We thank the reviewer for reviewing our manuscript and providing his/her valuable feedbacks. We have now addressed all of his/her comments and discussed them in the following. The comments were very helpful to identify some unclear issues with regard to the scope, methodologies and conclusions of the paper. We have revised the manuscript to resolve these issues and make our approach and conclusions more clear-cut. Thanks to the reviewer's feedback, the paper is now much improved.

General Comments

- This paper compared different rainfall datasets over East Africa that have 30+ year record, to station data in Ethiopia, Kenya and Tanzania. The authors found that the CHIRPS rainfall and the ORH Tmin/Tmax are the best products to use for long-term climate studies (trend, variability, and extreme indices) and input for climate or hydrological models. While I think this paper is a good start to necessary analysis of daily rainfall products, I have concerns with the lack of independent station data and the narrow scope of the research (results are only regionally relevant, not very generalizable). I think there are ways to work around this problem of data validation in sparse-regions but the authors would need to reframe the paper and consider how to address the greater challenge of evaluating the quality of satellite rainfall in data sparse regions. The authors also need to be more transparent/detailed about their methods (metrics and data sets).

Authors´ response:

- We appreciate acknowledgement that our paper provides a good start for analysing rainfall products. We would rather propose to speak of scarcity of station data, not a general lack of independent station data. We are convinced that this kind of validation of climate data products can only be done at a regional level and indeed we feel that it is impossible to do this at a significantly larger scale than provided here (3 countries in East Africa). We rephrased respective sections (abstract, lines 6-8), introduction (page 5, lines 18-25) and summary and conclusion (page 20, lines 6-12) to make this point clearer. In fact, earlier studies with a similar goal of validating climate data (e.g., CHIRPS) products did so based on much more limited data basis in terms of temporal resolution or spatial dimension (e.g., *Duan et al., (2016) and Dembélé and Zwart (2016)*). We also think that our general approach is well generalizable and can and should be used for validating these climate data products in other areas as well.
  We added some explanations in the methodology part (section 3.2, page 9, lines 19-21) to be more transparent and clear.

Specific Comments (Major)

- The focus on daily rainfall is useful/novel as the authors state that this has not been done before. In general this is a regional study that has limited applicability to studies beyond Ethiopia, Tanzania and Kenya, and to the extent that this is generalizable in country in questionable judging from Figure. 1. Not to say that research can't be done in data sparse regions but it has to be framed appropriately, and think that the authors could improve in this respect.

Authors´ response:

- Thanks for pointing out the novelty of our study. It is true that the study is a regional study with a main focus in Ethiopia, Kenya and Tanzania. Following this we will modify sections of the paper (e.g., abstract (lines 6-8), introduction (page 5, lines 18-25) and summary and conclusion (page 20, lines 6-12)) to emphasize the regional focus – while pointing out that the approach is applicable in other regions, too.

Concerning framing we not only revised sections explaining the regional focus, but also improved the emphasis on the temporal dimension (daily resolution) and the different modes of data validation, going beyond the typical point to pixel comparison (see below).

- In fact, "how to evaluate rainfall and temperature in a data sparse region?" is a good question, although i don't think comparing to a handful of stations that are not independent is necessarily the answer. Major concern is the use of the EMA and GSOD data for evaluation and the conclusion that CHIRPS is the best performing product.

Authors´ response:

- We felt the same and this is why we posed the question – in this case for East Africa. Concerning the number of stations we did our best to get data from more than just a handful of stations: At least for Ethiopia the data set we used represents the most comprehensive to date and the most comprehensive possible (based on a quality controlled dataset from the National Meteorological Agency of Ethiopia). For Kenya and Tanzania data availability is admittedly thinner, partly due to quite restrictive data sharing policies (as explained in the paper), but the available data in our view do allow making the case that we validated the considered data sources at a regional scale. Adding data from two more countries provides additional insights about the accuracy of these data products. The concern with regard to the independency of data is taken up below. Indeed, the inclusion of station data in CHIRPS may raise such concerns. However, the station data used in CHIRPS is mainly a monthly total from a limited number of stations. For the sake of independence it would have been advisable to exclude those stations from our validation data set. However, observed station data were not included consistently through the study period and even in a single year (e.g., stations used in the first month are not all included in second month). For example, in Ethiopia, in Jan/1983 monthly data from 140 stations are included in CHIRPS and decreased to 133 in Feb/1983. In addition, in Aug/2005 data from 213 stations are included in CHIRPS and decreased to 169 in Dec/2005. In addition to the publicly available data from Ethiopia, more data from hardcopies had been added to the available stations during a working visit at the National Meteorological Agency of Ethiopia. Therefore, for our validation procedure we used quality-controlled and improved/extended daily data from as many stations as possible. For this validation data set we argued that the "dependency (is) rather weak and indirect" due to the much higher number of stations and the higher temporal resolution.

- I do think that CHIRPS is a very good product (from prior monthly/season scale evaluations and performance in hydrologic models & compared to other remotely sensed data), and it does need to be more carefully evaluated at the daily time-step.

Authors´ response:

- Thanks for pointing out the need for more careful evaluation at a daily time step. This is a very important point and that is why we tried to use the maximum possible validation areas and to find other methods than the typical point-pixel comparison. Using non-aggregated daily data resulted in a comparably weak correlation, but our 3rd method, stations average to area grid-cell average produces a good correlation. Therefore, our recommendation for hydrological (or other impact-) modelling is to use the area average (grid-cell average) instead of point/pixel information. We are using the same approach to model a water balance in one of the biggest rivers basins

in Ethiopia (area > 62,100 KM^2) and we found a good preliminary model performance of an r2=0.74, NSE=0.73 (final results will submitted to a journal).

- The station data that goes into the different rainfall products needs to be described in the methods/data.

Authors´ response:

- In Ethiopia, monthly data from 140 (January 1983) to 169 (December 2005) stations are included in CHIRPS. Additionally, in Kenya and Tanzania monthly data from 142 (Kenya) and 171 (Tanzania) in January 1983 and 62 (Kenya) and 55 (Tanzania) in December 2005 are included, respectively. We will include this information in the revised version as far as possible in section 2.2 (data sets) under CHIRPS page 7 line 21.

- In addition to the discussion. CHIRPS includes stations from several sources including GTS and GSOD, ARC includes GTS. Please include information on what stations the other products blend in.

Authors´ response:

- Thanks, ORH also used quality-controlled and gap-filled Global Summary of the Day (GSOD). This is already included in the paper (page 8, line 2), but to highlight the use of gap filled GSOD data we will modify the text as: *ORH is corrected for temporal inhomogeneity and biases and random errors are omitted through assimilation with quality checked GSOD data (Chaney et al., 2014).*

- The authors indicate that GSOD is only used in the CHIRPS monthly totals making the "dependency rather weak and indirect" Seems to me incorporating GSOD would contribute to the strong monthly correlations in Figure 4.

Authors´ response:

- As described in the paper, *"The inclusion of monthly station data can be assumed to improve CHIRPS´ performance compared to other rainfall products"*, so indeed we agree that inclusion of observed monthly totals will contribute to strong correlation on monthly time scale. While the correlated data are not fully independent, even at monthly resolution they are only partially related given that we re-calculated monthly means after quality-controlling daily values and used a much higher number of stations with more data added from hardcopies.

- From my interpretation of Funk et al. (2015) the GSOD data is included for pentad-totals as well. You may want to ask the data producers to clarify (and then include that information in the data/methods here.

Authors´ response:

- It is true that sparse GSOD data is also included in pentad-totals globally and we will add this information in section 2.2 (data sets) page 7 lines 9-21 as recommended.

- I* think* Ethiopia NMA stations are included in CHIRPS. Check with the date providers Funk et al. 2015 says: "Addi-tional observations have been provided by national meteorological agencies, primarily in Mexico, Central America, South America, and sub-Saharan Africa".

Authors´ response:

- Yes multiple stations, particularly monthly data, from Ethiopia are included in CHIRPS. But, as explained above, not all stations used in this study are included in

CHIRPS and the stations are not consistently used in the development of CHIRPS due to missing values. For example, in Jan/1983 monthly data from 140 stations are used and decreased to 133 in Feb/1983 and in Aug/2005 213 stations are used and decreased to 169 in Dec/2005. We will add more information about this in the revised version as described above.

- apparently ORH also uses GSOD "assimilating quality-controlled and gap-filled Global Summary of the Day (GSOD) in situ measurement"

Authors´ response:

- Yes GSOD data is used in ORH and this is included in the paper (page 8, line 2) to be clearer we will modify the text as described above.

- ...what spatial interpolation method do they use?

Authors´ response:

- A bilinear interpolation method is used and this will be added in (page 7, line 24) as: *ORH is developed by a spatial downscaling of the NCEP–NCAR reanalysis to a spatial resolution of 0.1º using a bilinear interpolation…..*

- Evaluation of daily rainfall for trend/variability/extremes/hydro model input is a worthwhile goal. Also not sure if the authors accomplished this given that i have questions about their metrics. I understand that you are comparing to stations but... Daily rainfall intensity: intensity is depth per unit time. How are you getting this when you just have daily totals? And then how does the "intensity" metric differ from what you describe as daily totals? Please include your definition of intensity.

Authors´ response:

- The analysis of trends/variability/extremes will be subject of another paper. We used the term average daily intensity to indicate average daily totals (mm/day) and we will change this into average daily rainfall (mm/day) in the revised version.

- Number of wet/dry days: is this just a count that does/does not match the stations? Or are you using something like probability of detection and false alarm rate? These metrics need to be defined in the methods.

Authors´ response:

- It is true that the number of wet/dry days is a count of daily records and we will add the definition as recommended in the revised version in section 3.2, page 11.

- I can't really tell what you did to come up with the results on page 13. Not obvious what "point to area-grid-cell" average means. I gather that its the average over the polygons shown in Figure 1, but this needs more explanation in the methods.

Authors´ response:

- Yes it is true "area-grid-cell average" means the basin/polygon average we will explain this in detail as recommended in section 3.2 page 9 line 21.

- Where do these polygons come from? Is there a reason why this level of basin was used to define the watersheds for a country?

Authors´ response:

- The polygons are basins retrieved from the global river basins available at the WaterBase hosted by the United Nations University (UNU-INWEH:

http://www.waterbase.org) and we will add few lines about the polygons and the data source in the revised version in section 3.1, page 8 lines 23-24.

- Since you're not comparing to hydrological/ streamflow data why not just average from 0.05 to 0.25 degree – essentially producing the same results as what you discuss with the coarser CHIRPS data?

Authors´ response:

- That is possible, but our objective is to find finer spatial resolutions that can be used later for hydrological and climate modelling in areas of the region with no ground observation. As we showed in section 4.2, page 15 lines 1-6, the improved version of CHIRPS (0.05 degree) is more accurate than the coarser resolution of CHIRPS (0.25 degree).

Specific comments (Minor)

- Additional information on how the data is produced should help explain your results (e.g. why is point to area-average best, does this have to do with the interpolation schemes that ARC and CHIRPS and the other product use?

Authors´ response:

- This is not because of the interpolation schemes that the products used but due to the method we used to compare the products. Compared to the point to pixel method, area averaging produces higher correlation and lower errors. During area averaging extremely high rainfall events obtained for a location from the various data products are levelled off by averaging and this makes the product much more accurate. In most of the rainfall products, including CHIRPS, there are occasionally higher daily rainfall values recorded and the averaging removes those extremes, which are much higher than the observed data in the area. We will include this information in the revised version in page 17 lines 1-3.

- Do CHIRPS results improve at 0.25deg because that is its original resolution, before being downcaled to 0.05deg with the CHPclim? This kind information will be useful for the other products as well.

Authors´ response:

- NO, there seems to be a misunderstanding. As we showed in section 4.2 page 15 lines 1-6, the higher resolution of CHIRPS (0.05 degree) showed an improvement in correlation with station data by up to 3.2% compared to the coarse resolution of CHIRPS (0.25 degree).

- Is only a historic record needed for env. Management? ORH isn't updated regularly (2012?) This should be clear in the paper, pls include in methods.

Authors´ response:

- We will include more detailed explanations and discussions on data availability and usage in the revised version. And yes, for environmental management real-time observations and projections of future climatic conditions are more important than historic records. We are working on this issue and added a respective section in the discussion.

- Meanwhile, ARC & CHIRPS are updated regularly. It will help contextualize the metrics if you discuss the products strengths and weaknesses more with some example of an environmental management application that they might be used for. I

am sure there are some that would benefit from ORH long record, or ARC's 1-day latency.

Authors´ response:

- Thank you very much for this remark and proposal. We will include the strength and weakness of the products in terms of resolution, length of time period and progress (regular updating) in the data set (section 2.2) and discussion part of the revised version.

- If you are including OHR why not include what they use routinely in the Africa Flood and Drought monitor? 3B42RT...

Authors´ response:

- We agree that the 3B42RT is a very good rainfall product and lots of papers are already available, but for our project we are interested in products with an observation period of more than 30 years, which is not the case for 3B42RT.

- how does blending datasets impact the application to environmental management? With respect to hydrologic modeling GLDAS (Rodell et al. 2004) uses ORH/Princeton+other, Africa flood and drought monitor (Sheffield et al. 2014) uses ORH/TRMM-RT, FLDAS (McNally et al. 2017) uses CHIRPS. How does all this relate to the climate models?

Authors´ response:

- This is a very good point and the products have shown a positive impact in environmental management in data sparse regions such as Africa. Compared to climate models, satellite based rainfall products, based on our findings, show higher accuracy (see figure 4) and we believe that satellite based products can be more accurate for environmental management than output from climate models. The role of data blending, including the recommended papers, in environmental management will be highlighted in the introduction part of the revised version.

- intro was vague and too focused on data scarcity - we have lots of data (models, remote sensing, some in situ)...just not lots of dense rainfall stations.

Authors´ response:

- Thank you very much; we will specify this in the revised version, pointing out that data scarcity mainly refers to station data.

- …there are lots of datasets to get temperature (e.g. MERRA-2, CFS-R). Why weren't these included?

Authors´ response:

- We screened multiple data sources for temperature and our selection is based on their spatial resolution, length of time period (> 30 year) and recommendations from previous papers.

- Technical corrections > fix citation (also 2017): Kimani, M., Hoedjes, J. and Su, Z.: Uncertainty Assessments of Satellite Derived Rainfall Products, , 15. doi:10.20944/preprints201611.0019.v1, 2016. Don't cite the pre-print use this one: Kimani, Margaret Wambui, Joost CB Hoedjes, and Zhongbo Su. "An Assessment of Satellite-Derived Rainfall Products Relative to Ground Observations over East Africa." Remote Sensing 9.5 (2017): 430.

Authors´ response:

- Thank you very much and this will be fixed in the revised version.

- Typo RFE pg 6...its RFE Rainfall Estimation Version 2 (REF 2.0) (Novella et al., 2013)

Authors´ response:

- Thank you very much and this will be fixed in the revised version.

References:

Chaney, N. W., Sheffield, J., