Short Comment 1: General Figure Improvement

From Reviewer #1 major comment #2:

> "A bigger point around the figures and indeed the results in general, is that I question whether the authors have chosen the right things to show. Their main point (if I have understood correctly what they have done, which might not be the case) seems to be that in humid regions, heat stress thresholds are reached at lower temperatures and heat stress is less variable than in arid regions. These points are obvious (and have been quantified in several of the papers cited), and you don't need a GCM, let alone online diagnostics, to show it."

From Reviewer #2 major comment #4:

> "The joint distribution analysis is not really convincing, it is not clear what research question it addresses and thus it does not add any novel understanding. Do you want to understand which indices give more weight to temperature or to humidity? If so you could basically do that in an xy-plot showing temperature on one axis and relative humidity on the other axis, and then add the isolines for the individual indices. The slopes would then tell you which indices give more weight to temperature or to humidity. Another approach would be to produce a QQ-plot of temperature and humidity versus each index or correlate their time series. If the emphasis is more on the spatial pattern I would like to see a more quantitative analysis like a pattern correlation of the contributing variables and the indices. But again, emphasis in the results section should be put on demonstrating the added value of the new code implementation."

We apologize we were not clear. As noted by both reviewers, our figures did not adequately illustrate our main point of this section. The results, shown below, show that a GCM is required to determine what causes extreme high heat stress events.

An open question is what drives extreme high heat stress events, which are, by definition, rare events. For example, we cannot determine from the mean climate state or from theory, in a warm and humid climate, if abnormally high temperature, abnormally high moisture, or a combination of the two, caused a heat stress event. This is a question of the covariance of perturbations of temperature and humidity, not a statement of mean conditions, and there is no theory to explain these situations. For example, we may apply Reynolds averaging to the NWS Heat Index equation:

$$\begin{aligned}\overline{HI} = a + \overline{b\overline{T}} + \overline{c\overline{RH}} + \overline{d\overline{T}\,\overline{RH}} + \overline{e\overline{T^2}} + \overline{f\overline{RH^2}} + \overline{g\overline{T^2}\,\overline{RH}} + \overline{h\overline{T}\,\overline{RH^2}} + \overline{\imath\overline{T^2RH^2}} \\ + \left[\overline{dRH'T'} + \overline{eT'^2} + \overline{fRH'^2} + \overline{gT'^2RH'} + \overline{hT'RH'^2} + \overline{\imath T'^2RH'^2}\right] \quad (1)\end{aligned}$$

where *a*, *b*, *c*, *d*, *e*, *f*, *g*, *h*, and *i* are constants in the polynomial. *RH* and *T* are relative humidity and temperature, respectively. We are not concerned with the terms outside the brackets, as they are the means. The terms within the bracket are representative of turbulent effects on the Heat Index, which we are discussing. It is these turbulent states where a GCM is able to determine these individual factors, by calculating the heat stress metrics and thermodynamic quantities at every model time step. Furthermore, each heat stress metric has different assumptions (such as body size, or physical fitness) that weigh temperature and humidity differently. A high heat stress event indicated by one metric does not necessarily transfer onto another metric.

Since we obviously did not present this clearly, we present our revised Figure 3, heat stress metrics and *T-Q* regime maps. We do not use xy-plots or QQ plots, because we cannot determine from theory the behavior of the extremes. Figures 3a-c are the 99th percentile exceedance values of the metrics sWBGT, HI, and T_w, respectively. The spatial patterns are similar to work previously cited in our manuscript.

The T-Q regime plots (Figure 3d-f) are expected rank values derived from our new Figure 2. We calculate the values in a series of steps. 1) we take the conditional distribution of T and Q that represent ≥99th percentile of the source heat stress or moist thermodynamic metric. 2) we take the expected value (median) of the conditional distributions of T and Q and determine what percentile they come from in their respective time series. 3) we condition these values on each other to create the expected rank values (Figure 2).

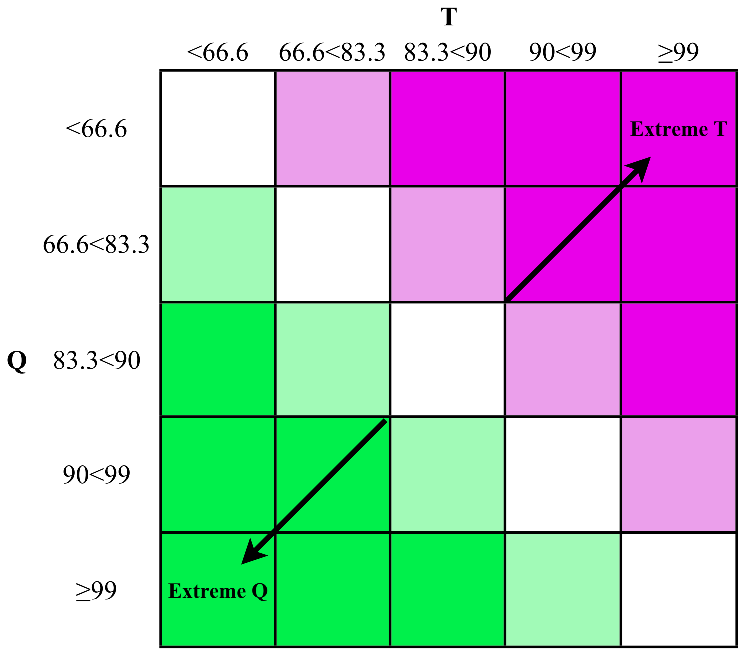
The 99th percentile of T_w (Figure 3c) is dominated by extreme Q (Figure 3f). However, this is not necessarily the same for 99th percentile of heat stress metrics, specifically, HI (Figure 3b). We show that HI is dominated by extremes in T, globally (Figure 3e). The 99th percentile of sWBGT (Figure 3a) is in-between the responses of HI and T_w (Figure 3d), however, at the lower latitudes Q mostly dominates the metric.

Hopefully, it is more clear with this analysis that the results:

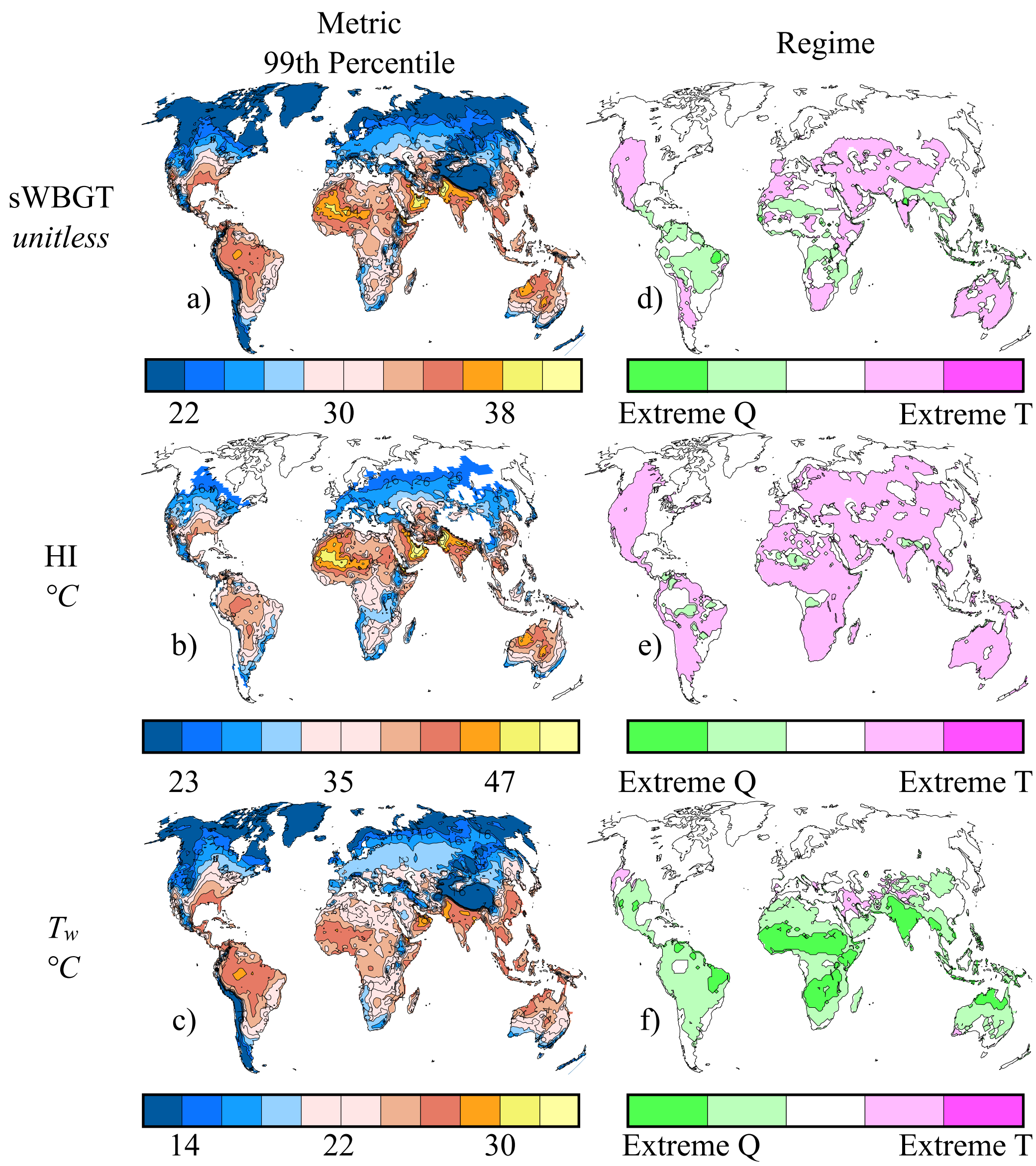
1) are not obvious, as they arise from turbulent covariances.

2) have never been described previously.

3) show that there is no general theory that describes the behavior of these various indices, and that a GCM is necessary.



Preliminary Figure 2. Expected value rank. *T* and *Q* conditioned upon ≥99th percentile of a heat stress or moist thermodynamic metric. The *T* and *Q* values are compared to their respective time series as a percentile. These *T* and *Q* percentiles are binned and are compared to each other. Extreme *Q* are greens and extreme *T* are magentas.



Preliminary Figure 3. 99th percentile exceedance value of 3 metrics for a) sWBGT, b) HI, and c) T_w (left). Expected rank value T-Q regime maps d), e), and f) (right) conditioned by a), b) and c), respectively. Each plot uses the color bar below the plot. Rank values for d)-f) are described in Figure 2.