Reviewer #2

This paper presents a decomposition of precipitation scaling in CMIP6 simulations, in order to determine the source of uncertainty under climate change.

The paper is interesting but suffers from a rather nonlinear structure that makes it hard to follow.

Major issues

1. **It takes a while to understand that essential information appears in the Appendix, and one has to guess many details, that become eventually discussed later in the text. Please make the reading more linear, so that the reader does not have to go back and forth between the Appendix and the text. For example (but not limited to), the list of CMIP6 models and how they are processed is not alluded to before l. 140 (p. 7!) and Fig. 3, while this information would have been necessary since Fig. 1. My recommendation is to use a more traditional structure, with a "methods and data" section right after the introduction.**

   Response: Thank you for pointing this out. We have moved the Methods from the Appendix into a new Section 2, "Data and Methods," directly after the Introduction (L87). This section now introduces the CMIP6 data used, the processing steps for Rx1day and its future changes, the diagnostic scaling and decomposition, and the uncertainty analysis, etc. This makes the manuscript easier to follow and ensures that the information needed to interpret the figures and subsequent results is provided earlier in the text.

2. **The methods section in the Appendix leaves too many important to the imagination. The decompositions are based on different evaluations of Eq. (1), but the corresponding equations, and their precise interpretation are not provided.**

   Response: In the revised manuscript, we have expanded Sect. 2.4, "Decomposition into thermodynamic and dynamic contributions" (L120). This section now rewrites Eq. 1 in terms of ω_e and $\Gamma_e = dq_s/dp|_{\theta^*}$, and explicitly defines the thermodynamic, dynamic, mean-state, and interaction terms. We also clarify the interpretation of these terms and explain how their corresponding fractional changes are calculated as the thermodynamic, dynamic, and residual contributions.

3. **Generally speaking, we never know if RX1day in CMIP6 is even close to observations or reanalyses. Therefore, what is called RX1day in some**

models could actually be a drizzle, and not be so extreme from the physical points of view. Is there a way to consider only CMIP6 models for which RX1day is more than a drizzle?

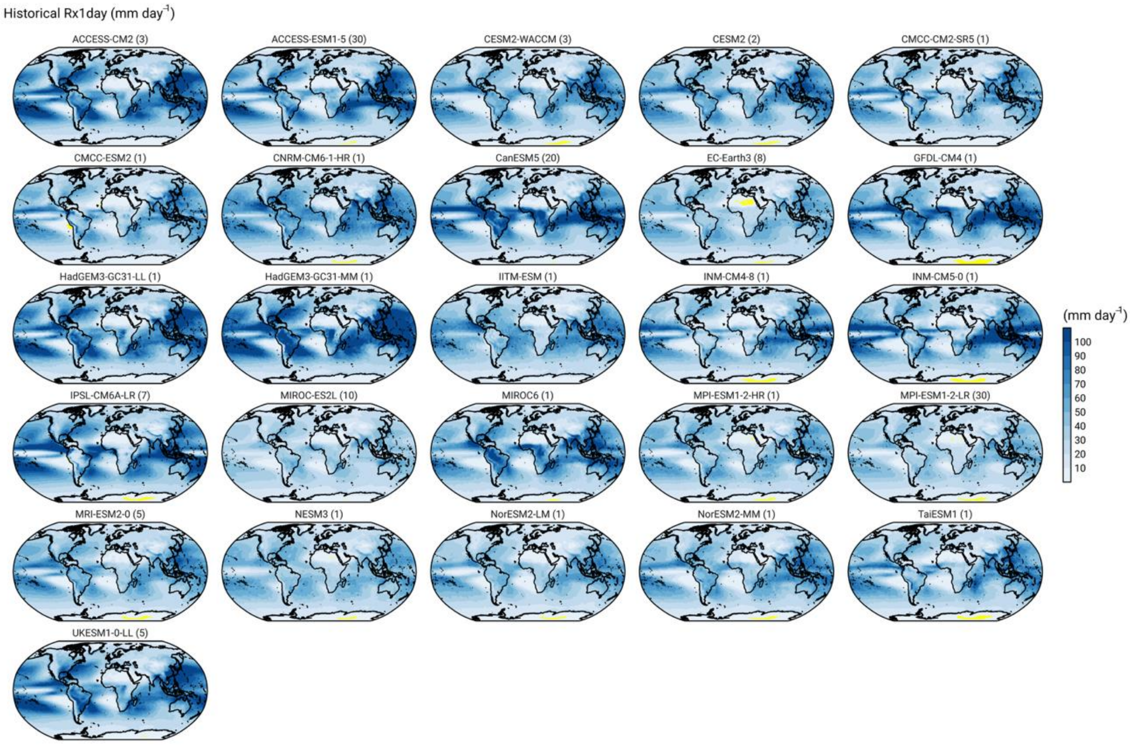

Figure 3 Historical Rx1day. *Ensemble means of the historical 50-year mean of Rx1day over 1950-1999 for each model (in mm d^{-1}). Grid cells with Rx1day below 1 mm d^{-1} are masked in yellow.*

