Response to reviewers

We are pleased to see that the reviewers value the content of our study. We appreciate their feedback and suggestions. Below, we provide a detailed, point-by-point response to the comments from the reviewers.

Our responses to reviewer comments are organized by category. Each response is labeled with a code in the specified range. The response categories are given below.

Reviewer Comments	Author Responses
CC1	A1-A2
RC1	B1-B26
RC2	C1-C30
CC2	D1-D5

RC1:

Review on "Global relevance of atmospheric and land surface drivers for hot temperature extremes" by Yigit Uckan and colleagues.
The manuscript investigates different atmospheric and land surface drivers of hot extremes on two time-scales in the period 2001-2020. The authors find geopotential height to be by far the most important driver of 1 day events, while for longer 7 day events land surface drivers become more important.
The manuscript is well written and well organized and supported by meaningful figures. The topic covered is a timely one and nicely supplements the existing literature, for example, a recent study by Röthlisberger et al. (2023; 10.1038/s41561-023-01126-1) which the authors should consider discussing as it investigates a very similar question but using a quite different approach for atmospheric drivers.
B1: We thank the reviewer for pointing us to this reference and have included it in section 3.1 where we interpret figure 2 (See response B16).

While I do not have any major comments, I see several open questions that should be addressed before publication:

- independence of the variables used as drivers: This is briefly discussed buy should be quantified in some way, in particular to show that atmospheric and land surface drivers are indeed independent as assumed by the authors (line 186).

B2: That is a valid point, thank you for this comment. We believe a cross-correlation matrix can address this. This can be computed by taking 5x5 grid cells to have enough data from the example regions (e.g. central Europe, Amazon, Australia etc.). Since we have 3 hot extreme events per grid cell, this would result with 5x5x3 = 75 data points to compute the correlation between considered driver variables.

Related results will be added to the manuscript as a supplementary figure.

- effect of analogue quality: What role does the 'closeness' of the found analogues to the observed event have on the results and could this influence, e.g., the differences in 1 day and 7 day events (as it might be harder to find good analogues for 7 day events)?

B3: The analogue quality will be assessed by plotting maps of the difference between mean analogue values from the actual values. The variables we will investigate are geopotential, EVI and radiation (as most important predictor variables).

Related maps will be added to the manuscript as a supplementary figure.

- the metric 'degree of relevance' and its interpretation is not quite clear to me as mentioned in the specific comments below

B4: The degree of relevance metric shows to what extent a variable can explain hot extremes in a certain location. It serves as a basis of comparison to determine how much of the observed extreme temperature anomaly can be explained by the analogue temperature conditions of our variables. A detailed explanation is given in our responses C13 and B20.

- the discussion of changes due to climate change (3.3) is very short could benefit from some more analysis and contexualization. In particular since the two time periods investigated are quite short, I'm wondering if any of the effects are statistically significant.

B5: Thank you for this suggestion. We agree that further analysis and contextualization would strengthen this section. Currently, Figure 5 does not indicate whether the observed changes are statistically significant. To address this, we will conduct a bootstrapping analysis for each variable to assess statistical significance.

In this analysis, we will use the original 15 analogues (3 hot extremes x 5 analogues per hot extremes = 15 analogues) per grid cell and per time period (2001-2010 and 2011-2020), and draw a random sample of 15 from them (with replacement). Resampling will be done 1000 times. Then, we can compare the mean degrees of relevance of the resampled 15 analogues from both periods per grid cell such that we can infer significance from analyzing whether a substantial fraction (e.g. 950 out of the 1000) of the mean degrees of relevance between the two time periods is consistently above or below zero.

Identifying statistically significant shifts in relevance will allow a more insightful interpretation of our findings. We will compare these findings with the literature to contextualize changes in relevance. The significance results will be added to Figure 5.

Minor comments

Given that the manuscript is well organized and written this is not a large issue but the authors could consider focusing data section (2.1) better. Currently it reads like a mixture of data and method section, with sentences like: "In addition, we compute the geopotential height

differences at 500 hPa pressure level for each grid cell with respect to the values in adjacent grid cells in the northern, eastern, southern and western directions."

B6: Thank you for your feedback, indeed some sentences belong to the methodology or introduction section. We will revise section 2.1 by moving some sentences to the introduction or methodology parts. Moreover, we will introduce additional subsections in section 2. Specifically, the sections will look like this:

1. Introduction (problem definition, motivation, variables considered in this study, novelty)
2. Data and methods
   2.1 Data (datasets, sources, and preprocessing methods.)
   2.2 Definition of hot extreme events
   2.3 Description of the analogue approach
   2.4 Explanation of the 'relevance index'
   2.5 Description of the attribution analysis
   2.6 Description of the trend analysis

line 79: Could the authors elaborate why the use only 2001-2020?

