Final author comments on "Global characterization of the vertical temperature anomaly structure of heat extremes over land in ERA5"

by Belinda Hotz, Heini Wernli, Matthias Röthlisberger, and Robin Noyelle

General comments to the reviewers:

We thank both reviewers for their thoughtful and positive feedback on our manuscript. The reviewers' comments are indicated in black, and our responses are in blue.

Reviewer 1:

Summary:

The authors examine vertical temperature profiles in the ERA5 reanalysis occurring during (and just before, and just after) annual maximum 2-m temperature events ("TXx"). After suitable normalization, a clustering algorithm is used to better understand the main categories of temperature profiles. Their clustering procedure identifies three distinct clusters, that approximately correspond to the tropics, midlatitudes and high latitudes.

Major comments:

This is an interesting study. The paper is written very well. I thought that the insight, presented towards the end of the paper, that especially extreme heat waves have a more "typical" structure, in line with large deviation theory, was especially interesting. The manuscript appears suitable for publication after some minor revisions.

We thank the reviewer for the overall very positive assessment of our work and the valuable comments below.

I have one main critique of the study. There is a good deal of discussion on the physical mechanisms underlying the cluster distributions, but the discussion is also mainly qualitative and a bit vague. There is a level of subjectivity inherent to any clustering analysis. Some clearer physical insight would help the reader understand whether or not the clustering algorithm is identifying something physically meaningful or not. For example, the fact that parts of Uruguay and southern Brazil are in the polar cluster (Fig. 2a, discussed on lines 297-304) could be potentially very interesting if there is some deeper physical meaning to that classification; alternatively, it might just be a boring artifact of the classification (given this result disappears on using six clusters, rather than three, I suspect the latter). If clearer physical explanations can't be provided (which may well be a challenging task beyond the scope of this manuscript) then it would be good to see a little more reflection on this limitation in the manuscript itself.

Thank you for your comments. The temperature anomalies are quantitative, even though only indirectly. Based on the existing research, we proposed to relate our results to known physical mechanisms. A direct quantification of the processes associated with the clusters, for example by employing the Lagrangian temperature anomaly decomposition, is beyond the scope of this manuscript.

However, we looked in more detail at the mentioned "clustering artifacts" in South America. The region around Uruguay and southern Brazil has a different vertical structure compared to the surrounding regions (Fig. C3). In the literature, this region is often referred to as subtropical southeastern South America (SESA) (e.g. Gelbrecht et al., 2021). SESA is located just south of the South American Monsoon region (e.g. Gelbrecht et al., 2021). Therefore, north of SESA, TXx events tend to occur in autumn (during the monsoon), whereas in the SESA region, the extremes occur rather in boreal winter (see Fig. S1). The SESA region is still subtropical and less influenced by blockings than the regions further south (Rodrigues and Woollings, 2017). Therefore, the SESA region features relatively deep PBLs and relatively weak upper tropospheric temperature anomalies (T'_{scaled}). The relatively weak upper-tropospheric values of T'_{scaled} are similar to the anomalies in the polar cluster, leading to the polar classification. Although this region does not feature the same physical mechanisms as the typical polar cluster grid points, the clustering correctly points out the difference in the vertical structure of the SESA region compared to the surrounding regions. We acknowledge the limitations of the clustering when it comes to distinguishing between physical processes. Those limitations are mitigated when using 6 clusters, as pointed out by the reviewer.

We will add in L301: "The different characteristics of this region likely come from the different seasonality of the TXx events compared to the monsoon region further north (see Fig. S1) and the reduced co-location of upper-level blocking anticyclones during these events compared to the region further south (Rodrigues and Woollings, 2017)."

Furthermore, in L 494: "... TXx events *indirectly* highlight distinct physical mechanisms."

Specific comments:

Line 60: clarify the term "moisture availability", as this can have several different meanings. What specific water flux or storage is being referred to here?

Thank you for pointing this out. We meant the soil moisture storage and will adapt it to "... reflect variation in the soil moisture availability and ..."

Line 107: I think it's reasonable to use the ERA5 reanalysis in this study, but there needs to be more discussion of its limitations here. Yes, much of the vertical structure is probably fine, but the boundary layer and near-surface vertical structure is probably rather poor, and those aspects of the profile are arguably most relevant because that is where people live.

We agree with the limitations raised by the reviewer. The ERA5 reanalysis has important limitations in its representation of the boundary layer and near-surface vertical structure. The near-surface maximum temperatures are generally underestimated in ERA5 (Lopez et al., 2024), which might lead to a weaker vertical scaling (i.e. higher T_{scaled} in the mid- and upper troposphere). Furthermore, Nicolas and Hotz (2026) have shown that ERA5 can have biases in super-adiabatic layers in elevated arid areas compared to observational soundings, which could alter the vertical near-surface structure. Additionally, ERA5 tends to overestimate the surface latent heat flux over land, indicating that the partitioning of surface energy to heat and drought stress is biased (Martens et al., 2020), which can similarly change the boundary layer structure. However, based on regional studies, the correspondence of ERA5 to observational soundings is relatively good (e.g. Cao et al., 2026).

We will add this discussion to L108.

Line 122: typo, “1’533” should be “1,533”

Thanks for pointing this out. We will change it in the revised manuscript.

Line 130: While this normalization procedure all sounds reasonable enough, it would be good to more clearly motivate some of the more arbitrary choices made here. For example, why not calculate \sigma(i,j,k) in a manner analogous to the calculation of T_clim, using a moving window?

For the climatological temperature, we used the methods for detrending and de-seasonalising following Röthlisberger and Papritz (2023) and Hotz et al. (2024). By doing so, we ensure the comparability of the TXx events despite climate change and seasonality. We then use the normalised temperature anomalies to compare events across the globe (i.e. different variability, see Fig. R5). In general, the estimation of a variance is more variable over time (and thus more uncertain) than the estimation of a mean. If we compute the standard deviation over the same sample of data as the climatological temperature, the standard deviation is not as robust because we do not use enough data points (see day-to-day variability in Fig. R1). Therefore, we decided to increase the sample size by using a 21-day centred window across all 73 years. Furthermore, the change of the variance $\sigma(i, j, k)$ with climate change is relatively small compared to the change of the mean (Fischer et al., 2012). We will revise this section in the revised manuscript.

