Dear Editor of NHESS December 2017

Thank you for your on-going help with this manuscript, and we are grateful that the two reviewers were prepared to look again at our Brief Communication:

Drought Likelihood for East Africa.

The main technical change is the request to include in the analysis two additional measurement datasets. These are the CRU-TS and a member of the CenTrends dataset family (CHIRPS). This now leads to two additional panels in Figure 2, where we normalize climate models against different rainfall datasets. From these, we calculate future risk of drought.

We have also adopted the request to present all of Figure 1, which captures the magnitude of the Autumn 2016 East Africa drought, in terms of the CenTrends data (CHIRPS).

All other technical requests we implemented in full, and outlined in our responses below. The one exception is the query as to whether our normalization of GCMs to each alternative measurement dataset is a like-for-like comparison. We believe that it is, but recognize that the wording of the paper in the original version needed to be improved to remove any ambiguity.

Our responses are described below, and including presentation of adjusted text. We realise that both reviewers have requested significant changes through the progression of this paper. However we hope that the manuscript is now acceptable as a short summary document capturing current projections of autumn rainfall by GCMs for East Africa. This is where the severe drought of 2016 occurred, enabling assessment of future likelihood of reoccurrence.

Thank you for all the help so far.
With kind regards,

Hui Yang

Hui Yang (and on behalf of Chris Huntingford).
Email: yang_hui@pku.edu.cn

Response to the reviewers

To Reviewer #1:

***Reviewer #1* General Comments:**

The analysis and description of the method in this study has been significantly improved during the revision. The additional figures and tables help the reader understand the method and the results. The conclusion has been significantly improved and it now puts the study and its results in context. This study can be published after several small technical corrections, listed below.

[Response] We are pleased that the reviewer likes our new paper version. Thank for all your comments and suggestions that have helped us generate a better manuscript.

LANGUAGE: *The English could still be improved in several places.*

Abstract: The abstract still has to be improved in terms of clarity and language, especially the second part.

[Response] We have polished the language in large parts of manuscript. We hope the new paper version has a level of clarity as to be useful to a relatively broad audience. The uploaded version to the NHESS website is with "track-changes" so our additional minor edits can also be seen.

***Reviewer #1* Detailed Comments:**

Page 1, line 31: an equation explaining how this is computed would be helpful, e.g. is daily data used?

[Response 1.1] We have followed the reviewer's advice and now put the bias-corrected equations in the main text. We use indexing and superscript notation to now make it unambiguous as to exactly the correction we apply. We write: "The GCM precipitation mean ASO estimates, both past and future, are corrected for each model year by a GCM-specific mean correction factor. This factor is a ratio of the climatological mean of each GCM to that of the CHIRPS product as:

$$x^{\mu}_{corr,i,j} = x_{model,i,j} \times \frac{\mu_{obs}}{\mu_{model,j}} \quad . \quad (1)$$

Here $x_{model,i,j}$ and $x^{\mu}_{corr.i.j}$ are, respectively, model simulated and mean bias-corrected ASO precipitation data of the ith year (i=1,2,…,31) for the jth GCM (j=1,2,…,37). μ_{obs} and $\mu_{model,j}$ are the observed and GCM-specific time-mean (i.e. average across indices i) of ASO rainfall estimates during the period 1981-2015. Second, we then adjust the mean-corrected data from

Eqn. (1), such that they further are corrected to have identical standard deviation (STD) to the CHIRPS product whilst maintaining the bias correction for the mean. This gives bias-corrected estimates $x^{\mu,\sigma}_{corr.i.j}$ as:

$$x^{\mu,\sigma}_{corr,i,j} = (x^{\mu}_{corr.i,j} - \overline{x^{\mu}_{corr,j}}) \times (\frac{\sigma_{obs}}{\sigma^{\mu}_{corr,j}}) + \overline{x^{\mu}_{corr,j}} \quad . \quad (2)$$

Here $\overline{x^{\mu}_{corr,j}}$ (= μ_{obs}) is the 31-year average of mean bias-corrected data from Eqn. (1). σ_{obs} and $\sigma^{\mu}_{corr,j}$ are standard deviation of the ASO rainfall estimates during the period 1979-2015 from observations and from the mean bias-corrected precipitation data created by Eqn. (1)." (Page: 2, Lines: 1-14)

Page 1, line 29 / page 2, lines 19 and 25: change "the east Africa" to "East Africa" or "eastern Africa"

[Response 1.2] Done at the noted points in the manuscript. We thank the referee for pointing this out.

