

Historical model biases in monthly high temperature anomalies indicate under-projection of future temperature extremes

Corresponding Author: Dr Lei Duan

This file contains all editorial decision letters in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Decision Letter:

**** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors ****

Dear Dr Duan,

Your manuscript titled "Historical model biases in monthly high temperature anomalies indicate under-projection of future temperature extremes" has now been seen by 2 reviewers, and we include their comments at the end of this message. They find your work of interest, but some important points are raised. We are interested in the possibility of publishing your study in Communications Earth & Environment, but would like to consider your responses to these concerns and assess a revised manuscript before we make a final decision on publication. Specifically, we ask you to:

1. Address the potential contradiction in your adjusted projections based on model underestimation to the previous 'hot model' view based on model underestimation.
2. Revise your interpretation of the bias corrected projections and also check the validity of your assumption about the stability of model biases in the future.

We therefore invite you to revise and resubmit your manuscript, along with a point-by-point response that takes into account the points raised. Please highlight all changes in the manuscript text file.

Please submit your point-by-point responses as a separate file, distinct from your cover letter where you can add responses to the Editors' comments that you do not want to be made available to the reviewers. Word files are preferred. We recommend that any figures, tables or graphs that are included in the response to reviewers are also included in the main article or Supplementary Information.

We are committed to providing a fair and constructive peer-review process. Please don't hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the referees' comments (which should be in a separate document to any cover letter), a tracked-changes version of the manuscript (as a PDF file) and the completed checklist:

Link Redacted

**** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first ****

We hope to receive your revised paper within six weeks; please let us know if you aren't able to submit it within this time so that we can discuss how best to proceed. If we don't hear from you, and the revision process takes significantly longer, we may close your file. In this event, we will still be happy to reconsider your paper at a later date, as long as nothing similar has been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

Please do not hesitate to contact us if you have any questions or would like to discuss these revisions further. We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Best regards,

Prof Seung-Ki Min
Editorial Board Member
Communications Earth & Environment
0000-0002-6749-010X

Aliénor Lavergne, PhD
Senior Editor, Communications Earth & Environment
Consulting Editor, Communications Sustainability
ORCID: 0000-0002-4591-1217

EDITORIAL POLICIES AND FORMATTING

We ask that you ensure your manuscript complies with our editorial policies. Please ensure that the following formatting requirements are met, and any checklist relevant to your research is completed and uploaded as a Related Manuscript file type with the revised article.

Editorial Policy: [Policy requirements](https://www.nature.com/documents/nr-editorial-policy-checklist.pdf) (Download the link to your computer as a PDF.)

For Manuscripts that fall into the following fields:

- Behavioural and social science
- Ecological, evolutionary & environmental sciences
- Life sciences

An updated and completed version of our Reporting Summary must be uploaded with the revised manuscript

You can download the form here:

<https://www.nature.com/documents/nr-reporting-summary.zip>

Furthermore, please align your manuscript with our format requirements, which are summarized on the following checklist: [Communications Earth & Environment formatting checklist](https://www.nature.com/documents/commsj-phys-style-formatting-checklist-article.pdf)

and also in our style and formatting guide [Communications Earth & Environment formatting guide](https://www.nature.com/documents/commsj-phys-style-formatting-guide-accept.pdf).

*** DATA: Communications Earth & Environment endorses the principles of the Enabling FAIR data project (<http://www.copdess.org/enabling-fair-data-project/>). We ask authors to make the data that support their conclusions available in permanent, publically accessible data repositories. (Please contact the editor if you are unable to make your data available).

All Communications Earth & Environment manuscripts must include a section titled "Data Availability" at the end of the Methods section or main text (if no Methods). More information on this policy, is available at <http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf>.

In particular, the Data availability statement should include:

- Unique identifiers (such as DOIs and hyperlinks for datasets in public repositories)
- Accession codes where appropriate
- If applicable, a statement regarding data available with restrictions
- If a dataset has a Digital Object Identifier (DOI) as its unique identifier, we strongly encourage including this in the Reference list and citing the dataset in the Data Availability Statement.

DATA SOURCES: All new data associated with the paper should be placed in a persistent repository where they can be freely and enduringly accessed. We recommend submitting the data to discipline-specific, community-recognized repositories, where possible and a list of recommended repositories is provided at <http://www.nature.com/sdata/policies/repositories>.

If a community resource is unavailable, data can be submitted to generalist repositories such as [figshare](https://figshare.com/) or [Dryad Digital Repository](http://datadryad.org/). Please provide a unique identifier for the data (for example a DOI or a permanent URL) in the data availability statement, if possible. If the repository does not provide identifiers, we encourage authors to supply the search terms that will return the data. For data that have been obtained from publically available sources, please provide a URL and the specific data product name in the data availability statement. Data with a DOI should be further cited in the methods reference section.

Please refer to our data policies at <http://www.nature.com/authors/policies/availability.html>.

