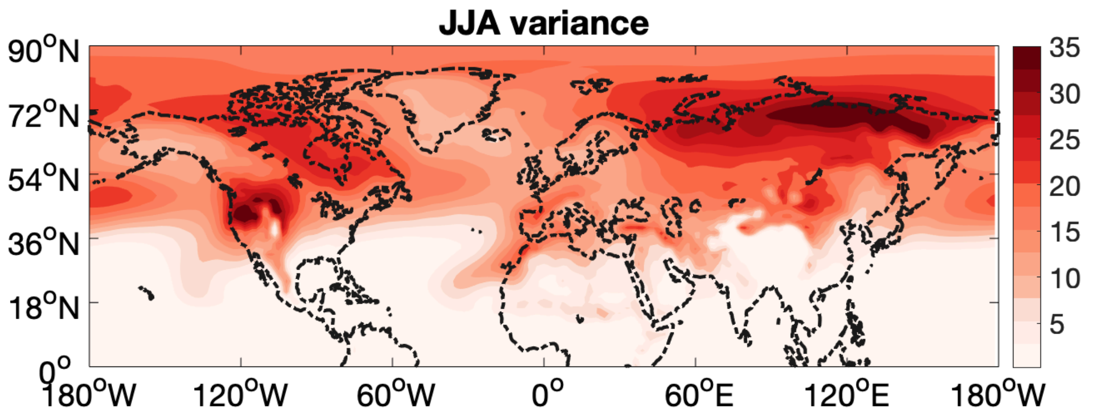
Review of “The substructure of extremely hot summers in the Northern Hemisphere”

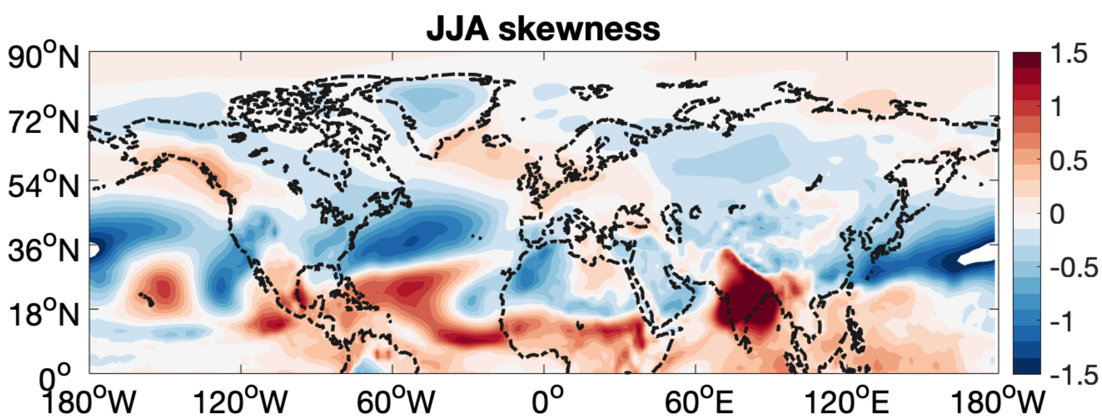
In this paper, the authors introduce the method of calculating rank day anomalies for each summer in order to characterize the distribution of temperatures during extreme summers. The method, as I understand it, is to sort the 92 daily mean temperature values at each location and then calculate the average at each rank. Then for each summer, the deviation from this climatological mean is taken. They find that in the arctic, extreme summers occur when cold days are warmer than usual and in India, the hottest days drive the anomalously extreme summers. A point that I think is particularly important that is made somewhat in passing is that the characteristics of the extreme summers are consistent with the characteristics of the underlying temperature distributions—there is no obvious regime shift or equivalent for the hottest summers. From this perspective, I think this is a useful tool to verify that we can understand extreme seasons by understanding the underlying temperature distributions.

Overall, I find this study to be worthwhile, but a bit confusing. As the authors state, this is a novel method for looking at extreme summers. They do not spend much time justifying the introduction of such a method, and the advantages it has over examining the local temperature distributions themselves or over methods such as looking at compound heatwaves (Baldwin et al. 2019). Indeed, one of my main takeaway messages from this paper was that extreme summers can be relatively well described by understanding the variance and skewness of the underlying temperature distribution (more below). This method proved that particular point quite nicely. If there are other advantages or conclusions that can be drawn uniquely from these metrics, the authors should highlight them. I believe this paper will be suitable for publication after it addresses the following concerns:

Major:

1) As mentioned above, what is the advantage of this method over more typical examinations of temperature distributions? How does the calculation of RDA differ from quantile analysis? How does the comparison of the contributions of the top 33% and the bottom 33% differ from examining skewness? How does the spatial pattern of XA compare to the spatial pattern of temperature variance? I have included plots based on the ERA-I data I had handy (850 hPa, 1980-2014, 4xdaily), but I think the inclusion of ERA-I surface temperature variance and skewness plots is essential. The comparison with Loikith et al. 2018 is pretty impossible given the size of the panels in their Fig 4.





