We thank this reviewer for their useful review of our manuscript. Below we provide a point-by-point response.

Major issues.

2 major methodological choices that need to be addressed because the processes destroy physical relationships or modeling is directly outside the scopes of calibrations.

I. Bias Correction of non-linear, thus non-stationary variables.

1) Bias Correction Motivation

The authors do not motivate why they are bias correcting their data. They cite a handful of manuscripts that show research groups bias correcting data. But they don't actually explain why they need to bias correct their own data. Bias corrections are necessary when the output being used is incompatible with a tool that it is being applied for. For example: precipitation from a GCM is on a 100km x 100km grid and the rainfall fields produced are often a constant drizzle. This output is required to drive a hydrological catchment model, however, the rainfall from the GCM does not represent any catchment scale stochastic processes. Therefore a correction is required to be able to continue the research. Within the context of this manuscript, I don't see any motivation for requiring some sort of bias correction. Population data is interpolated to the GCM grid, or diagnostics are executed on the the GCM outputs. Nothing that warrants utilizing a bias correction that would be imperative for interpreting the results.

We thank the reviewer for pointing out this issue. We agree with the reviewer's opinion, and the bias correction has been deleted from the manuscript. The entire analysis has been re-done with the original output from the model without bias-correction, including the reviewer's further suggestion on calculating the wet-bulb temperature.

We have replaced the bias correction with a sensitivity test that evaluates potential uncertainty in our results (described in Sect. 2.1 and throughout the paper). Overall, we find that the biases lead to changes much smaller than the changes in temperature and wet-bulb temperature due to climate change.

2) Bias correction methods.

Bias correction of a covariance of temperature-humidity is extremely difficult to produce reliable results that are not physics breaking …

As discussed above, we have removed this from the paper.

II. Choice of heat stress algorithm.

The authors use wet bulb temperatures as their primary heat stress indicator. There are various reasons why this is good and bad, and a battery of metrics would probably be a better approach (see Buzan et al., 2015). I think adding multiple more metrics to the manuscript would reduce the clear language and systematic approaches in the analysis. However, what is of major concern is the use of Stull 2011 for wet bulb temperatures. Much like how the statistical bias correction methods are only valid for modern climate, Stull wet bulb temperature, too was specifically calibrated for modern climate, which limit its capacity in global warming applications (Buzan et al., 2015). Figure 1 Buzan et al., 2015 demonstrates the increasing growing errors that occur as temperature increases. A better method is the Davies-Jones 2008 wet bulb temperatures. Specifically equations, 4.8-4.11 using Bolton 1980 eqn. 39. for equivalent potential temperature inputs (Davies-Jones 2009 evaluates various different equivalent potential temperature calculation methods and demonstrates that bolton eqn. 39 is the best). The easiest way to calculate all of these variables is with the HumanIndexMod (Buzan et al., 2015). Python enabled:

We have replaced the Stull calculation in the paper with the Davies-Jones calculation.

It is difficult to determine if the wet bulb temperature errors are coming from the Stull or the Bias correction (likely both). But these errors have serious consequences for the results: I am suspicious that line 327 states that 5% of the Earth's population is exposed to 180 deadly days and 302 tropical nights. Just a quick peak at 4x daily JRA55 shows the 1986-2005 climatology the value of Tw 25°C does not appear until the ~60th Percentile, i.e. less than half of the available deadly days (if am understanding the definition of days properly). I am not sure how the authors were able to generate 302 tropical nights deadly for modern climate, which is more than 9 months a year. My JRA55 climatology only starts to have 25°C appearing at the 25th percentile. I recommend using the Davies-jones eqn. 4.8-4.11 as in the HumanIndexMod

As mentioned in previous points, we have replaced our calculation with Davis-Jones 2009. However, we still find many Mora et al deadly days (daily maximum w2m over 24°C) and tropical nights even in our reference period (Fig. R1). While we have not confirmed the results with the JRA data, it is certainly possible that the difference is due to different time periods — the reviewer is looking at data from 1986-2005, while our reference period covers 2003-2017.

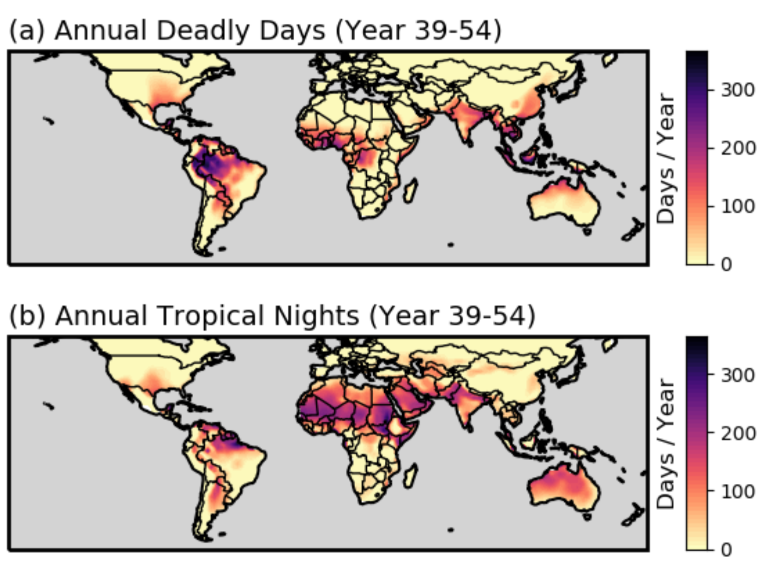

Figure R1. Number of ensemble-averaged annual (a) deadly days (daily maximum over 24°C) and (b) tropical nights (daily minimum t2m over 25°C) in 1% CO_2 experiment, in same global average temperature as present day (0.87°C, year 39-54)

To investigate this in more detail, we plotted the 15-year median value of daily maximum w2m (metric for deadly days) and daily minimum t2m (metric for tropical nights) for present day (2003-2017 in ERA-I, year 39-54 in MPI) as in figure R2.

As seen in the figure, in both ERA-I and 1% CO_2, we find median values of daily maximum being over 24°C (threshold for deadly days) in central Africa, Southeast Asia, and Northern part of South America. This translates to over 180 days of deadly days per year. As for 15-year median values of daily minimum t2m, we find values over 25°C (threshold for tropical nights) in Southeast Asia, India, central Africa, Northern Australia, and Northern South America. This also translates to over 180 days of tropical nights per year.

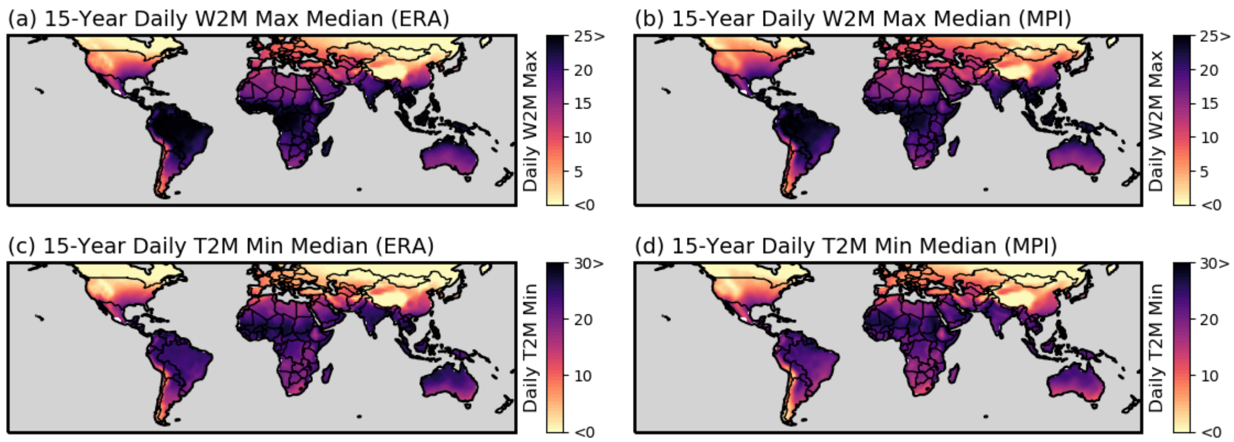

Figure R2. 15-year median values of daily maximum w2m and daily minimum t2m for ERA-Interim and 1% CO_2 experiment in present day warming (2003-2017 for ERA Interim, year 39-54 for 1% CO_2).