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

This manuscript presents a valuable analysis of biases in CMIP6 climate models, specifically focusing on how models simulate high temperature anomalies relative to monthly mean temperatures, a topic that has been largely overlooked in the literature. The study highlights a substantial underestimation of extreme temperature anomalies across regions and seasons, suggesting potential biases in the projections of future extreme events. While the results are intriguing, they raise concerns about their consistency with existing literature. Specifically, various lines of evidence suggest that CMIP6 models are generally over-sensitive to external forcings, often resulting in more pronounced projections of warming and impacts. Historical simulations are better constrained by observations, whereas the influence of "hot models" becomes more evident as projections evolve, particularly under high emissions scenarios. It would be helpful if the authors addressed this contradiction and provided additional context or explanations for why their findings differ from this established view.

Extrapolating historical extreme temperature anomaly biases into the future relies on the evaluation that biases remain relatively stable between two close periods with similar external forcings, assuming that the near-invariance of these biases persists over time, regardless of changes in external forcings and emissions scenarios. To enhance the compelling nature of this study, it would be important to assess whether these biases remain stable under stronger external forcings. The authors could consider using simulations from each climate model as pseudo-observations for both past and future climates, thereby evaluating the robustness of this assumption. This approach would provide a more thorough evaluation of the reliability of the results and strengthen the conclusions drawn from bias extrapolation.

Reviewer #2 (Remarks to the Author):

GENERAL COMMENTS

The manuscript evaluates monthly maximum temperature extremes in 19 CMIP6 models, mainly using the multi-model mean results. The focus is on the maximum temperature anomalies, defined as the difference between the highest maximum temperature of the month and the monthly mean temperature. Compared to reanalysis data, the models show a slight tendency to underestimate the average magnitude of these anomalies, although this varies by region and from month to month. There is a larger and more systematic tendency to underestimate extreme temperature anomalies, defined for each calendar month as the difference between the highest maximum temperature in a 22-year period and the monthly mean temperature.

Similar biases in the maximum temperature anomalies are found in 1980-2001 and 2002-2023. This suggests that the highest maximum temperatures simulated by the models are affected by similar biases also in the end of the 21st century. Therefore, the authors devise a multiplicative bias adjustment for the difference between the monthly maximum and mean temperatures. These bias adjustments suggest that, in most areas, the highest temperatures (and their deviations from the monthly mean temperature) in the end of the century will be higher than the raw model output would suggest. In some cases, this difference reaches about 3 K.

It is well known in the climate science community that bias correction should nearly always be used when generating projections of future climate in absolute terms (e.g., temperatures in the late 21st century) rather than just in climate change terms (e.g., temperature differences between future and present). Still, this manuscript serves as a useful reminder of this. In addition, there is some novelty in decomposing the monthly maximum temperatures (TXx) to the monthly mean component and the anomaly component, as the latter has not been studied much explicitly. Furthermore, the manuscript is mostly well written (with a few exceptions noted in the detailed comments), and the figures are generally well prepared (again with some exceptions, see below). For these reasons, I judge it to be publishable in Environmental Research Communications after relatively minor revisions.

DETAILED COMMENTS

1. L75. The words "enhanced extreme warming" are misleading. Your results indicate that the highest temperatures in the future will be warmer than would be inferred directly from the simulation results, but they do not indicate that the warming from the present to the future climate would be significantly larger than the warming from the model-simulated present climate to the model-simulated future climate.
2. L82. increase over time?
3. L107. Why would drying of the surface reduce $\Delta(TXx)$? If the surface becomes dry enough to restrict evaporation, weaker evaporative damping should rather increase than reduce temperature variability.
4. L142-143. This is most likely due to the larger effect of internal variability on once-in-22-years than yearly extreme temperatures.
5. L167 and 169. "with additional warming" is misleading, because the future-minus-recent-past difference in the bias-corrected temperatures differs only slightly from that in the model-simulated temperatures (this slight difference only

occurring due to the multiplicative nature of the bias correction). It would be more correct to use "with an upward correction/adjustment" instead.

6. L192-194. Is this also true if the biases in monthly mean temperature are given as model-minus-reanalysis differences and not as ratios (which, in my opinion, make no sense for the mean temperature)?

7. L308. area-weighted

8. L333. What do you mean by "with a few registration"?

9. L492. Rates ... are averaged (or: Rate ... is averaged)

10. Figure 4. Please include additional panels for the smallest (most negative) bias adjustment across all 12 months, to avoid the impression of cherry-picking.

11. Figure S14. Please include additional panels for the smallest (most negative) bias adjustment across all 12 months.

12. Figure S15. Why do the values in E and F start from a higher level than those in the other panels? The description on L294-300 suggests that, for all cases, the beginning of the time series should represent models 1980-2001 minus reanalysis 1980-2001.