Page 3, line 23: the papers listed below might be helpful references

Gleixner, S., Keenlyside, N., Viste, E., & Korecha, D. (2016). The El Niño effect on Ethiopian summer rainfall. Climate Dynamics, 49(5-6), 1865–1883. http://doi.org/10.1007/s00382-016-3421-z

Gleixner, S., Keenlyside, N. S., Dimissie, T., Counillon, F., Wang, Y., & Viste, E. (2017). Seasonal predictability of Kiremt rainfall in CGCMs. Environmental Research Letters. http://doi.org/10.1088/1748-9326/aa8cfa

[Response 1.4] Thank for this. We have cited this paper quoted by the reviewer in the line **. We now write: "Strong teleconnections are known to exist between El Niño Southern Oscillation (ENSO) and East African rainfall (Segele et al., 2009; Gissila et al., 2004; Gleixner et al., 2016), and with longer-term fluctuations in Pacific SSTs either increasing or decreasing rainfall (Funk et al., 2014; Liebmann et al., 2014; Gleixner et al., 2017)." (Page: 5; Lines: 1-3)

Page 7, Figure 2, caption: please explain better

[Response 1.5] We re-write the caption of Figure 2, and the revised text now reads: "**Figure 2**: CMIP5 GCM-based histograms of probabilities of mean ASO rainfall falling below year 2016-based threshold values. This is for different time periods and for different observation-based precipitation products of CHIRPS, CRU-TS, ERA-interim, GPCP, PREC/L, CPC and TRMM.

Shown for years 1861-1891 (blue), 2001-2031 (black), 2035-2065 (orange) and 2070-2100 (red), and using GCMs simulations corresponding to historical and RCP8.5 estimates. Individual GCM projections are bias-corrected by the (panel-specific) precipitation product. This data is combined to give single overall probabilities across the 37 GCMs sampled. The histogram bars without horizontal hatching (left) are for the mean-corrected GCM precipitation estimates. The bars with hatching (right) are for the mean- and variance-corrected GCM estimates. The error bars are two standard deviations (estimated via bootstrapping 80% replications from the 37 GCM precipitation data for the 31-year periods). Data in the CHIRPS panels repeats that of the insets of Figs 1c, d." (Page: 9)

To Reviewer #2:

***Reviewer #2* 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.

[Response] We thank the reviewer for re-reviewing the manuscript and providing us with an additional set of helpful comments. This insightful advice has helped us to improve the manuscript further. Our responses are below, and revised text in the paper repeated in red font.

***Reviewer #2* 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.

[Response 2.1] The threshold is different for each observational record. It is based on that dataset and its estimate of ASO rainfall in 2016. These values are given in Table S2. The number 46mm is for ERA-interim, however your request (below) to use CHIRPS data, the threshold in that instance is 40mm. (Hence we use the second option the reviewer notes). We state this more clearly in the paper, now writing "The probability of drought occurrence is based on estimates of ASO rainfall in 2016 and for each individual dataset (values in Table S2)."

Each measurement dataset does have different distance between the year 2016 ASO rainfall value and the ASO climatological mean, different climatological mean and different climatological standard deviation (see Table below, which also appears as Table S2). That is, based on different product, the severity of year 2016 ASO drought are different. We then correct both mean and standard deviation of combined GCMs estimates, and for climatological period (1981-2015), to be identical to each dataset. With both these statistics (μ and σ) identical, we believe that for the historical period, the model estimates corrected by different precipitation data are like-for-like. From this, we assess future model projections in the context of how rainfall below the year 2016 ASO threshold is expected to change. Hence such changes are GCM and dataset specific. We have considered at length this, and think this is an appropriate like-for-like ("apples-for-apples"), but recognize the original paper wording could have been made clearer in this regard.

We now write "Large uncertainty in the observation-based precipitation products has been well reported (Angélil et al., 2016), we additionally use six other precipitation estimates (CRU-TS, ERA-interim, GPCP, PREC/L, CPC and TRMM) to bias-correct GCM estimates. The probability of drought occurrence is based on estimates of ASO rainfall in 2016 and for each individual dataset (values in Table S2). There are substantial differences between these values. We use each of these extra datasets to repeat the bias-correction of every GCM by same Eqn. (1) and Eqn. (2), but now with new data-specific μ and σ values. These μ and σ quantities are also given in Table S2." (Page: 3, Lines: 6-11)

Table S2. The mean August-to-October (ASO) rainfall (mm month^{-1}) of year 2016, multi-year mean (not including 2016) and multi-year standard deviation (STD, not including 2016) over east Africa for years 1981 to 2015. The seven global precipitation data sets used are listed. Six products of CHIRPS, CRU-TS, ERA-interim, GPCP, PREC/L, CPC and TRMM are available from 1981 to 2016. These six precipitation data sets are either interpolated gauge observations only (i.e. CHIRPS, CRU-TS, PREC/L and CPC), gauge observations combined with satellite measurements (i.e. GPCP), or reanalysis data (i.e. ERA-interim). The TRMM satellite observations are available from 2001 to 2016.

