nature portfolio

Peer Review File

Combined emergent constraints on future extreme precipitation changes

Corresponding Author: Dr Hideo Shiogama

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Review of manuscript #NCOMMS-25-06567-T entitled "Combined emergent constraints on future extreme precipitation changes"

Shiogama et al seek to reduce uncertainty in global climate model projections of future extreme precipitation through a new method that combines constraints on global warming and the sensitivity of heavy precipitation to warming. While the constraint does not suggest a central estimate for future extreme rainfall that is very different from the ensemble mean, it does manage to reduce uncertainty in projections by a substantial amount. The authors also apply this methodology on a grid-box level, which could be a valuable tool if shown to be reliable. There are several studies that have sought to tackle this important problem, and it is difficult to say if the proposed method is superior to the others. The paper is well written and features interesting analysis that is appropriate for this journal. However, I have some concerns about overconfidence and the methods which should be addressed prior to publication.

Major comments:
Methodology: How do you ensure that this combined approach is not overconfident? The authors note the method of Bretherton and Caldwell (2020) in the text, which seems very relevant. While their application is different (they combine various constraints on the same y-axis metric instead of constraints on warming and a sensitivity term), Bretherton and Caldwell (2020) note that combining constraints is only useful if those constraints provide some independent information. I believe that the same principles should apply to this framework since the two future change metrics in this case are likely correlated. Please elaborate on why you think this type of consideration can be ignored for your approach. Bretherton and Caldwell (2020) also include a method for adjusting for overconfidence, a topic which should be discussed in more detail here.

Baseline period: I do not see the purpose in calculating future change variables in relation to a 1851-1900 baseline period. For one, the change signal includes an extended period for which we already have observed data (which GCMs are often somewhat tuned to replicate in the case of global mean temperature change). Secondly, a portion of the change has already occurred (i.e., the warming from 1850 to 2022 is included in the y-axis change), making the y-axis metric not just a future change metric. Revise to make use of a recent past baseline.

Interpretation of results: The paper often lacks interpretation around its findings and quantification of results in the main text (as pointed out in several minor comments below). These should be expanded on to better match the format of this journal.

Minor comments:
L22: change "the present" to "historical"
L24-25: change "whereas the ΔTgm-related EC reduces it by 26%" with "an improvement from only using ΔTgm (26% reduction)".
Abstract: It is useful for the community when EC studies list the unconstrained and constrained estimates (e.g., central +/- 1 standard deviation) in the abstract. I would prefer that to the material on L27-29.
L41: "last 15 years" it has been nearly 20 years since Hall and Qu (2006).
L44: remove "of each ESM" to make this sentence less awkward.
L46: remove "until the end of the 21st century"

L49: state roughly how much reduction in variance there was.
L67: when is ΔX not going to be correlated to ΔTgm? Should there be some threshold for meaningful correlation if it is advisable for this method to be used by others?
L71: rx1day would be a more standard definition for annual maximum daily precipitation.
L85: quantify these correlations
L89: Set the stage by first stating what the intermodel spread is in this term.
L96-97: Not sure what this means. Double the variance and redo the Bowman calculation?
L101: Are there any resolution dependencies with this metric and its spread across GCMs or observational datasets?
L104: it is important to state that the EC does not drastically alter the central estimate from the unconstrained ensemble. I recommend stating both of these values and their +/-1 standard deviation in the text.
L110: Expand on this humidity analysis in the main text. The word limit is far from being reached so no need to be so brief.
L114: Can you think of a better name than "the ΔTgm-related EC on ΔRgm"?
L134-135: Even if this is the case, you should still state what the central estimate for change is and how it relates to the climatological value.
L147: How robust are these spatial patterns? Does it make physical sense for the global value for say global warming to be related to heavy rainfall change in all places or should it be a localized relationship (e.g. local warming)?
L149: state that the R climatology is a global mean.
L150: One way to show robustness in these spatial patterns would be to split CMIP5/6 and stipple areas where there are significant correlations in both.
L151-153: Expand on this relationship here.
L156-169: Make this paragraph more quantitative.
L202-203: This statement is wrong. Ref #43 is not an emergent constraint.
L205-207: Show spatial maps of the observational uncertainty (in the supplement) so the reader can understand how this component of the EC varies in space.
L210-212: Elaborate on how the results of these studies compare (including Dai et al 2024 and Li et al. 2024)
L418: replace 'of' with 'from'
L426: replace "in the" with "stemming from"
L434-436: I found this confusing. Better explain how you get to a value of 0.0254.
L439-440: Comment on some of the differences between these products. Also, comment on the large variability across datasets.
L440-441: How much does this quantity vary by year in these datasets? Calculate the mean Rgm each year to determine the standard deviation associated with each mean estimate shown here.
L517: The 1851-1900 R isn't defined as R_overbar anywhere.
Figure 1: A mean Rgm of nearly 80 mm/day seems like a major outlier. Can these values be listed in a table somewhere? The inclusion of this one model makes the model spread much larger, and thus the ratio of observational to model uncertainty smaller (a factor in the Bowman calculation), which directly impacts the RRV value.
Figure 4: Provide some more insight into why the combined EC is so effective in certain areas. Is this just where the outlier models are out to lunch?
Extended Data Fig 3: Show this for both CMIP5 and CMIP6 separately.
Extended Data Fig 5: More detailed descriptions of these terms in the caption would help the reader.

