nature portfolio

Peer Review File

Worst-Case European Heat Storylines generated using Ensemble Boosting

Corresponding Author: Dr Laura Suarez-Gutierrez

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

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 Suarez-Gutierrez,

Your manuscript titled "Worst-Case European Heat and Drought Storylines generated using Ensemble Boosting" has now been seen by 3 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.

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,

Alice Drinkwater, PhD
Associate Editor
Communications Earth & Environment
Consulting Editor
Communications Sustainability

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.

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.pdf

Furthermore, please align your manuscript with our format requirements, which are summarized on the following checklist: Communications Earth & Environment formatting checklist

and also in our style and formatting guide Communications Earth & Environment formatting guide .

*** 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 or Dryad Digital Repository. 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):

Dear authors,

I've genuinely enjoyed reading your manuscript and have only minor comments. The figures are beautiful and analyses are convincing. Here are a few remarks & questions in no particular order:

1) I'm a bit confused as to why P4 is considered to showcase 'extreme heat' (p. 7 onwards, e.g., table 1); in Fig. 1, P3 is more extreme with regards to both heatwave area and intensity. In Fig. 2, the peak of P4 looks slightly more extreme than P3, but even there, I do not see a sufficiently clear distinction. Therefore, also because I don't think any conclusions depend on this choice, it might make more sense to only consider P1 & P5 as parent heatwaves with extreme heat.

2) Fig. 1: You could consider showing what the grey dots and black circles with numbers correspond to in the legend, perhaps splitting between a) and b) to make room for the additional legend entries.

3) 100-member CESM2-LE from NCAR vs own CESM2 simulations with 35 ensemble members: since you use both, it is a bit confusing that "CESM2-LE" seems to refer to both ensembles. Clearly distinguishing throughout the paper would help.

4) p. 13, "In sum, the worst-case ... the most extreme heatwaves plausible today": it is still only plausible in the model world (specifically the CESM2 world), as explained in the introduction -- you have no means of asserting that your potential tail overestimation does not fundamentally bias this assessment. Considering the overall length of the paper, I'm a bit surprised that you never highlight this caveat again after the introduction. Similarly, on p. 15, "..., successive heatwaves of the magnitude presented in these boosted storylines would indeed be unprecedented": Well, yes, but isn't this more of a demonstration what 'hunting' for the most extreme heatwaves can reveal rather than an actionable statement linking to climate change? Since your approach has no probabilistic interpretation, and without showing just how bad heatwaves could become even in the absence of human-induced climate change, one could argue that there is an extremely tiny (but non-zero) probability of something similar occurring entirely "naturally". And, as a consequence, you cannot be sure that such an event is truly "unprecedented" -- it might not be on the historical record, but that is a different statement.

5) p.14, 18.2% successive heatwaves in ERA5: isn't this just 2 out of 11 samples? I don't think it makes sense to provide this level of accuracy and conclude that "CESM2-LE" and ERA5 show "remarkably good agreement"; you can, however, assert that based on the limited sample, the behaviour seems at least broadly consistent.

6) Fig. 5, specifically 5d: I find the map inlet more confusing than helpful because the colors are not sufficiently distinct from the parent heatwaves. Why not show 3 inlet maps for a or b, c or d, and e or f, showing only the relevant domain?

7) The text after the initial description of Fig. 6 on p.18 & 19 feels rather lengthy and could (partly) be moved to the Supplementary. At this point, the reader has been confronted with a lot of information already, and a condensed version would probably not take anything away from the main story.

8) p.23, "particularly the first heatwave in July 2019", is this a typo? According to Sousa et al. (2020), there was a June and a July heatwave in that region during 2019, with the latter potentially being enhanced due to soil moisture depletion during the former.

9) Finally, though the text is generally pleasant to read, there seem to be a few typos (e.g., "stotylines"), sometimes words are missing ("We also how comparisons..."), "The/These findings highlight the risk of a potentially devastating high-impact hazard" seems to be featured twice (not sure if that is intentional), but I won't list more here and instead recommend running a spellcheck.

References:
Sousa, P.M., Barriopedro, D., García-Herrera, R. et al. Distinct influences of large-scale circulation and regional feedbacks in two exceptional 2019 European heatwaves. Commun Earth Environ 1, 48 (2020). https://doi.org/10.1038/s43247-020-00048-9

Reviewer #2 (Remarks to the Author):

Manuscript Title: Worst-Case European Heat and Drought Storylines generated using Ensemble Boosting
Overall Assessment:
This is a highly ambitious and timely study that successfully leverages the ensemble boosting methodology to generate plausible worst-case heat and drought storylines for Europe. The core finding—that the most extreme heatwaves in the current climate are not only possible but also highly likely to occur successively—is significant and has important implications for climate risk assessment. The experimental design is sophisticated, and the manuscript is generally well-structured.
While the paper excels at risk discovery, its mechanistic analysis aimed at explaining the drivers behind these successive extremes suffers from several methodological and interpretive shortcomings.

Major Points of Concern:
1. One hypothesis is that a first heatwave "primes" the system for a more extreme second event via soil moisture depletion and ocean heat accumulation. However, the study fails to present the most direct evidence: the evolution of key variables (e.g., soil moisture, SSTs) during the inter-period between the first and second heatwaves. Without demonstrating this temporal chain of events, the claim of causation is weak. The observed succession could merely be a statistical byproduct of selecting extremely hot summers where exceedances of a fixed threshold are naturally more frequent and clustered. Analyzing this inter-period evolution can also help to explain the study's own finding that the most persistent successive heatwaves last longer than their associated atmospheric blocking events (Fig. 5a). It is possible that an initially dynamic-driven heatwave rapidly depletes soil moisture, and this newly established dry state then becomes the dominant driver, maintaining extreme temperatures even as the blocking pattern weakens.

2. The composite analyses (e.g., Fig. 6) comparing conditions during successive and non-successive heatwaves are potentially confounded. According to SI Figure S2-S5, the parent events that produce the successive heatwaves (P4, P5, P6, P7) are inherently different from those that produce most non-successive ones (P1, P2, P4, P6, P7). Therefore, differences in composites (e.g., higher SSTs in successive events) can be driven by the inherent properties of the parent cases rather than by the "successive" attribute itself. For example, when comparing successive vs. non-successive events stemming from the same parent, like P4, warm SST anomalies are observed in both types (Figure S5).
3. The use of different methods to calculate anomalies for different variables (T, SM, SST, etc) creates an inconsistency. This makes it difficult to compare the "extremeness" across variables and might add noise and obscure the physical relationships the study seeks to uncover.
4. The use of total column soil moisture (integrated over 100 cm) is appropriate for assessing ecological drought but might not be the optimal metric for controlling land-atmosphere feedbacks during heatwaves. The partitioning of sensible and latent heat fluxes that governs near-surface temperature maxima is primarily controlled by moisture availability in the top few centimeters of soil. Using a deep-layer integral dilutes the strong signal of rapid surface drying that is likely the key mediator of the "priming effect" and heatwave intensification.