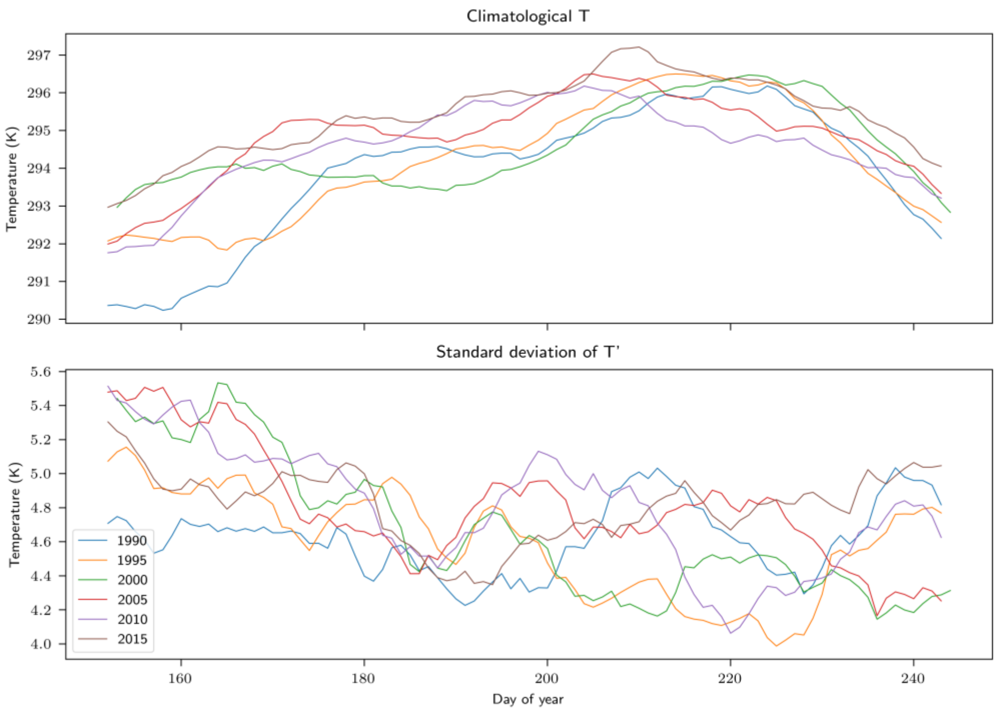

Fig. R1: Day-to-day variability of (top row) the climatological temperature and (bottom row) standard deviation, computed over a 21-day and 9-year centered window, during June-July-August at an example grid point (48° N, 12°E) for the years 1990, 1995, 2000, 2005, 2010, and 2015.

Section 2.3: this is very clearly explained. Well done.

Thank you! We are glad that the section came through clearly and appreciate the encouraging words.

Figure 2: caption has typos, “26’325” should be “26,325” and same problem for other numbers reported in the caption. Please check for and fix this issue throughout the manuscript.

Thanks for pointing this out. We will check the manuscript and will make the changes accordingly.

Line 245: some discussion of the literature on the “weak temperature gradient” approximation would be appropriate here.

Rereading this paragraph, we realised that we can further streamline this section and thereby also add a discussion of the “weak temperature gradient” approximation. From L240 to L245 we will change the order of some of our arguments and add “The absence of an upper-level temperature anomaly is consistent with the weak horizontal temperature gradient in the tropical free troposphere. This gradient is typically weak since gravity waves efficiently adjust the free-tropospheric temperature profile to a moist adiabatic profile throughout the tropics (Nicholls et al., 1991; Sobel and Bretherton, 2000; Holloway and Neelin, 2007).”.

Section 3.2 and Figure 3 is very nicely done.

Thank you!

Reviewer 2: Aaron Donohoe

This manuscript presents global analyses of the vertical structure of the warmest days of the global land masses and argues that; i. the characteristic vertical structure of warm extremes fall into three general clusters organized by climatic zone; ii. Variability of the vertical structure of warm events primarily reflects inter-event variability at each grid point as opposed to departure of the local median profile from the cluster median and iii. the most extreme events are well (if not better than the moderate event) represented by the cluster median profile. I really enjoyed reading this paper. The questions asked are novel as are the analysis techniques used and I think the analysis generally (though not completely) supports the conclusion and interpretation. I have a bunch of comments but think this paper is nearly publishable as is without changing the analyses but clarifying why the methodological choices were made and how to interpret them. However, there are certain methodological decisions that, if I have understood the techniques correctly, I think could be altered to better support the physical questions being asked. The length of my review mostly reflects the novelty of the analysis techniques bringing up questions of how to interpret things and not the quality of the manuscript.

We also thank our second reviewer, Aaron Donohoe, for his thoughtful and positive feedback. We genuinely appreciate the time and care you put into engaging with our methodological decisions.

Major comments:

Normalization by vertically resolved standard deviation and the implication for interpretation of extreme warm events in physical units. If I understand the order of operations correctly, the temperature at each grid point is first normalized by the ***vertically resolved standard deviation*** and then scaled by the surface standardized anomaly. As such, the TXx events should not be interpreted as the vertical profile of events in temperature units but in units of vertical resolved standard deviations. For example, if all TXx events were vertically homogenous (heating the whole column uniformly) the TXx vertical profile would not be homogenous if the temperature variance had a vertical structure (would likely be amplified aloft assuming temperature variance is highest at the surface). Perhaps I've misunderstood or thought through this wrong and perhaps this was choice was intentional – I think many parts of the text are consistent with this definition (especially the text centered on line 165). However, there are times (including the abstract and discussion of the profiles in lines centered by 250) where the profiles are referred to as "bottom heavy" and "amplified" etc. implying that the vertical profiles reflect the magnitude of the temperature anomalies during warm events in temperature units and not vertically resolved standardized anomalies.

My personal preference would be to convert the profiles back to temperature units by multiplying by the vertically resolved standard deviation (to be interpreted as the temperature anomaly per standard deviation surface event). Alternatively, if the Author's wish to retain dimensionless profiles, they could multiply by the ratio of the vertically resolved standard deviation to the surface standard deviation. I think these choices would be more consistent with the physical questions being asked and that, as written, most readers will interpret the profiles as the relative temperature anomalies at each level during a warm event.

We thank the reviewer for this thoughtful comment and agree that the normalisation procedure has important implications for the interpretation of the vertical profiles.