(Remarks on code availability)
I cannot comment on the code accompanying this manuscript as it is in IDL, a language that I do not use. There does appear to be an appropriate amount of information for one to try to replicate the results if they are familiar with these files types though.

Reviewer #2

(Remarks to the Author)
Review of "Combined emergent constraints on future extreme precipitation changes"

This paper addresses an important topic—the projected response of extreme precipitation to climate change. The study introduces a novel combined EC method, which integrates constraints on global mean temperature changes and extreme precipitation sensitivity (EPS) to effectively reduce the uncertainty in regional extreme precipitation projections. These findings contribute valuable insights into the interpretation of GCM simulations of extreme precipitation and offer useful implications for impact researchers and policymakers. I find the study promising and recommend acceptance after the authors address the following concerns.

Major comments:
1. Clarification on the "combined EC" approach
I find the methodology of the combined emergent constraints approach somewhat unclear. Specifically, when constraining ΔR_gm using Ttrend, the authors apply the hierarchical EC framework, while the constraint on extreme precipitation sensitivity (also a single variable) is derived using the combined EC approach. Could the authors clarify the distinction between these two methods? For example, could an approach similar to Eq. (8) be used to estimate the ΔT_gm-related EC on ΔR_gm, or could a method akin to Figure 1b be employed to Figure 1c?
Since extreme precipitation sensitivity and global mean temperature may also exhibit inter-model correlations, does this imply that using Ttrend to constrain ΔT could inherently provide some constraint on extreme precipitation sensitivity as well? How should this be interpreted in the combined EC?
According to line 479, the 10,000 random samples represent the constrained range. Would this mean that the original inter-

model standard deviation is approximately equivalent to the standard deviation obtained from the 10,000 random samples using the raw model values in Eq. (8) for both extreme precipitation sensitivity and global mean temperature? Is this understanding correct?

2. The paper examines the influence of internal variability in global mean temperature but does not assess the impact of internal variability in precipitation, which could be large. As mentioned in lines 442–443, the internal variability of global mean precipitation is much smaller than the discrepancies among different observational datasets. However, in the regional EC, the authors use climate-state precipitation at individual grid points. At the grid-point scale, is the influence of internal variability still negligible? How significant is its impact on the results of the regional EC?

3. In Figure 2, the changes in the 50th percentile values appear minimal for all EC methods. In the regional EC, do the 50th percentile values differ among different EC methods?

Minor comments：
Line 54: Should the cited reference here be [9] instead?
Lines 117–118: It might be helpful to visualize the joint distribution or correlation coefficient between the two variables to better illustrate their relationship.

(Remarks on code availability)
Yes, the code provide a README file with enough instructions for installing and running the application.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have done an excellent job of responding to my prior comments and adding more detail to the main text. I have nothing further to add prior to publication. I will just note a typo on L111, where "50 percentile values" should say "50th percentile values".

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
The authors have made a good effort in revising the paper and have addressed most of the issues I raised.

Reviewer 1 raised a concern regarding the choice of the 1851–1900 baseline period. Although the authors have provided some justification, this reasoning primarily supports the ΔT-related EC approach and may not directly apply to the ΔR/ΔT-based method. So, it would be helpful to assess whether the results are sensitive to the choice of baseline years. For example, how would the results change if a more recent baseline period were used? I would be interested to see how the EC method performs when the constraint is weakened.

In addition, it is interesting that the 1997–2019 mean Rgm values are close to the 1851–1900 mean values (Supplementary Fig. 1). This supports the use of 1997–2019 precipitation to constrain projections at the global mean scale. However, does this relationship also hold at the regional scale? The authors should consider providing a figure showing the regional correspondence between these two baseline periods.

(Remarks on code availability)
I did not run the code, but it includes a README file and appears to contain the necessary scripts for reproducing the key calculations.

Reply to Reviewer #1

Shiogama et al seek to reduce uncertainty in global climate model projections of future extreme precipitation through a new method that combines constraints on global warming and the sensitivity of heavy precipitation to warming. While the constraint does not suggest a central estimate for future extreme rainfall that is very different from the ensemble mean, it does manage to reduce uncertainty in projections by a substantial amount. The authors also apply this methodology on a grid-box level, which could be a valuable tool if shown to be reliable. There are several studies that have sought to tackle this important problem, and it is difficult to say if the proposed method is superior to the others. The paper is well written and features interesting analysis that is appropriate for this journal. However, I have some concerns about overconfidence and the methods which should be addressed prior to publication.

Thank you very much for your useful comments. We did some additional analyses and revised the manuscript. We found small mistakes in the GPCP's climatological value (changed it from 33.7 mm/day to 33.8 mm/day in the global mean) and the code used for Fig. 4f. We corrected the errors and found no apparent changes in the results. Some changes were made to the manuscript to conform to the journal's format. Approximately 50 words had to be omitted from the abstract. We moved the humidity analysis from the supplementary material to the main text.

Major comments:

Methodology: How do you ensure that this combined approach is not overconfident? The authors note the method of Bretherton and Caldwell (2020) in the text, which seems very relevant. While their application is different (they combine various constraints on the same y-axis metric instead of constraints on warming and a sensitivity term), Bretherton and Caldwell (2020) note that combining constraints is only useful if those constraints provide some independent information. I believe that the same principles should apply to this framework since the two future change metrics in this case are likely correlated. Please elaborate on why you think this type of consideration can be ignored for your approach. Bretherton and Caldwell (2020) also include a method for adjusting for overconfidence, a topic which should be discussed in more detail here.

By considering the correlation (r_0) between ΔT_{gm} and $\frac{\Delta R_{gm}}{\Delta T_{gm}}$ (or $\frac{\Delta R}{\Delta T_{gm}}$) in the Eqs. 9 and 11 of Methods, we avoided the overconfident estimation. For the global mean analyses, $r_0 = 0.28$. If we set r_0=0 (i.e., ignoring the dependence), the RRV of the combined-EC on the global mean ΔR_{gm} could be 51%, while the actual RRV is 42% (L153-158). For the regional analyses, significant positive correlations between $\frac{\Delta R}{\Delta T_{gm}}$ and ΔT_{gm} are found over, for example, the North Pacific Ocean, the North Atlantic Ocean and the South Pacific Convergence Zone (Supplementary Fig. 5). The understanding of the mechanism for these relationships remains for future work. If we ignored the correlations between $\frac{\Delta R}{\Delta T_{gm}}$ and ΔT_{gm} in our $\frac{\Delta R}{\Delta T_{gm}}$-related and combined EC methods, we could overestimate RRVs by 5-25% (L229-233).

Baseline period: I do not see the purpose in calculating future change variables in relation to a 1851-1900 baseline period. For one, the change signal includes an extended period for which we already have observed data (which GCMs are often somewhat tuned to replicate in the case of global mean temperature change). Secondly, a portion of the change has already occurred (i.e., the warming from 1850 to 2022 is included in the y-axis change), making the y-axis metric not just a future change metric. Revise to make use of a recent past baseline.

To apply the ΔT_{gm}-related EC, we have to use the preindustrial period as the baseline period. Because anthropogenic aerosol emissions are small in both the periods of 2051-2100 and 1851-1900, ΔT_{gm} and ΔR_{gm} (2051-2100 minus 1851-1900) are mainly forced by increases in greenhouse gas concentrations. We define trT_{gm} as the recent past (1970-2022) global mean temperature trends. Because global aerosol emissions were nearly constant in this period, the effects of the aerosol forcing trends on trT_{gm} are small. ΔR_{gm} and ΔT_{gm} are significantly correlated with trT_{gm}, because all ΔR_{gm}, ΔT_{gm} and trT_{gm} are mainly driven by increases of greenhouse gas concentrations (L83-90). This idea is the heart of the ΔT_{gm}-related EC approach. If we use the present climate as the baseline, the baseline-period-mean values involve significant aerosol influences, and ΔT_{gm} and ΔR_{gm} include responses to the declines of aerosol emissions from the present to the future. In that case, we have to develop another method to constrain the aerosol effects and combine it with our methods. Therefore we use 1851-1900 as the baseline period here.

Interpretation of results: The paper often lacks interpretation around its findings and quantification of results in the main text (as pointed out in several minor comments below). These should be expanded on to better match the format of this journal.

Thank you very much for your many useful comments. We added some interpretation around the findings and quantification of results in the revised manuscript. Please see the following responses to your comments.

Minor comments:
L22: change “the present” to “historical”

We changed it (L23).

L24-25: change “whereas the ΔTgm-related EC reduces it by 26%” with “an improvement from only using ΔTgm (26% reduction)”.

Changed (L25).

Abstract: It is useful for the community when EC studies list the unconstrained and constrained estimates (e.g., central +/- 1 standard deviation) in the abstract. I would prefer that to the material on L27-29.

We are sorry that we could not include it in the abstract due to the tight limit of the abstract length. Instead, we added two new tables indicating the unconstrained and constrained estimates (Table 1 and Supplementary Table 2) and discussed those values in L139-152 and L168-177.

L41: “last 15 years” it has been nearly 20 years since Hall and Qu (2006).
We cited Hall and Qu (2006) and changed it to “nearly 20 years” (L39).

L44: remove "of each ESM" to make this sentence less awkward.

Removed. (L42)

L46: remove "until the end of the 21st century"

Removed. (L44)

L49: state roughly how much reduction in variance there was.

It was 8-30% reduction of the variance. (L48)

L67: when is ΔX not going to be correlated to ΔTgm? Should there be some threshold for meaningful correlation if it is advisable for this method to be used by others?

If the correlation is small, RRV of Bowman et al. (2018) becomes close to zero as the definitions of Eqs. 7-8 (L417). While there is no clear threshold for meaningful correlation, we usually consider "statistically significant correlation" is meaningful.

L71: rx1day would be a more standard definition for annual maximum daily precipitation.

Yes. However, if we use "rx1day" instead of R, our equations become longer, and the manuscript would be difficult to read. Therefore we use R here.

L85: quantify these correlations

Those correlations (0.55 and 0.76) are denoted on the top of the panels (Figs. 1a-b). We also stated those in L88.

L89: Set the stage by first stating what the intermodel spread is in this term.

We changed "reduce the variance" to "reduce the inter-model variance". (L92)

L96-97: Not sure what this means. Double the variance and redo the Bowman calculation?

Yes. We made clear it as "Therefore, as a sensitivity test, we double the variance in the internal climate variability added to the observed trT_{gm}[28] and redo the EC calculations. Our results remain robust (Figs. 1a-b).". (L99-101)

L101: Are there any resolution dependencies with this metric and its spread across GCMs or observational datasets?

Although it is an interesting issue, it is difficult to investigate resolution dependencies. Many ESMs provided only the regrided output data (e.g., 1° lon x 1 ° lat) but not the original resolution one. Although the modelling centers reported the nominal resolutions of their models, the numbers are very roughly delimited (50km, 100km, 250km ..). With only three observational datasets, we cannot discuss any resolution dependency. Therefore, we did not investigate the resolution dependency issue.

L104: it is important to state that the EC does not drastically alter the central estimate from the unconstrained ensemble. I recommend stating both of these values and their +/- 1 standard deviation in the text.

We added the following sentence in L108-112: "The central value and the lower and upper bounds of $\frac{\Delta R_{gm}}{\Delta T_{gm}}$ (mm/day/°C), estimated as the 50th [2.5th, 97.5th] percentile values, are 2.07 [0.316, 3.82] for the raw ensembles and 2.03 [0.632, 3.44] for the constrained range. The EC does not drastically alter the 50 percentile values (only −0.04), but changes lower (+0.316) and upper bounds (−0.38).”

L110: Expand on this humidity analysis in the main text. The word limit is far from being reached so no need to be so brief.

We moved the humidity analysis from the supplementary material to the main text. (L117-118, L190-192, L247-285)

L114: Can you think of a better name than “the ΔTgm-related EC on ΔRgm”?

We thought about it a lot, but couldn't come up with any better terms.

L134-135: Even if this is the case, you should still state what the central estimate for change is and how it relates to the climatological value.

We added the following sentences in L147-152: “The changes in the 50th percentile values are slight for all the EC methods (changes from 5.49 (the raw) to 5.15 (the ΔT_{gm}-related EC), 5.13 (the $\frac{\Delta R_{gm}}{\Delta T_{gm}}$-related EC) and 5.02 (the combined EC)), because the ESM averaged values of trT_{gm} (0.220 °C/10-years) and the 1997-2019 mean R_{gm} (43.1 mm/day) are close to the mean values of the observations (0.208 °C/10-years and 42.3 mm/day) (also see Eq. 6 of Method).”