13. Caption of Figure S16, last three lines. A caveat in this apparent correlation is that the internal variability in the 1980-2001 values affects the 2079-2100 minus 1980-2001 differences in exactly the same way as the 2002-2023 minus 1980-2001 differences. To eliminate this caveat, it would be better to compare 2079-2100 with the whole 1980-2023 period.

14. Figure S17. It does not make sense to express the biases in monthly mean temperature on the ratio scale, because the result depends on the unit used (e.g., Celsius vs. Kelvin, although the actual values of ~0.5-2 seem inconsistent with both alternatives). Absolute differences (models minus reanalysis) would be much more informative.

15. Figure S19. Are the values on the percentage scale or on the ratio scale (on which, e.g., 0.4 means 40 %)?

** Visit Nature Research's author and referees' website at www.nature.com/authors for information about policies, services and author benefits**

Communications Earth & Environment is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the Manuscript Tracking System by clicking on 'Modify my Springer Nature account' and following the instructions in the link below. Please also inform all co-authors that they can add their ORCIDs to their accounts and that they must do so prior to acceptance.

<https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research>

For more information please visit <http://www.springernature.com/orcid>

If you experience problems in linking your ORCID, please contact the Platform Support Helpdesk.

Version 1:

Decision Letter:

** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors **

Dear Dr Duan,

Your manuscript titled "Historical model biases in monthly high temperature anomalies indicate under-projection of future temperature extremes" has now been seen by our reviewers, whose comments appear below. In light of their advice we are delighted to say that we are happy, in principle, to publish a suitably revised version in Communications Earth & Environment.

We therefore invite you to revise your paper one last time to address the remaining concerns of our reviewers. At the same time we ask that you edit your manuscript to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table".

*****Please take care to match our formatting and policy requirements. We will check revised manuscript and return manuscripts that do not comply. Such requests will lead to delays. *****

Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed at the end of the Editorial Requests Table; the list of required files is also available at <https://www.nature.com/documents/commsj-file-checklist.pdf>.

OPEN ACCESS:

Communications Earth & Environment is a fully open access journal. Articles are made freely accessible on publication. For further information about article processing charges, open access funding, and advice and support from Nature Research, please visit <https://www.nature.com/commsenv/open-access>

At acceptance, you will be provided with instructions for completing the open access licence agreement on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

Please use the following link to submit the above items:

Link Redacted

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Prof Seung-Ki Min
Editorial Board Member
Communications Earth & Environment
0000-0002-6749-010X

Alienor Lavergne, PhD
Senior Editor, Communications Earth & Environment
Consulting Editor, Communications Sustainability
ORCID: 0000-0002-4591-1217

Springer Nature
The Campus, 4 Crinan Street, London N1 9XW, UK
www.nature.com/commsenv
[@commsearth.bsky.social](https://twitter.com/commsearth.bsky.social)

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

I appreciate the authors for adequately addressing my concerns and comments. I have no further questions regarding the revised version.

Reviewer #2 (Remarks to the Author):

General comments

The revisions made to the manuscript address my previous comments in a satisfactory manner, with one slight exception (Comment 1 below). My other comments are purely editorial.

Detailed comments

1. P7, L1-3. You seem to have misunderstood my earlier comment. What I meant was that the larger inter-period differences

and weaker inter-period correlation for the 22-year maximum metrics, compared with the yearly maximum metrics, is probably due to internal variability. However, as internal variability is a stochastic process, I don't see how it could explain the systematically larger negative bias in the very extreme tail (22-year maximum anomaly) of the distribution. This systematically larger underestimation most probably has some other cause. It suggests shorter extreme tails in the temperature anomaly distribution in the CMIP6 models than in the reanalyses, even though the underlying mechanism is not obvious.

2. P9, paragraph 2, L8. ... sensitivity of regional climate extremes?
3. P12, L5 from bottom. Replace "to be" with "being".
4. P15, Data availability, L6. The same link as for CMIP6 is given for ERA5. Please correct.
5. P15, Data availability, L7. MERRA2
6. Supplementary Text S1, L2. reliability of models?
7. Figure S2A. 5th to 95th (not 90th) percentile?
8. Caption of Figure S24, L5. random separations?
9. Caption of Table S1. Please specify the meaning of the values given in parentheses.

** Visit Nature Research's author and referees' website at <http://www.nature.com/authors> for information about policies, services and author benefits**

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Definitions of context type in this response:

1. Reviewer comments
2. [Author response](#)
3. *[“Context from the text”](#)*

Reviewer #1:

This manuscript presents a valuable analysis of biases in CMIP6 climate models, specifically focusing on how models simulate high temperature anomalies relative to monthly mean temperatures, a topic that has been largely overlooked in the literature. The study highlights a substantial underestimation of extreme temperature anomalies across regions and seasons, suggesting potential biases in the projections of future extreme events. While the results are intriguing, they raise concerns about their consistency with existing literature. Specifically, various lines of evidence suggest that CMIP6 models are generally over-sensitive to external forcings, often resulting in more pronounced projections of warming and impacts. Historical simulations are better constrained by observations, whereas the influence of "hot models" becomes more evident as projections evolve, particularly under high emissions scenarios. It would be helpful if the authors addressed this contradiction and provided additional context or explanations for why their findings differ from this established view.