2) The authors need to better justify not somehow accounting for the trend in summertime temperatures in ERA-I (or better yet, they need to account for the trend). The current justification, i.e., "as we are interested in extreme summers exhibiting the largest absolute T2m anomalies and not the largest T2m anomalies relative to a long-term trend" does not make sense in the context of the later discussion. The analysis as currently presented naturally conflates factors associated with global warming with the dynamics associated with internal modes of climate variability.

    e.g. The point in Nevada has 2016, 2017, and 2018 all included in its five most "extreme" summers. The earliest "extreme" summer there is 2007. Surely, then the signal in RDA is one of global warming. And indeed, if we compare this to the results of McKinnon et al. (2016a) Figure 4, we see a warming of the whole distribution and the largest warming in the bottom quantiles. This then seems to be an examination of the forced response rather than internal variability. Meanwhile the authors argue, quite convincingly, that the extreme summers in India are related to the timing of the monsoon onset, a signal too strong to be dominated by global warming.

3) "Substructure" is not really an appropriate representation of what is studied in this paper. This study is not detailing the relative timing and duration of heatwaves—indeed all temporal ordering is lost in the novel method introduced here. Substructure as I would typically understand it is considered in Fig. 1 and Fig. 8 only. This isn't such a major point about the importance of the paper, but it will require some thought as to a more appropriate term and then significant rewriting.

Other points:
The use of "d" in the equations in combination with the term "substructure" made me mistake "d" for day instead of rank. Consider a different variable name, perhaps? Or explicitly mention that ordering is lost?

l. 144 rewrite for clarity. Perhaps just "allows assessment of"? Consider adding a specific example here.

Consider mentioning which "third" is 30 days so that this calculation is perfectly reproducible
l. 181 Normally → normal

Consider changing the figures so that it is easier to compare ERA-I and CESM. E.g. put Fig 3 a and 4 a together.

Paragraph beginning l. 243: the quantitative spatial correlation value would be helpful here.

Fig. 6: The yellow contours are really difficult to read. Consider having thin dotted lines for continents so that you could use thicker black lines in place of the yellow? Or some other change to make this more readable. Magenta might be better than yellow.

l. 359 normal

Fig. 9: Label lines within the panel b

Analysis of Nevada. Consider work by McKinnon et al. (2016b), which is primarily looking at Eastern US, but their conclusions still seem relevant.

l. 381 Why … the troughs associated with cold anomalies (black contours in 10a) did not occur…

l. 388 It seems like the goal (c.f. Hoskins and Woollings 2015) is to explain the full shape of the temperature PDF, since extreme summers seem consistent with the underlying distribution. But you are correct that a combined approach is necessary for that as well. So maybe just add "… to fully reveal the physical causes of the full shape of the temperature distribution, including extreme summers" or something along those lines?

Paragraph beginning l. 406: This seems like perhaps the major conclusion of this work. Emphasize this more at the beginning.

l. 425 This phrasing is not appropriate. “Often” cannot be determined from these three case studies, and one of the three case studies (US) is in fact a clear case of temperature advection’s importance due to an anomalously zonal jet stream.

Paragraph beginning l. 436: This is completely consistent with the eddy advection argument of Garfinkel and Harnik 2017, Tamarin-Brodsky et al. 2019, and Linz et al. 2018

l. 443 New paragraph

l. 445 Not convinced of this (esp. the coherent regions aspect, since mostly this has looked at individual points) by this particular study.

l. 455 A more zonally symmetric/less wavy flow is still a pattern, so this phrasing doesn’t really make sense.

References:
Baldwin et al. 2019 Temporally Compound Heat Wave Events and Global Warming: An Emerging Hazard. Doi: 10.1029/2018EF000989

Hoskins and Woollings 2015 Persistent Extratropical Regimes and Climate Extremes. DOI 10.1007/s40641-015-0020-8

McKinnon et al. 2016a, The changing shape of Northern Hemisphere summer temperature distributions doi: 10.1002/2016JD025292

McKinnon et al. 2016b Long-lead predictions of eastern United States hot days from Pacific sea surface temperatures doi:10.1038/ngeo2687

Tamarin-Brodsky et al. 2019 A Dynamical Perspective on Atmospheric Temperature Variability and Its Response to Climate Change. DOI: 10.1175/JCLI-D-18-0462.1