ASO rainfall (mm month^{-1})	CHIRPS	CRU-TS	ERA-interim	GPCP	PREC/L	CPC	TRMM
2016	39.97	45.93	46.10	46.56	57.16	35.78	32.05
Climatological mean, μ (1981-2015)	53.05	55.49	70.81	62.94	62.01	44.59	60.69*
Climatological STD, σ (1981-2015)	7.09	10.20	11.55	10.68	11.72	13.33	11.83*

* TRMM satellite precipitation data is only available from 2001 to 2016. The climatological ASO rainfall averages of the period 2001-2015 is computed.

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)?*

[Response 2.2] Thank you for this. We perform this analysis as the reviewer suggests i.e. instead of normalizing the models individually, we instead normalize just once the full ensemble of GCM estimates. That is, we use μ and σ from a single array that contains all years and from all GCMs. This generates a diagram identical in format to Figure 2, which we put in SI, and as repeated below.

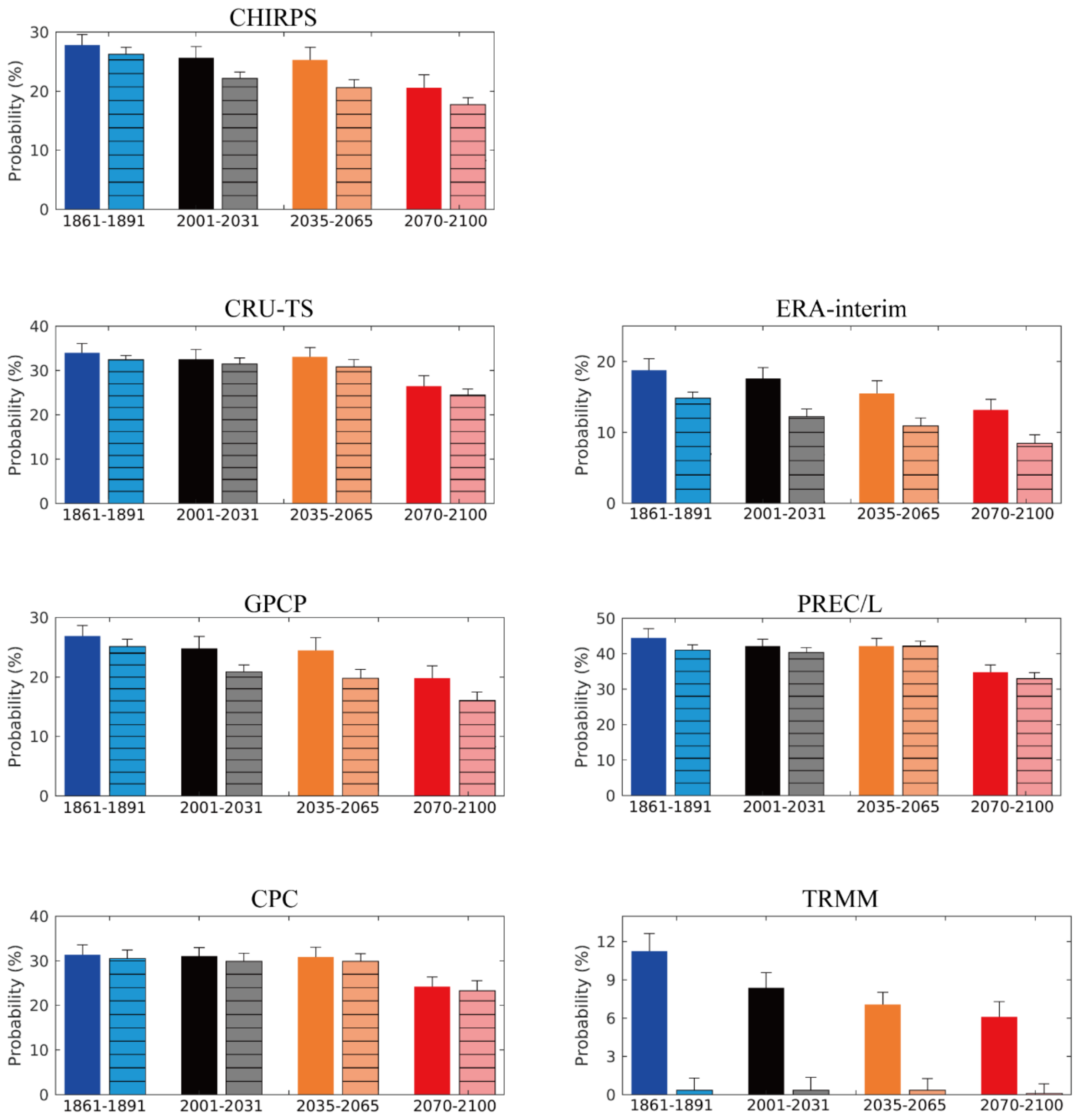


Figure S1: Same as Figure 2, but 37 GCM estimates are combined into single multi-model ensemble, and then this ensemble is bias-corrected only once by using each precipitation product.

We do know that the models have very divergent behavior of precipitation projections for East Africa (also highlighted by Figure 3 of the main paper). As information to Reviewer 2, using ACCESS1-0 and IPSL-CM5A-LR as examples, we plot three PDFs of ASO rainfall for the modelled future period 2070-2100 (Figure R1). That is, without bias-corrected, with individual bias-corrected μ and σ, and using the multi-model bias-corrected GCM rainfall estimates (μ and σ). Figure R1 shows that not only whether using bias-correction (against historical measure) or not, but also the choice of bias-correction method (i.e. individual bias-correct or all model large ensemble single correction) can have a very large impact on future estimates of rainfall.