We thank the reviewer for raising this question so that we can add text to clarify this point in the revised manuscript.

Many studies have suggested an unrealistically high climate sensitivity in the latest generation of models participated in CMIP6, e.g., *Zelinka et al. (2019)* and *Zhu et al. (2020)*, due to processes such as a stronger positive cloud feedbacks. However, these findings are mostly associated with the mean climate condition, such as the response of global annual mean temperature, and do not necessarily hold for the regional extreme metrics. For example, many previous studies (e.g., *Kim et al. (2020)*, *Fan et al. (2020)*, *Luca et al. (2020)*, and *Thorarinsdottir et al. (2020)*), by comparing CMIP5 and CMIP6 models with observations and reanalysis datasets, have also found that both the absolute value and biases of extreme indices are largely comparable in these two generations of models, with CMIP6 models generally showing better performance. Results from *Seneviratne and Hauser (2020)* have shown that in contrast to global climate sensitivity, the regional climate sensitivity in climate extremes (including yearly hot days and other extreme indices) is very similar in CMIP5 and CMIP6 ensembles, which “highlights regional climate sensitivity as a distinct feature compared with global climate sensitivity”.

In fact, our results are found to be consistent with results of *Freychet et al. (2021)*, which also found that CMIP6 models have underestimated future extreme hot days based on the relationship of daily mean temperature and daily hot temperature threshold. Therefore, our study does not contradict previous work. To clearly reflect this point, we have added a new paragraph of discussion in the revised manuscript that reads:

“Studies have suggested that some of the latest generation of models participating in CMIP6 have an unrealistically high climate sensitivity^{1,2}. However, these findings are mostly associated with the global mean metrics, such as the response of global annual mean temperature to a doubling of CO₂ concentration, and do not necessarily hold for extreme metrics at regional scales. Studies^{3–6} have shown that both the absolute value and biases of extreme indices are similar in CMIP5 and CMIP6, with CMIP6 models generally showing better performances relative to observations and reanalysis datasets. A recent study found that in contrast to the global climate sensitivity, sensitivities of regional climate extreme (including yearly hot days and other extreme indices) is very similar in these two generations of models, highlighting regional climate sensitivity as a distinct feature compared to global climate sensitivity⁷. Our results are consistent with work showing an under-estimation of future extreme hot days in CMIP6 models based on the relationship of daily mean temperature and daily hot temperature thresholds⁸.”

Not all regions and months in our study exhibit a positive bias adjustment. For example, we have revised sentences that read:

“Ratios for the 22-year average ΔTXx range between 0.7 and 1.2 for different regions-month combinations, with 64% of region-month combinations exhibiting an underestimation of the ΔTXx (i.e., ratios < 1) and others exhibiting an overestimation of the ΔTXx (i.e., ratios > 1). Ratios for the 22-year maximum ΔTXx range between 0.6 and 1.1, with ~ 87% of region-month combinations exhibiting an underestimation of the ΔTXx and others exhibiting an overestimation of the ΔTXx .”

Consideration of differences between model and reanalysis results could lead to both positive and negative bias adjustments. In the revised manuscript, we have revised sentences to highlight this, and we have also revised our figures to show results for the warmest vs. coolest months, and maximum vs. minimum bias adjustment across different regions for results under both SSP245 and SSP585 scenarios:

“We use the 22-year average ΔTXx to estimate what the typical high temperature anomaly would be over the 2079 to 2100 period (the last 22 years of this century) after considering historical biases. Under the SSP245 scenario, for months with the warmest monthly mean temperature during future periods, bias adjustments vary among regions (Fig. 4A, S12, and S13), from more than 1 K cooler than model results in low latitude regions such as West Asia and South America to about 2 K warmer than model results in high latitude regions such as Russian-Arctic. For months with the coolest monthly mean temperature, bias adjustments for the 22-year average ΔTXx vary from 1.6 K cooler to 2.6 K warmer than model results (Fig. 4B). Warming biases are more substantial for the 22-year maximum ΔTXx , i.e., the models have less skill at predicting rare, very extreme, events than they do at predicting more common, somewhat extreme, events. For months with the warmest monthly mean temperature

(Fig. 4C), thirty-three regions have adjusted 22-year high temperatures larger than the raw model outputs, with bias adjustment reaching 3 K in North Europe and Russian-Arctic, and 2 K in North America and New Zealand. For months with the coolest monthly mean temperature, bias adjustments for the 22-year maximum ΔTX_x are positive for almost all regions (Fig. 4D). The range of bias adjustment for different regions is shown in Fig. S14. For the 22-year average ΔTX_x , the largest bias adjustments across region-month combinations exceed 3 K and the smallest bias adjustments reach about -2 K; for the 22-year maximum ΔTX_x , the largest bias adjustments across region-month combinations are more than 5 K relative to raw model outputs for five regions. Similar patterns can be seen under the SSP585 scenario (Fig. S15 and S16)."