Our analysis first normalises temperature anomalies at each level by their level-specific standard deviation and subsequently scales the profiles by the standardised surface anomaly. Consequently, the resulting profiles should indeed be interpreted as dimensionless, standardized temperature anomalies rather than as temperature anomalies in physical units. Our primary motivation for retaining this framework is that it allows a direct comparison of profile shapes across regions with very different climatological variability and TXx intensities. We considered the suggestion of transforming the profiles back to temperature units. However, such a conversion would also require additional assumptions regarding both the representative surface-normalised temperature anomaly and the vertically resolved standard deviations. We therefore viewed it as methodologically cleaner and more transparent to retain the dimensionless framework.

As the reviewer correctly noted, if all TXx events were vertically homogeneous (heating the whole column uniformly), the TXx vertical profile of T'_{scaled} would not be homogeneous (with the given standard deviation profile). Furthermore, it is also correct that temperature variance is generally largest near the surface and decreases with height, except in polar regions (see Fig. R2). Therefore, the normalisation affects the apparent vertical structure of the profiles. Nevertheless, profiles exhibiting large values near the surface in the scaled framework generally also correspond to "bottom-heavy" structures in both the underlying scaled temperature anomalies and in temperature anomalies. We acknowledge, however, that terms such as *bottom-heavy* and *amplified* can be misleading if interpreted as referring directly to temperature anomalies in physical units. We will rephrase these statements.

To further aid interpretation, we will add Fig. R2 in this reply document to the Appendix, which shows the vertical structure of the standard deviation of the temperature anomalies. Furthermore, we will add in L127 "Note that $\sigma(i, j, k)$ is largest near the surface and decreases with altitude, except for polar regions (Fig. R2 in this reply document)." and in L253 "The temperature anomaly profile is bottom-heavy, with the largest temperature anomalies in terms of T'_{scaled} and T' near the surface and in the lower troposphere since the standard deviation decreases from the surface to free troposphere (Fig. R2 in this reply document)." and in L265 "As discussed in the introduction, climatologically, polar regions feature strong temperature inversions over the cold surface, leading to thin stable boundary layers with relatively small temperature variability (see Figs. A1b and R2 in this reply document)." We hope this additional information will clarify the relationship between the standardised profiles and the underlying temperature variability.

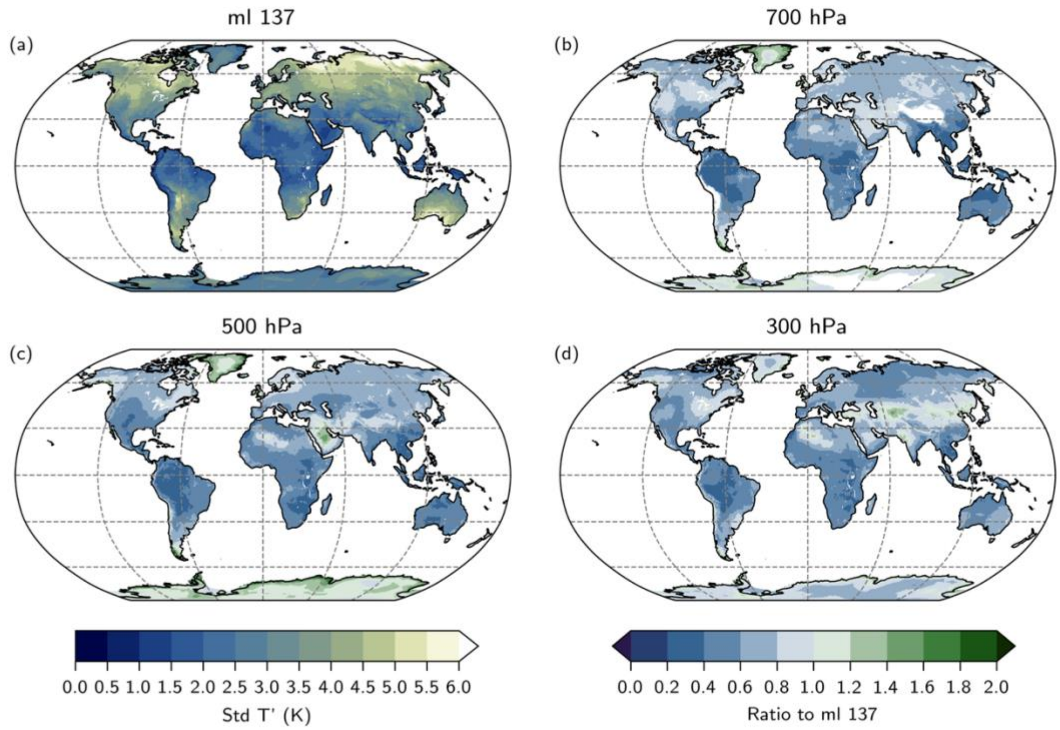

Fig. R2: Global distribution of the mean standard deviation of the temperature anomaly during TXx events (a) at the lowest model level (ml 137) and the ratio of the anomaly at (b) 700 hPa, (c) 500 hPa, and (d) 300 hPa to the one at the lowest model level. The white areas over land in (b) are where the surface pressure is below 700 hPa.

Impact of seasonality and diurnal cycle on the vertical structure of TXx events. The first step of the analysis is to normalize the temperature anomaly by a seasonally (but not interannually) varying standard deviation. This is logical for certain purposes but has some important impacts (I think, but may have misunderstood): i. TXx events at a given gridpoint represent a blend events from different seasons, with different climatological mean state atmospheric thermal structures; ii. extreme events from different seasons may be labeled as a TXx event, even if events in the same year have larger absolute magnitudes. I would have thought these factors have important implications in the Arctic particularly: i. a winter TXx event which erodes the climatological inversion is likely more surface amplified than a summer event and ii. a summer TXx event may be very small absolute magnitude, especially in the marginal ice zone where the standard deviation is very small since temperature is locked to the freezing temperature of ice. I would have thought that seasonality is contributing substantially to the inter TXx variability of the vertical structure and, if grouped by seasonality, the vertical structure is less varied across events than stated in the manuscript.

The Authors state (line 140) that TXx events primarily occur in the summer which, if true (I believe the Authors) can only occur if the summer T2m distribution is positively skewed (since the data is normalized by a seasonally varying standard deviation, unless I missed something). This in itself is interesting, but I wonder if it is universally true, or if some of the spatial inhomegeneity reflects the underlying skewness projecting onto whether summer or winter events are identified. I don't think the Author's need to redo the analysis to address these points but should simply mention some of the drawbacks of generalizing the analyses across seasons.