Response: Thank you for raising this important point. We agree that the historical magnitude of Rx1day should be evaluated before interpreting projected fractional changes. To address this, Figure 3 in this response document shows the ensemble-mean historical 50-year mean Rx1day for each of the 26 models over 1950-1999. Grid cells with Rx1day below 1 mm d^{-1} cover only a very small area fraction and are masked in yellow.

The figure shows that historical Rx1day magnitudes vary across models, in some regions by up to approximately a factor of two. Comparisons between the available lower- and higher-resolution versions within the same model families suggest that horizontal resolution can influence simulated Rx1day intensity, with higher-resolution versions generally producing more intense extremes at their native resolution. However, differences in spatial patterns across model families are generally larger than those between resolution versions of the same model family, indicating that model structure and parameterizations also contribute substantially to the spread.

Very low historical Rx1day values below 1 mm d^{-1} are very rare and occur predominantly over Antarctica and some very dry grid cells. These regions do not affect our main conclusions. In the revised manuscript, we retain the comparison between historical Rx1day and the diagnostic scaling in Figure A1 and discuss the main regional mismatches in Sect. 3. This provides context for regions where percentage changes may be less reliable because of very low baseline precipitation or reduced performance of the diagnostic scaling.

Specific comments

4. **29: I do not understand what is meant by a "global intensification of Rx1day". RX1day is not defined globally but locally.**

   Response: Thank you for pointing this out. By "global intensification of Rx1day", we mean an intensification that occurs across much of the globe in the spatial distribution of locally defined Rx1day. To make this clearer, we have changed the formulation to "widespread intensification".

5. **32: What is a robust thermodynamic signal? How can it be robust if dynamic processes can potentially amplify, suppress or reverse it?**

   Response: The "robust thermodynamic signal" refers to the consistently positive and spatially widespread thermodynamic intensification associated with increased atmospheric moisture-holding capacity. This does not imply that the total Rx1day response is consistent across models everywhere, as dynamic changes can amplify or weaken the thermodynamic signal, or even locally reverse the sign of the total Rx1day response. To avoid ambiguity, we have clarified this wording in the revised manuscript (L30).

6. **47: "A common strategy for reducing the influence of internal variability is to [...]". This formulation is strange: internal variability is part of the system. One could argue that only god(s) could reduce its influence. You might mean the influence of internal variability on the estimate of some quantity.**

   Response: Thank you for this clarification. We agree that internal variability itself is an inherent component of the climate system and is not reduced by our analysis. We have revised the text to state that these approaches reduce the imprint of internal variability on the estimates of the forced model response (L49).

7. **57: What is the fractional summer response?**

Response: Thank you for pointing this out. We have clarified that the "fractional summer response" refers to the fractional change in precipitation extremes during summer (L60).

8. **58: The weakened ascent of what?**

   Response: We have revised the text to specify that weakened ascent occurs during extreme precipitation events (L61).

9. **62: A recent study is mentioned in this paragraph, but the cited references are in 2020 and 2018. The adjective recent might not be the most appropriate here.**

   Response: We have replaced "a recent study" with "a previous study" in the revised manuscript.

10. **Eq. (1) is not very precise. What is the coefficient of proportionality? What is \theta^\star? Things would be clearer if the explicit integral was written down.**

    Response: Thank you for pointing this out. We have revised the accompanying text to make the diagnostic scaling more explicit. We define θ^* as saturated equivalent potential temperature and state that the derivative is evaluated along a moist adiabat during extremes (L112).

    Eq.1 does not contain an unspecified fitted coefficient of proportionality. The braces denote the mass-weighted pressure integral, and the numerical implementation additionally converts the resulting rate from s^{-1} to d^{-1}. Thus, the factor (86400/g) is included in the definition of the vertical integral.

11. **87: I do not understand "It further assumes that [...]". Where do we see this?**

    Response: Thank you for pointing this out. We have revised the text to clarify that the right-hand side of Eq.1 estimates condensation associated with updrafts during extremes. Interpreting this quantity as surface precipitation requires the assumption that a sufficiently large fraction of the condensate reaches the surface as rainfall, ignoring the losses due to re-evaporation, detrainment, or condensate storage during the days of annual maximum 1-day precipitation (L113).

