

Atmos. Chem. Phys. Discuss., referee comment RC2  
<https://doi.org/10.5194/acp-2021-109-RC2>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Review of acp-2021-109

Sabine Undorf (Referee)

---

Referee comment on "The effect of forced change and unforced variability in heat waves, temperature extremes, and associated population risk in a CO<sub>2</sub>-warmed world" by Jangho Lee et al., Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2021-109-RC2>, 2021

---

### Review of "The effect of forced and unforced variability on heat waves, temperature extremes, and associated population risk in a CO<sub>2</sub>-warmed world"

The authors analyse changes in high-temperature climate extremes over global land regions as projected in a large ensemble of 1%-CO<sub>2</sub> runs with MPI-ESM. Focussing on a suite of standard metrics, they address various questions regarding the role of internal climate variability as well as how the projected changes vary by geographical region and by wealth, and discuss them as a function of global-mean warming levels. In particular, they find that, in the model world, the internal variability is smaller than ensemble-mean model bias; that the co-variability of the occurrence of single days with high absolute temperatures across a range of large coastal cities mainly at lower latitudes resembles modes of internal variability, with ENSO's influence in particular tending to make them to occur simultaneously; that different regions on Earth show distinct characteristics with regard to climatology and projected changes in the high-temperature indices, with the Tropical regions showing the largest increase in heatwave days and disproportionately larger changes above 1.5degC global-mean warming; that very large and growing proportions of the global population experience heat extremes; that the largest present-day values of, projected changes in, and differences between global-mean warming levels in both hot days and warm nights happen in regions of lowest GDP. They also show the projected changes in the need for cooling and heating in the cities considered. The authors highlight that the poorest and hence most vulnerable areas of Earth will hence be exposed most to (changes in) high-temperature extremes.

The paper addresses relevant scientific questions of societal importance, using a suite of innovative methods and an interesting combination of different ways to adress related issues. They use only one but state of the art model and study that comprehensively. No all-surprising conclusions are reached, but a set of aspects better illuminated and, in the model, quantified, and important messages identified. The interpretations and conclusions are sufficiently supported by the results, and the methods and assumptions generally valid, with the caveat of some unclear methodological aspects (see major point #3 below). Overall the methods seem clear/standard enough to allow reproduction attempts by a fellow scientist, but their presentation is not well enough structured yet (#7 below), and

there are many instances of incorrect, imprecise, or otherwise unclear language. Some figure labels are missing units, and the figure captions do not provide enough information for the figures to be understood without reading the text.

## Major points

#1 The authors motivate much of their work by referring to climate impacts as a function of a population's vulnerability. Given this context, it would be good if they could acknowledge the vast amount of literature (see e.g. IPCC WGII's work) there is that conceptualises climate risk and vulnerability, use terms such as >risk<, >exposure<, and >vulnerability< accordingly with more care, and back up their claims on the strong dependency of vulnerability of GDP with references.

Related, minor text comments: 67 that is only one factor. Please avoid misunderstand by saying that the impact of climate extremes on different populations depends on a range of factors, including... 116 "risk" – likelihood? 351-2 "it is well-known that not everyone is equally vulnerable to extreme weather, with richer developed countries having more resources to deal with extreme events" – 408-9 "given underdeveloped countries' lack of ability to endure climate extremes" – problematic phrasing? Firstly – 'least developed' maybe? 'underdeveloped' may sound like the country is deficient and you're judging it for that but check e.g. ipcc wgII terminology or united nations. Secondly, it seems that many developing countries are already enduring more climate extremes than developing countries, and showing much more endurance than any developed country has in recent decades had to show. Thirdly it's too unspecific anyway – are you referring to resilience or adaptive capacity maybe? See general point.