L147: How robust are these spatial patterns? Does it make physical sense for the global value for say global warming to be related to heavy rainfall change in all places or should it be a localized relationship (e.g. local warming)?

L150: One way to show robustness in these spatial patterns would be to split CMIP5/6 and stipple areas where there are significant correlations in both.

Extended Data Fig 3: Show this for both CMIP5 and CMIP6 separately.

We showed correlation maps for CMIP5-only and CMIP6-only as well as that of CMIP6+CMIP5 (Supplementary Fig. 4). We also denoted the grids where correlations are significant for both CMIP5-only and CMIP6-only. The spatial patterns of the correlations between ΔR and trT_{gm} are similar between CMIP5 and CMIP6, but significant correlations are found in larger areas for CMIP6 because the CMIP6 ensemble involves more 'hot ESMs' than the CMIP5 ensemble. The correlations between the historical R and $\frac{\Delta R}{\Delta T_{gm}}$ are similar between CMIP5 and CMIP6 (L194-198).

Shiogama et al. (2024) found that regional temperature changes and regional specific humidity changes are well related to ΔT_{gm} (and trT_{gm}) in most of the world. Therefore the thermodynamic components of regional ΔR (ΔR due to changes in humidity) are correlated well with ΔT_{gm} (and trT_{gm}). We explained this point in L185-188. We discussed the mechanism underlying the correlations between the past R and $\frac{\Delta R}{\Delta T_{gm}}$ in L190-192 and L247-285.

L149: state that the R climatology is a global mean.

Here, R is regional. We insert "regional" in L183 and L189.

L151-153: Expand on this relationship here.

We discussed this relationship in L247-285.

L156-169: Make this paragraph more quantitative.

We made this paragraph more quantitative. (L199-219)

L202-203: This statement is wrong. Ref #43 is not an emergent constraint.

Thank you. We omitted Ref #43. (L300)

L205-207: Show spatial maps of the observational uncertainty (in the supplement) so the reader can understand how this component of the EC varies in space.

We showed spatial maps of the observational uncertainty and internal variability in Supplementary Fig. 7 and explained it in L394-396.

L210-212: Elaborate on how the results of these studies compare (including Dai et al 2024 and Li et al. 2024)

There are differences in indicators analyzed by us (changes in the global mean and regional Rx1day), Thackeray et al. (2022) (precipitation ≥99th percentile), Paik et al. (2023) (*percentage* changes in the global mean or sub-continental mean Rx1day), Li et al. (2024) (*percentage* changes in the intensity of 50-year event over the global land) and Dai et al. (2024) (*percentage* changes in the global mean or sub-continental mean Rx1day). Therefore, we could not quantitatively compare the results between us and them. Instead, we elaborated on the comparison of the methods in L307-328.

L418: replace ‘of’ with ‘from’

Replaced. (L354)

L426: replace “in the” with “stemming from”

Replaced. (L362)

L434-436: I found this confusing. Better explain how you get to a value of 0.0254.

It was calculated by $\sqrt{0.00267^2 + 0.0179^2 \times 2}$, where 0.00267^2 and 0.0179^2 were the variances of the HadCRUT5 realizations and the internal variability. (L361, 366, 368 and

372)

L439-440: Comment on some of the differences between these products. Also, comment on the large variability across datasets.

We added the following sentences.

"GPCP is a merged dataset of gauge stations, satellites and sounding observations. MSWEP2 is a merged dataset of gauge-, satellite- and reanalysis-based data. GSWP3-W5E5 is the combined dataset of GSWP3 (a merged dataset of dynamically downscaled twentieth-century reanalysis data and global observations of precipitation) and W5E5 (a merged dataset of bias-adjusted and raw reanalysis datasets). The differences between the observational datasets of precipitation are large due to many sources of uncertainty[48,49], e.g., limited number and spatial coverage of surface stations, and differences in satellite algorithms, data assimilation methods, reanalysis datasets and bias correction methods." (L377-385)

"In most parts of the world, the inter-observational-datasets variance dominate the total variance (inter-observational-datasets + internal climate variability) of the 1997-2019 mean R (Supplementary Fig. 7)." (L394-396)

L440-441: How much does this quantity vary by year in these datasets? Calculate the mean Rgm each year to determine the standard deviation associated with each mean estimate shown here.