We provide new text to place the reviewer's sensitivity request in the paper, i.e. leading in the new diagram S1. We add the following sentences to the paper "As a sensitivity study, we also perform a bias-correction based on each precipitation product but for the full ensemble of 37 GCM estimates together. That is, we combine all GCM present-day estimates in to one single vector, and calculate single overall μ and σ values. All seven precipitation datasets are used to repeat the bias-correction, by similar methods to Eqn. (1) and Eqn. (2). This approach implies that the probability of drought occurrence in East Africa has decreased slightly from pre-industrial to the end of 21st century, regardless of whether variance has been corrected (Fig. S1). However this approach should be viewed with caution, as making single bias-corrections for all the GCMs combined neglects model differences, which are known to be large in precipitation projections (Collins et al., 2013)." (Page: 3, Lines: 25-31)

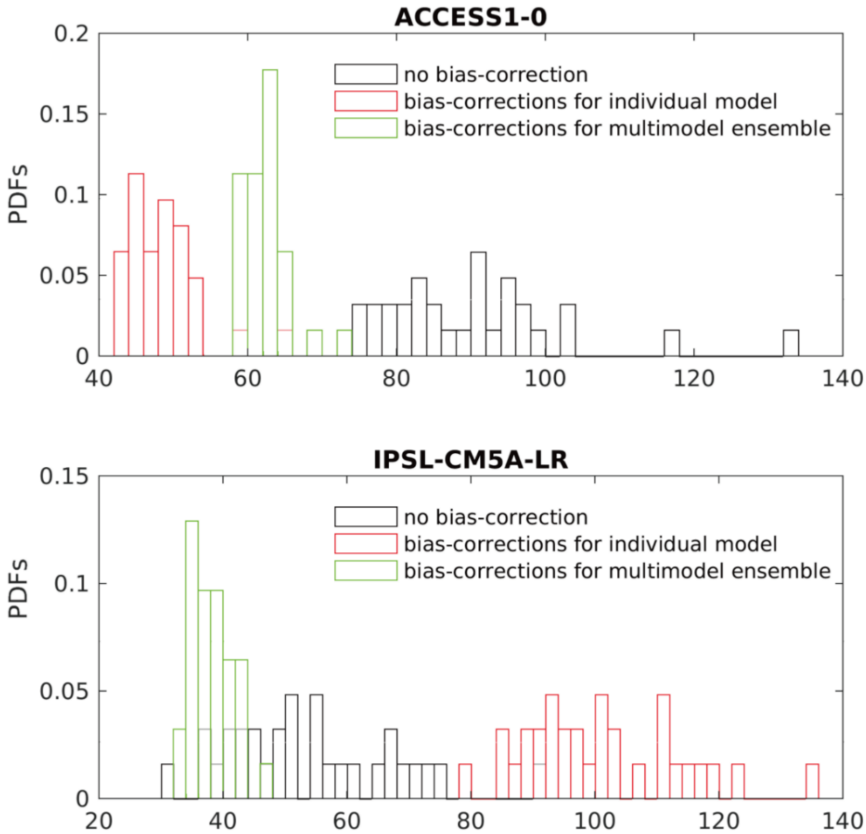


Figure R1: As an example, ACCESS1-0- and IPSL-CM5A-LR-based PDFs of mean ASO rainfall in East Africa for the period 2071-2100 under RCP8.5 scenario. Black curve is based on modelled rainfall outputs without any bias-correction. Red and green curve is based on mean- and variance-corrected GCM estimates. The former is corrected for each individual model, wherein the latter is corrected for multi-model ensemble.

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?

[Response 2.3] We thank the reviewer for pointing us to the recently published paper about Ethiopian drought of 2015, and apologies that we missed that paper. We now add several sentences in the main text to compare to this already-published manuscript, as follows: "Our result is broadly consistent with the recent analysis of Ethiopian drought projections by Philip et al. (2017), who also use observations to reduce the model uncertainty in GCM projections. They project future changes in drought by the use of only models, which can reproduce well the observed distribution of February to September climatological rainfall. They find that under RCP8.5 scenario there is no significant change in the likelihood of 2015 Ethiopian drought event. Although it is in many regards logical to exclude models that do not perform well for modelling the contemporary period, our approach is possibly more cautious. This is because there always remains a concern that a rejected model may hold important information about expected future changes, even if having strong biases in modelling the present day." (Page: 3-4, Lines: 33-5)

Following the reviewer's advice, we do take the approach of Philip et al. (2017) and apply it to our area of interest. This generates a new Figure S2 in SI, of equivalent format to Figure 1d. Based on this, we further write: "we also apply the same method as Philip et al. (2017) for our study region. This is with the CHIRPS dataset, and we place our findings in Fig. S2. The probability of 2016 ASO drought are based on rainfall projections from three models (i.e. CMCC-CM, GFDL-ESM2G and MPI-ESM-MR). They are the only models that match the climatology from CHIRPS product when using a KS test, and at a significance level of 0.1. The results are generally consistent with the both mean- and variance-corrected GCM results of Fig. 1d. That is, they indicate that the probability of drought occurrence in the East Africa may increase slightly from present towards the end of 21st century." (Page: 4, Lines: 6-11)

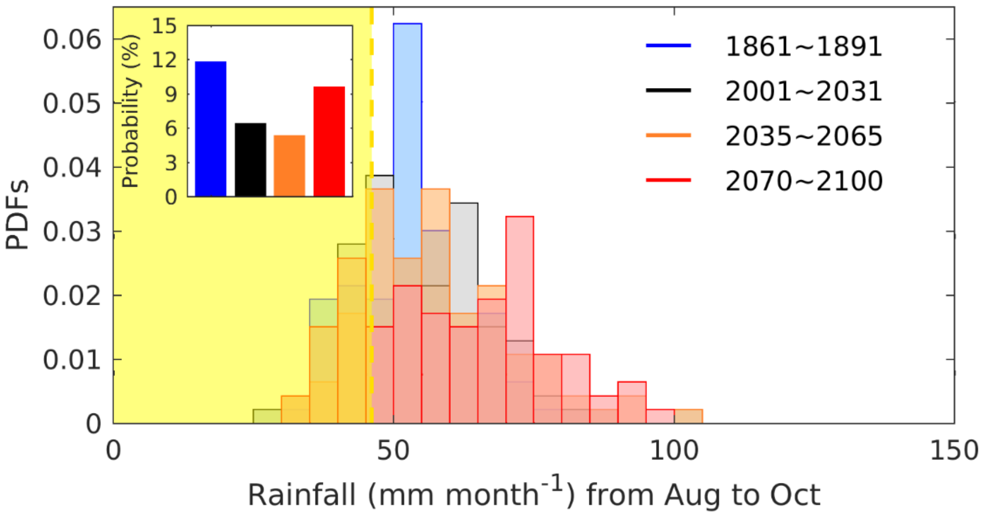


Figure S2: Following the method in Philip et al. (2017), CMIP5-based PDFs of mean ASO rainfall for periods 1861-1891 (blue), 2001-2031 (black), 2035-2065 (orange) and 2070-2100 (red). Each curve corresponds to the modelled outputs from 3 CMIP5 models (CMCC-CM, GFDL-ESM2G and MPI-ESM-MR) forced by historical emissions and RCP8.5 future scenario. GCM selection is based on the CHIRPS precipitation product. Yellow shading is mean ASO rainfall less than 40 mm month^{-1}, which is the CHIRPS 2016-based threshold (mean of ASO). Inset shows probabilities of mean rainfall of ASO falling below the threshold for the same modelled periods (colours match those of curves).

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.