12. **Fig. 1: what is fractional change? Fig. 1 and Fig.A1 are not comparable (not the same units).**

    Response: Fractional changes (or also percentage changes) in Rx1day are estimated from the linear regression of annual Rx1day against GMST over 1950-2099. The resulting fractional changes are expressed relative to historical Rx1day and expressed in % K^{-1} of global warming. We have clarified the definition of fractional change in the revised method section 2.2 “Historical precipitation extremes and their future changes” (L99). Figure 1shows the fractional changes in Rx1day and diagnostic scaling, and their difference in % K^{-1}. Figure A1 shows the historical 50-year mean of Rx1day and diagnostic scaling in mm d^{-1}, and their absolute difference, defined as historical diagnostic scaling minus historical Rx1day, as well as the fractional difference, defined as the absolute difference divided by historical Rx1day (in %).

13. **105: why does “this approach reduce the influence of internal variability”?**

    Response: Internal variability leads to random climate variations and their effect becomes increasingly smaller the longer the period considered. To account for differences in historical Rx1day intensity and climate sensitivity among models, we focus on the sensitivity of extreme precipitation to global warming. Specifically, we estimate the fractional change in precipitation extremes per K of GMST increase over 1950-2099 using linear regression. Using the full 150-year time series provides a more robust estimate of the forced response than a difference between two shorter periods (see detailed description of this effect in Fischer et al. 2014, Geophys. Res. Lett.). Interannual and decadal internal variability is treated as residual variability around the regression line, while the regression coefficient captures the component of Rx1day changes that covaries with GMST. This approach therefore reduces the imprint of internal variability on the estimated precipitation response to global warming. We have included this discussion in the revised manuscript (L256).

14. **128: What is technically challenging? and so what? If you obviously coped with this difficulty, why do something that you consider a simpler metric? And why did not you start with the simple metric?**

    Response: Thank you for raising this point. The full diagnostic scaling is required as the starting point because it provides the physically consistent and quantitative decomposition of Rx1day changes into thermodynamic, dynamic, and residual contributions. It depends on the vertical integral of the product of event-conditioned vertical motion and the pressure derivative of saturation specific humidity. Consequently, it requires daily three-dimensional temperature and updraft fields at multiple pressure levels, event selection at every grid point, and model-specific vertical integration. This implies that high resolution model output is needed (that for many experiments is not available), and the implementation of the calculation in a programming language is not straight-forward, the calculation is computationally demanding and less straightforward for interpretation.

To this end, we test whether these contributions can be approximated with less model output but still retain the dominant physical information. Specifically, we compare the thermodynamic contribution with changes in vertically integrated saturation specific humidity q_s and the dynamic contribution with changes in vertically averaged updraft ω_e (Figure 2b and d).

We have revised the text to clarify that the full diagnostic scaling remains the basis for the quantitative decomposition, whereas the simpler variables are introduced only to interpret the physical drivers of the decomposed contributions and to identify potentially more accessible constraints (L281).

15. **Fig. 2: what period is considered?**

    Response: Figure 2 shows projected changes estimated from linear regressions against GMST over the full 150-year period, 1950-2099.

16. **188: Why are q_s and \omega_e called "diagnostics"? As far as I understood, they are climate variables. The next sentence calls them "proxies". Proxies of what?**

    Response: Thank you for pointing this out. We have replaced these terms with "variables" and clarified that changes in vertically integrated q_s and vertically averaged ω_e during extremes are examined because they summarize the dominant thermodynamic and dynamic information in the scaling decomposition.

17. **Section 4: I have always wondered why (and not only in this paper) surface temperature is considered for the Clausius-Clapeyron scaling of precipitation, because precipitation almost never forms at the surface of the planet, but often above the boundary layer. This is why Eq. (1) works remarkably well (because it integrates information on the vertical). Why not consider temperature at the cloud altitude, to be consistent with Eq. (1)?**

    Response: Thank you for this thoughtful comment. We agree that the thermodynamic term in Eq. (1) depends on the full vertical temperature structure, and that surface temperature alone cannot fully represent the thermodynamic conditions relevant to condensation aloft. In our analysis, surface temperature is therefore not intended to represent the temperature at which precipitation forms, nor to replace the vertically integrated diagnostic in Eq. (1). Rather, it is used as a readily observable and widely used proxy for the lower-tropospheric thermodynamic state.

    Following the reviewer's suggestion, we have added a sensitivity analysis using event-conditioned temperature changes at 850, 700, and 500 hPa as alternative predictors of the thermodynamic contribution (Figure A6). The explained variance is higher for temperature changes at 700 and 500 hPa, indicating that mid-tropospheric temperature changes are more strongly associated with the thermodynamic contribution than surface temperature in some regions.