Extrapolating historical extreme temperature anomaly biases into the future relies on the evaluation that biases remain relatively stable between two close periods with similar external forcings, assuming that the near-invariance of these biases persists over time, regardless of changes in external forcings and emissions scenarios. To enhance the compelling nature of this study, it would be important to assess whether these biases remain stable under stronger external forcings. The authors could consider using simulations from each climate model as pseudo-observations for both past and future climates, thereby evaluating the robustness of this assumption. This approach would provide a more thorough evaluation of the reliability of the results and strengthen the conclusions drawn from bias extrapolation.

This comment by the reviewer was very thoughtful and creative, and helped our manuscript substantially. To address the reviewer's concern, we have separated models used in this analysis into two groups. The ensemble means of one group are used as reference cases (i.e., the assumed reality) and the ensemble means of the other group are used as testing cases. Similar to processes shown in the main text, we have applied the bootstrapping approach to estimate the on-average biases between these two groups during the historical periods (e.g., 1980 to 2023), and then adjust results from the testing group for year 2079 to 2100. Comparisons of adjusted values from the testing group with raw model outputs from the reference group suggest that the adjusted monthly high temperature anomalies agree well with each other. Therefore, using model outputs alone confirm the robustness of our conclusions. In the revised text, we have added corresponding discussions in Methods that read:

"Extrapolating biases of the historical extreme high temperature anomaly into the future relies on assumptions that similar biases persist over time, regardless of changes in external forcings or emissions scenarios. To further test the robustness of our conclusions, we randomly separated models used here into two groups, one acting as the reference condition (i.e., the reality) and the other group acting as the testing cases. We have then performed the same adjusting processes described above to these two groups. Comparisons of the adjusted and unadjusted results from the testing group with raw

model outputs from the reference group suggest that adjusting monthly high temperature anomalies reduces the root mean square error from 1.05 K to 0.20 K for the 22-year average and from 1.13 K to 0.36 K for the 22-year maximum ΔTX_x (Fig. S24 and S25). The success of the methods applied here in improving prediction results of models for the end of the century adds confidence to the robustness of our conclusions.”

And we have also added a sentence in the main paper to highlight this:

“Tests of this method using climate model outputs alone demonstrate substantial improved predictive skill (See Methods).”

Reviewer #2:**GENERAL COMMENTS**

The manuscript evaluates monthly maximum temperature extremes in 19 CMIP6 models, mainly using the multi-model mean results. The focus is on the maximum temperature anomalies, defined as the difference between the highest maximum temperature of the month and the monthly mean temperature. Compared to reanalysis data, the models show a slight tendency to underestimate the average magnitude of these anomalies, although this varies by region and from month to month. There is a larger and more systematic tendency to underestimate extreme temperature anomalies, defined for each calendar month as the difference between the highest maximum temperature in a 22-year period and the monthly mean temperature.

Similar biases in the maximum temperature anomalies are found in 1980-2001 and 2002-2023. This suggests that the highest maximum temperatures simulated by the models are affected by similar biases also in the end of the 21st century. Therefore, the authors devise a multiplicative bias adjustment for the difference between the monthly maximum and mean temperatures. These bias adjustments suggest that, in most areas, the highest temperatures (and their deviations from the monthly mean temperature) in the end of the century will be higher than the raw model output would suggest. In some cases, this difference reaches about 3 K.

It is well known in the climate science community that bias correction should nearly always be used when generating projections of future climate in absolute terms (e.g., temperatures in the late 21st century) rather than just in climate change terms (e.g., temperature differences between future and present). Still, this manuscript serves as a useful reminder of this. In addition, there is some novelty in decomposing the monthly maximum temperatures (TXx) to the monthly mean component and the anomaly component, as the latter has not been studied much explicitly. Furthermore, the manuscript is mostly well written (with a few exceptions noted in the detailed comments), and the figures are generally well prepared (again with some exceptions, see below). For these reasons, I judge it to be publishable in Environmental Research Communications after relatively minor revisions.

We thank the reviewer for positive comments of our results. We have addressed the reviewer's comments point-by-point below.

DETAILED COMMENTS

1. L75. The words “enhanced extreme warming” are misleading. Your results indicate that the highest temperatures in the future will be warmer than would be inferred directly from the simulation results, but

they do not indicate that the warming from the present to the future climate would be significantly larger than the warming from the model-simulated present climate to the model-simulated future climate.

We thank the reviewer for this thoughtful comment. In the revised manuscript, we have changed this sentence to:

“When historical biases are used to constrain future model projections of temperature anomalies, we find both increased and decreased extreme high temperature across regions and months.”

We have also gone through the draft and corrected any potential misleading sentences.

2. L82. increase over time?

Typo. We have changed the word in this sentence as suggested by the reviewer.