B7: Evapotranspiration data from the X-base dataset that is used in the calculation of EF is only available between 2001 and 2020. The sentence in section 2.1 has been updated to:

"The spatial and temporal resolutions considered are 0.25 degrees and daily intervals, respectively, for the study period from 2001 to 2020. This period was selected because the evapotranspiration data from the X-base dataset used to calculate EF are only available during these years."

table 1: some of these drivers are probably quite correlated (e.g. GPH and surface radiation), could the authors comment on how this might influence the analysis and interpretation of the results?

B8: We will address this issue by computing the cross-correlations of the variables as we have mentioned in response B2. The correlations will be presented in section 2.1 Data, and we will adapt our discussion in the light of the cross-correlation results.

90: "For the 7-day time scale we apply a moving average" Could the authors state the window size explicitly here? (I'm guessing its 7 days?)

B9: Yes, the window size for the moving average is indeed 7 days. We will revise the sentence in section 2.2 to explicitly state this for clarity.

"For the 7-day time scale, we apply a 7-day moving average to smooth out daily variability."

91: "For each type we select the three hottest events". I'm assuming the authors refer to the 1 day and 7 day events here? However, this is not really clear form the context of the last few sentences.

B10: Yes indeed we refer to the 1 day and 7 day events there. Here we explicitly mention this in section 2.2:

"For the 1-day events, we select three individual days with the next highest temperatures, ensuring that each selected day is at least 15 days apart from the others to maintain independence. For the 7-day time scale, we apply a 7-day moving average to remove variability from shorter time scales and then select the three 7-day periods with the next highest average temperatures, also ensuring they are at least 15 days apart from each other for independence."

To make sure I understand correctly: for the 1 day events these would be 3 individual days and for the 7 day event three 7 day periods?
B11: Yes exactly. We've changed the relevant section in the light of your feedback as mentioned in our response B10.

In general I think this section might benefit from a concrete example. I'm not a 100% sure I understood the approach.
For example: for the 1 day events, this would give 3 individual days which are all separated by at least 15 days? Meaning if there's a 7 day heatwave only the hottest day of it would be selected for the 1 day event?
Its also never mentioned if this is done on an annual basis (as seems to be indicated in figure 1) or for the entire dataset at once.
B12: Thank you for your comments. The event selection is performed over the entire study period (2001 to 2020), rather than on an annual basis. To ensure clarity, we have updated Figure 1 to better illustrate the process as can be seen in our response B13. We also have added an example to facilitate the understanding of the methodology in section 2.2 in addition to our response B10:

"For example, consider event selection for 1-day hot extremes within the study period (2001-2020). Let's say in a specific grid cell, the hottest day recorded during this period is July 15, 2012. After selecting this day, we mask July 15 and the 30 days surrounding it (July 1 to July 30) to prevent selecting any overlapping or consecutive days. We then identify the second hottest day from the remaining time series after masking, which could be August 5, 2010, and apply the same 15-day masking around this date (July 21 to August 20). This process is repeated to find the third hottest day, ensuring that all three selected days are at least 15 days apart, maintaining their independence.

For 7-day events, the procedure is similar. Suppose the highest 7-day average temperature in the grid cell occurs from July 10 to July 16, 2015. We mask this period and the surrounding 30 days (June 26 to August 1) before selecting the next highest 7-day period, such as August 20 to August 26, 2013. This ensures that each selected 7-day event is independent and focuses on the warm season."

Figure 1:

- could mention that "Hottest Period x" does also refer to a single day in the case of 1 day events?

- increase font size

B13: Thank you for the suggestions. We have updated figure 1 by following your suggestions and the explanation of "Hottest Period x" has been included in the figure caption.

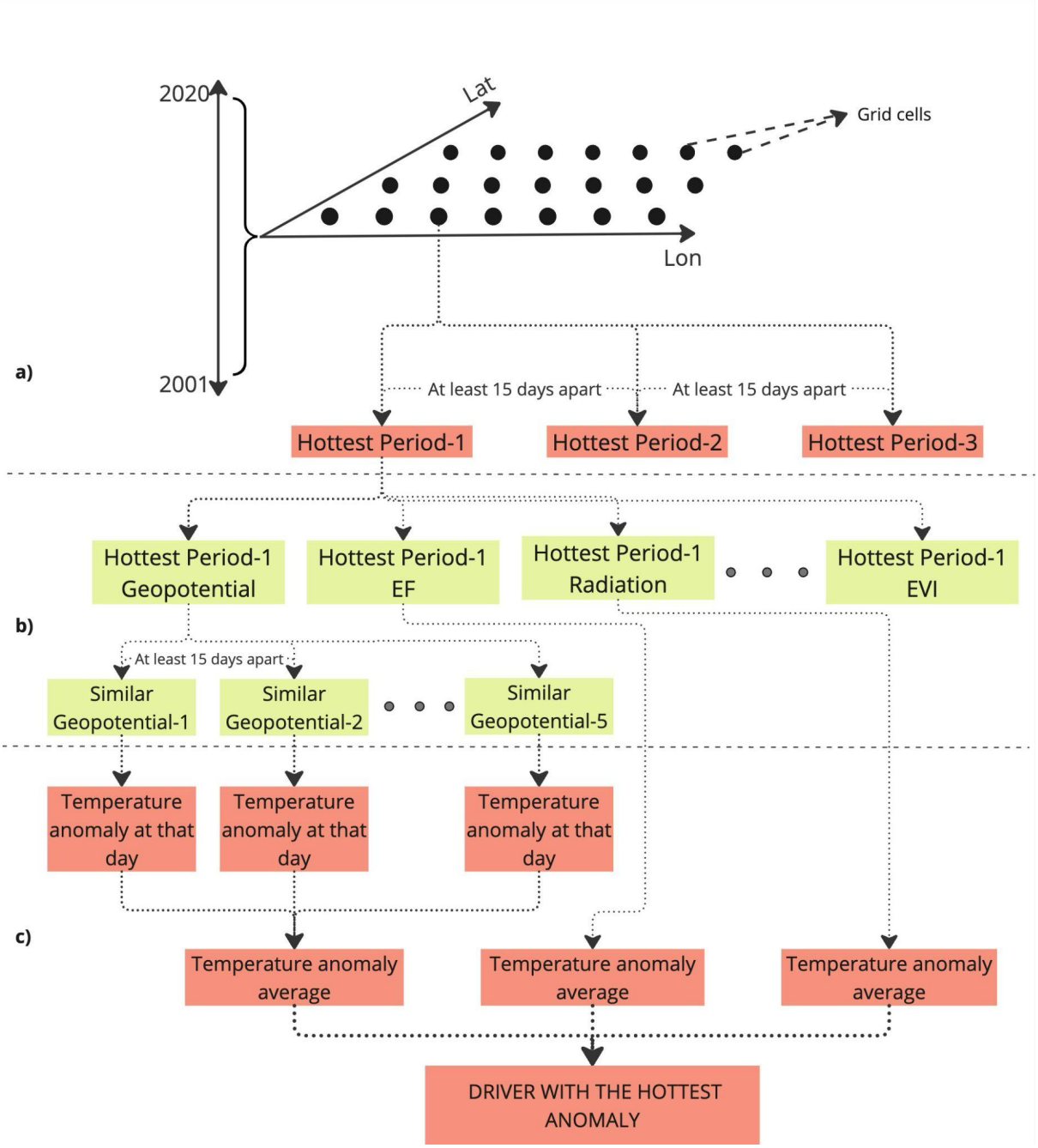


Figure 1 Workflow for determining main drivers of hot temperature extremes. Hottest periods refer to 1-day or 7-day hot extreme events.

"2.5 Effect of the increasing trend" For me 'increasing' indicates an acceleration of a trend (which is already a change measure) so it is probably not what the authors want to say here? 'increasing number' or 'positive trend' instead?

B14: Thank you for pointing this out. We agree that 'increasing trend' might imply an acceleration, which was not our intention. We will revise the section title to 'Effect of the positive trend in hot temperature extremes on the relevance of driver variables' for clarity.

Figure 2:
- the authors could consider using ‘more different’ colors to make the separation of the different drivers easier? In particular, for the separated view in figure A1 it is almost impossible to really separate drivers due to the chosen colormaps.
B15: Thank you for your suggestion. We understand the challenge you've pointed out regarding the visibility of the drivers in Figure A1. However, due to the high number of variables and the high spatial resolution, achieving clear visual differentiation is quite difficult. Even with adjustments to the colormap, it would be still difficult to clearly distinguish the drivers completely. In Figure 3, we address this issue by highlighting the dominant factors in different climate zones, helping to interpret the results presented in Figure 2. For Figure A1, we will provide a table to summarize the results as we have mentioned in response C26. We appreciate your understanding of these constraints.

- it could also be interesting to compare these results to the work from Roethlisberger et al. 2023 (10.1038/s41561-023-01126-1) at least for the atmospheric drivers?

B16: Thank you for this suggestion. We agree that comparing our results to the findings from Röthlisberger et al. (2023) could provide additional context. Röthlisberger et al. use a Lagrangian approach to decompose temperature anomalies into contributions from advection, adiabatic warming, and diabatic heating, highlighting the regional variability of these processes in forming hot extremes. Findings show that diabatic heating is a factor affecting the temperature anomalies especially in regions where the soil moisture is limited. This finding aligns with our results regarding EF dominant regions. We’ve added the following sentence to section 3.1 where we interpret Figure 2:

“Similarly, the findings of Röthlisberger et al. (2023) show that diabatic heating is a factor affecting the temperature anomalies especially in regions where the soil moisture is limited. This finding aligns with our results regarding EF dominant regions ”

- Röthlisberger, M., Papritz, L.: (2023). Quantifying the physical processes leading to atmospheric hot extremes at a global scale. Nat. Geosci. 16, 210–216. https://doi.org/10.1038/s41561-023-01126-1

153: could this be partly due to the fact that it is (presumably) harder to find good analogues for 7 day events for GPH compared to 1 day events and hence the temperature anomaly is less pronounced for the 7 day case?

B17: We thank the reviewer for raising this point. We will analyze this (see response B3 on the planned approach) and consider these findings also at this place in the results section.

186: “Furthermore, our main goal is to disentangle land surface and atmospheric drivers of hot extremes which are not expected to be strongly related to each other.” As commented earlier: could the authors quantify the cross-dependence of the drivers is some way?

B18: We will address this issue by computing the cross-correlations of the variables as also mentioned in response B2. The correlation methodology will be in section 2.5 and the results will be presented in section 3.1.

192: another limitation might be the quality of the analogues, which seems to be crucial for the quantification of the contribution?

B19: Depending on the outcome of the analysis described in response B3, we will update the limitations section if necessary. Furthermore, our rationale for analogue selection and the updated relevant text in section 2.3 is given in C11.

Figure 4: “The degree of relevance is computed as the ratio between the respective analogue temperature anomalies and the observed temperature anomalies during hot extremes.” This could be explained in a bit more details in the methods section? For example: this seems to mean that the degrees of relevance from different drivers can sum up to more than 100% percent, right?

B20: Thank you for your comment. We agree and will clarify this in the methods section. The degrees of relevance for different drivers can indeed exceed 100% in sum due to collinearity among the drivers, as also shown in Figure A5.

For the methodology part in section 2.4 we’ve added the following sentence:

“....The expected degree of relevance ranges between 0 and 1. Values closer to 1 can be interpreted as hot extremes are better explained by the relevant variables.”

We’ve added the following sentence to section 3.2:

“....However, in some cases the cumulative degree of relevance of variables can exceed 1 due to collinearities among the variables. Typically, however, this is not the case as shown in Figures 4 and A5. This indicates that dependencies between driver variables are not critically affecting our analysis….”

203: “While EVI is the most relevant driver of hot extremes in more areas at longer time scales (Fig. 2), we find in the main driving variables of hot extremes summarized across climate classes that it also exhibits a higher relevance in these areas but also in other areas where other variables are even more important”
This sentence is somewhat convoluted.

B21: We have updated the sentence in L203:

"Not only the relevance of EVI extends to more regions at 7-day time scale than at 1-day time scale (Fig. 2), we also find that, when summarizing the main drivers of hot extremes across climate classes, EVI's degree of relevance increases in regions where other variables play a more important role."

200-210: This section reads a bit strange in general and seems to make the same point over and over?
"Notably, the relevance of EVI increases with the time scale, in contrast to that of geopotential height, probably due to the longer memory of land surface variables compared to the atmospheric variables"
"This finding highlight that the land surface generally affects hot extremes at longer time scales, as opposed to the more immediate influence of atmospheric drivers."
"This is related to the fact that land surface effects such as evaporative cooling or shading are comparatively smaller but more persistent."
"they are more influential at longer time scales and for hot extremes that build up during a time period without major changes in weather and air masses at a given location"

B22: We agree that the same message has been repeated a couple of times in different parts. We've tried to condensate this into a concise paragraph.