Reviewer #3 (Remarks to the Author):

This manuscript leverages an ensemble boosting technique to produce a very large ensemble of climate simulations for characterising European heat and drought extremes. This research provides a remarkable plethora of relevant results regarding the driving mechanisms of European heat and drought extremes, focusing on both large- and local-scale forcing dynamics.

Overall, while the manuscript provides a wealth of remarkable findings, the current length and level of detail risk diluting the impact of its key messages. This specifically refers to the Results section.

Besides this general comment, specifically referring to the first results section involving the CESM-LE evaluation. Driving HW mechanisms explored in the manuscript are essentially represented by 500 hPa geopotential anomalies in terms of blocking, soil moisture (SM) deficit and SST anomalies spatial distribution. Although the surprisingly good capability of CESM-LE in reproducing heat and drought statistics/metrics, it would strengthen the analysis to include a brief discussion of CESM-LE’s ability to reproduce SST anomaly variability and basic blocking statistics during the evaluation period. This would reinforce confidence that the good performance in excess-heat metrics is indeed driven by the mechanisms under investigation. To maintain a balanced section length, I would rather condense some of the current evaluation findings and include some evidence/consideration (also retrieved from published studies to limit the already vast number of analyses) regarding CESM-LE’s representation of basic blocking statistics, SST and SM anomaly variability. The latter point is especially important for SM, since at coarse GCM resolution (the resolution is not reported in the Methods section and should be clarified), models often struggle to capture SM variability and its feedbacks with the planetary boundary layer.

In the section “In the Worst-case ensemble boosting storylines for extreme heatwaves”, please elaborate a bit on what is possibly behind the mechanism reported in the following statement: “indicating that boosted simulations remain more similar to each other than an independent sample and exhibit above-average heat and drought conditions throughout the summer” This is relevant.
Regarding the “Background conditions during successive vs. non-successive boosted heatwaves” (that perhaps could be merged with a more concise version of the previous section), it is of particular relevance. However, I think that the statement: “However, the most intense and especially persistent successive heatwaves tend to last longer than their associated atmospheric blocking events” deserves some more explanation. In fact, this leads readers to expect somewhat of a relevant modulation introduced by the soil state in amplifying heat extremes beyond what is already provoked by favourable large-scale forcing.
I found the last results section a bit redundant and not introducing any relevant and not particularly informative.

Methodology.
Additional detail on the re-initialisation procedure of CESM ensemble members is needed. As currently described, it could be misinterpreted as a cold start of CESM components, which is likely not the case. Please specify the reinitialisation strategy employed for each CESM component beyond the atmosphere. In addition, clarify and justify the application of instantaneous random perturbations applied exclusively to the specific humidity variable, and specify at which pressure levels.

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 Suarez-Gutierrez,

Your manuscript titled "Worst-Case European Heat Storylines generated using Ensemble Boosting" 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 Portfolio, 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,

Alice Drinkwater, PhD
Associate Editor
Communications Earth & Environment
Consulting Editor
Communications Sustainability

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

The authors have done a thorough job addressing the comments from all three reviewers. The revised manuscript is more focused now that the drought storyline has been removed, the evaluation against ERA5 has been meaningfully expanded, and the analysis of conditions preceding successive vs. non-successive heatwaves is substantially improved. The switch to shallow soil moisture and the refined 15-day pre-heatwave window are well-motivated changes. I am largely satisfied with the revisions and believe the paper is close to publication.
I only have a few remaining minor points:

1. My main concern from the first round, i.e. the plausibility of these storylines is established only within the CESM2 model world, has been addressed well in the discussion (e.g., lines 697–698, 761–766) but is still largely absent from the abstract. I suggest the authors add a brief qualifier in the abstract, consistent with the careful language already used in the discussion (e.g., line 697), making clear that these storylines reflect plausibility within the CESM2 model world and carry no probabilistic interpretation for the real world. As it stands, phrases like "we reveal the risk for far more intense and unprecedented heatwaves" (line 13–14) could give a casual reader the impression of a real-world probability statement, which (I believe) is not what the authors intend.

2. Related to the above, the word "unprecedented" continues to be used somewhat loosely in places. Line 712 uses the phrase "extremely unprecedented," which is both awkward and imprecise. More importantly, "unprecedented" in this work means unprecedented in the historical record and in the model ensemble, rather than a claim about physical impossibility prior to anthropogenic forcing. The authors acknowledge this distinction in their rebuttal but could further sharpen the language in the manuscript itself in a few key instances (e.g., lines 483–484, 712).

Reviewer #2 (Remarks to the Author):

The authors have addressed all of my concerns. I would recommend it for publication.

Reviewer #3 (Remarks to the Author):

Dear authors,

I have reviewed the revised manuscript and believe that all my main comments have been adequately addressed.

I have just one final, minor, comment regarding the figure Fig. S1 in the revised manuscript:

"Figure C: Evaluation of CESM2-LE against ERA5 (as SI Fig. S1 in the revised manuscript). Probability density functions (PDF) for the 35-member CESM2-LE (gray; 2005–2023)".

Perhaps I am missing something, but I cannot see the colour “gray” mentioned in the caption.

Sincerely.

Response to the Reviewers' comments on manuscript COMMSENV-25-3411-T "Worst-Case European Heat Storylines generated using Ensemble Boosting" by Suarez-Gutierrez et al.

Please find enclosed our response to the Reviewers' remarks on our manuscript. Following their suggestions, we have revised our work by expanding the evaluation of CESM2 against ERA5 data, which now includes the evaluation of seasonal mean values and daily variability in atmospheric blocking, soil moisture deficits and sea surface temperature anomalies. We have also refined the analysis of conditions preceding successive vs. non-successive heatwaves, which has greatly helped clarify our conclusions regarding what conditions precede different types of heat extremes. Lastly, following the reviewers suggestions we have also streamlined the presentation of the results in our revised manuscript, and we have shortened the analysis to focus solely on extreme heat storylines, leaving the drought and atmospheric moisture demand analysis for a follow up publication.

We thank the reviewers for all their suggestions and comments, which we believe have greatly improved our work, and which we address individually more in depth below (in blue, with line numbers referring to the tracked-changes revised manuscript version).

Reviewer #1 (Remarks to the Author):

Dear authors,

I've genuinely enjoyed reading your manuscript and have only minor comments. The figures are beautiful and analyses are convincing. Here are a few remarks & questions in no particular order:

We thank the reviewer for their kind words and for their feedback on our paper, which we address point by point below.