#2 I wonder about the choice of cities (e.g. line 223) and how transferable your results regarding those are to other places. You talk of (line 244) representative cities – representative in which respect? And how do you know? Please specify or don't claim this. Further, a different choice of cities, including e.g. Australia with high GDP >and< high exposure to extremes, for instance, might give a slightly different picture; or including more continental cities might have impacted the impact of ENSO; etc. Also, I worry that these 15 cities give very few degrees of freedom for your EOF analysis. If the authors want to claim these to be representative, they should do a sensitivity study, repeating the respective analysis with 15 different cities, drawn randomly out of a suitable selection based on size and with some coverage. Or with all cities of a certain size, or the largest city in some region, or similar? If not, the authors should be careful not to overgeneralise the results. In any case, a description needs to be added as to how you have chosen these cities, so that the reader can assess the potential impact of selection bias on the results.

Related – you explain the low absolute spread of days of extreme heat in Moscow with low absolute values in the mean, which makes sense of course. However, Moscow is at the same time the only city with truly continental climate, as far as I can see. It's also the only city at a latitude higher than 45deg. And most of the cities you consider are in the Tropics. All of these aspects will also shape how much each of these cities will be impacted by internal variability and on which timescales (atmospheric vs. SSTs). I think some of

this should be discussed.

#3 I am unsure about the bias correction, and I am not sure I can assess the method's correctness given unclear text. Generally: Given the authors have historical model simulations available, why not compare those with observational(ly derived) data for the same period, they should be much more comparable than the CO2 runs. The authors might otherwise end up correcting a bias that is due to the model not having all the forcings (ghgs, lu, aerosols, solar, etc), not due to the model systematically over/underestimating temperature extremes. 15 years also seems rather short.

Minor/further - the authors show the area-averaged bias and show it's near zero - (line 140) unclear whether that is the bias of the averages or the average of the bias? T2m and d2m are corrected separately, and then combined, I'm not sure that is physical, the separate bias-correction might break the physical link between the variables, but that might not be an issue for the subsequent analysis. The bias-correction should further apply the same correction to each member, calculated from the whole ensemble, since it's a single-model w unperturbed physics; I think the authors follow this but not all clear from line 155-6.

I don't understand what the authors mean in lines 177-8. Fig. 2e-f shows only data where the bias is greater than a threshold (lines 150-1); I don't see why? That will impact the average bias quoted. Can the authors further be kind enough to explain to me how the maximum value of what they show in those figures can be around a third. That means that the ensemble-mean bias is at every grid point (much) larger than the ensemble spread, right? Also for t2m; how does that go with their Fig. 1? That would mean really rather large bias? I think I am misunderstanding, so the manuscript could do with clarifications there.

#4 I think different baselines are used in the different parts of the study, w/o motivating this. e.g., 200-201 heatwaves defined based on PI. 380 CDD/HDD days also PI. Fig. 9 present-day baseline (description missing as to how the values compared to the present-day baseline is derived). I see that both baselines are useful, they are just pointing out different aspects (emphasis on all changes due to anthropogenic cc (well, co2 here) vs. changes still to come), but the differing usages should be noted & motivated in the text. Would the conclusions regarding CDD/HDD changes furthermore be more meaningful too for a PD rather than PI baseline?

#5 The conclusions section is not comprehensive enough in terms of the limitations, uncertainties, and caveats (e.g., line 430 etc). Firstly, the results are from 1 model only, and another model may give different results, and the reality be yet again different. An obvious uncertainty here is the sensitivity of the model to CO2 (measured e.g. by TCR); please discuss this see e.g. Mauritsen, T. & Roeckner, E. Tuning the MPI-ESM1.2 global climate model to improve the match With instrumental record warming by lowering Its climate sensitivity. J. Adv. Model. Earth Syst. 12, (2020). Other aspects will be the model-simulated internal variability, that also varies substantially between models (spectral characteristics, magnitude, couplings). Then there might be uncertainty that should be

discussed in how much the real 1.5/2/3/4degC given other climate forcings will be different from the 1.5/2/3/4-CO2 world, which might also be scenario-dependent. Then the paper would gain too from a discussion on what the assumption of unchanged socioeconomic and population (e.g. 113-4) distribution means for the results (I guess for instance that Fig. 8 will show even larger share of the global population exposed to heat extremes considering projected population changes. Please discuss this in the text? Of course there might be feedbacks too but they also will have important implications).

#6 It's a rather trivial and much documented fact in the literature that larger-scale aggregation reduces variability, and I don't see why/how this is worth as much highlighting as done by the authors (in the text, e.g. key point #1; 312 "Notably" – Unsurprisingly? 334 "notable".) More importantly, though, the authors conclude (line 339) that "this [int variability averages out over large averages] indicates that int var will play a minor role in determining global exposure to temp thresholds", but they write earlier that 'no one lives in the average' as a motivation for their study, emphasising hence the need to look at smaller scales. Can they please explain (in the manuscript) how that goes together?

#7 Many methods come as a surprise in the results section w/o being mentioned in the earlier methods section, and are then not sufficiently explained. Some variables, regions, indices, are not introduced either (see specific comments below). Please can the authors make sure to remedy this. Examples are 223/Fig 3a, info on how the values for cities are calculated (surrounding 3x3 grid boxes) should not just be in the caption but also in the text; 242, EOF should be in the methods too; 257-9, it is not explained in the text (partially in the figure caption only) how you calculate ENSO, PDO, and AMO. Etc.

### **Minor/other/text**

There are very many typos/minor language errors that need to be fixed (e.g., line 42 "efforts and to", 45 "value(H", 47 "various assessments", 48 "assessments and those by the IPCC", 63 "links", 65 "the PDO" and "the AMO" (following conventions in the literature; also should not be "the ERA-Interim" but in turn "the ECMWF", "the CDF-t method" etc), 66 "investigated in" etc.) Some language is a bit too colloquial too (e.g., 72 "run", 73 "runs", 91 "go", 106 better "provide" than "has", 398 "stats" etc.). The tenses should be consistent too (e.g., 136 compare 137 compared), also 130 "Figure 1a and 1b" 143 "Figures 2a and 2b" etc.

- Title. I think w/ 'forced variability' you mean the long-term change in response to CO2? The variability itself too could change in response to forcing, so I find this not ideal

- Key Points. Need to add 'in one model' or 'in the MPI-ESM model' to all. #1 w/o reading the paper, I think that sounds like it's the int var rather than changes in co2 that are driving extreme events. #2 'by using the large ensembles' confusing here – that refers to 'is shown', not 'to reduce', I think, but everything you do is with a large ensemble anyway.

- 46 regional extreme heat events and heat waves – isn't a heatwave also an heat event?

- 52 risk ratio – of what? I think it can't be listed like this, it's a comparison method not an index in itself

- 55 probably unclear to the reader less familiar to heatwave indices how the mean of a heat wave is different from its amplitude and how the mentioning of consecutive-day differs from duration; so maybe this can be clarified

- 60-62 I don't follow this reasoning, can that be clearer please. Is it that the future change in risk is due to forcing, so to assess that it's important to tell how much was due to forcing in the first place, and not due to variability?

- 66 side note - there is a large body of literature discussing to degree to which the observed AMO is actually unforced or not...

- 72 install and operate? It also needs buildings in the first place.. not sure everyone has them everywhere, thinking of the most vulnerable in a society

- 73/74 Which model? What model experiment? This information is not complete

- 77/78 please clarify, could also mean the economic status within societies or the economic status >during< extreme events. Similar line 116 – sounds like the wealth defines the extreme event hazard like the global-mean temperature levels do

- 94 was that the model resolution? Then specify that please, that is what matters more I think that what data your analysis starts with

- 98 climate projections

- 100/101 can you please explain to the ignorant reader why using the ensemble allows you to estimate the effect of unforced variability (initial-condition ensemble, a range of realisations of internal variability, etc)

- 103 how many RCP runs?
  
- 111 average or sum – confusing. 118-120 confusing too, total or per capita in the end?
  
- 127 should repeat annual her (or why repeat global?) - important since it makes of course a difference for the timing of reaching a warming level whether you look at monthly or annual exceedences
  
- 128 why not cite 3degC too? That's shown in Fig 3 etc
  
- 129 4.6. 236, 54, 53, 57, 50,50,51,52 etc. (be consistent)
  
- 140 near-land needs to be defined
  
- 143-5 syntax; difference in what?
  