Because we use the climatology, we showed the standard errors of the mean ($s/\sqrt{n}$) rather than the standard deviations (s). The 1997-2019 mean R_{gm} values of these datasets are 33.8, 48.2 and 44.9 mm/day, and their standard errors of the mean values are 0.479, 0.307 and 0.329 mm/day, respectively (L386-387). The differences between the datasets are much larger than the standard errors.

L517: The 1851-1900 R isn't defined as R_overbar anywhere.

We defined it in L260.

Figure 1: A mean Rgm of nearly 80 mm/day seems like a major outlier. Can these values be listed in a table somewhere? The inclusion of this one model makes the model spread much larger, and thus the ratio of observational to model uncertainty smaller (a factor in the Bowman calculation), which directly impacts the RRV value.

By applying "1 ESM removed test", we confirmed that our results were not sensitive to the ESM sampling:

"To assess the influences of outliers, we redo the EC calculations with each of the ESMs omitted[40], confirming that the RRVs are not sensitive to the ESM sampling (Supplementary Fig. 3e). " (L164-166)

Figure 4: Provide some more insight into why the combined EC is so effective in certain areas. Is this just where the outlier models are out to lunch?

We confirmed that our results are not sensitive to the outliers: "By redoing the EC calculations with each of the ESMs omitted[40], it is confirmed that the fractions of area are not sensitive to the ESM sampling (Supplementary Fig. 6e)." (L243-244)

We discussed why the combined EC is so effective:

"The ΔT_{gm}-related EC can reduce the uncertainties in the thermodynamic components of ΔR, but not the dynamic components of ΔR[23,24]. Therefore the RRVs for the ΔT_{gm}-related EC on ΔR are small over the tropical ocean where the contributions of dynamic components to ΔR are large[23,24] (Fig. 4a). By contrast, the RRVs for the $\frac{\Delta R}{\Delta T_{gm}}$-related EC on ΔR are large in the major precipitation regions in the world, e.g., the ITCZ in the tropics, the storm-track regions in the middle latitude and the Asian monsoon region (Fig. 4c), because the sensitivities of extreme precipitation to humidity are the important factor for the intensities of R worldwide (Figs. 5a and 5d). The $\frac{\Delta R}{\Delta T_{gm}}$-related EC can reduce the uncertainty of ΔR in the tropical ocean regions, where the ΔT_{gm}-related EC is not effective (Figs. 4a and 4c). In the middle and high latitudes, both the ΔT_{gm}-related and $\frac{\Delta R}{\Delta T_{gm}}$-related ECs can reduce the uncertainties of ΔR. The combined EC is effective because the two ECs complement and strengthen each other (Figs. 4d-e)." (L274-285)

Extended Data Fig 5: More detailed descriptions of these terms in the caption would help the reader.

We added more detailed descriptions in the caption of Fig. 5.

Reviewer #1 (Remarks on code availability):
I cannot comment on the code accompanying this manuscript as it is in IDL, a language that I do not use. There does appear to be an appropriate amount of information for one to try to replicate the results if they are familiar with these files types though.

Thank you.

Reviewer #2 (Remarks to the Author):

Review of “Combined emergent constraints on future extreme precipitation changes”

This paper addresses an important topic—the projected response of extreme precipitation to climate change. The study introduces a novel combined EC method, which integrates constraints on global mean temperature changes and extreme precipitation sensitivity (EPS) to effectively reduce the uncertainty in regional extreme precipitation projections. These findings contribute valuable insights into the interpretation of GCM simulations of extreme precipitation and offer useful implications for impact researchers and policymakers. I find the study promising and recommend acceptance after the authors address the following concerns.

Thank you very much for your useful comments. We did some additional analyses and revised the manuscript. We found small mistakes in the GPCP’s climatological value (changed it from 33.7 mm/day to 33.8 mm/day in the global mean) and the code used for Fig. 4f. We corrected the errors and found no apparent changes in the results. Some changes were made to the manuscript to conform to the journal's format. Approximately 50 words had to be omitted from the abstract. We moved the humidity analysis from the supplementary material to the main text.

Major comments:

1. Clarification on the "combined EC" approach

I find the methodology of the combined emergent constraints approach somewhat unclear. Specifically, when constraining ΔR_gm using Ttrend, the authors apply the hierarchical EC framework, while the constraint on extreme precipitation sensitivity (also a single variable) is derived using the combined EC approach. Could the authors clarify the distinction between these two methods? For example, could an approach similar to Eq. (8) be used to estimate the ΔT_gm-related EC on ΔR_gm, or could a method akin to Figure 1b be employed to Figure 1c?