Similarly, my understanding is that the same considerations above apply the diurnal cycle, since the normalization is done by dividing by the standard deviation at each 3 hourly time step independently. If I understand this correctly, some TXx events are warm nights and potentially are less anomalous in the absolute sense than daytime highs if the nighttime temperature variance is smaller than the daytime temperature variance (I believe this to be the case in most locations). Perhaps the temporal evolution in Figure 3 already addresses this point though its unclear if those results are daily average or instantaneous only at the 3-hour instant of the TXx event. A sanity check on the impact of diurnally varying normalization would be to plot the Tnorm temporal evolution of the median event a 3 hourly resolution to see how jumpy the structure is. I've seen some preliminary analysis suggesting that daily max and daily min temperature are poorly correlated and this gives me pause. I think in general, the discussion in the manuscript is framed in terms of daytime warm events controlled by surface heating and if the TXx profiles are evening events, this might have to be rethought. I don't know enough about the skewness of daytime versus evening temperature to speculate on whether TXx events are primarily daytime (I would guess daytime temperature is more positively skewed and, thus, TXx events are during the day). Again, I don't think any analysis needs to be redone but perhaps some discussion could guide the reader.

Thank you for your comments. We believe that there is a misunderstanding of our definition of the TXx extremes. The extremes that we investigate at each grid point are the highest 2-meter temperature in **absolute** terms and not in terms of anomalies of each year (so-called TXx events). Even though the normalisation section comes before the heat-extreme definition section, we identify TXx events using absolute temperature. For the days of the TXx events, we then extract the normalised temperature anomaly profiles. Therefore, the TXx events usually occur roughly on the same day of the year, with a small standard deviation of about 5 days (see Supplementary Fig. S1 in our manuscript). Based on Fig. S1b, we argue that at a given grid point, TXx events tend to occur in the same season, implying that the scaled profiles we produce tend to have similar climatologies and standard deviations.

Generally, in the extratropics and polar regions, we identify events during the respective summer season (see Supplementary Fig. S1a). However, there are a few regions, mostly monsoon regions, that differ from this by having their TXx events more in the respective spring season. We will change the text in L132 to "In this study, we define heat extremes based on the annual maximum three-hourly 2-m absolute temperatures at each grid point, so-called TXx events." and in L139, we add "For those events, we then extract the temperature anomaly and the normalised temperature anomaly." The sentence "TXx events usually occur during summer in their respective hemispheres, except in monsoon regions and near the equator (Fig. S1)" will be removed and we only refer to the seasonality in L147: "... a season of approximately 90 days (see Fig. S1)."

We agree that mixing daytime and nighttime events would be problematic because it would involve different vertical structures. However, we expect that TXx events occur during the day. We added a figure showing the time of day when the TXx events occur, indicating that land TXx events occur in the afternoon/early evening (Fig. R3). As raised by the reviewer, the nighttime normalised temperature structure can look quite differently. Therefore, we decided to only show the daytime structure: -1/3/5 days and +1/2/3 days relative to the exact TXx event time.

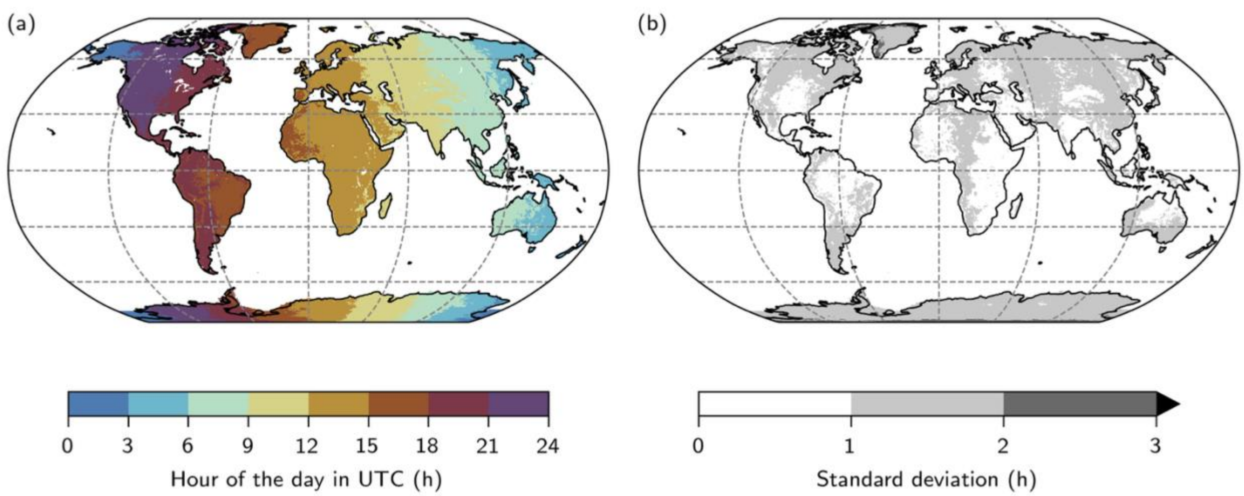

Fig. R3: Global distribution of (a) the mean hour in UTC time of TXx events from 1951 to 2023 at each grid point, and (b) the corresponding standard deviation. The hour of the day for the 73 TXx events at each grid point is computed using a circular mean over 24 hours.

I'm not convinced that the analysis presented supports the idea that the most extreme events have vertical structures that are more similar to the median structure than moderate events. Though I anecdotally believe this is true, I'm not sure the two figures (6 and 7) in this section demonstrate that point. If I understood what was plotted correctly, the analysis in Fig. 6 suggests that the vertical structure of extreme events are more similar to each other (not to the median) as compared to the similarity of moderate events to each other. Additionally, I think normalization by a bigger surface Tnorm in the most extreme events inherhently mutes the variability in the 10 most extreme events. Stated otherwise, if the two samples (extreme versus moderate) had identical absolute temperature variability in the free troposphere, the extreme sample would have reduced Tscaled due to dividing by a greater Tnorm_surface. Perhaps this is just a matter of how the question being asked is framed. It seems that Figure 7 suffers from the same problem. If you started from a null hypothesis that there is similar noise (comparable absolute temperature variability) unrelated to the surface event in the upper atmosphere, one would expect the vertically averaged variance to scale as one over the square root of the normalized temperature anomaly – which seems to be a good fit to the distribution in Fig 7A. Again, I might have missed something here. It seems like the right test is to take some measure of the error between (the vertically resolved) Tscaled in the extreme events compared to the median profile and then compare to the same analysis in the moderate event case.

