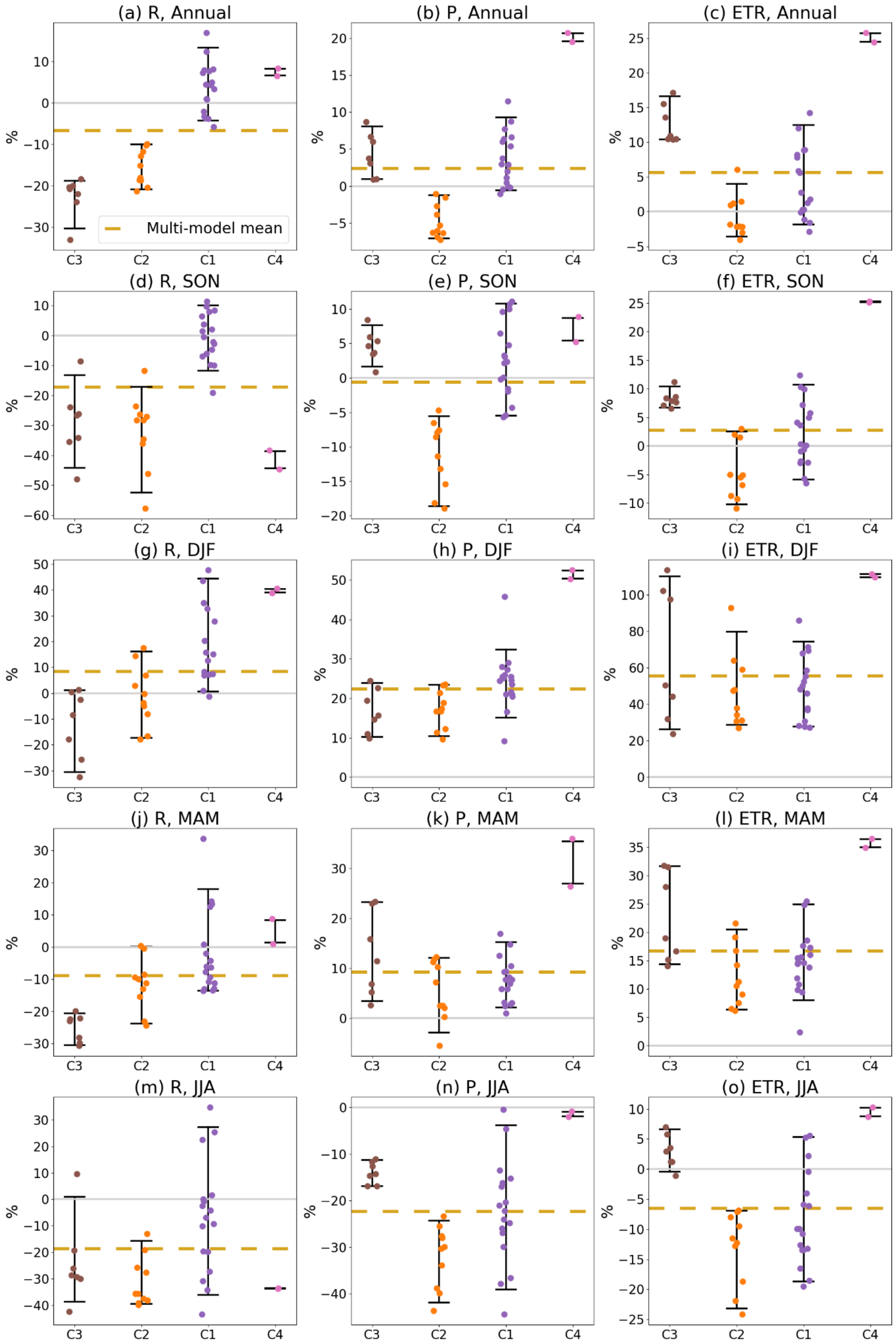

Figure R4: Same as Figure 4 but with four clusters of models. The associated dendrogram of the clustering is given Figure R1.