1) I'm a bit confused as to why P4 is considered to showcase 'extreme heat' (p. 7 onwards, e.g., table 1); in Fig. 1, P3 is more extreme with regards to both heatwave area and intensity. In Fig. 2, the peak of P4 looks slightly more extreme than P3, but even there, I do not see a sufficiently clear distinction. Therefore, also because I don't think any conclusions depend on this choice, it might make more sense to only consider P1 & P5 as parent heatwaves with extreme heat.

In this case, we are referring to extreme heat conditions over the whole extended summer season, not over the specific heatwave itself. While all heatwaves selected for boosting are extreme (e.g., Fig 1a), for the May-Sept. accumulated excess heat metric (Fig. 1b) P4 and P5 stand out indeed as substantially more extreme than the other parent summers, and are therefore highlighted as showcasing extreme heat. P1 is the next-most extreme parent according to this metric, and we had originally also labelled it as a parent showcasing extreme heat. We agree with the reviewer that this label is somewhat subjective, and does not influence the conclusions of the paper in any way. However, later on in the analysis parents P4 and P5 exhibit some common behaviours, so to aid the readers to identify P4 and P5 as somewhat similarly behaving parents and for concise writing purposes, we choose to label these two parents as showcasing extreme heat in the revised version of the paper, and have removed P1 from this label. In any case, as we state above, this decision does not influence the conclusions of our work in any way. We have also modified Figure 1 in the revised manuscript to bring these points across more clearly.

2) Fig. 1: You could consider showing what the grey dots and black circles with numbers correspond to in the legend, perhaps splitting between a) and b) to make room for the additional legend entries.

Thanks, this has been changed as suggested.

3) 100-member CESM2-LE from NCAR vs own CESM2 simulations with 35 ensemble members: since you use both, it is a bit confusing that "CESM2-LE" seems to refer to both ensembles. Clearly distinguishing throughout the paper would help.

Thanks for bringing this to our attention. The 100-member CESM2-LE from NCAR is used exclusively for the 3˚C warmer world comparisons, as the 35-member ensemble does not span beyond 2035. We have stated this more clearly in the methods section (Lines 1894-1897) and have renamed all instances of the former as '100-member NCAR CESM2-LE' through the manuscript.

4) p. 13, "In sum, the worst-case … the most extreme heatwaves plausible today": it is still only plausible in the model world (specifically the CESM2 world), as explained in the introduction -- you have no means of asserting that your potential tail overestimation does not fundamentally bias this assessment. Considering the overall length of the paper, I'm a bit surprised that you never highlighted this caveat again after the introduction. Similarly, on p. 15, "…, successive heatwaves of the magnitude presented in these boosted storylines would indeed be unprecedented": Well, yes, but isn't this more of a demonstration what 'hunting' for the most extreme heatwaves can reveal rather than an actionable statement linking to climate change? Since your approach has no probabilistic interpretation, and without showing just how bad heatwaves could become even in the absence of human-induced climate change, one could

argue that there is an extremely tiny (but non-zero) probability of something similar occurring entirely "naturally". And, as a consequence, you cannot be sure that such an event is truly "unprecedented" -- it might not be on the historical record, but that is a different statement.

We thank the reviewer for bringing these points to our attention. Indeed, the extreme storylines presented here reflect conditions unprecedented exclusively in the historical record, and their plausibility cannot yet be explored beyond the CESM2-LE model world. Furthermore, when referring to the potential of such events to occur, we are not arguing that they must occur under climate change nor that they would not have been possible without it. As we state throughout the manuscript, we are interested in exploring potential worst-case conditions that would be plausible at current global warming levels. We agree with the reviewer's arguments and have expanded the discussion on several instances of our revised manuscript to better reflect these issues (e.g., Lines 1062-1068, 1672-1674, 1872-1876).

5) p.14, 18.2% successive heatwaves in ERA5: isn't this just 2 out of 11 samples? I don't think it makes sense to provide this level of accuracy and conclude that "CESM2-LE" and ERA5 show "remarkably good agreement"; you can, however, assert that based on the limited sample, the behaviour seems at least broadly consistent.

This has been changed as suggested, and now reads as follows (Lines 1119-1121):

We find that, based on the limited sample in the ERA5 record, the frequency of occurrence of successive heatwaves in CESM2-LE seems broadly consistent with ERA5 (23%, successive heatwaves in CESM2-LE; 18%, 2 out of 11, in ERA5).

6) Fig. 5, specifically 5d: I find the map inlet more confusing than helpful because the colors are not sufficiently distinct from the parent heatwaves. Why not show 3 inlet maps for a or b, c or d, and e or f, showing only the relevant domain?

This has been changed as suggested.

7) The text after the initial description of Fig. 6 on p.18 & 19 feels rather lengthy and could (partly) be moved to the Supplementary. At this point, the reader has been confronted with a lot of information already, and a condensed version would probably not take anything away from the main story.

This section, including the original Fig. 6 and related discussion has been changed substantially based on the feedback from the other reviewers, and we believe is now more concise and streamlined with the rest of the story in the manuscript.

8) p.23, "particularly the first heatwave in July 2019", is this a typo? According to Sousa et al. (2020), there was a June and a July heatwave in that region during 2019, with the latter potentially being enhanced due to soil moisture depletion during the former.

We thank the reviewer for bringing this reference to our attention. This is indeed not a typo, as according to our metrics and for the central Europe region used in our study, 2019 exhibits two explicitly defined heatwaves, in July and in late August to early September, as well as a longer period of more moderate heat in June and a sharp yet short peak at the beginning of August (see Fig. A below). This stands slightly in contrast with the heatwave timing in Sousa et al. 2020, likely due to the somewhat different region used in their study (e.g., [44–60 N, 10W–20E] latitude-longitude domain in our case, [43-53N, 0-10 E] for Sousa et al. 2020), as well as the different threshold-based heatwave metrics and definition criteria used in our work (i.e., a minimum of 5 days where at least 30% of the land grid cells of the Central Europe region of study exhibit positive Excess Heat values, meaning daily maximum temperatures above their 2000-2023 95th percentile). Lastly, both studies use different datasets in their analysis (ERA5 regridded to match the coarser resolution of CESM2-LE in our case, E-Obs in its original resolution, roughly 4 times finer than in our case, for Sousa et al. 2020).

Regardless of these technical differences, which make a direct comparison of the results in both studies somewhat complex, we have added their findings relating to the successive heatwaves in 2019 and potential mechanisms to our discussion (e.g., Lines 1199-1202, 1338-1340, 1391-1394).