[Response 2.4] We have extensively re-worked our manuscript based on these two requests. First, the new version (v4.01) of CRU-TS, recently released 20th Sep (after our last revision submission of 21st Aug) now includes data for year 2016. Hence we can now add that to our set of datasets used. We have also included the CHIRPS dataset. These two extra datasets give two new panels of our Figure 2 (repeated below).

For the second request, the ERA-interim based data results of Figure 1 have now been replaced by the observations-based high resolution precipitation product, CHIRPS. The new Figure 1 is also given below. In general terms, the original findings using ERA-interim as similar, although we appreciate now to be using a more data-based product.

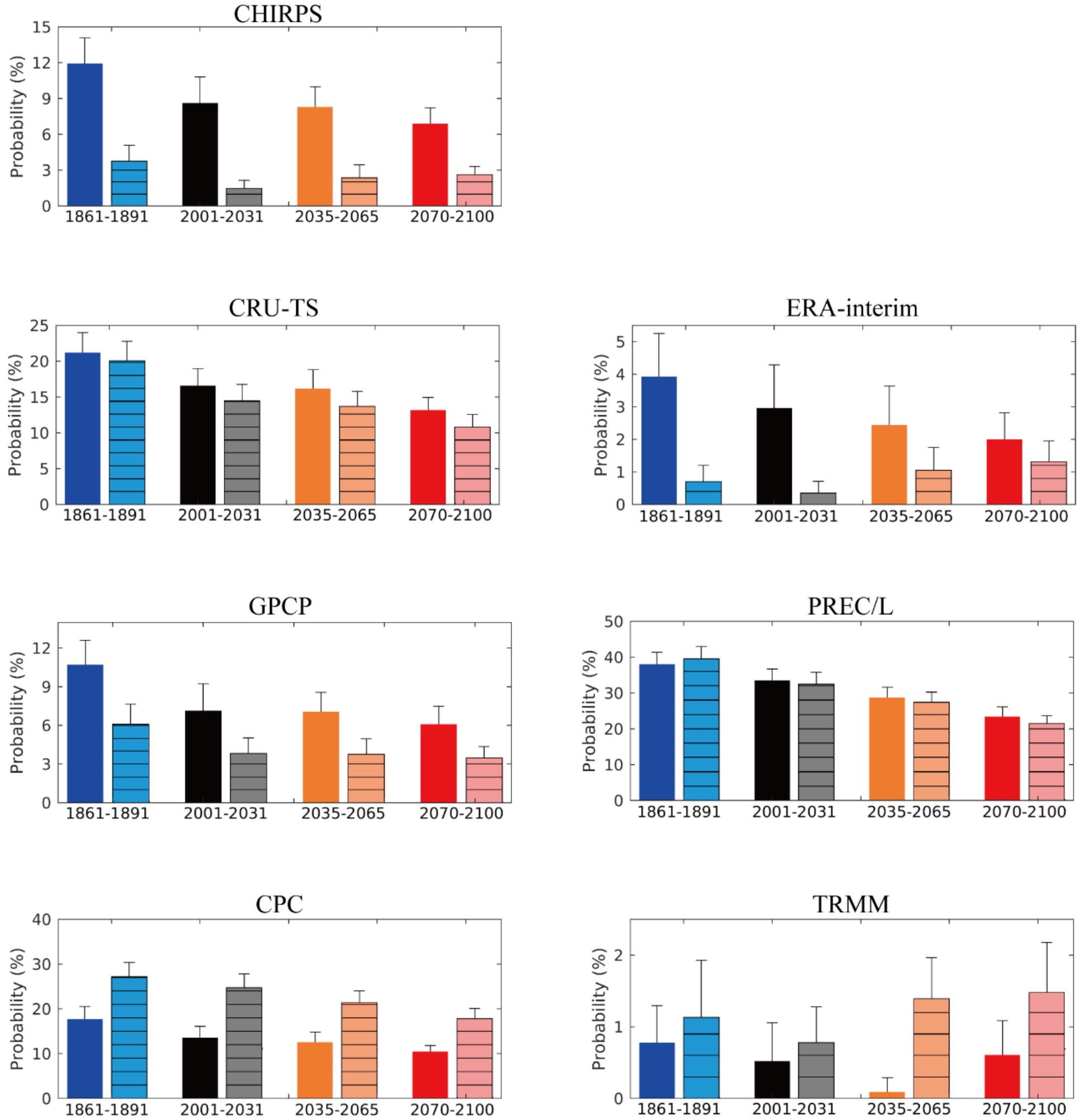


Figure 2: CMIP5-based histograms of probabilities of mean ASO rainfall falling below year 2016-based threshold values. Shown for periods 1861-1891 (blue), 2001-2031 (black), 2035-2065 (orange) and 2070-2100 (red). In each panel, GCM projections are bias-corrected by seven different observation-based precipitation products: CHIRP, CRU-TS, ERA, GPCP, PREC/L, CPC and TRMM. Each bar corresponds to merged normalized outputs from 37 CMIP5 models forced by historical emissions and RCP8.5 future scenario. The bars without horizontal hatching (left) are for the mean-corrected GCM precipitation estimates. The bars with hatching (right) are for the mean- and variance-corrected GCM estimates. The error bars are two standard deviations (estimated via bootstrapping 80% replications from the 37 GCM precipitation data for the 31-year periods)