3. L107. Why would drying of the surface reduce delta (TXx)? If the surface becomes dry enough to restrict evaporation, weaker evaporative damping should rather increase than reduce temperature variability.

Yes, as the reviewer has said, drying of the surface will increase temperature variability and thus the high temperature anomaly. This is consistent with what we have in the paper, where we have shown that high latitude regions are featured with a decreasing trend of high temperature anomaly due to the loss of sea ice and snow, and low latitude regions are featured with an increasing trend of high temperature anomaly due possibility to the surface drying. We have revised the sentences to clarify:

“Regions located in low and middle latitudes, such as South Africa, Australia, and South America, tend to exhibit an increasing trend. The increased high temperature anomaly in these regions is likely associated with transitions of the surface condition (e.g., drier surface under warming) and shifts in atmospheric structure^{9,10}.”

4. L142-143. This is most likely due to the larger effect of internal variability on once-in-22-years than yearly extreme temperatures.

Here, we want to highlight the distribution of ratio for both 22-year average and 22-year maximum high temperature anomaly stays very similar during the historical periods. We have revised this sentence which now reads:

“The distribution of ratios for both 22-year average and 22-year maximum ΔTXx across regions and months is very similar over the two historical periods (Fig. S9 and S10) indicating the stability of these biases over time.”

We agree with the reviewer in explaining the larger underestimation of the 22-year maximum metrics, and we have added one sentence in the revised manuscript that reads:

“The larger underestimation of the 22-year maximum metric is likely due to a bigger impact of internal variability on once-in-22-year temperatures than yearly extreme temperatures.”

5. L167 and 169. "with additional warming" is misleading, because the future-minus-recent-past difference in the bias-corrected temperatures differs only slightly from that in the model-simulated temperatures (this slight difference only occurring due to the multiplicative nature of the bias correction). It would be more correct to use "with an upward correction/adjustment" instead.

We thank reviewer for pointing this out. We have revised this paragraph to avoid potential confusions:

“Under the SSP245 scenario, for months with the warmest monthly mean temperature during future periods, bias adjustments vary among regions (Fig. 4A, S12, and S13), from more than 1 K cooler than model results in low latitude regions such as West Asia and South America to about 2 K warmer than model results in high latitude regions such as Russian-Arctic. For months with the coolest monthly mean temperature, bias adjustments for the 22-year average ΔTX_x vary from 1.6 K cooler to 2.6 K warmer than model results (Fig. 4B). Warming biases are more substantial for the 22-year maximum ΔTX_x , i.e., the models have less skill at predicting rare, very extreme, events than they do at predicting more common, somewhat extreme, events. For months with the warmest monthly mean temperature (Fig. 4C), thirty-three regions have adjusted 22-year high temperatures larger than the raw model outputs, with bias adjustment reaching 3 K in North Europe and Russian-Arctic, and 2 K in North America and New Zealand. For months with the coolest monthly mean temperature, bias adjustments for the 22-year maximum ΔTX_x are positive for almost all regions (Fig. 4D). The range of bias adjustment for different regions is shown in Fig. S14. For the 22-year average ΔTX_x , the largest bias adjustments across region-month combinations exceed 3 K and the smallest bias adjustments reach about -2 K; for the 22-year maximum ΔTX_x , the largest bias adjustments across region-month combinations are more than 5 K relative to raw model outputs for five regions. Similar patterns can be seen under the SSP585 scenario (Fig. S15 and S16).”

6. L192-194. Is this also true if the biases in monthly mean temperature are given as model-minus-reanalysis differences and not as ratios (which, in my opinion, make no sense for the mean temperature)?

The reviewer raised a great point that it might not be appropriate to directly apply the same ratio metric to monthly mean temperature. In the revised manuscript, we have examined bias defined as difference between models and reanalysis datasets for monthly mean temperature and monthly high temperature anomaly. We have found that biases of the anomaly component shows a better proportional relationship

with the baseline condition, whereas biases of the monthly mean temperature follow an additive relationship with the baseline condition. Such findings further highlight the importance of assessing the anomaly components and mean state separately when examining model simulated extreme temperature events. In the revised manuscript, we have added a new figure and revised corresponding sentences to reflect this point:

“In contrast to ΔTX_x , where biases of CMIP6 models exhibit a better proportional relationship with the baseline condition, biases of the monthly mean temperature (TMm) show a better additive relationship with the baseline condition (Fig. S19). Our results thus underscore the importance of independently assessing both anomalies and the mean states when evaluating model simulated extreme temperature events.”

In **Methods**, we have added a paragraph that reads:

“Using model-reanalysis ratio to quantify biases lead to a larger correlation coefficient between the two different historical periods compared to that defined as model-reanalysis difference, especially for the 22-year maximum metric (Fig. S23). Using biases defined as model-reanalysis difference to constrain model projections leads to similar or even larger magnitudes of future monthly high temperature anomalies. For example, for months with the warmest monthly mean temperature (averaged between 2079 and 2100) under the SSP245 scenario, bias adjustments for the 22-year average ΔTX_x vary from -1.5 K to 1.9 K, with the maximum values among region-month combinations to be ~ 3.1 K; the maximum region-month combination of bias adjustment reaches over 7 K for the 22-year maximum ΔTX_x .”