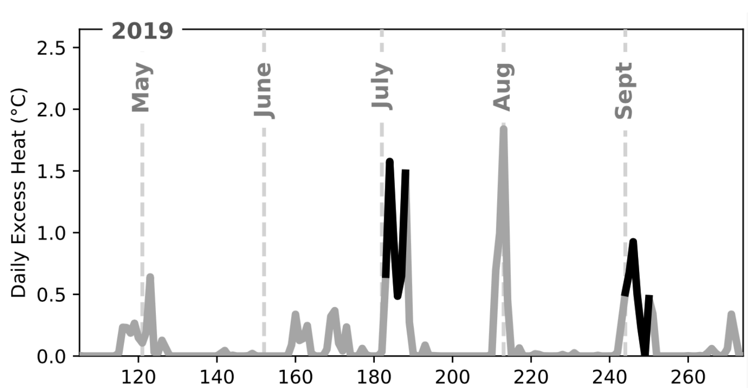

Figure A: *Daily Excess Heat 2019 time series for ERA5 data averaged over the Central Europe region used in our main analysis, [44–60 N, 10W–20E]. Excess Heat is defined, for each grid cell and day, as the difference between the actual daily maximum 2m air temperature minus the given heat threshold. Heat thresholds are defined as the 95th percentile in daily maximum 2m air temperature for each day and grid cell, pooled over a 15-day window centered on each day, for the period of 2005-2023 to allow a comparison between CESM2-LE and ERA5. Black lines mark explicitly defined heatwave periods according to our criteria: a minimum of 5 days where at least 30% of the land grid cells of the Central Europe region of study exhibit positive Excess Heat values.*

9) Finally, though the text is generally pleasant to read, there seem to be a few typos (e.g., "stotylines"), sometimes words are missing ("We also how comparisons…"), "The/These findings highlight the risk of a potentially devastating high-impact hazard" seems to be featured twice (not sure if that is intentional), but I won't list more here and instead recommend running a spellcheck.

We thank the reviewer for bringing these errors to our attention, which we have corrected as suggested.

References:

Sousa, P.M., Barriopedro, D., García-Herrera, R. et al. Distinct influences of large-scale circulation and regional feedbacks in two exceptional 2019 European heatwaves. Commun Earth Environ 1, 48 (2020). https://doi.org/10.1038/s43247-020-00048-9

Reviewer #2 (Remarks to the Author):

Manuscript Title: Worst-Case European Heat and Drought Storylines generated using Ensemble Boosting

Overall Assessment:

This is a highly ambitious and timely study that successfully leverages the ensemble boosting methodology to generate plausible worst-case heat and drought storylines for Europe. The core finding—that the most extreme heatwaves in the current climate are not only possible but also highly likely to occur successively—is significant and has important implications for climate risk assessment. The experimental design is sophisticated, and the manuscript is generally well-structured.

While the paper excels at risk discovery, its mechanistic analysis aimed at explaining the drivers behind these successive extremes suffers from several methodological and interpretive shortcomings.

We thank the reviewer for their feedback and suggestions, which we address point by point below.

Major Points of Concern:

1. One hypothesis is that a first heatwave "primes" the system for a more extreme second event via soil moisture depletion and ocean heat accumulation. However, the study fails to

present the most direct evidence: the evolution of key variables (e.g., soil moisture, SSTs) during the inter-period between the first and second heatwaves. Without demonstrating this temporal chain of events, the claim of causation is weak. The observed succession could merely be a statistical byproduct of selecting extremely hot summers where exceedances of a fixed threshold are naturally more frequent and clustered.

Analyzing this inter-period evolution can also help to explain the study's own finding that the most persistent successive heatwaves last longer than their associated atmospheric blocking events (Fig. 5a). It is possible that an initially dynamic-driven heatwave rapidly depletes soil moisture, and this newly established dry state then becomes the dominant driver, maintaining extreme temperatures even as the blocking pattern weakens.

We thank the reviewer for their feedback. We have revised our analysis to better address these issues, including redesigning the last part of our analysis (pertaining to revised Figures 5, 6 and 7 and related discussion, as well as Supporting Information Figures S2, S3 and S4) to better visualize the inter-heatwave period and conditions preceding the most extreme heatwaves successive and non-successive. The additional analyses include:

1. Additional time series of soil moisture deficits in sea surface temperature anomalies for the simulations with the topmost extreme successive heatwaves, highlighting heatwave and non-heatwave periods and shown against conditions in the renaming simulations stemming from the same parent events (SI Fig. S2).

2. Refinement of Fig. 5c-f to explicitly address conditions the 15 days prior to the heatwave start, as opposed to our original analysis assessing 15 days prior to the peak in heat and potentially confounding pre-heatwave behaviour with conditions developing after the heatwave onset for longer events.

3. Redesigning of former Fig. 6, now expanded into Figs. 6 and 7, to address soil moisture and sea surface temperature conditions explicitly the 15 days prior to the start of the most extreme successive heatwaves per parent (left columns) and the difference or uniqueness of these conditions with respect to the whole sample of simulations from the same parent event at the same point in time (right columns). For comparison we also include the same analysis for non-successive heatwaves in SI figures S3 and S4.

This additional analysis reveals, firstly, that the TOP50 most extreme successive heatwaves do not start from systematically more extreme spatially aggregated local SM and SST conditions than their non-successive counterparts (Fig. 5c-f; SI Fig. S2; Fig. 6 and 7 vs. S3 and S4). Second, that SM deficits and SST anomalies follow comparable evolutions for TOP50 successive heatwaves across different parents: a sharp intensification from conditions at the start of the heatwaves already extreme yet not unseen in other same-parent simulations (close or exceeding the 90th percentile across the same-parent boosted ensemble) that as the successive heatwave progresses evolve into some of the most extreme SST and SM levels in each parent boosted ensemble. Furthermore, in several cases across the TOP50 successive heatwaves, these extreme SST or SM conditions plateau at these extreme levels for days or weeks as the heatwaves continue or even beyond the end of the heatwave period (SI Fig. S2). This indicates potential heat advection or moisture depletion-temperature feedbacks that may intensify heat conditions at these plateau stages. Third, that beyond extreme local conditions, worst-case successive heatwaves also present common non-local patterns, such as preceding extreme drought conditions upstream of the area where heat develops, and strong warm SST anomalies close to the continent and cold SST anomalies in the farther North Atlantic region. With this revised analysis we also find that these patterns are also either more intense or unique for the most extreme successive heatwaves with respect to other simulations from the same parent conditions. For the readers' reference, we include in our discussion references to existing work discussing potential mechanisms of how such conditions could intensify extreme heat as found in past observed extreme heat events.