Analogously to Sect. 4 in the manuscript and Eq. 9, we define $\mathrm{var}(T'_{scaled})$ as

$\mathrm{var}(T'_{scaled}) := \frac{1}{N_n}\sum_n^{N_n}\left(T'^{n}_{scaled}(i,j,k) - \overline{T'}_{scaled}(i,j,k)\right)^2$, where $\overline{T'}_{scaled}(i,j,k)$ is the median of all the TXx profiles at each grid point and $T'^{n}_{scaled}(i,j,k)$ is the TXx profile. Therefore, Figs. 6 and 7 show that the most extreme TXx events follow more closely the median profile of all TXx events at that grid point and, thus, do not necessarily need to be closer to each other (as they can be equally distributed around the median profile). Figure R4 shows the vertically resolved $\mathrm{var}(T'_{scaled})$ for the hottest TXx event and the ratio of $\mathrm{var}(T'_{scaled})$ between the hottest 10 TXx events and moderate TXx events. The ratio is constant across the vertical layers, indicating that the hottest TXx events are systematically closer to the median throughout the vertical profile than more moderate TXx events.

We agree with your null hypothesis that if there is a similar noise in the normalised temperature anomaly, the scaled temperature anomaly would decrease. Therefore, the relative noise decreases with σ^2 (slope of -2 in a log-log phase space of Fig. R5c). For example, both a 2-σ-event and a 4-σ-event have an error of 0.5 σ in the normalised temperature anomalies. However, the relative error for the 4-σ-event is actually four times as low as for the 2-σ-event. We therefore interpret the result as evidence that the relative importance of profile-to-profile variability decreases as event magnitude increases. We will mention this example in the text in L468. Furthermore, we replace Fig. 7 of the originally submitted manuscript with Fig. R5 in this reply document, which has an additional subpanel (c), showing this relationship in the log-log phase space, revealing a slope of -2.01.

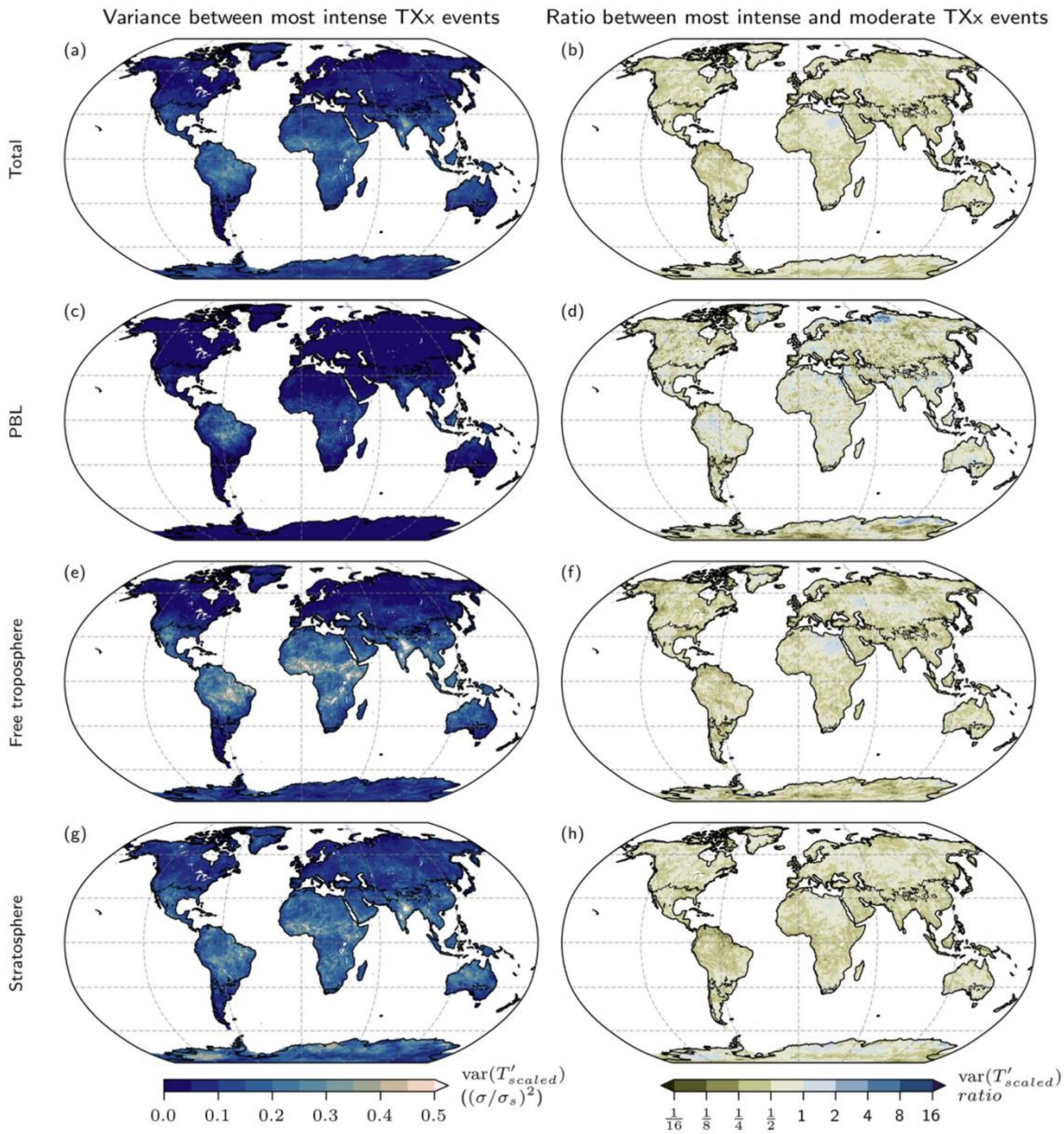

Figure R4: (Left) Variance of T'_{scaled} profiles for the 10 hottest TXx events relative to the median profile of all TXx events ($var(T'_{scaled})$), and (right) the ratio of $var(T'_{scaled})$ between the hottest 10 TXx events and moderate TXx events (which exclude the hottest and coldest 10 TXx events). (a, b) show all hybrid sigma–pressure levels, (c, d) only the PBL, (e, f) only the free troposphere, and (g, h) only the stratosphere. The brown lines denote the boundaries between the three clusters. The brown lines denote the boundaries between the three clusters.

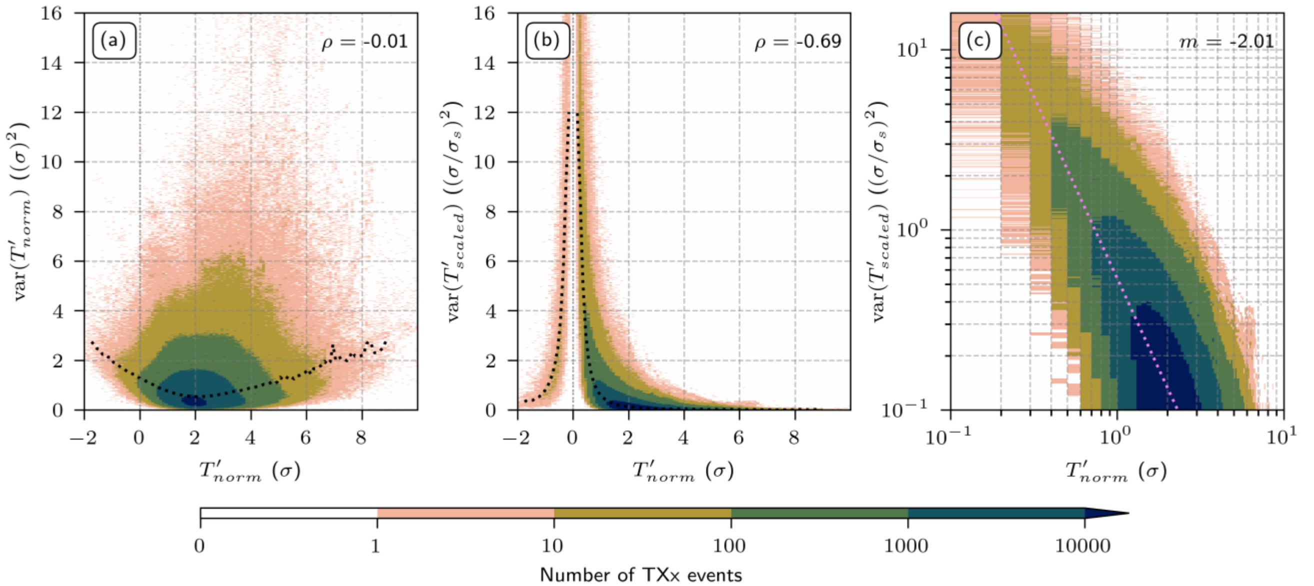

Figure R5: Joint distributions of the mean vertical variance of individual TXx event profiles relative to their median profile and near-surface $T'_{scaled}(k=0)$ (lowest model level), (a) for T'_{norm}, (b) T'_{scaled}, and (c) T'_{scaled} in log-log-scale. The values of ρ in the upper right corner in (a, b) denote the Spearman correlation (both p-values $< 10^{-3}$). Bin widths are 0.1 σ(x-axis) and 0.02 σ or 0.02 σ/σ_s (y-axis). The black dotted line marks the weighted median. The violet dotted line and the m value in the upper right corner in (c) indicate the regression line and slope, respectively. Note that the negative $T'_{scaled}(k=0)$ values arise from our event definition and make up 0.03 % of the TXx events.

The role of EOF truncation in the spatial homogeneity of the clusters. This is a pretty technical point related to applying PCA to global Tscaled and then retaining 8 PC/EOF pairs for the analysis. Wouldn't this inherently produce more spatially homogeneity of the vertical profiles of TXx events than is in the raw data? If a gridpoint or small region has a vertical structure of TXx that isn't seen elsewhere, it will not be retained by the 8 PC/EOFs moving forward. This isn't a problem if the analysis after the clustering references the raw Tscaled data (I couldn't tell if it did, or used the truncated data). I'm not really worried about this point since 95% of the variance is retained. But again, this is likely worth mentioning.

Thank you for bringing up this point. Applying a PCA means that we lose some information. We did a sensitivity test by alternating the retaining variance to 90%, leading to 5 principal components, and to 99%, leading to 23 principal components. Performing k-means clustering with n=3 and n=6 clusters yielded a nearly identical classification of the grid points (see Fig. R6 for the six-cluster classification). Therefore, we don't get more spatial heterogeneity by changing the PCA. Our reasoning for using PCA before clustering is as follows: distances between data points become large in high-dimensional phase spaces, due to the so-called curse of dimensionality (e.g. Assent, 2012), so that it is beneficial for clustering methods to decrease the number of dimensions. Since the k-means clustering method uses distance-based computations, having a high-dimensional phase space makes the clustering method less reliable.

It is true that we take the raw T'_{scaled} profiles after computing the clusters since the clustering remains close to identical despite varying the retaining variance of the PCA.

In the revised method section, we will add in L175 "... data set, as high-dimensional phase spaces can make the k-means clustering unreliable (e.g. Assent, 2012)." and in L184: "..., based on the full vertical profile:"

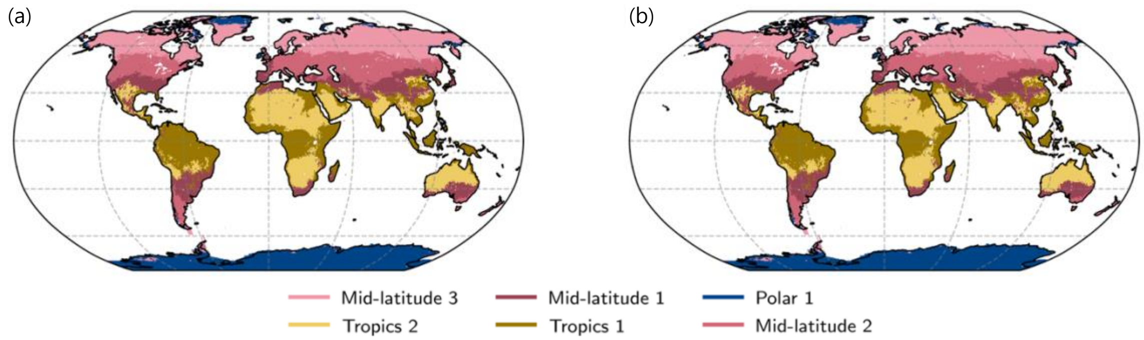

Fig. R6: K-means clustering with six clusters over land (a) with area-weighting five principal components of the full vertical profiles, explaining 90 % of the variance, and (b) with area-weighting 23 principal components, explaining 99 % of the variance. The colours show the distribution of the clusters.