7. L308. area-weighted

Typo. We have corrected this in the revised manuscript.

8. L333. What do you mean by "with a few registration"?

We have corrected this sentence in the revised manuscript:

“An account is required to download the data, which can be obtained with free registration.”

9. L492. Rates ... are averaged (or: Rate ... is averaged)

Typo. We have corrected this in the revised manuscript.

10. Figure 4. Please include additional panels for the smallest (most negative) bias adjustment across all 12 months, to avoid the impression of cherry-picking.

As suggested, we have revised Figure 4 to show bias adjustments for the warmest and coolest months. Similarly, we have revised Figure S14 to show the maximum and minimum bias adjustments across all 12 months. Corresponding text is updated that now read:

“We use the 22-year average ΔTXx to estimate what the typical high temperature anomaly would be over the 2079 to 2100 period (the last 22 years of this century) after considering historical biases. Under the SSP245 scenario, for months with the warmest monthly mean temperature during future periods, bias adjustments vary among regions (Fig. 4A, S12, and S13), from more than 1 K cooler than model results in low latitude regions such as West Asia and South America to about 2 K warmer than model results in high latitude regions such as Russian-Arctic. For months with the coolest monthly mean temperature, bias adjustments for the 22-year average ΔTXx vary from 1.6 K cooler to 2.6 K warmer than model results (Fig. 4B). Warming biases are more substantial for the 22-year maximum ΔTXx , i.e., the models have less skill at predicting rare, very extreme, events than they do at predicting more common, somewhat extreme, events. For months with the warmest monthly mean temperature (Fig. 4C), thirty-three regions have adjusted 22-year high temperatures larger than the raw model outputs, with bias adjustment reaching 3 K in North Europe and Russian-Arctic, and 2 K in North America and New Zealand. For months with the coolest monthly mean temperature, bias adjustments for the 22-year maximum ΔTXx are positive for almost all regions (Fig. 4D). The range of bias adjustment for different regions is shown in Fig. S14. For the 22-year average ΔTXx , the largest bias adjustments across region-month combinations exceed 3 K and the smallest bias adjustments reach about -2 K; for the 22-year maximum ΔTXx , the largest bias adjustments across region-month combinations are more than 5 K relative to raw model outputs for five regions.”

11. Figure S14. Please include additional panels for the smallest (most negative) bias adjustment across all 12 months.

We have added new figures in SI to show results under the SSP585 scenario:

“Similar patterns can be seen under the SSP585 scenario (Fig. S15 and S16).”

12. Figure S15. Why do the values in E and F start from a higher level than those in the other panels? The description on L294-300 suggests that, for all cases, the beginning of the time series should represent models 1980-2001 minus reanalysis 1980-2001.

Thanks for spotting this. This is actually a bug in the code where the wrong model periods are used for panel E and F. In the revised manuscript, we have corrected this figure.

13. Caption of Figure S16, last three lines. A caveat in this apparent correlation is that the internal variability in the 1980-2001 values affects the 2079-2100 minus 1980-2001 differences in exactly the

same way as the 2002-2023 minus 1980-2001 differences. To eliminate this caveat, it would be better to compare 2079-2100 with the whole 1980-2023 period.

In response to the reviewer's comment, we have compared the end-of-century results with the whole historical periods (i.e., 1980 to 2023) and updated corresponding figures in the revised manuscript. The revised sentence now reads:

“ ΔTX_x projected by the models for the 2079-2100 period under the same SSP245 scenario range from 78% to 111% of that projected for the whole 1980-2023 period.”

14. Figure S17. It does not make sense to express the biases in monthly mean temperature on the ratio scale, because the result depends on the unit used (e.g., Celsius vs. Kelvin, although the actual values of ~0.5-2 seem inconsistent with both alternatives). Absolute differences (models minus reanalysis) would be much more informative.

We agree with the reviewer that express biases in monthly mean temperature on the ratio scale might not be appropriate. We have corrected this issue in the discussions above.

15. Figure S19. Are the values on the percentage scale or on the ratio scale (on which, e.g., 0.4 means 40 %)?

Thanks for the reviewer. We have now corrected the units for this figure now.

