# Reviewer’s comment on the manuscript
“Rapid attribution analysis of the extraordinary heatwave on the Pacific Coast of the US and Canada June 2021” Philip Sjoukje et al.

## General comments

The authors try to assess in this manuscript to what extent the June 2021 heatwave affecting the Pacific Northwestern coast of US and Western Canada is attributable to climate change. They use multi-method multi-model attribution, and rely on the output of several (around 20) earth system models, reanalysis data, and meteorological observations. They also analyse qualitatively the meteorological conditions before and during the event, and discuss aspects of vulnerability and exposure.

Except the relevance of the topic, I don’t see any reason for the publication of this manuscript. The quality of the study and of the manuscript is not good enough for publication in a peer reviewed scientific journal. In the following I summarise my main concerns.

## The scientific methods are not adequate.

The Generalized Extreme Value (GEV) distribution is a limiting distribution. This means that the block size has to be large enough to allow for the convergence of the estimated GEV distribution to the asymptotic GEV. This should be checked based on the convergence of the shape parameter as function of increasing block size (see for example Coles, 2001, Springer). **Annual maxima are not automatically GEV distributed**. In fact, looking for example at Fig. 6b, it looks like the curvature of the fitted line would change if one would start the fit at larger return periods (e.g. 5 yr or 10 yr. This suggests that the shape parameter is not converged in case of annual maxima, thus the extrapolation towards unobserved events is not underpinned by the theory. One needs to select maxima over larger blocks, assuming that there is a convergence at all. However, if we would take for example block sizes of 10 years, we would not have enough maxima to estimate the GEV parameters properly, furthermore the uncertainty would substantially increase. Thus, I’m afraid that larger blocks will not solve the problem either, because of the lack of sufficient data. The slow convergence and the “waste” of data by considering only block maxima are common application issues in case of this method.

Now, let’s leave the convergence issue beside and concentrate on the GEV fit. Fig 6b, 7b, and 8b show:
1) if one omits the 2021 value, the fit seems to be appropriate for the rest of the annual maxima but the 2021 value is substantially above the upper limit of the estimated distribution (Fig. 6b);
2) things don’t get better if one considers the 2021 event as well, the fit is much worse generally and the 2021 event still lies outside the confidence intervals (Fig. 8b).

This can have following reasons:
1) the 2021 event belongs to a different parent distribution – this would be a violation of the condition of identically distributed variables, as required by the extreme value law;
2) the empirical return level is underestimated and the event is much rarer then it seems. Indeed, the maxima of a 72 years long times series can be a 1 in 72 years event, but also a 1 one in 100 or maybe a 1 in 1000 year event or even rarer.

The authors account for nonstationarity of extremes only based on a linear trend in the location parameter. However, there might be other sources of non-stationarity as well, leading to unrealistic estimates.

I don't understand at all the point of the method used to produce Fig. 7b. The authors write that "we still assume that the data up to 2020 can be described by a GEV with constant scale and shape parameters, but we reject all GEV models in which the upper bound is below the value observed in 2021". You have one time series based on which you estimate the GEV parameters and their confidence intervals. Where do all these GEV models come from? Besides, the advantage of fitting a GEV distribution is that the data itself "decides" about the optimal parameters. The way you proceed, you lose this advantage and it seems like you take a subjective decision to obtain specific results.

**Based on the above reasons, I think that the block maxima method at the selected block size does not provide a reliable statistical model for the analysed temperature extremes, thus the conclusions of the manuscript are not reliable as well.**

I also have concerns related to the definition of the heatwaves. Actually by taking the annual maxima we define the event at one point in time, thus we lose the information about the duration of the event, which is crucial in case of persistent extreme events, like heatwaves. Heatwaves with very different duration, let's say 3 days vs. 3 weeks, could produce the same annual maxima, and still be caused by different circulation anomalies and physical processes. However, the used definition cannot differentiate between several types of events. Another issue is that the annual maxima are then averaged over the study area. But we don't know if the maxima at different grid points refer to the same event. It is indeed very probable that they do not and in that case we would compute spatial averages of events happening at different time steps, which makes no sense.

<u>The results provide almost no valuable information, thus the conclusions make no sense.</u>

Considering the inadequate methodology and the huge uncertainties in the return periods and probability ratios, I'm afraid that the main conclusions are not trustworthy.

The authors do not state clearly which GEV fit they follow, but based on the final conclusion of "1 in 1000 year" event, I assume that they decide for the one shown in Fig. 8b. One of the final conclusions is that "the event is estimated to be about a 1 in 1000 year event in today's climate". I point out that one cannot obtain correct return level estimates for an event occurring on average once every 1000 years based on **72** values of **annual** maxima.

Another main conclusion is that "an event … as rare as 1 in a 1000 years would have been at least 150 times rarer without human-induced climate change", meaning that the uncertainty bounds for the probability ratio are 150 and infinity. And finally "in a world with 2 °C of global warming … a 1000-year event … would occur roughly every 5 to 10 years". This last statement is obtain based on the best estimate of the probability ratio of 175, with the uncertainty ranging from **3 to infinity**, giving a return period range of from 1 year to 333 years. There is too much uncertainty in these results making them not very useful.

I am puzzled as well by the infinite probability ratios. I assume that they come because of division by 0, which would be problematic since it does not say anything about the magnitude of the numerator.

However, the authors do not define the probability ratio in the paper, thus one can just assume that this is the case.

<u>The language is not clear enough, the formulation not precise enough, and the structure confusing.</u>

The quality of the scientific text is not good enough, and for me it is very strange that it was submitted in this form, considering also the number of authors of this paper.

There are many sentences/paragraphs where I don't get the message mainly because of wrong and/or inaccurate phrasing. Section 2, 3, 6 are particularly critical. I do not understand the meaning of Sec. 7.1 at all, and sentences like "This implies that the return time of an event as rare as this one or rarer, somewhere over land, is about 60 times smaller than the O(1000 yr) that it occurred at the specific location that it did." do not help.

The quality of many figures is not good enough:

- Several figures are direct output of a climate explorer tool with a poor choice of colours, and confusing axis labels. For example, what is "max Tmax"? Tmax is used also in the text, however it is not defined. It took me a while to find out the difference between TXx and Tmax. It is very confusing and it is not explained.
- Fig. 12 and Fig 13 are not self-explanatory. The meaning of the colours is explained in the text, but not in the figure caption.
- Fig 12 shows the synthesis results of the **past** climate compared with **today's** climate. Why is "historical-ssp" written in the figure legend?
- Fig. 15: I do not see the circulation changes described in the text. Assuming that the study area is the same as defined at the beginning of the manuscript, I cannot see at all a "southerly to southwesterly near surface flow" in the study area in panel B.
- Fig. 17: The figure label contains: "Fit: KNMI Climate Explorer". What is fitted in this case?
- In case of the maps, it would be good to mark the study area.

The structure of the paper is confusing:

- There are many sections and some of them are very short. For example, section 5 has 4 lines. It could be incorporated in section 6.
- The paper has no summary joining the main results from the different sections. The different topics (for example, the "statistical estimation", "synthesis results", "meteorological analysis") are like puzzle pieces which are not merged at the end.
- Line 136: "As discussed in section 1.2, we analyse the annual maximum of daily maximum temperatures (TXx)". This is discussed in section 1.1, actually.
- You mention all those different experiments in Sec. 2.2. The CM6A ensemble "is used to explore the influence of climate variability", you use high resolution GFDL and AMIP runs, but you do not explain what we learn from these runs, you just include them, together with all the other models, in the synthesis analysis.
- The second part of the abstract is not clear and precise enough.
- The probability ratio PR is not defined.