We retain surface temperature in the main text because it is more directly observable and facilitates comparison with previous Clausius-Clapeyron scaling studies. We now explicitly clarify its interpretation and refer readers to the pressure-level sensitivity analysis in the main text (L412).

We also added a decomposition of total warming at 700 hPa into shift-of-day warming/cooling and same-day warming (L423, Figure A7). These results support our main conclusion: over extratropical land, inter-model differences in the thermodynamic contribution are predominantly associated with a future shift of Rx1day events towards a different time of year (Figure A8). Same-day warming explains an additional part of the inter-model spread, particularly over tropical regions and oceans (Figure A8).

18. **The diagram in Fig. 6 is hard to understand. The notion of "shift-of-day" is not clear. The explanation should appear in the core of the manuscript, not in the figure legend.**

    Response: We have added a description of total warming and its decomposition in Section 2.7 "Total warming and its decomposition" (L217). Total warming is defined as the change in local surface temperature on Rx1day event days per K of global mean surface warming. It is estimated by regressing event-day surface temperature against GMST over 1950-2099.

    We decompose total warming into a shift-of-day component and a same-day component. Shift-of-day warming or cooling is the temperature change associated with a change in the mean calendar day on which Rx1day occurs; for example, a shift from summer toward a cooler shoulder season produces negative shift-of-day warming. Same-day warming is the change in the residual temperature on the corresponding calendar day relative to the 150-year daily climatology.

19. **Figure 8: a panel showing the vertical structure of the atmosphere could be useful, since this seems to be the main explanation.**

    Response: Thank you for this suggestion. We agree that a depiction of the vertical structure could help interpret the level dependence shown in Figure 9. However, the vertical structure of event-conditioned ascent differs substantially across years, regions and models. For example, the magnitude and vertical extent of ascent differs among tropical deep convection, extratropical frontal precipitation, high-elevation regions, and dry regions. A mean vertical profile across regions or across days would therefore not provide a clear physical explanation for the spatially aggregated relationships shown in Figure 9.

Nevertheless, the strong global agreement across models that 700 hPa is a key level for explaining uncertainty in the dynamic contribution to projected Rx1day changes is informative. A more complete interpretation of the underlying vertical structures would require region-specific analyses of event-conditioned updraft profiles. This is beyond the scope of the present global study but is being pursued in our ongoing regional work.

20. **312: The Methods section is not always clear, and needs more justifications.**

    Response: Thank you for this comment. To improve clarity, we moved the Methods from the Appendix to a new Section 2 “Data and Methods” immediately after the Introduction (L87). This section now provides the information needed to interpret the analysis and results before they are presented.

21. **341: What is a semi-partial explained variance?**

    Response: Thank you for noting this. We now define semi-partial explained variance explicitly in the method section 2.6 “Model variances explained by the dominant factors” (L179). For a given predictor, the semi-partial R^2 is the additional fraction of inter-model variance in the thermodynamic contribution explained by that predictor after accounting for the other predictor. Thus, the semi-partial R^2 for shift-of-day warming measures its unique explanatory contribution after same-day warming is included in the regression, and vice versa. This allows us to separate their individual contributions despite possible correlation between the two predictors.

22. **357: “Same-day warming is defined as [...]” Without an equation, this is hard to figure out.**

    Response: We have added a description of total warming and its decomposition in Section 2.7, “Total warming and its decomposition” (L217). Same-day warming is calculated with the linear regression between residual event-day temperature against annual GMST over 1950-2099, expressed in K K^{-1}. The residual event-day temperature is calculated by subtracting the 150-year daily climatological temperature of the corresponding calendar day from the event-day temperature.

23. **Table A1: many of the models share the same dynamical core (and hence biases). The horizontal resolutions vary from the very low (CanESM5) to very high (MPI-ESM1-2-HR or EC-Earth3). Therefore, the meaning of \omega does vary from one model to the other. And the overall choice is debatable.**

Response: Thank you for the comment. We agree that the CMIP6 models are not fully independent, as several share components or development heritage, and their horizontal resolutions vary substantially.

However, all models use the same formal definition of ω_e as pressure vertical velocity, expressed in Pa s^{-1}. Their event-conditioned ω_e represents grid-box-scale updrafts during extreme precipitation events; its magnitude and vertical structure can be influenced by model resolution, dynamical cores, and parameterized convection. We therefore focus on projected fractional changes in the relevant quantities relative to each model's historical intensity, normalized by its projected global warming, rather than directly comparing their absolute magnitudes.

Our model selection was determined by the availability of the required daily precipitation, temperature, and vertical velocity fields. We have clarified these limitations in the revised manuscript and added a discussion of the potential influence of model resolution on the decomposition (L492). We also refer to our response to Comment 8 of Reviewer #1, where this issue is discussed in more detail.