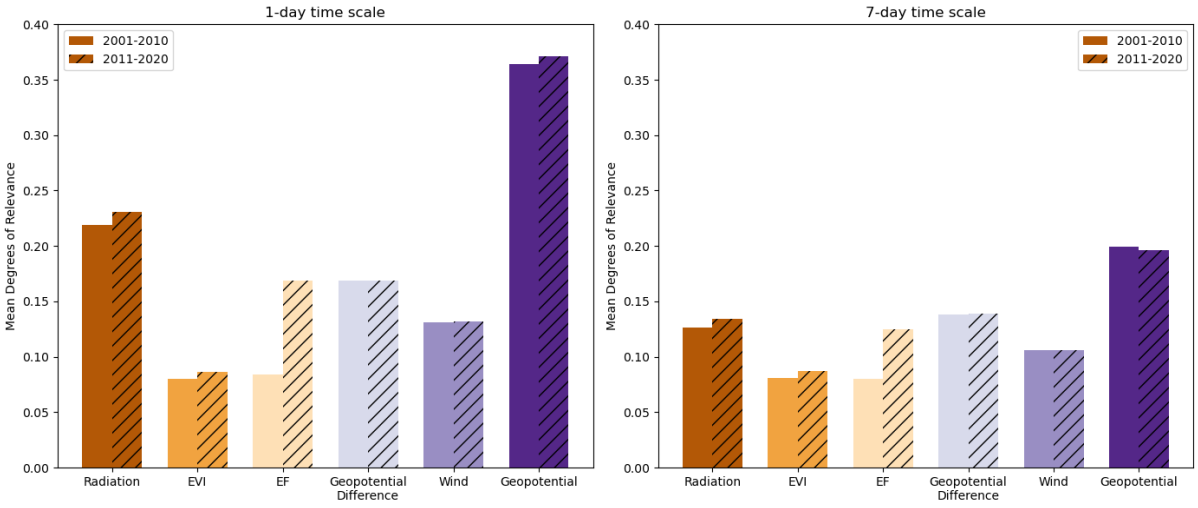
"Notably, the relevance of EVI increases from daily to weekly timescales, likely due to the land surface's 'memory' effects, which allow variables like evaporative cooling and vegetation shading to persist over time. In contrast, atmospheric drivers, such as geopotential height, have a more immediate but shorter-lived influence on hot extremes. This suggests that land surface processes play an important role in driving hot extremes that build up over prolonged periods without major changes in weather or air masses."

215: "Moreover we calculate the sum of the degree of relevance of the three most influential variables at each grid cell (Fig. A5). This shows which part of the observed hot temperature anomalies can be explained with our approach" I think I might misunderstand something here (see also my earlier comment on this). The temperature anomalies from the analogues of different drivers could sum up to more than the observed anomaly, right? So I'm not sure about the interpretation of this.

B23: The reviewer is correct; the anomalies could sum up to more than the observed temperature anomaly (see also response B20). Typically, however, this is not the case as shown in Figures 4 and A5. This indicates that dependencies between driver variables are not critically affecting our analysis. Furthermore, the idea of Figure A5 is to show spatial variations in the explained fraction of the temperature anomalies. We will clarify these points in the revised manuscript in section 3.2 (response B20).

Figure 5: set the maximum of the y-axis to ~.4 to avoid large empty spaces?

B24: Thanks for the suggestion. We have implemented it:



I'd like to see some kind of significance measure for these changes. It seems like apart from EF, none of them are significant even though the authors seem to indicate the opposite in line 233: "At the same time, the relevance of geopotential height, radiation and wind slightly decrease."

B25: We will update Figure 5 with the significance measures for these changes as we mentioned in our response B5.

247: "This finding underscores the significant role of atmospheric blocking mechanisms in the formation of hot extremes" I would assume that most positive GPH anomalies are not blocks even at mid latitudes? In particular on a time-scale of 1 day?

B26: That is a good point, thank you. We can rephrase this sentence:
"This finding underscores the important role of atmospheric circulation anomalies, such as atmospheric blocking, in the formation of hot extremes (Pfahl & Wernli, 2012)."

- Pfahl, S., and H. Wernli (2012), Quantifying the relevance of atmospheric blocking for co-located temperature extremes in the Northern Hemisphere on (sub-)daily time scales, Geophys. Res. Lett., 39, L12807, doi:10.1029/2012GL052261.