Review of the manuscript: The Effect of Forced and Unforced Variability on Heat Waves, Temperature Extremes, and Associated Population Risk in a CO2- Warmed World

Authors: Jangho Lee, Jeffrey C. Mast, and Andrew E. Dessler

Summary.
This manuscript evaluates a 28 member on the order of 400 model years at 4x daily values to determine heat stress states for the 0°-4°C warming from an 1850 baseline. The models are set up with a historical period followed by a 1% annual increase co2 as the warming until each simulation reaches ~4°C global increased temperatures. After these simulations are executed, the states of each simulation at 1°C increments are evaluated. Extracted are human relevant variables in the context of heat stress and heat stress impacts (such as increased energy use from air conditioning).

First thoughts.
Overall, the approach is thorough and systematic. The manuscript divides the data analysis into relative problems, such as exposure, or degree cooling days, or GDP etc., and then evaluates the outcomes, followed by summarizing the results in the end. Straight forward, and in most cases, relatively easy to read. Text clarifications from a writing editor/center recommendation would fix minor clarity issues.

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.

2) Bias correction methods.
Bias correction of a covariance of temperature-humidity is extremely difficult to produce reliable results that are not physics breaking. Temperature humidity, or T-Q, are non-linear in their combination and are dependent on initial conditions (Buzan and Huber, 2020). Figure 7 in Buzan