Reference

1. Zelinka, M. D. *et al.* Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.* **47**, 1029 (2020).
2. Zhu, J., Poulsen, C. J. & Otto-Bliesner, B. L. High climate sensitivity in CMIP6 model not supported by paleoclimate. *Nat. Clim. Chang.* **10**, 378–379 (2020).
3. Di Luca, A., Pitman, A. J. & de Elía, R. Decomposing temperature extremes errors in CMIP5 and CMIP6 models. *Geophys. Res. Lett.* **47**, (2020).
4. Fan, X., Miao, C., Duan, Q., Shen, C. & Wu, Y. The performance of CMIP6 versus CMIP5 in simulating temperature extremes over the global land surface. *J. Geophys. Res.* **125**, (2020).
5. Kim, Y.-H., Min, S.-K., Zhang, X., Sillmann, J. & Sandstad, M. Evaluation of the CMIP6 multi-model ensemble for climate extreme indices. *Weather and Climate Extremes* **29**, 100269 (2020).
6. Thorarinsdottir, T. L., Sillmann, J., Haugen, M., Gissibl, N. & Sandstad, M. Evaluation of CMIP5 and CMIP6 simulations of historical surface air temperature extremes using proper evaluation methods. *Environ. Res. Lett.* **15**, 124041 (2020).
7. Seneviratne, S. I. & Hauser, M. Regional Climate Sensitivity of Climate Extremes in CMIP6 Versus CMIP5 Multimodel Ensembles. *Earths Future* **8**, e2019EF001474 (2020).
8. Freychet, N., Hegerl, G., Mitchell, D. & Collins, M. Future changes in the frequency of temperature extremes may be underestimated in tropical and subtropical regions. *Commun. Earth Environ.* **2**, (2021).
9. Bathiany, S., Dakos, V., Scheffer, M. & Lenton, T. M. Climate models predict increasing temperature variability in poor countries. *Sci Adv* **4**, eaar5809 (2018).

10. Olonscheck, D., Schurer, A. P., Lücke, L. & Hegerl, G. C. Large-scale emergence of regional changes in year-to-year temperature variability by the end of the 21st century. *Nat. Commun.* **12**, 7237 (2021).

Definitions of context type in this response:

1. Reviewer comments
2. [Author response](#)
3. *[“Content from the text”](#)*

Reviewer #1:

I appreciate the authors for adequately addressing my concerns and comments. I have no further questions regarding the revised version.

We thank the reviewer for approving our revisions.

Reviewer #2:

GENERAL COMMENTS

The revisions made to the manuscript address my previous comments in a satisfactory manner, with one slight exception (Comment 1 below). My other comments are purely editorial.

We thank the reviewer for clarifications. We agree with the reviewer's point which we did not fully appreciate earlier. We have further revised the manuscript below.

DETAILED COMMENTS

1. P7, L1-3. You seem to have misunderstood my earlier comment. What I meant was that the larger inter-period differences and weaker inter-period correlation for the 22-year maximum metrics, compared with the yearly maximum metrics, is probably due to internal variability. However, as internal variability is a stochastic process, I don't see how it could explain the systematically larger negative bias in the very extreme tail (22-year maximum anomaly) of the distribution. This systematically larger underestimation most probably has some other cause. It suggests shorter extreme tails in the temperature anomaly distribution in the CMIP6 models than in the reanalyses, even though the underlying mechanism is not obvious.

We agree with the reviewer that the weaker inter-period correlation for the 22-year maximum metric is mostly due to internal variability instead of external forcings. We have made this clear in the revised paper by modifying the following sentence:

“The smaller correlation coefficient estimated for the 22-year maximum metric is primarily due to internal variability, which also contributes to inter-period differences.”

We also agree with the reviewer that “the shorter extreme tails in the temperature anomaly distribution in the CMIP6 models than in the reanalyses”. Our conjecture is that, because the reanalysis model is updated with real-world weather data, it has information relevant to finer temporal and spatial scales than do the CMIP6 models, which are limited by their time steps and grid scale. The introduction of information from finer temporal and spatial scales results in more variability in reanalysis results. We have added sentences in Discussions that reads:

“Our results suggest that, globally, models tend to underestimate the monthly high temperature anomaly observed in reanalysis datasets. While we do not investigate mechanisms that cause such bias, the underestimation of temperature variability in the climate models relative to ERA5, especially for the ratio of 22-year maximum ΔTX_x , could perhaps be a consequence of the ingestion of weather data by ERA5⁴²,

which brings in information from finer temporal and spatial scales not available to the climate models.”

2. P9, paragraph 2, L8. ... sensitivity of regional climate extremes?

Thanks. We have corrected this typo.

3. P12, L5 from bottom. Replace "to be" with "being".

We have changed this word as the reviewer suggested.

4. P15, Data availability, L6. The same link as for CMIP6 is given for ERA5. Please correct.

Thanks for noting this. We have corrected the link now.

5. P15, Data availability, L7. MERRA2

We have corrected this typo.

6. Supplementary Text S1, L2. reliability of models?

We have changed this word as the reviewer suggested.

7. Figure S2A. 5th to 95th (not 90th) percentile?

We have corrected the figure label.

8. Caption of Figure S24, L5. random separations?

We have corrected the word.

9. Caption of Table S1. Please specify the meaning of the values given in parentheses.

We have added explanations for numbers in parentheses.