We want to thank the reviewer again for their suggestions to expand this analysis, as we believe this has substantially refined our work and conclusions. However, we also want to reiterate that fully disentangling the chain of causation and the mechanisms behind these successive heat extremes would require a careful and detailed case-specific analysis that we believe is beyond the scope of this current work at this stage. In particular, quantifying the role of large-scale dynamics in driving these heat successions, and disentangling the role that dynamic versus thermodynamic processes may have in either producing or sustaining these extremes would require a more in-depth, and case-specific assessment of the meteorological evolution of the event successions and the inter-event periods, as the reviewer suggests. While we believe this analysis would be both interesting and of great importance, we also believe it would be too extensive to be presented properly in our current manuscript, and is currently being prepared as follow-up work. To reflect this, we reiterate that these are indeed remaining questions and open lines of research throughout the discussion in the revised manuscript (e.g., Lines 1592-1596, 1750-1759).

Lastly, and in particular referring to the statement ‘successive heatwaves last longer than their associated atmospheric blocking events’, we agree with the reviewers that this finding is particularly striking and warrants further attention, which we are planning to address in more depth in future work. However we also want to highlight that this behaviour, present most prominently for the most extreme event, may be somewhat dependent on our choice of metrics and region of study. For this particular case, persistent anticyclonic conditions may not be fully captured by our blocking index, or may not perfectly overlap with the region of study. A preliminary analysis seems to indicate that this boosted summer exhibits sequences of less-stable, potentially non-blocked yet recurring anticyclonic conditions. We are currently investigating this issue in more depth, but we believe fully addressing these processes at this stage is again beyond the scope of the current work. We have however revised our wording to clarify these remaining uncertainties (e.g., Lines 1162-1182).

2. The composite analyses (e.g., Fig. 6) comparing conditions during successive and non-successive heatwaves are potentially confounded. According to SI Figure S2-S5, the parent events that produce the successive heatwaves (P4, P5, P6, P7) are inherently different from those that produce most non-successive ones (P1, P2, P4, P6, P7). Therefore, differences in composites (e.g., higher SSTs in successive events) can be driven by the inherent properties of the parent cases rather than by the "successive" attribute itself. For example, when comparing successive vs. non-successive events stemming from the same parent, like P4, warm SST anomalies are observed in both types (Figure S5).

We thank the reviewer for highlighting this, and we agree completely with their arguments. As mentioned in the response to comment 1 above, we have refined our analysis to better visualize how unique to the worst-case successive and non-successive heatwaves specific conditions are, and how different those conditions are both across events stemming from different parents, successive versus non-successive events, and most extreme events versus the rest of the same-parent boosted simulations. By refining this analysis to address specific pre-heatwave conditions for the most extreme events (versus the average july-august conditions in our original analysis) we are indeed able to find more robust SM and SST patterns common to the most extreme events, as discussed above. And by distinguishing between different parents and by how unique these conditions are to the most extreme events among the same-parent ensemble we are indeed able to better quantify what conditions are uniquely present for the most intense successive and non-successive events.

We agree with the reviewer that our original analysis was not sufficiently clear and confounded different factors, and we have revised this analysis as per the suggestion to better capture these nuances. We have revised the results and discussion sections extensively to reflect these new and refined findings, and believe this revised analysis and associated discussion (e.g., last results section, Lines 1186-1607) has greatly improved our manuscript and the clarity of our conclusions.

3. The use of different methods to calculate anomalies for different variables (T, SM, SST, etc) creates an inconsistency. This makes it difficult to compare the "extremeness" across variables and might add noise and obscure the physical relationships the study seeks to uncover.

We thank the reviewer for this comment. We try to calculate anomalies and process variables as consistently as possible. For the case of temperature and heat metrics, we use threshold-based metrics, which is a standard and commonly used approach in heatwave research, as opposed to anomalies with respect to a mean state. However, we use the same periods for establishing this climatological threshold reference as we do for establishing climatological reference periods for the other variables. We also use cumulative heat-metrics, which are emerging as widely used in the literature in recent years, and arguably preferred to assess the intensity of heat extremes particularly for events of varying lengths (e.g., Russo & Domeisen 2023, https://doi.org/10.1029/2023GL103540). Therefore, while we agree that the use of cumulative, threshold-based heat metrics introduces some inconsistency with the other variables, we believe both points are crucial, first for capturing heat extremes consistently, and second to make our results comparable to other heatwave studies.

Besides this, all variables with the exception of temperature are processed as anomalies with respect to a common multi-year daily climatological reference period. For simulated data, this reference period is calculated for each simulation separately. However, due to data limitations for CESM2 simulated soil moisture, this multi-year climatological reference period is not available for the parent simulations nor for the majority of the non-boosted ensemble, as the majority of the simulations do not have daily soil moisture data for the multi-year climatological reference period. To overcome this limitation and not restrict our analysis to monthly-mean soil moisture deficits, we calculate a multi-year, multi-member daily mean climatological reference from members 31 to 35, which are the only members with daily soil moisture for the 2005-2023 reference period. We agree this is not an ideal setup, yet we argue that it does not substantially affect our conclusions, as we find relatively low variance across the climatological references across these 5 members (not shown). We would also argue that, while the anomaly calculation limitations expressed above introduce some inconsistency, the added value of using daily soil moisture for the boosted event analysis outweighs this inconsistency. Furthermore, we agree this anomaly calculation method is inferior, and we maintain the standard in-member climatological reference period calculation for the remaining variables with available data.

Lastly, to compensate for these potential inconsistencies across anomaly-calculation for different variables we also often present our results either as standardized anomalies or give standard deviation values as a visual guide, to provide some context with regards to the 'extremeness' of values with respect to their distributions.

4. The use of total column soil moisture (integrated over 100 cm) is appropriate for assessing ecological drought but might not be the optimal metric for controlling land-atmosphere feedbacks during heatwaves. The partitioning of sensible and latent heat fluxes that governs near-surface temperature maxima is primarily controlled by moisture availability in the top few centimeters of soil. Using a deep-layer integral dilutes the strong signal of rapid surface drying that is likely the key mediator of the "priming effect" and heatwave intensification.

We thank the reviewer for bringing this to our attention. Originally, the focus of our study was on compound heat and drought storylines, therefore the preferred deeper layer soil moisture analysis to address ecologically-relevant drought levels. However, following the reviewers' recommendations, we have narrowed the focus of this current paper to address solely extreme heat storylines. Therefore, we agree with the reviewer that a more shallow layer would be preferred to capture soil-atmosphere interactions, and we have modified our analysis to use a shallower upper soil layer of roughly 10 cm to assess soil moisture deficits. We also include here a summary of the changes in our results introduced by using a shallower layer, exemplified by Fig. 5 c-d (see below as Figure B).

The use of standardized values in our analysis (standardized by dividing by the st. dev. taken across periods of the same length at the same time of the year) makes the distinction between the deeper and shallower layer metrics more nuanced than for absolute SM values. However, using a shallow SM layer we see better agreement between the ERA5 distribution and CESM2-LE (e.g., SI Fig. S1-g in the revised manuscript for shallow SM, not shown for deep SM), as well as larger variability in SM changes before and after extreme heatwaves (Fig. B below, Fig 5c-d in revised manuscript), in line with the reviewer's intuition that the heatwave effect on the SM presented in our original work was somewhat 'diluted' by the use of a deeper soil layer.

We thank the reviewer for this suggestion, and we agree that the shallow SM layer is superior at representing the soil-atmosphere interactions and feedbacks during extreme heat events. Additionally, our findings also suggest that CESM2-LE may represent real-world behavior in shallow SM across the extended summer season mean deficits and variability with more fidelity than for our original deeper SM layer, and we have therefore revised our analysis accordingly to use this metric instead.

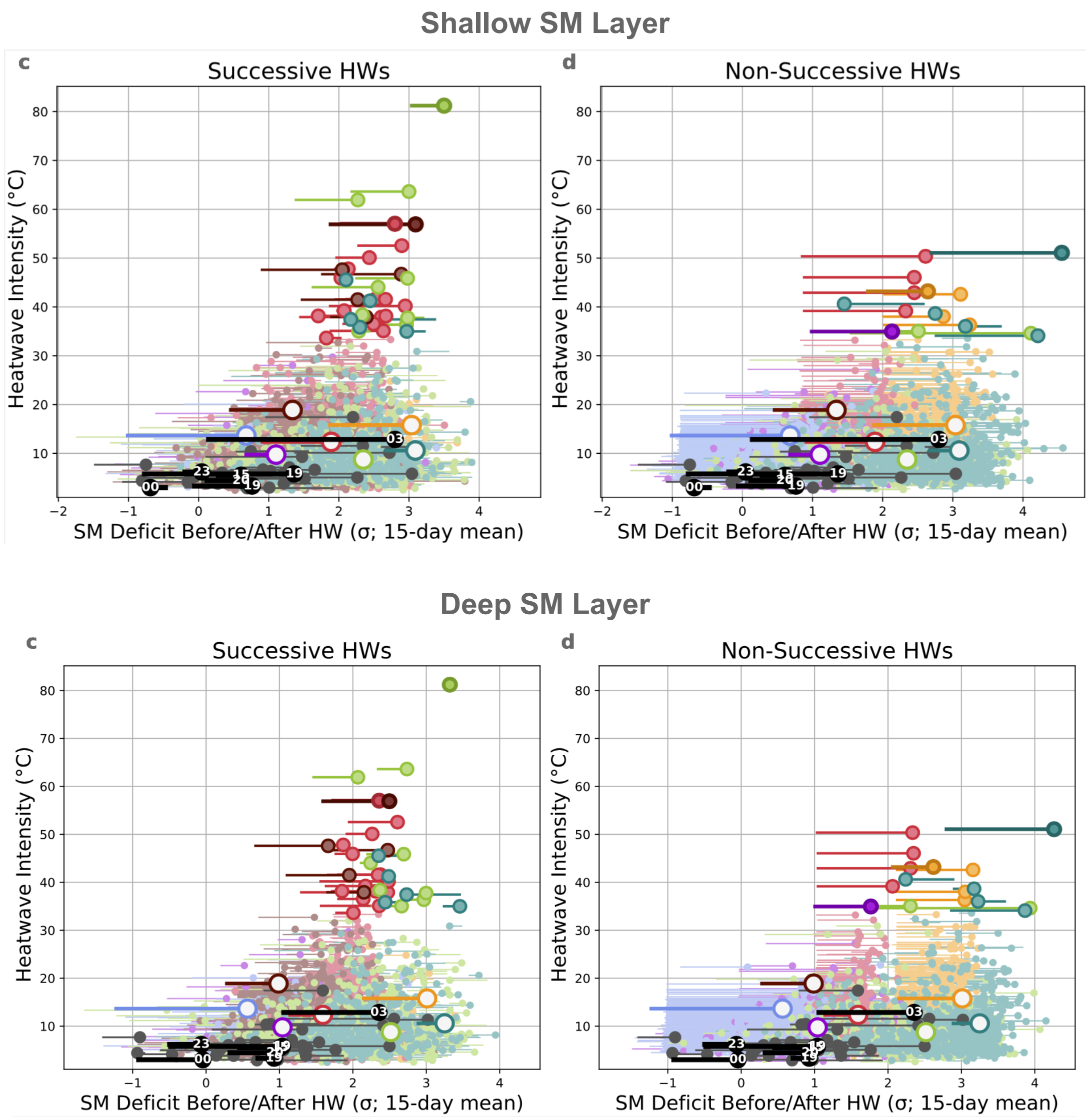

Figure B: *As Figures 5 c and in the revised manuscript, but for shallow (top row, same as in revised manuscript) and deep (bottom row, using same deeper soil layer data as in original manuscript) soil moisture layers.*

Reviewer #3 (Remarks to the Author):

This manuscript leverages an ensemble boosting technique to produce a very large ensemble of climate simulations for characterising European heat and drought extremes. This research provides a remarkable plethora of relevant results regarding the driving mechanisms of European heat and drought extremes, focusing on both large- and local-scale forcing dynamics.

Overall, while the manuscript provides a wealth of remarkable findings, the current length and level of detail risk diluting the impact of its key messages. This specifically refers to the Results section.

We thank the reviewer for their feedback. Following their suggestions and the recommendations of the other reviewers, we have revised our analysis and results sections to make our manuscript more concise and our findings and conclusions more streamlined. We address this and their others comments more in depth point by point below.

Besides this general comment, specifically referring to the first results section involving the CESM-LE evaluation. Driving HW mechanisms explored in the manuscript are essentially represented by 500 hPa geopotential anomalies in terms of blocking, soil moisture (SM) deficit and SST anomalies spatial distribution. Although the surprisingly good capability of CESM-LE in reproducing heat and drought statistics/metrics, it would strengthen the analysis to include a brief discussion of CESM-LE's ability to reproduce SST anomaly variability and basic blocking statistics during the evaluation period. This would reinforce confidence that the good performance in excess-heat metrics is indeed driven by the mechanisms under investigation. To maintain a balanced section length, I would rather condense some of the current evaluation findings and include some evidence/consideration (also retrieved from published studies to limit the already vast number of analyses) regarding CESM-LE's representation of basic blocking statistics, SST and SM anomaly variability. The latter point is especially important for SM, since at coarse GCM resolution (the resolution is not reported in the Methods section and should be clarified), models often struggle to capture SM variability and its feedbacks with the planetary boundary layer.