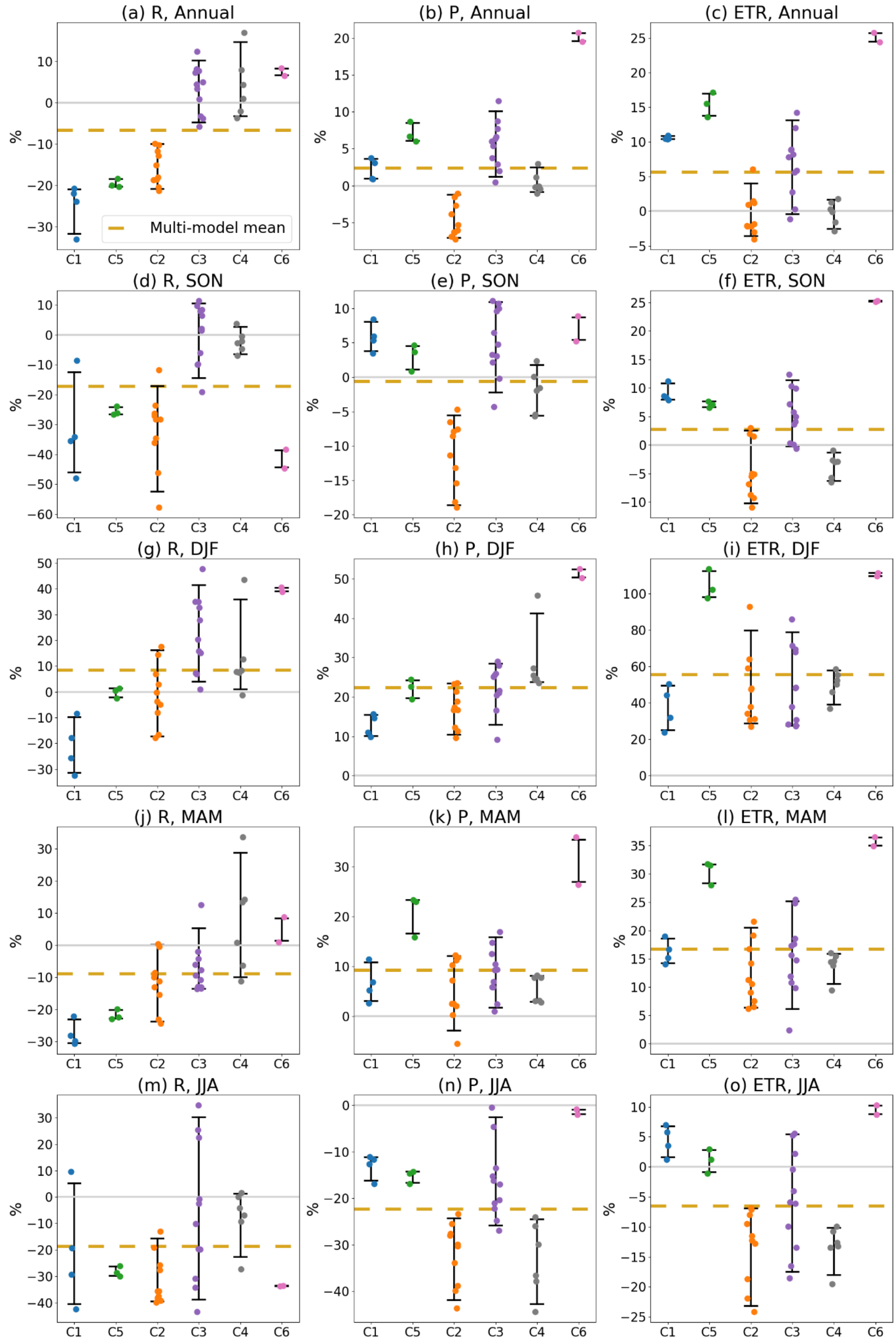

Figure R5: Same as Figure 4 but with six clusters of models. The associated dendrogram of the clustering is given Figure R2.

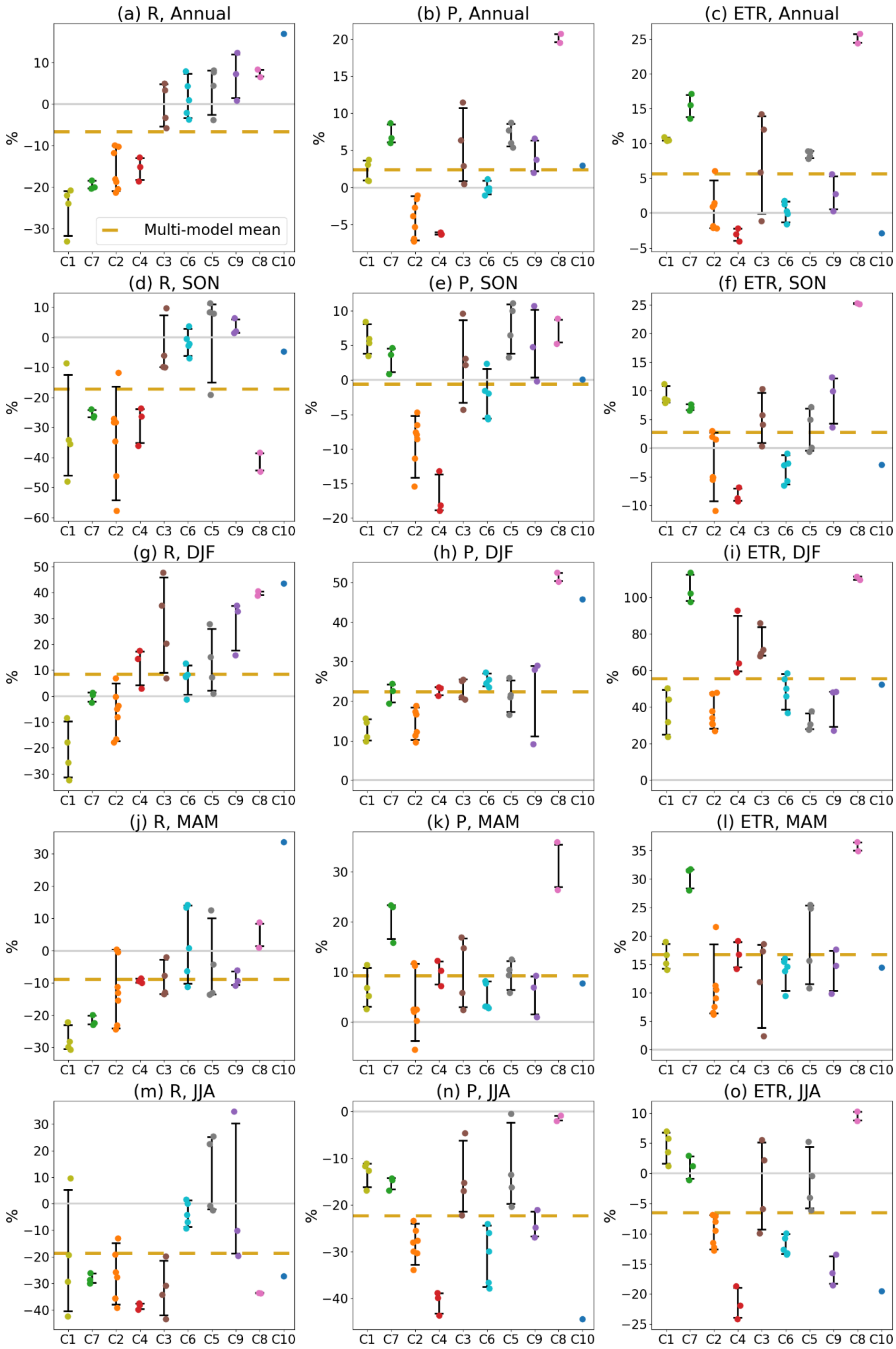

Figure R6: Same as Figure 4 but with ten clusters of models. The associated dendrogram of the clustering is given Figure R3.

My second comment relates to a point the authors flag in their discussion on ll. 399-403: the CMIP6 models display a wide range of climate sensitivities, and as such warm by very different amounts over WCE in ssp585 by the end-of-century period the authors examine. Accordingly, it would make more analytical sense to normalize the hydrologic changes the authors examine by the temperature change over that period (i.e., %/K). This would help identify situations in which models might have a similar hydrologic response to warming, but warm different amounts, and to more effectively partition how much of the model spread is due to dynamic vs. thermodynamic differences. Additionally, this normalization would make for more interpretable expectations of future change that are benchmarked to warming levels and thus more scenario-independent.

This is a very interesting point that we thought about when writing the article. We think there are multiple ways to look at the sensitivity of hydrological changes to warming. We therefore have done three sensitivity tests:

First, we have normalized the hydrological changes by local temperature for the clusters with the same classification as in the paper, i.e. on raw hydrological changes (Figure R7). Then we have performed the classification on hydrological changes normalized by local temperature change (Figures R8 and R9). Finally, we have performed the classification on hydrological changes normalized by the ECS (Figure R10 and R11).

The normalization of hydrological changes by local temperature change after classification (Figure R7) does not modify the two types of behaviors highlighted in Figure 4 (half of the clusters give a decrease and the other half an increase). The main differences are some differences of intra-cluster variability seen after normalization by local temperature change. Also, after normalizing by temperature changes, the CCCma models are still outliers but the difference between them and the rest of the models is smaller. The stronger than usual changes seen in CCCma models can therefore be attributed in part to their very large ECS, as discussed in the conclusion of the paper.

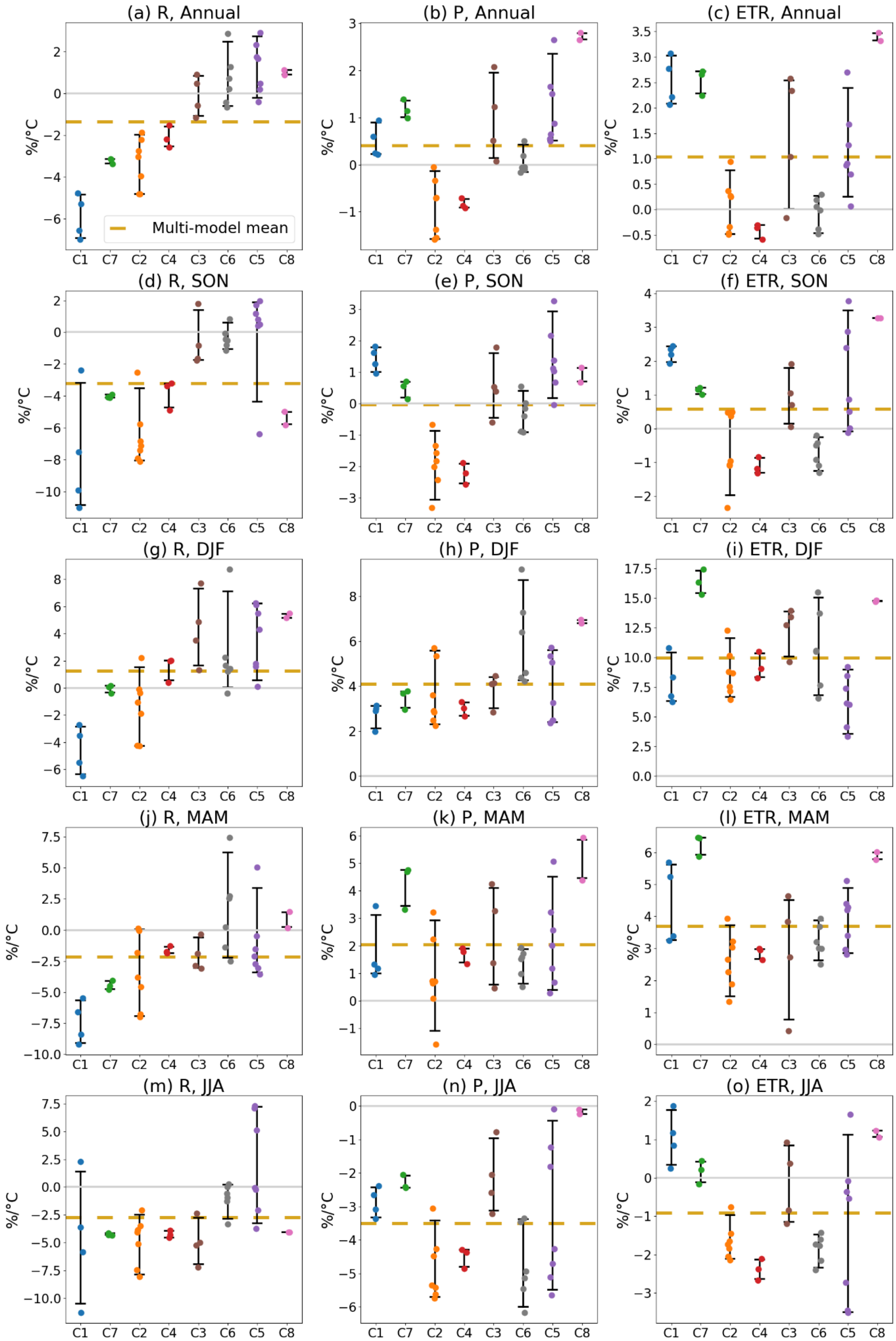

Figure R7: Same as Figure 4 but the changes in each season are divided by the changes of temperature in this season. The associated dendrogram of the clustering is given Figure 1.

Overall, the classification of hydrological changes (Figure 1 in the paper) and hydrological changes normalized by local temperature changes (Figure R8) shows important similarities, but also some differences. The Earth System model and high resolution climate model of CNRM/Cerfacs switches from cluster C3 of the standard classification to the clusters C7 and C8 of this new classification. Former C2 and C4 are grouped in the same cluster, therefore we can assume that the difference between their changes is of thermodynamic origin, or least correlated with temperature. The normalization by local temperature changes also highlights the outlier behavior of CAMS-CSM1-0, which shows strong annual hydrological changes despite small temperature changes (C4 in Figure R9 a, b, c). This model presents an unusual runoff increase in summer despite a decrease in precipitation and an increase in evapotranspiration (Figure R9 m, n, o).

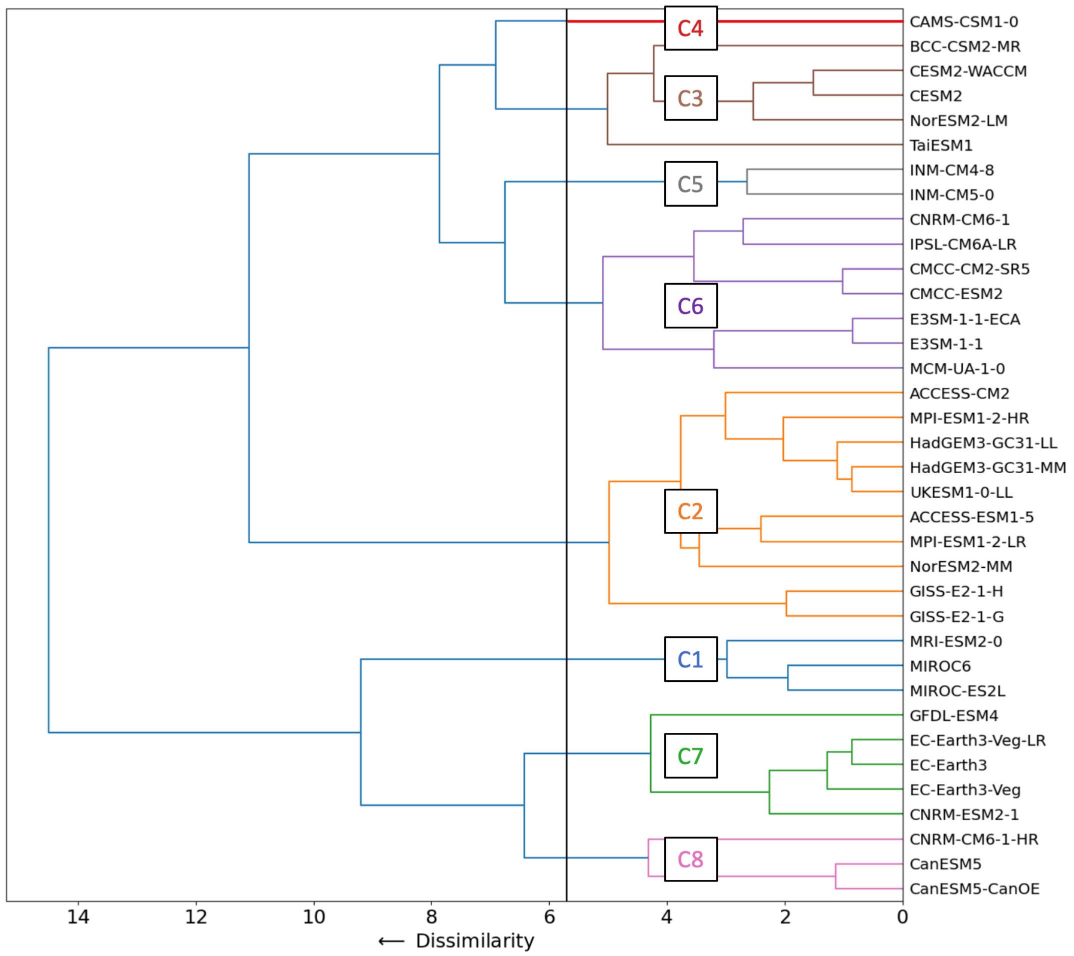

Figure R8: Same as Figure 1 but the hydrological changes in each season are normalized by local temperature changes in this season.

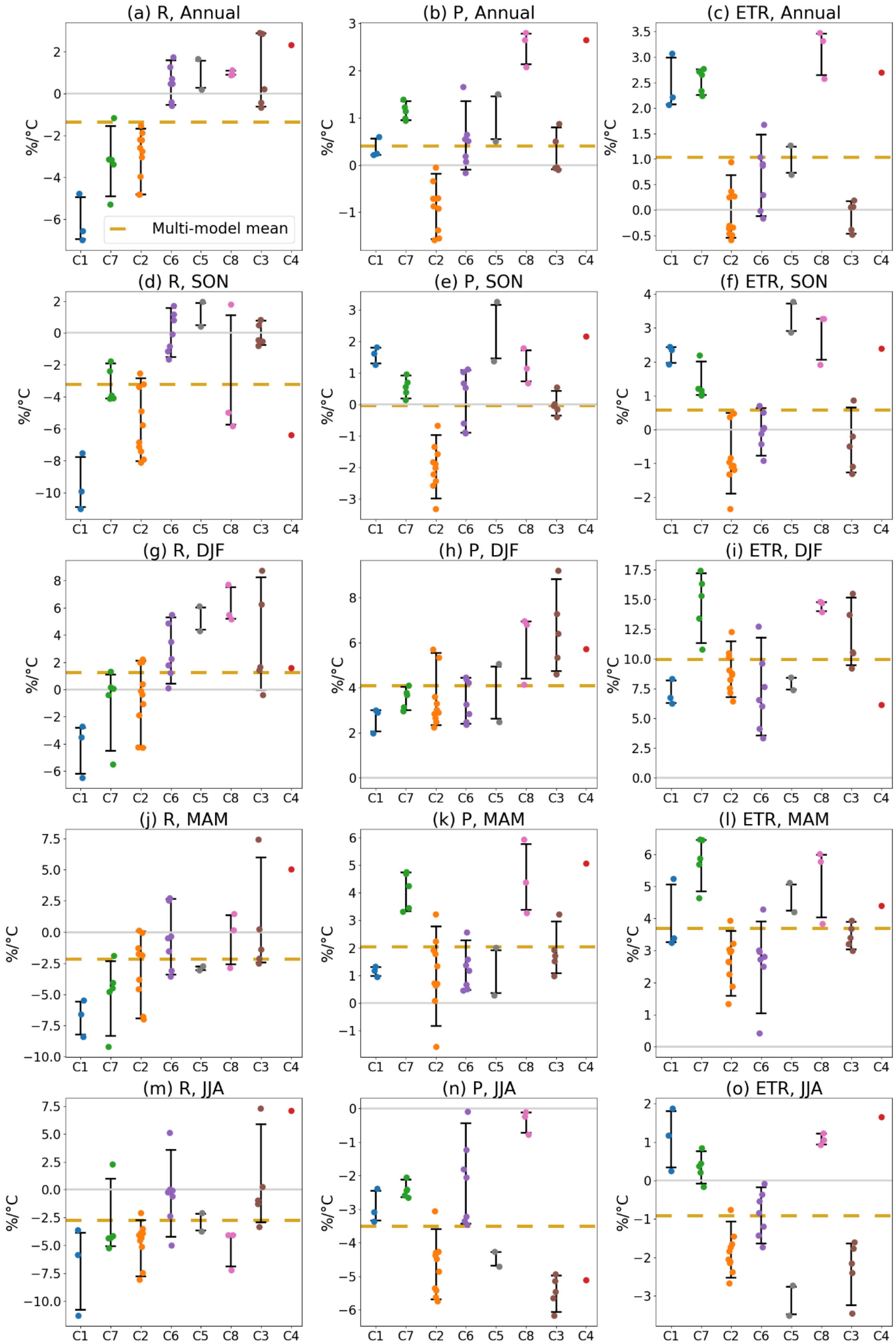

Figure R9: Same as Figure 4 but the changes in each season are divided by the local changes of temperature changes in this season. The associated dendrogram of the clustering is given Figure R8.