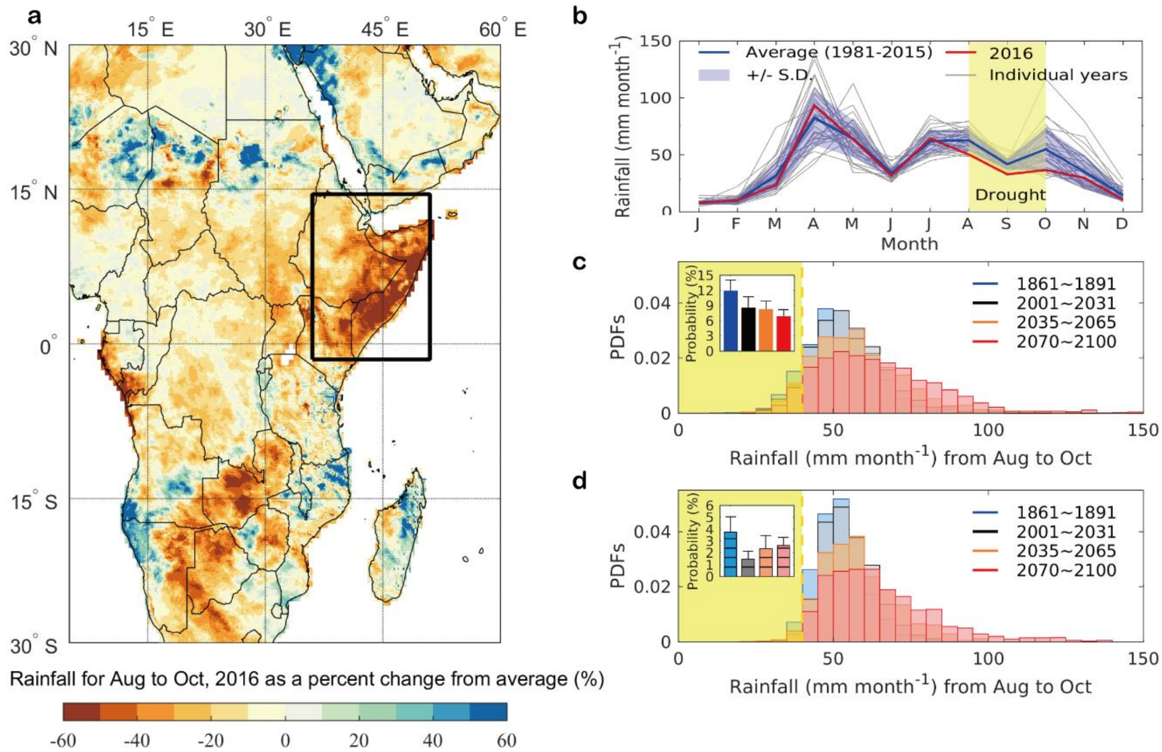


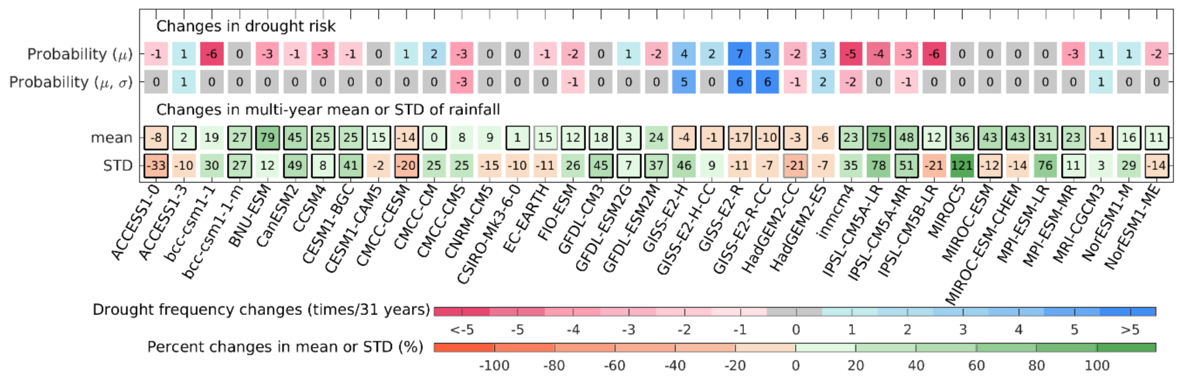
Figure 1: (**a**) Black rectangle is location of study region (14.5°N~1.5°S, 36°E~51°E). Plotted is mean rainfall for 2016 and months August to October inclusive (ASO), presented relative changes (as %) to long-term average ASO values (1981-2015). Values based on CHIRPS precipitation product. (**b**) CHIRPS-based monthly total rainfall (mm $month^{-1}$) over study region (panel a; land within black rectangle) for years 1981 to 2016. Year 2016 is red, other years are individual grey lines, and multi-year average (not including 2016) is blue line. Blue shading is ± one standard deviation of monthly rainfall across years 1981-2015. The drought event (shaded in yellow) is defined as the three consecutive months of ASO, and when rainfall in year 2016 is below blue shading. (**c**) CMIP5-based PDFs of mean ASO rainfall for periods 1861-1891 (blue), 2001-2031 (black), 2035-2065 (orange) and 2070-2100 (red). Each curve corresponds to the mean-corrected combined outputs from 37 CMIP5 models forced by historical emissions and RCP8.5 future scenario. Individual GCM bias correction is based on the CHIRPS precipitation product. Yellow shading is mean ASO rainfall less than 40 mm $month^{-1}$, which is the CHIRPS 2016-based threshold (mean of ASO, red curve in panel b). Inset shows probabilities of mean rainfall of ASO falling below the threshold for the same modelled periods (colours match those of curves). The error bars are two standard deviations (estimated via bootstrapping 80% replications from the 37 GCM precipitation data for the 31-year periods). (**d**) same as (**c**), but based on the mean- and variance-corrected GCM rainfall estimates.

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.

[Response 2.5] We recognize in light of this comment that the caption to Figure 3 is too vague. We confirm that the percentage changes in multi-year STD are based on two numbers, calculated from the 31-year precipitation data for period 2001-2031 and 2070-2100. In the Figure 3, for rainfall estimates from MIROC model corrected by CHIRPS data, the STD of annual rainfall for the period 2001-2031 is 6.38 mm yr^{-1} while