- 151 why median, not mean? should make a negligible difference, but i think the mean makes more sense when it's about averaging out random variability
  
- 152 Do the authors mean grid-point with 'region'? Please avoid, that's confusing. Or clarify otherwise
  
- 153 13% - average across what? Which line in the plot is that referring to? unclear
  
- 161 realistically – 'reasonably'? 'reliably'? will be more correct
  
- 163 rh is not introduced. w2m not introduced.
  
- 167 i think the 'runs' is in common usage the 'experiment' itself, not the data created. And you don't correct the experiments so I would think 'bias-corrected data from the xx run' is more correct. But I also deem the word 'run' rather colloquial.

- 169 estimated by these >runs<, not by these >models< - the model can be run in other configurations with those other forcings. (I know anything can be a 'model' but here I think it will be understood to refer to the MPI-ESM.) Also, what about other forcings like land use change and GHGs other than CO2?

- 175 regions? Which regions? Not explained anywhere. Also - a difference of 0.5degC seems >very much< compared to 1.5, 2, 3 degC, not "very small" as you write!

- 176 the reanalysis data contains some >response to< aerosol forcing, not the forcing itself, I think. And 178 the effect of aerosols fixed at/to >that of the> 2003-2017 period

- 181-186 I find this more confusing than necessary. Please clarify. Also, your definition means that you have not only a grid-point dependence, but also a seasonal dependence, so a heatwave in summer must be hotter than a heatwave in winter to be a heatwave, right? Worth noting I'd say.

- 201-203 this is not very well explained. I think you mean that heat wave thresholds are different ... because they are based on the respective values in pre-industrial times, this means that heat waves with the same index value will refer, region- and season- dependent, to heatwaves with different absolute values, which is from an adaptation/impacts perspective meaningful because there will exist some degree of adaptation to the existing climate conditions, but that there are on the other hand also physical limits that depend on the absolutely values of heatwaves, which is why you additionally look at the other indices? Then maybe that could be clarified.

- 238 another conclusion would be that in in this metric there is no feedback of CC on int var, right? Maybe noteworthy too.

- 239-241 distinguish what from what? Origin from mechanisms (what's the difference anyway) or unforced from forced (not written) or temperature (mean) from temperature (extremes); really not clear what you want to say.

- 241 EOFs don't show physical mechanisms, please avoid reiterating this common misconception. 268 also, higher modes very unlikely to refer to a clear mode due to the orthogonality constraints etc

- 241 ensemble spread? 285, 342, 343, 344, 346, 367?, etc the lowest etc ensemble >member< (or 'run'). It's only one ensemble.

- 242 what does that 'separately' refer to?
  
- 244 is it a really driver or maybe rather a manifestation?
  
- 246 I gather you mean the bar charts with 'EOF patterns'? Slightly confusing here since traditionally one would expect to see a map (and you show maps too)
  
- 261 shown above the map plot in each panel ('lower panel' sounds like d,e,f to me)
  
- 281 so this classification groups based both on climatology (from the intercept) and magnitude of the response to CO<sub>2</sub> forcing (slope), right? I think that would be worth clarifying somewhere
  
- 282 observations? Isn't it model data? 290 'observed' too, please avoid to avoid confusion
  
- 284 have you tried numbers of clusters other than 6? not suggest you need to do a sensitivity study now, but if you have, you should report on the results
  
- 287 "as might be expected" – as a consequence of the methodological approach?.
  
- 292 how do you know that that is the only reason or whether there is a differing response to the co<sub>2</sub> forcing too?
  
- 294 3.5 and 2.2., what do those numbers refer to?
  
- 296-297 "and now observed (... , 2017)" – potentially misleading because Polar (Arctic) amplification has been observed since much longer than 2017
  
- 302 I struggle to see that. Isn't it rather cluster 6 that shows a more rapid increase beyond ~1.5degC, but not cluster 4, in Fig. 7j? 'these regions' would be 1-4 I think



- 306 could you briefly discuss how the clusters you find link to existing knowledge about climate zones; do they make sense from a physical point of view?