Minor comments

Line 55. Might be worth mentioning that this mechanism assumes weak temperature gradients aloft.

Thank you; we will adapt it in the revised version to "In equatorial regions, dry surface conditions over land enhance sensible heat fluxes, producing steeper lapse rates (larger dT/dp) and consequently higher surface temperatures than over wet surfaces due to the weak horizontal temperature gradients in the tropical free troposphere (Joshi et al., 2008)."

Line 60. What about the role of surface solar variability? Or is the assumption that hot days are clear and thus downwelling solar at the surface is set by clear sky atmospheric transmissivity? Even if the later is true, the magnitude of downwelling solar anomalies at the surface during the hot days depends on the climatological cloudiness. Lucas Vargas Zeppetello has some nice arguments about this in "The climatology of extremely high warm season temperatures...."

Thanks for pointing out the nice work of Vargas Zeppetello (2025). However, we think this work is not relevant in our context as it focuses on changes in the physical processes with warming. The literature on cloudiness during tropical heat extremes is scarce. In the Amazonia region, heat extremes occur under extreme dryness with increased downward solar radiation (Costa et al., 2022).

Line 252. “highly amplified” relative to what? I would call the surface warming in the mid-latitude band moderately amplified relative to the warming in the free troposphere.

Based on the T'_{scaled}, the warming is moderately amplified compared to the free troposphere, but as the temperature anomalies and the normalised temperature anomalies are the largest in the mid-latitudes, they are extremely high compared to the climatology. We agree that it is a bit misleading and will adapt the phrasing to “extremely high temperature anomalies”.

Lines 290. I wonder how defining events in the hybrid coordinate over high topography influences which cluster the location belongs to. If I have understood correctly, you are comparing the behavior of the Greenland surface pressure to that at near sea level pressure in the analysis.

For illustrative purposes, we convert the cluster profiles to pressure levels by using a reference pressure of 1000 hPa. During TXx events, even though the PBL top pressure is relatively low over high topography, the depth of the PBL is smaller than in the surrounding regions (see Fig. A1). This is not only the case in terms of pressure but also on model levels (however, to a smaller extent). Fig. C2 shows the vertical profiles over Greenland. In Fig. C2c and e, the profiles follow the polar cluster profile closely, while the profiles of the grid points in Fig. C2d and e are in the mid-latitude cluster. From Figs. C2 and R7, we argue that the difference in the clustering comes from the variation in the strength of the near-surface temperature inversion layer rather than topography.

We will add in L301: “Note that, aside from unexpected clusters of these regions, elevated regions generally cluster with their expected cluster, indicating that topography itself does not affect the clustering (Figs. 3a and C2).”

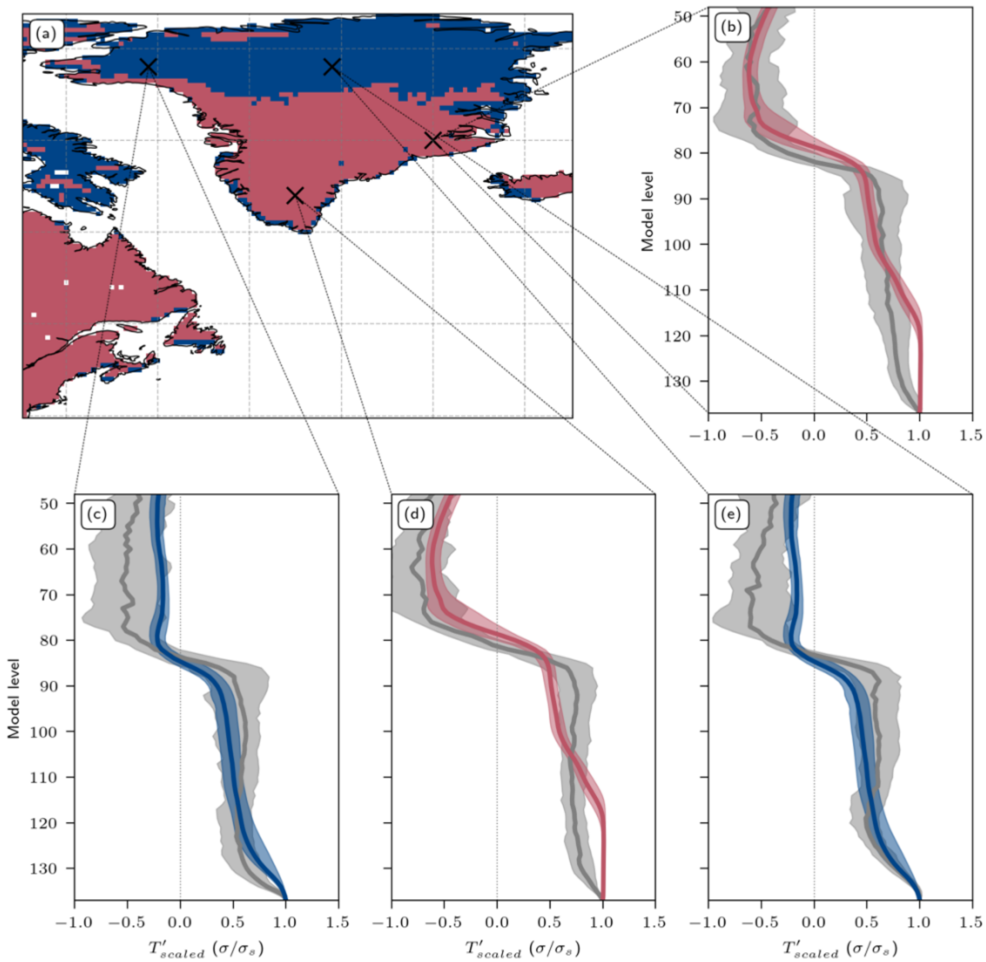

Fig. R7: Area-weighted k-means clustering with three clusters from Fig. 2a shown as (a) a zoomed in map of Greenland and (b-e) vertical T'_{scaled} profiles for individual grid points on model levels, (b) 70° N, 30° W, (c) 78° N, 61° W, (d) 64° N, 45° W, and (e) 78°N, 41°W. The grey line and its shading denote the median and interquartile range of all TXx events at that grid point. The red and blue lines and shading denote the clusters' median and interquartile range of all $\overline{T'}_{scaled}$ within the mid-latitude and polar clusters, respectively.