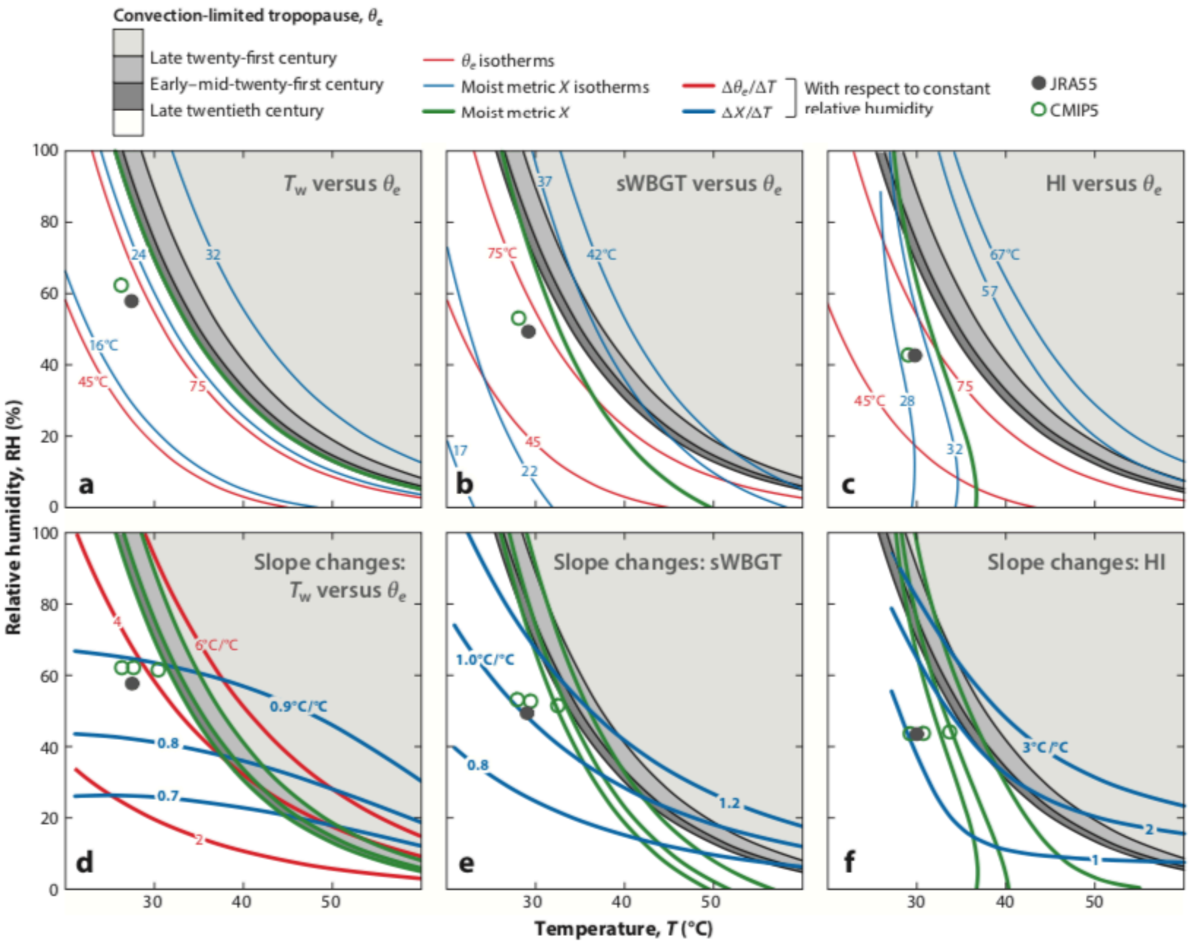


Figure 7

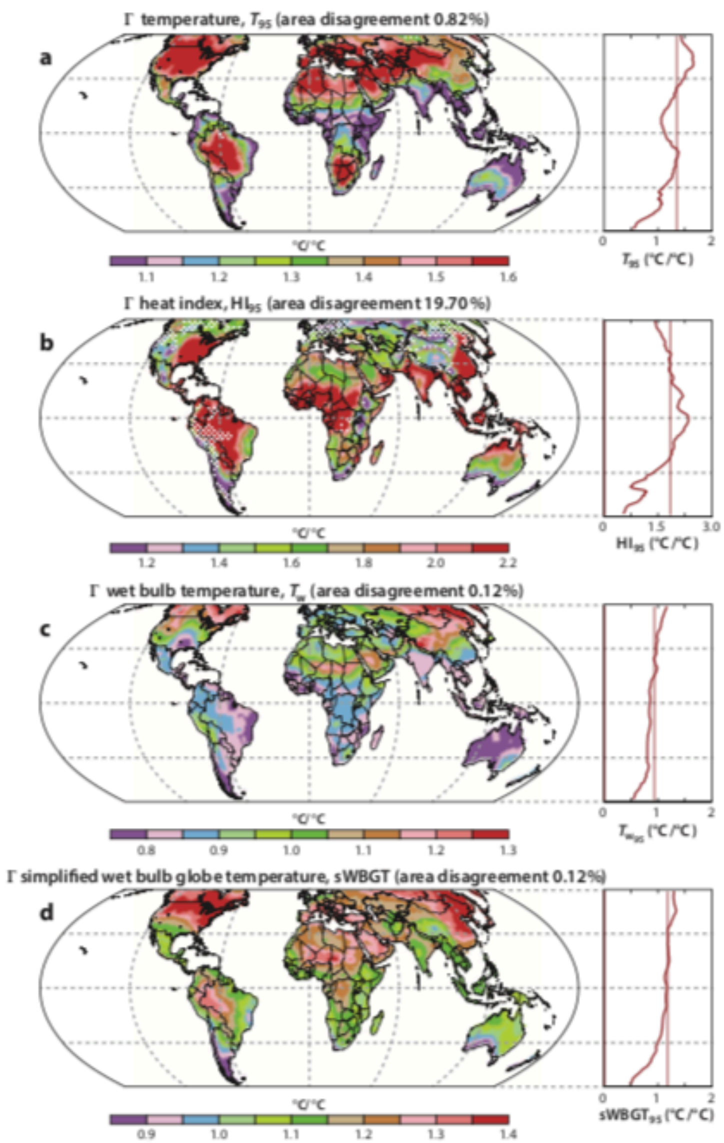
Moist thermodynamic state diagram where T is temperature, T_w is wet bulb temperature, and X is the heat stress metric. Convection-limited θ_e (*grays*) and moist metric X (*green lines*) are extracted from ninety-fifth-percentile spatial maximum CMIP5 ensemble climate states for the late twentieth century (1986–2005), early to mid-twenty-first century (2026–2045), and late twenty-first century (2081–2100). Average ninety-fifth-percentile temperature and RH are extracted from average ninety-fifth-percentile X (CMIP5, *green circles*; JRA55, *gray dots*). Abbreviations: CMIP, Coupled Model Intercomparison Project; HI, heat index; JRA55, Japanese 55-year Reanalysis; sWBGT, simplified wet bulb globe temperature.