We thank the reviewer for their feedback. Following their suggestion, we have added a more extensive evaluation of several key metrics, including mean soil moisture deficits and variability over the extended summer season, as well as blocking persistence and spatial variability over the extended summer season and mean SST and SST variability also over the extended summer season. These results can be now found in a revised version of Supplementary Information (SI) Figure S1 (shown below also as Figure C), and a summary of our evaluation has also been added to the main body of the article (Lines 152-449).

Our evaluation of these additional metrics shows that CESM2-LE captures extended summer mean values for SM deficits (Fig. Bd), blocking persistence (Fig. Be) and SST anomalies (Fig. Bf) remarkably well, both in terms of the whole distribution and upper level percentiles. In terms of the extended summer season variability (represented by standard deviations across daily data in the May-Sept. period; Fig Bg-i), we find somewhat larger discrepancies between the two samples, particularly between the upper level percentiles and for SM deficit variability (Fig. Bg). This suggests that CESM2-LE may overestimate the frequency of summers with large SM changes (e.g., intra-seasonal seesaw effects from dry to wet soil moisture conditions or vice versa). To compensate potential variability biases, we present most of our SM results as standardized values (see Methods section, Lines 2022-2030)

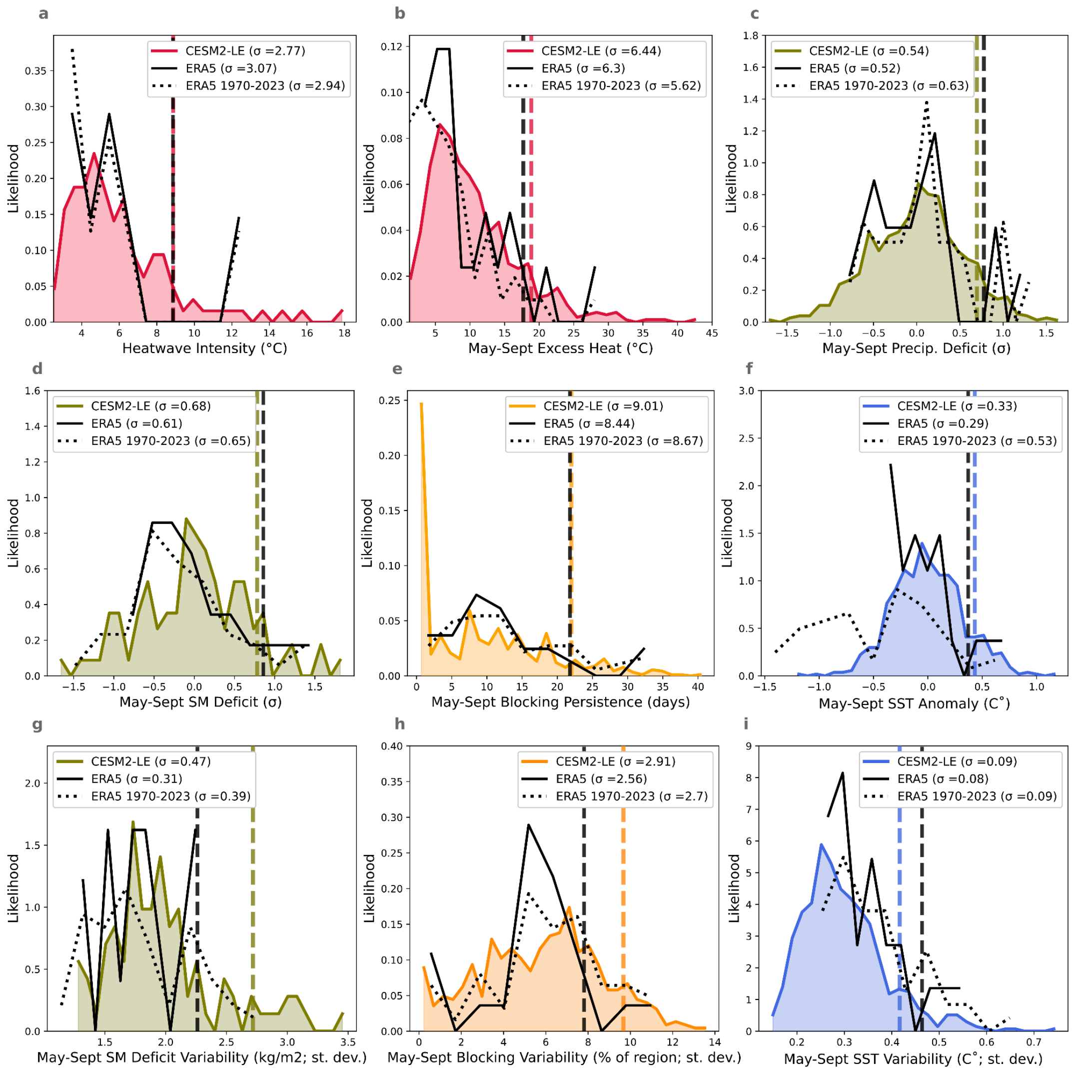

Figure C: Evaluation of CESM2-LE against ERA5 (as SI Fig. S1 in the revised manuscript). Probability density functions (PDF) for the 35-member CESM2-LE (gray; 2005–2023) and ERA5 (solid black line for 2000–2023 and dotted black line for 1970–2023). Vertical dashed lines mark the 90th percentile in ERA5 (black; 2000–2023) and CESM2-LE (gray; 2005–2023). σ values show the standard deviations of the different samples for Heatwave Intensity (a), accumulated Excess Heat (b) standardized accumulated Precipitation Deficits (c), mean soil moisture (SM) deficits (d), accumulated blocking persistence (e), mean sea surface temperature (SST) anomalies (f), SM variability (standard deviation across seasonal SM daily values; g), blocking coverage variability (standard deviation across seasonal daily values of blocking spatial coverage as percentage of grid cells; h), and SST variability (standard deviation across seasonal daily SST values; i). All metrics are assessed for the May-Sept. extended summer season

In the section "In the Worst-case ensemble boosting storylines for extreme heatwaves", please elaborate a bit on what is possibly behind the mechanism reported in the following statement: "indicating that boosted simulations remain more similar to each other than an independent sample and exhibit above-average heat and drought conditions throughout the summer" This is relevant.