Minor issues:

Mora et al., 2017 really puts a low threshold for deadly heat stress. The world exposure to these conditions is fairly high, yet we don't have people dying all over. More likely, there are epidemiological reasons for people suffering heat stress, i.e. health, socio-economic, etc (which Mora states). But it really makes it difficult to use these values as a realistic driver of impacts on humans. Harder limits, such as Tw 32°C, where all laborers cannot work anymore for sustained amounts of time (Brunt 1943; Liang et al., 2011), at least have clearer thresholds for populations being impacted. Something to consider in analysis…

We agree that Mora et al., 2017 suggested a low threshold, although it represents a temperature where mortality begins to occur. So, we revised our definition of "Deadly Days" to daily maximum wet bulb temperature above 26°C.

We have changed the text to read: [L188-196] We therefore also count the number of days each year with daily maximum w2m above 26°C, which we refer to as “deadly days”. We note that other values could be chosen, with higher values occurring less frequently but having more significant impacts. This value is based on the analysis of Mora et al. (2017), who demonstrated that w2m of about 24°C is the threshold which fatalities from heat-related illness occur. However, since we find that there are some regions that already experience over 9 months of 24°C w2m events per year, we increase this threshold to 26°C in our analysis. We could have chosen higher w2m values, but any choice in this range is associated with negative impacts, so we have chosen a value near the bottom of the range where mortality occurs in order to maximize the signal in the model runs.

The CDD methods are fine, but I worry about using the bias corrections. The 28 member ensemble should characterize the variability of that 18°C threshold without the need of bias correcting the results.

We have removed the bias-correction from the analysis and used the original output of the model.

Line 319 is stating rapid increase in wet-bulb extremes. I am not quite following the language. Is this due to large population living in the tropics? Some clarity here would be useful.

Explanation is added for clarity.
[L330-334] Given that the planet has already warmed about 1°C above pre-industrial, this suggests that the world should presently be experiencing a rapid increase of wet-bulb extreme frequency, particularly in the tropics. This is related to the increased slope in Figure 6, in which cluster 1 and 2’s values of HWD_{w2m} and HWF_{w2m} increase rapidly until 3.0°C and 2.0°C of global warming.

Figure 6 K means should be listed. a) is not described in the caption (no colorbar). I think the colors match figure 7, but I am unsure. Ah, I think it is in Table 2. I think there just needs to be a reference in Figure 6 and 7 to Table 2 for the descriptions.

Thanks for pointing this out. Figure captions have been revised for clarity.

Figure 5. (a) Clustered regions via K-means clustering. Characteristics of each cluster are listed in Table 2.
Figure 6. Evolution of each index averaged over each cluster. Colors are consistent with Figure 5 and Table 2.

Recommended citations:

Buzan JR, Oleson K, Huber M. 2015. Implementation and comparison of a suite of heat stress metrics within the Community Land Model version 4.5. Geosci. Model Dev. 8:151–70

Brunt D. 1943. The reactions of the human body to its physical environment. Q. J. R. Meteorol. Soc. 69:77– 114

Liang, C., Zheng, G., Zhu, N., Tian, Z., Lu, S., and Chen, Y.: A new environmental heat stress index for indoor hot and humid envi- ronments based on Cox regression, Build. Environ., 46, 2472– 2479, 2011.

Davies-Jones, R.: An efficient and accurate method for computing the wet-bulb temperature along pseudoadiabats, Mon. Weather Rev., 136, 2764–2785, 2008.

Davies-Jones, R.: On formulas for equivalent potential temperature, Mon. Weather Rev., 137, 3137–3148, 2009.

Zhang et al., 2021 Projections of Tropical heat stress constrained by atmospheric dynamics.

Schwingshackl et al., 2021 Heat Stress Indicators in CMIP6: Estimating Future Trends and Exceedances of Impact-Relevant Thresholds

Pierrehumbert RT. 1995. Thermostats, radiator fins, and the local runaway greenhouse. J. Atmos. Sci. 52:1784– 806

Williams IN, Pierrehumbert RT. 2017. Observational evidence against strongly stabilizing tropical cloud feed- backs. Geophys. Res. Lett. 44:1503–10

Williams IN, Pierrehumbert RT, Huber M. 2009. Global warming, convective threshold and false thermostats. Geophys. Res. Lett. 36:L21805

Mauren 2016 Bias Correcting Climate Change Simulations - a Critical Review