and Huber demonstrate this. A shift in the T-Q on the diagram, changes the behavior of outcomes from global warming. This is further complicated when examined in Figure 8 which shows that relative humidity and temperature are not constant with global changes. And is further complicated by Figure 9c showing that regionally each location has a different relationship of temperature and humidity in the context of global changes. Some locations are due to increasing humidity while others are constant or negative. So a bias correction method would need to consider this.

The authors use CDF-t method (Michelangeli et al., 2009). A critical statement in Michelangeli's paper on applying their methods: "Remark that this common assumption of stationarity made by

Figure 9

Maps and zonal means of multimodel mean extreme slope parameters: (*a*) $\Gamma_{T,95}$, (*b*) $\Gamma_{HI,95}$, (*c*) $\Gamma_{T_w,95}$ (repeated here from **Figure 5*a*** for comparison), and (*d*) $\Gamma_{sWBGT,95}$. The area of disagreement is calculated by the coefficient of variation, $c_{v,\Gamma,X,95}$, > 0.35 between CMIP5 simulations (visually represented by *white stipple* in panels *a* and *b*; area of disagreement is too small to see here for panels *c* and *d*). Abbreviation: CMIP, Coupled Model Intercomparison Project.



644 *Buzan • Huber*

most of the statistical downscaling approaches should be taken with care because it is not guaranteed." Furthermore, the Michelangeli method is supposed to be applied to downscaling approaches, which is not is undertaken here with the 28 member ensemble.

The authors don't show in their bias corrections that results of their adjustments remain physical. They only show that they can get their models to mimic the behavior of a heavily forced dataset, ERA-Interim. GCM strengths are energy balances constrained by radiative-convective balances. Quasi-radiative equilibrium balances are heavily dependent on the T-Q covariance (Pierrehumbert 1995; Williams et al., 2009). This is also observed (Williams and Pierrehumbert 2017). Furthermore, modern bias corrections cannot account for climate change trends (Maraun, 2016). To appropriately bias correct T-Q covariance in the 28 member ensemble for the future projections requires correcting the data against the same time series in observations, which do

