Review for NHESS submission, "Brief Communication: Drought Likelihood for East Africa" by Yang and Huntingford

General comments

I appreciate the efforts made by the authors to address my concerns, and I believe the paper has been improved significantly. However, there are still multiple issues arising from their analysis which I believe need to be taken into account, before this Brief Communication could be published.

Therefore, my recommendation is that *this article requires major revisions*. In accordance with the NHESS review criteria, my assessment is as follows:

Scientific Significance: 3 (fair)

Scientific Quality: 3 (fair)

Presentation Quality: 3 (fair)

Specific comments

1) I'm confused by the presentation of results in Figure 2. The authors adjust each model according to a different observational data set, and then proceed to look at the likelihood of the 2016 drought (or worse) in each new ensemble. But is the drought event threshold for each of these new ensembles still 46mm? Or is it the absolute rainfall total associated with ASO 2016 for each individual dataset? It is unclear based on the current text. Also, if it is the latter approach, this is equally problematic, since the sigma-anomaly associated with the ASO2016 event might differ dramatically between the different observational products, which would thus render Figure 2 as no longer an apples-for-apples approach.

My recommendation would be to identify the percentile anomaly associated with ASO2016 for each observational product, then take the mean of these answers, and use this average percentile anomaly as the event threshold employed for all panels in both Figures 1 and 2.

2) As it stands, the authors first bias-correct the mean, and sometimes then also variance, for each individual model, before combining all models into a single ensemble and calculating the relevant statistics. How do the answers differ if you instead combine all raw model results into a single ensemble first, and then proceed to correct the multi-model ensemble by a singular correction factor for mean (and then variance)?

3) A paper recently published has provided a multi-method assessment of changes to the likelihood of occurrence of the Ethiopian drought of 2015 (doi: 10.1175/JCLI-D-17-0274.1). The authors will therefore need to provide explicit justification of the *value added* by publishing their analysis, relative to this already-published paper. Especially given the significant level of overlap in both regions considered and topics discussed (specifically, changes in drought likelihood in the context of ongoing climate change).

Second, the CMIP5-relevant part of the 2015-relevant analysis employs a specific method of event attribution, whereby they exclude models from subsequent analysis of 'changes in drought likelihood' if the rainfall climatology of a given model did not match the climatology from observations (using a KS test). Given this CMIP5-based approach has also been used in multiple other attribution studies, I would like the authors to comment on how their approach is better or worse, and why. For example, if you employed this alternative approach to your analysis, how would the answers change?

4) There are two additional observation-based precipitation products which would be of considerable value to add to this analysis: CRU-TS (as I previously highlighted), and CenTrends dataset (doi:10.1038/sdata.2015.50), which has been specifically developed for analysis of seasonal rainfall anomalies over the Horn of Africa region. Further, based on these updated, observation-based products, I would strongly question the continued use of ERA-Interim as the primary product of consideration in Figure 1.

5) The bottom row of Figure 3 presents changes to standard deviations in model rainfall, based on running 31-year periods. It's not clear to me what this standard deviation represents: if it's based on '31-year rainfall mean' as stated, then this implies only one data point. Or is it the annual average for each of the 31 years, or the ASO-averaged rainfall per year? And besides, either of these latter suggestions would yield only 31 data points, hence I'm not sure it provides any information of real value. This is particularly true, given the fact that a purported 121% increase in rainfall SD (by MIROC5) is not considered a statistically significant increase.

My recommendation is to remove the row considering future changes in model variability, and just leave the row mentioning changes in the mean.

6) I have significant issues with the treatment of uncertainty estimates in Figures 1 and 2. The width of the error bars is an implicit estimate of whether a 'statistically significant' difference exists. By showing error bars with 1-standard deviation only, this implies a 68% confidence level (using the assumption of a normal distribution). Most studies tend to consider uncertainty estimates based on confidence levels of 90% or higher. I strongly recommend presenting all uncertainty ranges on the figures using two standard deviations – this will be more representative as to whether a statistically significant difference really exists or not.