Finally, we have performed the classification on the hydrological changes normalized by the ECS. This has been done using the 30 models out of the 36 for which the ECS is available in the literature.

With the classification, the CCCma model is no longer an outlier because of its very large ECS (Figures R10 and R11).

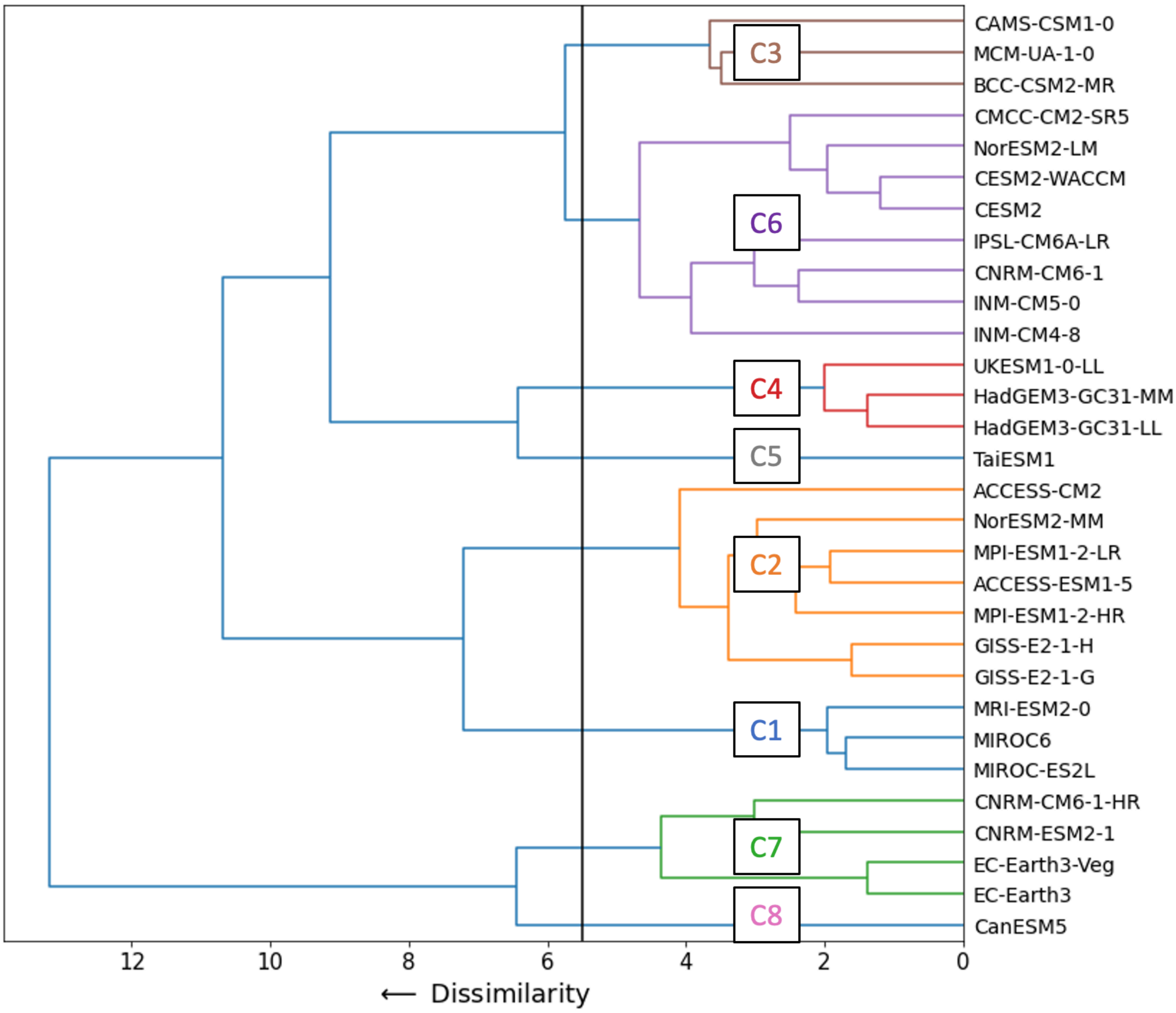

Figure R10: Same as Figure 1 but the hydrological changes are normalized by the ECS.