- 308 please consider adding 'global' to population so it's easier to understand you are now going from regions/zones to global

- 310 'heatwaves lasting 131 days' – not really, it's 131 heatwave days, isn't it, that's not the not the same as 1 or more heatwaves of 131-day length. 64- days same. Also, can you please give a range for these values? Lastly, 'shows that ... will experience' IN THE MODEL – are projected in the model to experience or similar

- 327 now you cite the number for 5%, elsewhere in the text you cite numbers for 10%. What motivates that? Doesn't make for good comparison and seems arbitrary

- 340 no one is exposed to thresholds

- 341 I don't understand. 'climate realisations' do you mean, depending on how internal variability will play out, people in different regions will >temporarily< be differently affected? Confusing as written

- 375, 384 "the" 15 cities (otherwise unclear whether you maybe looked at more and the statement applies to only 15 of them) or 'the cities considered', even better, not need to iterate it's 15. Lots of typos/missing articles/wrong prepositions on this page, please fix!

- 348 I think this statement is too broad and needs to be more specific to the context (that it includes the share of the global population etc). As it is it sounds like a mechanistic statement about physical limits the magnitude of extreme events.

- 383 "also large compared to other cities" – please add "(not shown)" to avoid confusion?

- 384 4062 days make for a very long year!?

- 391 20degCs

- 389 IN A MODEL. Needs to be added

- 397 significant regionally >and temporarily<? Int var impact won't change the cc-induced change in their likelihood profiles, will it?

- 399 the global population

- 411 investigate >some< economic impacts of increasing heat extremes (heating and cooling is only part of that)

- 417 the high relative cost? Or how are energy costs globally, would need a reference. And: high demand? The demand is much lower in the poorest countries than in the richer countries currently.

- 419 how does the 6-hourly climate model output contribute uncertainty? Which? What temporal resolution would be better? and aren't all the indices in tab. 1 for daily (>6h) min/max anyway, isn't that a model output (tmin, tmax), calculated by the model over its time step (that will be less than 6h perhaps)?

- 427 the model you used is already bias-corrected? In which way? I don't understand. Also, I think the model data is corrected, not the model.

- Fig. 1 Units missing x and y axis labels. the figure caption should be improved to include all the necessary information. Also: land and near-land ocean area: both together? How is near-land defined? And: why do you choose those areas?

- Fig. 2 unit x axis. Why now call it MPI run and in Fig. 1 only run? If anything, turn around. I think the label 'internal variability %' on the x axis is an incomplete name of your variable. a)-d) can you have the same y axis range?

- Figs. 3 & 10: I would consider plotting the ensemble spread around the ensemble mean, to allow inferences at least by eye about the relative variations implied by this (e.g., Moscow low).

- Fig. 4 Can you add the percentage of variance explained by each mode? Unit missing x axis barcharts. Please specify in the caption that you look at annual values.

- Fig. 5 Why is the AMO not shown? Caption should be in methods. Is the model lacking multidecadal variability or is that just not impacting enough deadly days in the model? Should be discussed in the text.

- Fig. 6 I would add that it's arbitrary colouring/see Tab. 2. "observed" – in the model I guess, so I'd avoid that term. Crossed?

- Fig. 8 Don't you want to show the 95% also for a-d? I think that would be interesting too. The purple lines in e-f carry IMO a very powerful message. Languagewise, the second sentence of the caption is particularly unclear.

- Fig. 9 absolute number .. above present day? Sounds like relative, but think you refer to 'warming'. Please clarify in the text.

- Fig. 10 Again, a more complete caption would be helpful. Like, change (in percentage) since pre-industrial in the model-simulated number of cooling degree days (CDD; blue) and heating degree days (HDD; red) in the 1%CO<sub>2</sub> experiments after the time they cross the global-mean temperature thresholds of (a) 1.5degC, (b), .., (d) 4degC, respectively. Error bars ...

2.11.0.0