We are sorry that we confused you. We apply the hierarchical EC framework for all the Figs. 1a-c. The hierarchical EC framework is used for the ECs on these single variables (L74-118).

In the next section, we combine the information about the raw ΔT_{gm} and the constrained $\frac{\Delta R_{gm}}{\Delta T_{gm}}$ (taken from the hierarchical EC framework) to apply the $\frac{\Delta R_{gm}}{\Delta T_{gm}}$-related EC on ΔR_{gm} (L123-133). After that we combine the information about the constrained ΔT_{gm} and the constrained $\frac{\Delta R_{gm}}{\Delta T_{gm}}$ (both taken from the hierarchical EC framework) to do the combined ECs on ΔR_{gm} (L134-138). Our new methods (the $\frac{\Delta R_{gm}}{\Delta T_{gm}}$-related and combined ECs) are used to integrate multiple information. To clarify this point, we revised the following sentence:

"By applying the hierarchical EC framework[32], we also constrained the ΔT_{gm}-unrelated uncertainty in $\frac{\Delta R_{gm}}{\Delta T_{gm}}$ in Fig. 1c. To investigate the effects of the constrained $\frac{\Delta R_{gm}}{\Delta T_{gm}}$ for the uncertainty reduction of ΔR_{gm}, we develop the following method that uses the information of the constrained $\frac{\Delta R_{gm}}{\Delta T_{gm}}$ and the raw ΔT_{gm}." (L122-125)

Since extreme precipitation sensitivity and global mean temperature may also exhibit inter-model correlations, does this imply that using Ttrend to constrain ΔT could inherently provide some constraint on extreme precipitation sensitivity as well? How should this be interpreted in the combined EC?

By considering the correlation (r_0) between ΔT_{gm} and $\frac{\Delta R_{gm}}{\Delta T_{gm}}$ (or $\frac{\Delta R}{\Delta T_{gm}}$) in the Eqs. 9 and 11 of Methods, we avoided an overconfident estimation. For the global mean analyses, $r_0 = 0.28$. If we set r_0=0 (i.e., ignoring the dependence), the RRV of the combined-EC on the global mean ΔR_{gm} could be 51%, while the actual RRV is 42% (L153-158). For the regional analyses, significant positive correlations between $\frac{\Delta R}{\Delta T_{gm}}$ and ΔT_{gm} are found over, for example, the North Pacific Ocean, the North Atlantic Ocean and the South Pacific Convergence Zone (Supplementary Fig. 5). The understanding of the mechanism for these relationships remains for future work. If we ignored the correlations between $\frac{\Delta R}{\Delta T_{gm}}$ and ΔT_{gm} in our $\frac{\Delta R}{\Delta T_{gm}}$-related and combined EC methods, we could overestimate

RRVs by 5-25% (L229-233).

According to line 479, the 10,000 random samples represent the constrained range. Would this mean that the original inter-model standard deviation is approximately equivalent to the standard deviation obtained from the 10,000 random samples using the raw model values in Eq. (8) for both extreme precipitation sensitivity and global mean temperature? Is this understanding correct?

Yes. When we randomly sample from the normal distribution using the original mean and standard deviation values as the input parameters, the original inter-model standard deviation is approximately equivalent to the standard deviation obtained from the 10,000 random samples in principle. In Eq. 10 of the revised manuscript, we use the mean and standard deviation values estimated from the raw model values of ΔT_{gm}, those from the constrained $\frac{\Delta R_{gm}}{\Delta T_{gm}}$ (taken from Eqs. 6-7 of the hierarchical EC framework) and the covariance terms.

2. The paper examines the influence of internal variability in global mean temperature but does not assess the impact of internal variability in precipitation, which could be large. As mentioned in lines 442–443, the internal variability of global mean precipitation is much smaller than the discrepancies among different observational datasets. However, in the regional EC, the authors use climate-state precipitation at individual grid points. At the grid-point scale, is the influence of internal variability still negligible? How significant is its impact on the results of the regional EC?

