Response to Reviewer Comments

We thank the reviewers for their thoughtful and constructive comments and for recognizing the methodological framework and significance of our work. We have revised the manuscript to improve the clarity of the presentation. Below, we respond to each comment point by point. Line numbers refer to the tracked-changes version of the revised manuscript. Many of the tracked changes reflect moving the Methods from the Appendix to the new Section 2 “Data and Methods” in response to the reviewers’ comments.

Reviewer #1

In this paper, the authors analyse changes in Rx1day precipitation across 23 climate models, including 4 models with 10+ ensemble members. They uncover sources of internal and inter-model uncertainty in Rx1day through a dynamical-thermodynamic decomposition, and analyse the extent to which various simplifying approximations to their decomposition can give equivalent results. They complement this with an analysis of local temperature anomalies, and reveal that inter-model uncertainties in thermodynamic changes are dominated by a shift in the seasonality of Rx1day.

This is a really fantastic paper. The authors’ methodological framework is excellently suited to the task they approach, allowing them to obtain a clean conceptual synthesis of the multi-model ensemble they analyse. I found it very enlightening, and have been pleased to review it. I do have a small number of comments aimed at clarifying certain aspects of the methodology and contextualising the findings, and so I am recommending minor revisions overall.

1. **The paper is laid out in a Nature-like format with methods in an appendix at the end. I would prefer to see these as Section 2. The various “(see methods)” comments in later sections could then be removed.**

   Response: Thank you for pointing this out. We have moved the Methods from the Appendix into a new Section 2 “Data and Methods” (L87).

2. **I would like to see a slightly clearer presentation of what exactly you show in Figure 5. Based on the discussion from L165 to L178 I do not feel confident I can write down the formula for what you’ve computed. This could be added into the methods, and referred to in the Figure discussion.**

Response: We have now added a detailed description of the quantities shown in Figure 5 in Section 2.4, “Decomposition into thermodynamic and dynamic contributions.” This section now defines the thermodynamic contribution, dynamic contribution, and modified thermodynamic contribution explicitly. We also clarify the explained variance in Section 2.6, “Model variances explained by the dominant variables.” Specifically, single-variable linear regressions are applied across the 26 model ensemble-mean fields at each grid point, and the explained variance is measured by R^2.

3. **In Equation 1 can you explicitly specify the meaning of theta_star?**

   Response: θ^* denotes the saturation equivalent potential temperature. Thus, $dq_s/dp|_{\theta^*}$ represents the pressure derivative of saturation specific humidity evaluated at each pressure level along a saturated moist adiabat (at constant θ^*). This information has been added to section 2.3 (L112).

4. **You refer to Figure 7 before Figure 6**

   Response: thank you for pointing out. We have switched the two figures.

5. **The evaluation of how well simpler, vertically integrated variables can reproduce the full scaling is a very valuable aspect of the paper. It would have been interesting to see a full decomposition of Eq. 1 in this line, and I’m just curious as to why you didn’t take that approach here. The scaling of Eq. 1 decomposes fully into 6 terms: contributions from the vertical mean and vertical deviation of the dynamical, thernodynamic and nonlinear eddy trends. I suppose from that alone you would be able to see that the vertical deviation terms are small?**

   Response: Thank you for this helpful suggestion. We agree that a decomposition into vertical-mean and vertical-deviation terms could provide an additional perspective on the role of vertical structure. We have clarified the decomposition used in this study in Section 2.4 by rewriting Eq. 1 in terms of the event-conditioned vertical velocity ω_e and thermodynamic factor $\Gamma_e = dq_s/dp|_{\theta^*}$ and by explicitly defining the thermodynamic, dynamic, mean-state, and interaction terms.

   We did not perform a further vertical-mean/deviation decomposition. It would introduce several additional covariance and interaction terms whose interpretation is less direct. Instead, we assess the role of vertical structure through simpler diagnostic comparisons of the decomposed contributions with column-integrated q_s, vertically averaged updraft ω_e and ω_e changes at individual pressure levels. These analyses identify how much of the uncertainty can be captured by simpler variables while retaining the full diagnostic scaling as the quantitative basis of the decomposition.

6. **In a similar vein, do you know how much of the difference between Rx1day changes and the diagnostic scaling do you think may be a result of time-evolving precipitation efficiency? Modelling Pe = [E0+E1*dT]*{omega dq/dp}, then (ΔPe,true - ΔPe,diagnostic)/Pe,diagnostic = E1 . Have you previously assessed if including such a second-order term explains much of Fig. 1c? Obviously you're interested in the linear scaling here, but there are interesting implications to whether the difference is actually just for the numerical reasons you discuss or from some nonlinearity in the response.**

   Response: Thank you for this very interesting suggestion. We agree that time-varying precipitation efficiency could, in principle, contribute to differences between projected changes in Rx1day and in the diagnostic scaling. However, we have not separately estimated a term of the form [E_0+E_1\d T], because the residual difference between the diagnostic scaling and Rx1day cannot be simply interpreted as a precipitation efficiency term.

   The diagnostic scaling is an estimate of condensation associated with local extreme precipitation events, but it does not fully capture the true condensation that contribute to surface precipitation. As shown in Fgure A1, the diagnostic scaling can both overestimate and underestimate historical Rx1day. The underestimation may partly result from using daily-mean atmospheric profiles, which can smooth the short-lived peak updrafts and thermodynamic conditions associated with the most intense precipitation. Therefore, ΔPe,true - ΔPe,diagnostic)/Pe,diagnostic does not have a simple interpretation. It may reflect changes in precipitation efficiency, sub-daily variability or other processes not represented in the diagnostic scaling. We therefore do not include it as an additional term. However, we acknowledge that this would be an interesting starting point for future research into temporal changes of precipitation efficiency.

   We further compare the historical and future 50-year mean Rx1day and diagnostic scaling (Figure A1 in the supplementary and Figure 1 in this response document). The spatial pattern of their fractional difference is broadly similar in the two periods, including the positive difference over the Sahel and the generally negative difference over much of the tropics. Thus, the mismatch is already present in the historical climate and does not show a strong, spatially coherent change in the future climate. This comparison provides no clear evidence that a warming-dependent factor dominates the difference between projected changes in the diagnostic scaling and Rx1day.

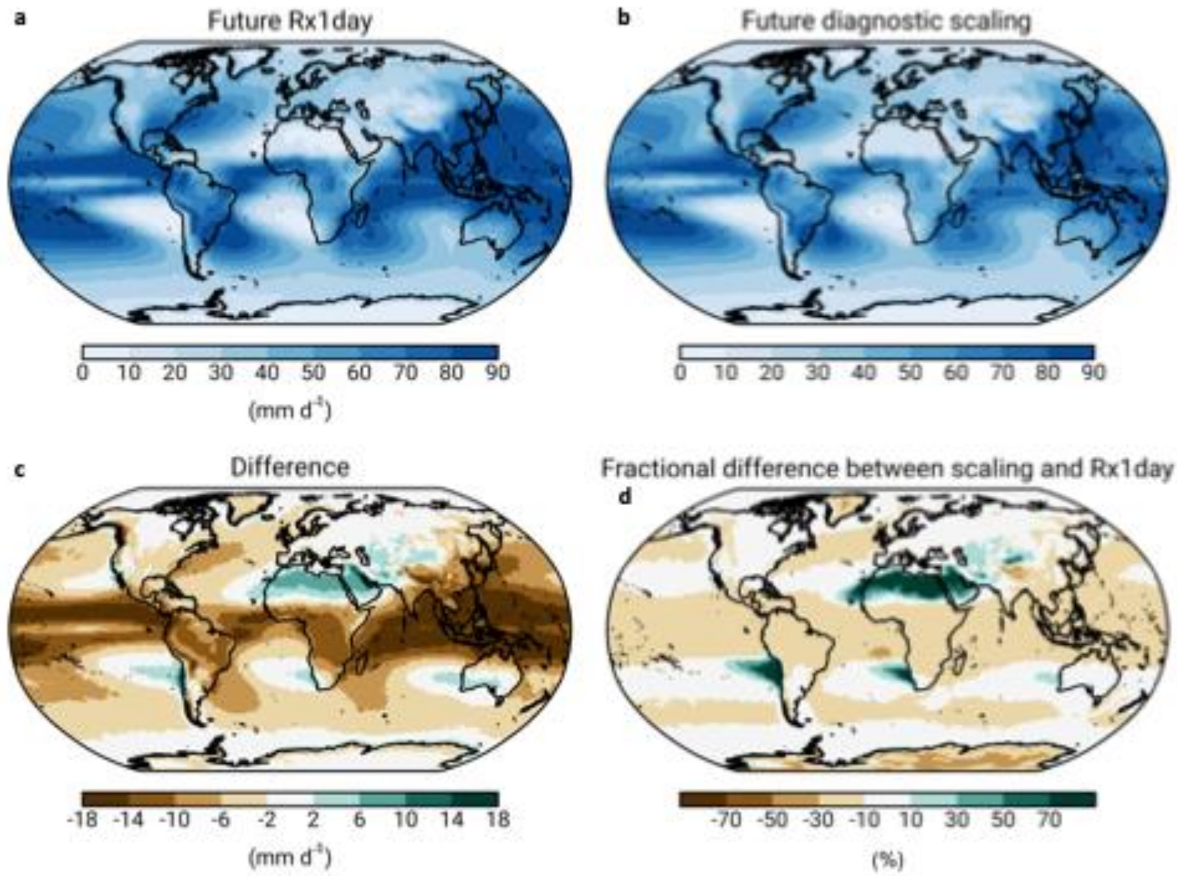

Figure 1 Future Rx1day and diagnostic scaling. *Multi-model mean of the future 50-year mean (a) Rx1day and (b) diagnostic scaling over the period 2050-2099 (in mm d^{-1}). Panel (c) shows the absolute difference between diagnostic scaling and Rx1day, defined as future diagnostic scaling minus future Rx1day. Panel (d) shows the fractional difference between diagnostic scaling and Rx1day, defined as the absolute difference divided by future Rx1day (in %).*

7. **In the conclusion could you briefly discuss/speculate how synoptic/large-scale dynamics might connect to some of your core results? To me it seems plausible that some of the seasonal shifts you see might be at least partly related to alterations in shoulder-season synoptic activity, for example. To the casual reader, it would be easy to misinterpret what your dynamic and thermodynamic terms represent and conclude that a predominately thermodynamic contribution to uncertainty means only small-scale processes are uncertain.**

   Response: Thank you for this helpful suggestion. We agree that the thermodynamic and dynamic terms in our study should not be interpreted simply as a separation between small-scale and large-scale processes. Rather, these terms arise from the diagnostic scaling framework and are evaluated on the days when local Rx1day events occur.

   Because we select the local Rx1day in each year over the period 1950-2099, the events sampled in the historical and future climates are not necessarily of the same type or associated with the same circulation regimes. For example, over Northern Hemisphere extratropical land regions, models project a weaker thermodynamic enhancement than over nearby oceans and little change in the dynamic contribution. This does not necessarily mean that large-scale dynamics are unimportant or unchanged. Instead, it indicates that the event-associated updraft intensity does not change substantially on average for the selected Rx1day events. At the same time, future Rx1day events occur at different times of the year and may therefore be associated with different dynamical regimes, or with similar regimes occurring preferentially in the shoulder seasons. Distinguishing between these possibilities would require targeted regional analyses, which are beyond the scope of the present global assessment. We are currently exploring this in ongoing work. We have added a short discussion in the Conclusion to clarify this point (L486).

8. **Also, could you speculate/comment in the conclusion as to how you think model resolution may impact your decomposition? I wonder how models with explicitly resolved mesoscale convective systems or tropical deep convection might look different/similar? Just to help add interpretative context.**

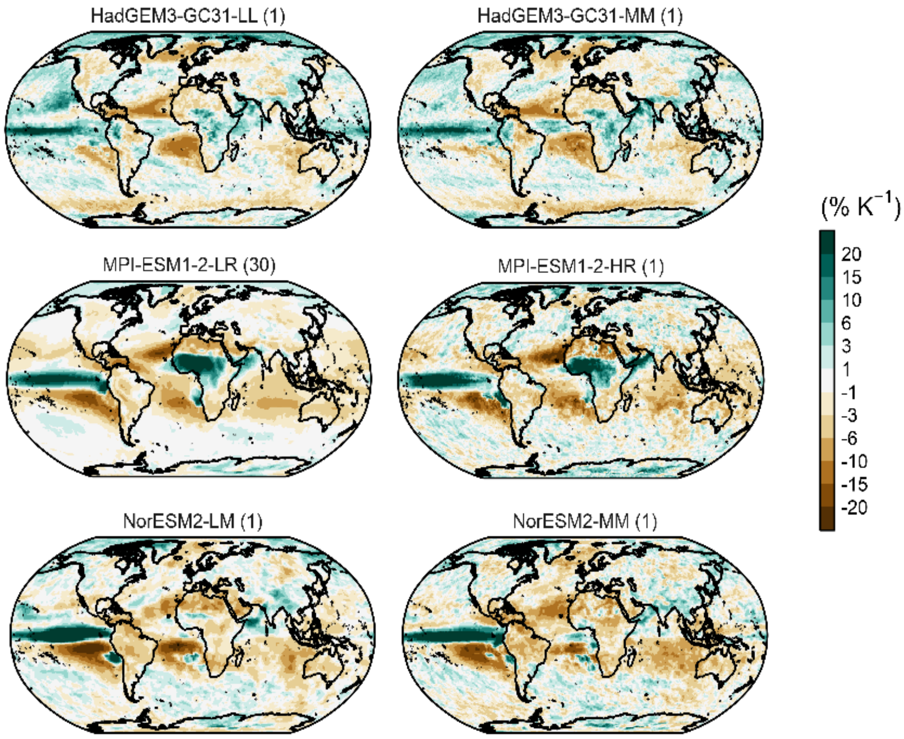

Figure 2 Comparison of the dynamic contribution to projected Rx1day changes in three CMIP6 model families with lower- and higher-resolution versions. *Numbers in parentheses after model names indicate the number of ensemble members used for the scaling analysis.*

Response: We agree that model resolution may influence the decomposition, particularly through the differences in the representation of mesoscale convective systems, tropical deep convection, and the vertical structure of updrafts during extreme precipitation events. It would therefore be very interesting to apply the diagnostic scaling framework to high-resolution or convection-permitting models and compare the results with those from the CMIP6 global climate models analyzed here. There is ongoing work in this direction using regional convective-permitting simulations over Europe.

However, within the CMIP6 models used in this study, we do not find a clear indication that model resolution strongly influences the decomposition. In particular, three model families provide both lower- and higher-resolution versions. For these model pairs, differences in the spatial patterns of the dynamic contribution are substantially larger across model families than between different resolution versions of the same model family (Figure 2 in this response document). We have added a paragraph in the Conclusion to discuss this point (L492).