We thank the reviewer for highlighting this. In this statement we mean that the same-parent boosted ensembles do not generally reach heat and soil moisture inter-member variability levels as high as those found from year to year or member to member in the climatological reference of the non-boosted CESM2-LE simulations. This indicates that, for example, moisture levels through this boosted summers do not revert back to wet or even neutral states with the same frequency than those states would happen in the climatology, and similarly than temperatures may remain more consistently above average levels due potentially to this increased drought level, with less colder-than-average periods and thus less variability. This is consistent with the following results of the boosted simulations presenting large numbers of successive heatwaves as well as extreme drought conditions. We have expanded our revised manuscript to better explain this (Lines 861-866).

Regarding the "Background conditions during successive vs. non-successive boosted heatwaves" (that perhaps could be merged with a more concise version of the previous section), it is of particular relevance. However, I think that the statement: "However, the most intense and especially persistent successive heatwaves tend to last longer than their associated atmospheric blocking events" deserves some more explanation. In fact, this leads readers to expect somewhat of a relevant modulation introduced by the soil state in amplifying heat extremes beyond what is already provoked by favourable large-scale forcing.

We thank the reviewer for highlighting this point, which we also address in our response to Reviewer's 2 comments above. We agree with the reviewers that this finding is particularly striking and warrants further attention, and we are planning to address this in more depth in future work. For the current manuscript, we have revised the wording of this statement and expanded the related discussion to clarify this point further. We now address that this behaviour, present most prominently for the most extreme event, may be somewhat dependent on our choice of metrics and region of study. For this particular case, persistent anticyclonic conditions may not be fully captured by our blocking index, or may span beyond the region of study. We also elaborate more specifically on the potential mechanisms at play in this behaviour as the reviewers' suggest, and on the associated uncertainty and metric-dependent issues (e.g., Lines 1162-1198)

I found the last results section a bit redundant and not introducing any relevant and not particularly informative.

We thank the reviewer for their feedback. Following their suggestion, we have removed this section from our revised manuscript, including original Figure 7 and related discussion, as well as the focus on drought storylines throughout the original manuscript. To achieve a more concise storyline flow for this revised manuscript, we will explore these aspects of compounding heat and drought in a separate publication.

Methodology.

Additional detail on the re-initialisation procedure of CESM ensemble members is needed. As currently described, it could be misinterpreted as a cold start of CESM components, which is likely not the case. Please specify the reinitialisation strategy employed for each CESM component beyond the atmosphere. In addition, clarify and justify the application of instantaneous random perturbations applied exclusively to the specific humidity variable, and specify at which pressure levels.

We thank the reviewer for bringing these unclear points to our attention. Indeed, the ensemble boosting technique used here does not apply a 'cold start' of CESM2 components. Boosted simulations are started from the parent conditions in the original non-boosted parent simulations. This means that, beyond the perturbation to the upper-atmosphere humidity field, all remaining model components are not reinitialized nor perturbed. We have modified the relevant method section to clarify this (Lines 1912-1940).

Furthermore, we also clarify the specifics of our perturbation setup (Lines 1879-1883). This perturbation is purposely introduced 1.) instantaneously, 2.) across all atmospheric pressure levels, and 3.) into a variable field that does not directly influence heatwave metrics not development to ensure that the variability introduced by this perturbation reflects the chaotic natural internal variability in the climate system as closely as possible. Lastly, we also clarify that these perturbations are introduced across all atmospheric grid cells and therefore all atmospheric levels (Line 1878).

Response to the Reviewers' comments on manuscript COMMSENV-25-3411A “Worst-Case European Heat Storylines generated using Ensemble Boosting” by Suarez-Gutierrez et al.

Please find enclosed our response to the final Reviewers' remarks on our manuscript. We have made minor final adjustments to our work, following their suggestions, which we address individually more in depth below (in blue, with line numbers referring to the tracked-changes revised manuscript version).

Reviewer #1 (Remarks to the Author):

The authors have done a thorough job addressing the comments from all three reviewers. The revised manuscript is more focused now that the drought storyline has been removed, the evaluation against ERA5 has been meaningfully expanded, and the analysis of conditions preceding successive vs. non-successive heatwaves is substantially improved. The switch to shallow soil moisture and the refined 15-day pre-heatwave window are well-motivated changes. I am largely satisfied with the revisions and believe the paper is close to publication.

I only have a few remaining minor points:

1. My main concern from the first round, i.e. the plausibility of these storylines is established only within the CESM2 model world, has been addressed well in the discussion (e.g., lines 697–698, 761–766) but is still largely absent from the abstract. I suggest the authors add a brief qualifier in the abstract, consistent with the careful language already used in the discussion (e.g., line 697), making clear that these storylines reflect plausibility within the CESM2 model world and carry no probabilistic interpretation for the real world. As it stands, phrases like "we reveal the risk for far more intense and unprecedented heatwaves" (line 13–14) could give a casual reader the impression of a real-world probability statement, which (I believe) is not what the authors intend.

2. Related to the above, the word "unprecedented" continues to be used somewhat loosely in places. Line 712 uses the phrase "extremely unprecedented," which is both awkward and imprecise. More importantly, "unprecedented" in this work means unprecedented in the historical record and in the model ensemble, rather than a claim about physical impossibility prior to anthropogenic forcing. The authors acknowledge this distinction in their rebuttal but could further sharpen the language in the manuscript itself in a few key instances (e.g., lines 483–484, 712).

We thank the reviewer for their time and for their feedback on our paper. As suggested, we have sharpened the language referring to ‘unprecedented’ events, here referring indeed to events that have not occurred previously in the historical record, on several occasions through the manuscript (e.g., lines 164, 436, 491, 694, 717-719, 736-737). We have also revised the wording regarding the plausibility of the presented storylines in the abstracts to align this wording with the rest of the manuscript (Lines 13-14, 20). We believe these changes have refined and sharpened our message, and we thank the reviewer for their suggestions.

Reviewer #2 (Remarks to the Author):

The authors have addressed all of my concerns. I would recommend it for publication.

We thank the reviewer for their suggestions and for the time they have dedicated to this review.

Reviewer #3 (Remarks to the Author):

Dear authors,

I have reviewed the revised manuscript and believe that all my main comments have been adequately addressed.

I have just one final, minor, comment regarding the figure Fig. S1 in the revised manuscript:

"Figure C: Evaluation of CESM2-LE against ERA5 (as SI Fig. S1 in the revised manuscript). Probability density functions (PDF) for the 35-member CESM2-LE (gray; 2005–2023)".

Perhaps I am missing something, but I cannot see the colour “gray” mentioned in the caption.

Sincerely.

We thank the reviewer for their feedback, and for catching this remaining error. Indeed, the mention to the gray color was incorrectly and a remanent from an earlier version of the figure, and should have been substituted by the word ‘color’. This has been corrected in the revised manuscript.