We considered the internal variability in precipitation as well as that of the observed datasets in our ECs (L375-396). We showed the maps of the internal variability variance and the inter-observation-datasets variance in Supplementary Fig. 7. In most parts of the world, the inter-observational-datasets variance dominate the total variance (inter-observation-datasets + internal climate variability) of the 1997-2019 mean R (L394-396). Please note that averaging over the 1979-2019 period reduced the variance due to the internal variability. We also added the following sentences about the sources of uncertainty in the observational datasets:

“GPCP is a merged dataset of gauge stations, satellites and sounding observations. MSWEP2 is a merged dataset of gauge-, satellite- and reanalysis-based data. GSWP3-W5E5 is the combined dataset of GSWP3 (a merged dataset of dynamically downscaled twentieth-century reanalysis data and global observations of precipitation) and W5E5 (a merged dataset of bias-adjusted and raw reanalysis datasets). The differences between the observational datasets of precipitation are large due to many sources of uncertainty[48,49], e.g., limited number and spatial coverage of surface stations, and differences in satellite algorithms, data assimilation methods, reanalysis datasets and bias correction methods.” (L377-385)

3. In Figure 2, the changes in the 50th percentile values appear minimal for all EC methods. In the regional EC, do the 50th percentile values differ among different EC methods?

In the tropical Pacific and Indican ocean, the changes in the 50th percentile values are large for the $\frac{\Delta R}{\Delta T_{gm}}$-related EC (Fig. 3h) and the combined EC (Fig. 3k), but not for the ΔT_{gm}-related EC (Fig. 3e). We discussed this point in L210-215:

“Although all the global mean ECs do not largely change the 50th percentile values of ΔR_{gm}, the $\frac{\Delta R}{\Delta T_{gm}}$-related EC cause increases of ≥ 3 mm/day in the ITCZ of the Pacific Ocean and India and decreases in the subtropics of the Pacific Ocean, the Maritime Continent, the Indian Ocean and the Central Africa (Fig. 3h). These opposite changes cancel each other out in the global mean analysis (Fig. 2). ”

Minor comments：

Line 54: Should the cited reference here be [9] instead?

We cited Shiogama et al. (2024) in L51.

Lines 117–118: It might be helpful to visualize the joint distribution or correlation coefficient between the two variables to better illustrate their relationship.

We showed the joint distribution and correlation in Supplementary Fig. 2 (L135 and L155).

Reviewer #2 (Remarks on code availability):
Yes, the code provide a README file with enough instructions for installing and running the application.

Thank you.

Reply to Reviewer #1

The authors have done an excellent job of responding to my prior comments and adding more detail to the main text. I have nothing further to add prior to publication. I will just note a typo on L111, where "50 percentile values" should say "50th percentile values".

Thank you. We corrected it. (L111)

Reply to Reviewer #2

Thank you for your useful comments. We added some additional analyses.

Reviewer 1 raised a concern regarding the choice of the 1851–1900 baseline period. Although the authors have provided some justification, this reasoning primarily supports the ΔT-related EC approach and may not directly apply to the ΔR/ΔT-based method. So, it would be helpful to assess whether the results are sensitive to the choice of baseline years. For example, how would the results change if a more recent baseline period were used? I would be interested to see how the EC method performs when the constraint is weakened.

If we use the recent past period (1970-2022) as the baseline instead of the 1851-1900 period, RRVs become smaller. The RRVs of the ΔT_{gm}-related, $\frac{\Delta R_{gm}}{\Delta T_{gm}}$-related and combined ECs on the global mean ΔR_{gm} change from 26%, 27% and 42% to 23%, 25% and 41%, respectively (Fig. 2 and Supplementary Fig. 3e). The combined EC decreases the variance of regional ΔR by ≥30% in 12% of the global area for the 1970-2022 baseline case, while it is 24% for the 1851-1900 baseline case (Fig. 4f and Supplementary Fig. 6e). These decreases of RRVs are caused by the two factors: (1) the declines of aerosol emissions from the recent past period to the future period affect ΔT_{gm}, ΔR_{gm} and ΔR in the 1970-2022 baseline case; (2) the magnitudes of ΔT_{gm}, ΔR_{gm} and ΔR relative to the 1970-2022 mean are smaller than that relative to the 1851-1900 mean. Because trT_{gm} is the metric for ECs on future climate responses to increases in greenhouse gas concentrations but not for climate responses to changes in aerosol emissions[18,19,22], we select the 1851-1900 period as the baseline.
(L457-470)

In addition, it is interesting that precipitation values (Supplementary Fig. 1). This supports the use of 1997–2019 precipitation to constrain projections at the global mean scale. However, does this relationship also hold at the regional scale? The authors should consider providing a figure showing the regional correspondence between these two baseline periods.

We show that differences between the 1997-2019 mean regional R and the 1851-1900 mean one are small (Supplementary Figs. 1b-c and L182-183).

Reviewer #2 (Remarks on code availability):

I did not run the code, but it includes a README file and appears to contain the necessary scripts for reproducing the key calculations.

Thank you.