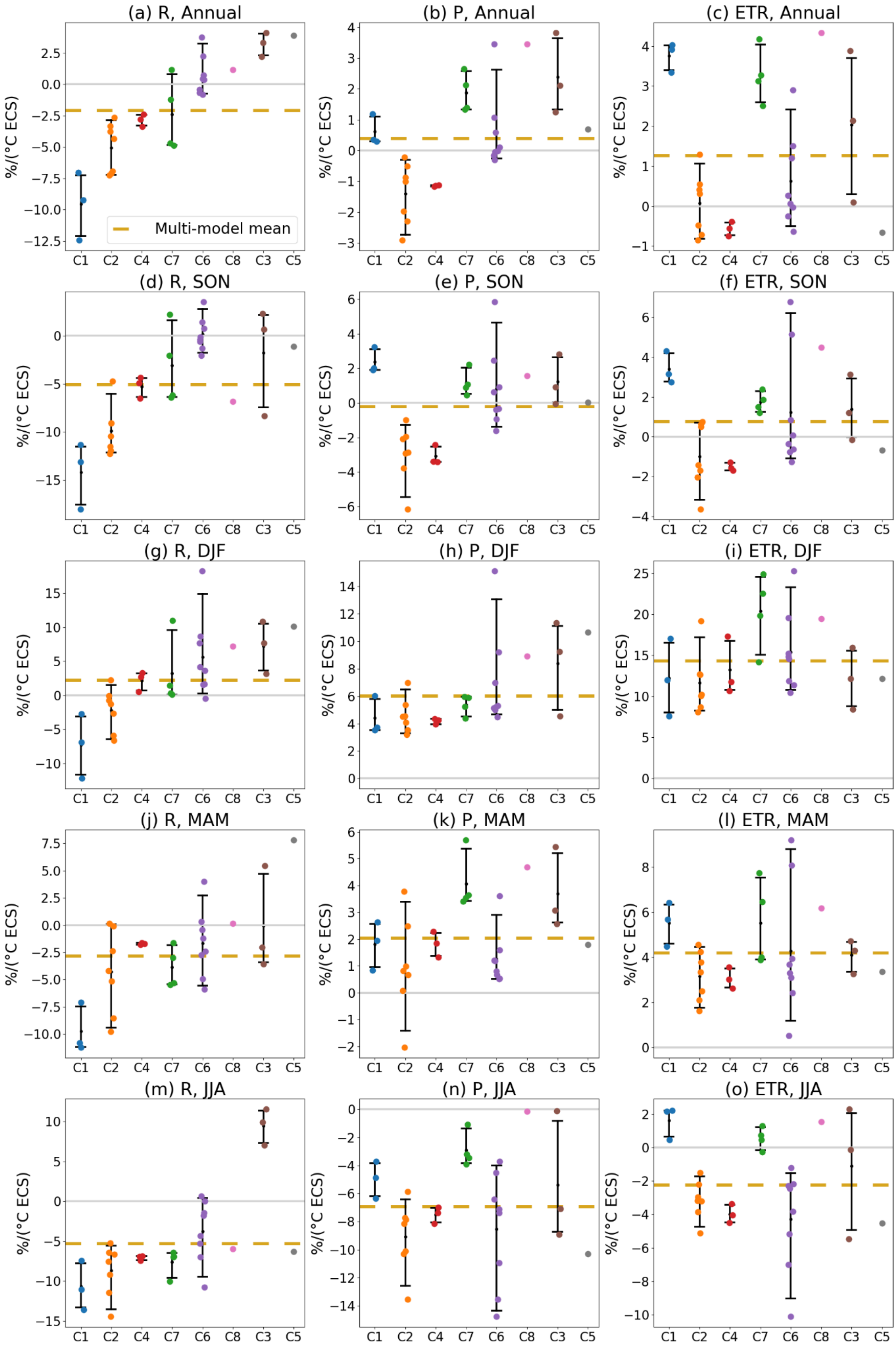

Figure R11: Same as Figure 4 but the changes are divided by the ECS. The associated dendrogram of the clustering is given Figure R10.

In conclusion, normalizing hydrological changes by temperature changes may lead to interesting discussions, as above, some of them already in the paper, but we are not sure that it is necessarily more interesting than discussing the raw hydrological changes. In addition, it leads to methodological questions: should we normalize by ECS, local annual temperature changes, local seasonal temperature changes? This choice has implications on the results and for the interpretation of the results. Moreover, even if normalizing by temperature seemingly reduces the hydrological difference between models, we are not sure that it should be necessarily interpreted as: "thermodynamic processes, scaling with temperature (e.g. the Clausius-Clapeyron relationship) explain the difference between models". This may not be causal or the causality could even be reversed. In some cases, the spread of temperature changes may be explained by hydrological changes rather than the other way around. For example, the physiological effect of CO_2 may reduce transpiration and therefore increase temperature.

Technical comments

I appreciated the authors' Table 2 and related discussion about the overlap in model components in their clusters. I'd encourage the authors to take a look at and cite the "model genealogy" literature (a few recommended papers below). to contextualize and strengthen this section.

We thank the reviewer for the references. We will add some discussion concerning model genealogy in the discussion related to table 2.

How much is the correlation between DJF SLP and precipitation change shown in Figure 6c driven by the outliers in C8 and TaiESM1? Given that all models show increased DJF precipitation regardless of the sign of SLP change, I'm not sure what conclusions to make about the importance of circulation changes for wintertime precipitation changes.

The correlation still holds but is less significant without the C8 models and TaiESM1 (Figures R12 a, b: $p<0.05$ instead of $p<0.01$). However we agree that other processes are likely to be important: the inter-model correlations with and without TaiESM1 and the C8 models are greater in magnitude for precipitation changes normalized by temperature change (Figures R12 c and d) pointing to an impact of temperature and thermodynamics processes in addition to the one of large scale circulation.

We will add a sentence on this topic in the text describing the results of Figure 6.

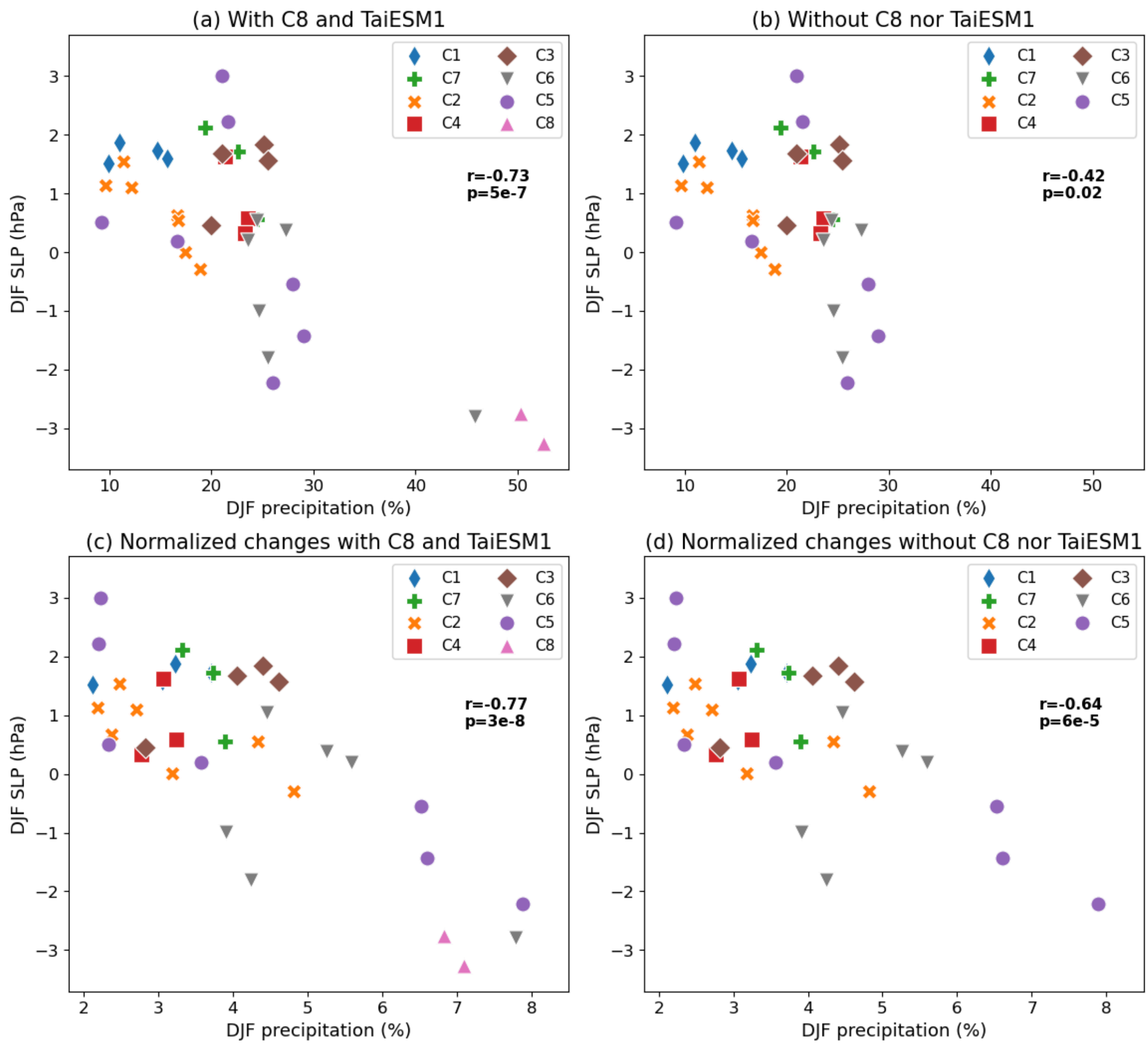

Figure R12: (a) Scatterplot between winter changes in precipitation (%) averaged over WCE (x-axis) and winter changes in sea level pressure (hPa) averaged over WCE (y-axis) of all CMIP6 models. The Pearson correlation coefficient ("r") and the corresponding p-value are given in the figure. (b) Same as (a) but without C8 models nor TaiESM1. (c) and (d) same as (a) and (b) but the DJF precipitation changes are normalized by the DJF temperature changes.

Knutti, R., D. Masson, and A. Gettelman (2013), Climate model genealogy: Generation CMIP5 and how we got there, *Geophys. Res. Lett.*, 40, 1194–1199, doi:10.1002/grl.50256.

Kuma, P., Bender, F. A.-M., & Jönsson, A. R. (2023). Climate model code genealogy and its relation to climate feedbacks and sensitivity. *Journal of Advances in Modeling Earth Systems*, 15, e2022MS003588. https://doi.org/10.1029/2022MS003588

Steinschneider, S., R. McCrary, L. O. Mearns, and C. Brown (2015), The effects of climate model similarity on probabilistic climate projections and the implications for local, risk-based adaptation planning. *Geophys. Res. Lett.*, 42, 5014–5044. doi: 10.1002/2015GL064529.