Line 370. Is there a role of the climatological stability in distinguishing between these regimes?

In the tropical free atmosphere, frequent convective events and mixing by gravity waves result in vertical temperature profiles that closely follow the moist adiabats with weak horizontal gradients (Nicholls et al., 1991; Emanuel et al., 1994; Vallis, 2017). Therefore, the stability in the free troposphere is roughly constant in the tropics. Over dry surfaces, sensible heat fluxes produce steeper lapse rates (dT/dp) and, thus, more stable profiles in the lower troposphere than over wet surfaces. Furthermore, higher temperatures are reached over dry surfaces compared to wet surfaces (Joshi et al., 2008).

Line 380. I wonder (speculate) if the distinction between clusters 2 and 3 is that atmospheric heat transport variability tends to be concentrated higher in the atmosphere towards the pole – I'm not sure this result is in the literature. The papers by Chris Cardinale and Brian Rose might show this.

Thanks for pointing out the interesting work of Cardinale and Rose. They showed that the vertical distribution of the transported heat determines how efficiently it warms the surface over polar regions during winter (Cardinale and Rose, 2022), which is an interesting aspect. However, Cardinale and Rose (2022) did not show the change in atmospheric heat transport variability with latitude.

Towards the pole, the vertical column starts to be more in an advective-radiative equilibrium (Cronin and Jansen, 2016). In Fig. R1, we show that in the mid-latitudes the standard deviation of T' tends to increase towards the poles on a given vertical level. The standard deviation of T' in the mid- and upper troposphere is rather constant (except at 300 hPa, where the standard deviation is smaller than in the free troposphere and near-surface layers). The change in standard deviation of T' does not directly align with the cluster boundaries. We rather argue that the lower tropopause and the deeper PBL dominate the difference between clusters 2 and 3.

References:

Assent, I.: Clustering high dimensional data, WIREs Data Min. Knowl. Discov., 2, 340-350, https://doi.org/10.1002/widm.1062, 2012.

Cao, Y., Yang, Y., Zhang, Y., Lv, X., Ma, Y., Liu, R., Ren, X., Hu, Y., and Xin, J.: Evaluation of FY-3E, CRA, and ERA5 temperature and humidity profiles over North China in summer, Remote Sens., 18, 1058, https://doi.org/10.3390/rs18071058, 2026.

Cardinale, C. J., and Rose, B. E. J.: The Arctic surface heating efficiency of tropospheric energy flux events. J. Climate, 35, 5897–5913, https://doi.org/10.1175/JCLI-D-21-0852.1, 2022.

Costa, D. F., Gomes, H. B., Silva, M. C. L., and Zhou, L.: The most extreme heat waves in Amazonia happened under extreme dryness, Clim. Dynam., 59, 281–295, https://doi.org/10.1007/s00382-021-06134-8, 2022.

Cronin, T. W., and Jansen, M. F.: Analytic radiative-advective equilibrium as a model for high-latitude climate, Geophys. Res. Lett., 43, 449–457, doi:10.1002/2015GL067172, 2016.

Emanuel, K. A., Neelin, J. D., and Bretherton, C. S.: On large-scale circulations in convecting atmospheres, Q. J. Roy. Meteorol. Soc., 120, 1111–1143, https://doi.org/10.1002/qj.49712051902, 1994.

Gelbrecht, M., Boers, N. and Kurths, J: Variability of the low-level circulation of the South American Monsoon analysed with complex networks, Eur. Phys. J.: Spec. Top., 230, 3101–3120, https://doi.org/10.1140/epjs/s11734-021-00187-w, 2021.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., De Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Q. J. Roy. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Holloway, C. E. and Neelin, J. D.: The convective cold top and quasi equilibrium, J. Atmos. Sci., 64, 1467–1487, https://doi.org/10.1175/jas3907.1, 2007.

Joshi, M. M., Gregory, J. M., Webb, M. J., Sexton, D. M. H., and Johns, T. C.: Mechanisms for the land/sea warming contrast exhibited by simulations of climate change, Clim. Dynam., 30, 455–465, https://doi.org/10.1007/s00382-007-0306-1, 2008.

Lopes, F.M., Dutra, E., and Boussetta, S.: Evaluation of daily temperature extremes in the ECMWF operational weather forecasts and ERA5 reanalysis, Atmos., 15, 93, https://doi.org/10.3390/atmos15010093, 2024.

Martens, B., Schumacher, D. L., Wouters, H., Muñoz-Sabater, J., Verhoest, N. E. C., and Miralles, D. G.: Evaluating the land-surface energy partitioning in ERA5, Geosci. Model Dev., 13, 4159–4181, https://doi.org/10.5194/gmd-13-4159-2020, 2020.

Nicholls, M. E., Pielke, R. A., and Cotton, W. R.: Thermally forced gravity waves in an atmosphere at rest, J. Atmos. Sci., 48, 1869–1884, https://doi.org/10.1175/1520-0469(1991)048<1869:tfgwia>2.0.co;2, 1991.

Nicolas, Q. and Hotz, B.: Dry and moist convective upper bounds for near-surface temperatures, Weather Clim. Dynam., 7, 843–856, https://doi.org/10.5194/wcd-7-843-2026, 2026.

Rodrigues, R. R., and Woollings, T.: Impact of atmospheric blocking on South America in austral summer, J. Climate, 30, 1821-1837, https://doi.org/10.1175/JCLI-D-16-0493.1, 2017.

Sobel, A. H. and Bretherton, C. S.: Modeling tropical precipitation in a single column, J. Climate, 13, 4378–4392, https://doi.org/10.1175/1520-0442(2000)013<4378:MTPIAS>2.0.CO;2, 2000.

Vallis, G. K.: Atmospheric and oceanic fluid dynamics: Fundamentals and large-scale circulation, in 2nd Edn., Cambridge University Press, Cambridge, ISBN 978-1-107-06550-5 978-1-107-58841-7, https://doi.org/10.1017/9781107588417, 2017.

Vargas Zeppetello, L. R.: The climatology of extremely high warm-season temperatures and their sensitivity to climate change, Q. J. Roy. Meteorol. Soc.,152, e70048. https://doi.org/10.1002/qj.70048, 2025.