that for the period 2070-2100 is 14.12 mm yr^{-1}. The percentage changes in STD is 121%, i.e. calculated as [(14.12/6.38)-1]×100%. A key enhancement based on this reviewer comment is that we now test for significance of changes in variance, using the *F*-test. These are marked with borders around its number, if changes are significant, thereby matching the similar presentation for mean changes.

In light of the original Reviewer 1 request to place more focus on STD changes, we would very much like to keep this extra row in Figure 3, but with the enhanced presentation. The caption now reads: "**Figure 3**: Rows 1 and 2 are changes in drought frequency (times per 31 years; top colourbar), for two methods of bias removal (mean-corrected only marked as "μ" and mean and standard deviation corrected as "μ, σ"). This is for each of 37 GCMs as labelled, and comparing the difference between the present period of 2001-2031 and period 2070-2100. GCM bias-correction and 2016 ASO rainfall threshold are from the CHIRPS rainfall product. Rows 3 and 4 show the GCM-based changes in multi-year mean and STD of ASO rainfall respectively, and between the same periods as top rows (bottom colourbar). Black borders indicate statistically significant differences in the 31-year rainfall mean between these two periods (row 3, *t*-test, with P < 0.05) and significant difference in STD of GCM projections (row 4, *F*-test, with P < 0.05). Light grey borders in row 3 and row 4 indicate statistically significant difference at 5%-10% significance level (0.05 ≤ P < 0.1). Values in the 3rd and 4th rows are the percentage changes in 31-year mean and STD of rainfall as $\left[\left(\overline{x^{\mu,\sigma}_{corr,2070-2100,j}} / \overline{x^{\mu,\sigma}_{corr,2001-2031,j}}\right)-1\right]\times 100\%$ and $\left[\left(\sigma^{\mu,\sigma}_{corr,2070-2100,j} / \sigma^{\mu,\sigma}_{corr,2001-2031,j}\right)-1\right]\times 100\%$, respectively. Here overbar is time-averaging over period of interest." (Page: 10)



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.*

[Response 2.6] Thank you for this, and we have done as suggested. The errorbars in Figure 1 and Figure 2 present uncertainty ranges using two standard deviations.