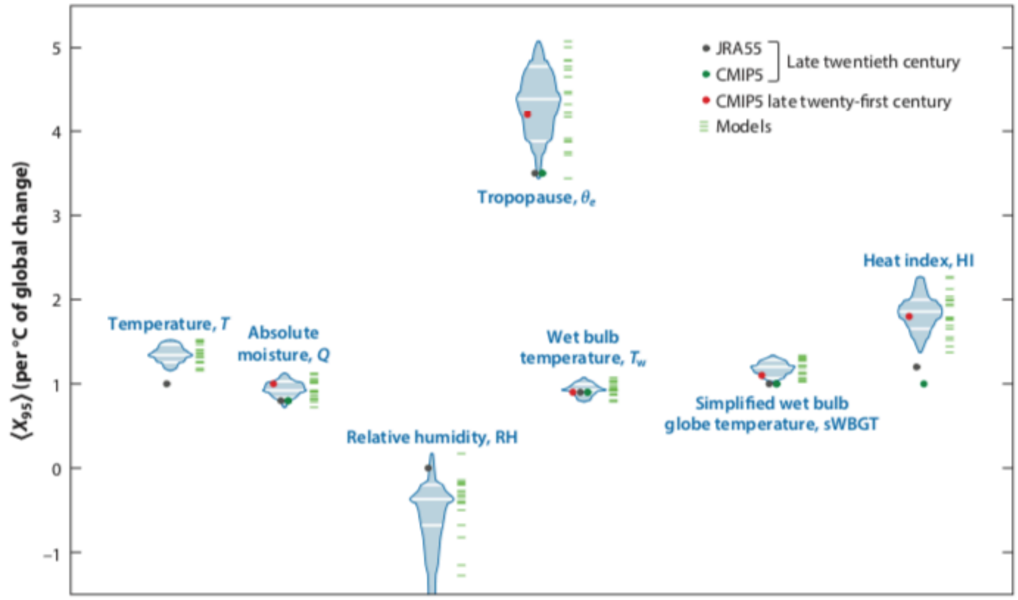


Figure 8

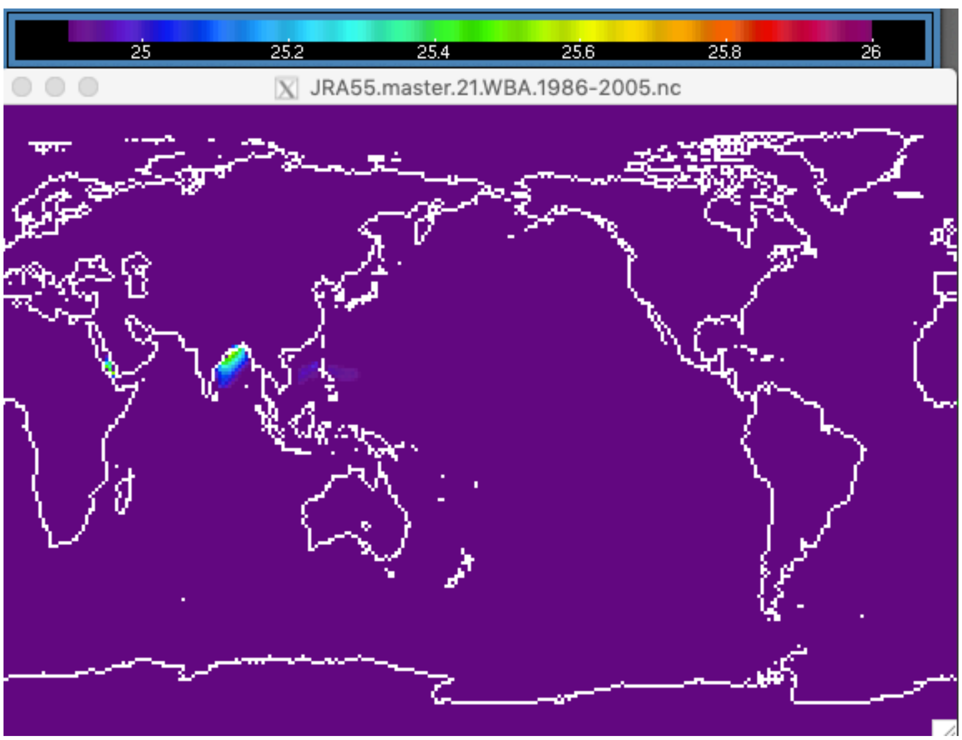
Intermodel spread 2091–1996 global mean ninety-fifth-percentile slopes (calculated over 57°S to 57°N). The spindle diagram contains the 18 CMIP5 simulations (width indicates concentration), with each individual model denoted by a horizontal green bar. The lower (*lower white bar*), median (*middle white bar*), and upper (*upper white bar*) quartiles represent 50% of the CMIP5 simulations; the tails are maximum and minimum, i.e., the range. The ninety-fifth-percentile slopes are late twentieth-century JRA55 (*gray*) and CMIP5 (*green*) and late twenty-first-century CMIP5 (*red*). Abbreviations: CMIP, Coupled Model Intercomparison Project; JRA55, Japanese 55-year Reanalysis.

not exist. Even with all the different parameterizations and scales of resolutions in CMIP5 and CMIP6 simulations, what is consistent, is that quasi-radiative equilibrium states are one of the most reliable aspects of GCMs (Buzan and Huber, 2020; Schwingshackl et al., 2021; Zhang et al., 2021). Contextualizing this, heat stress is derived from a T-Q covariance that dependent on quasi-radiative equilibrium. Maintaining this physical process is crucial for producing reliable heat stress impacts.

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:
https://github.com/jrbuzan/HumanIndexMod_2020
And NCL enabled version (attached).

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; example of JRA55 Tw exceedance threshold of 60th percentile.



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…

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.

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.

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.

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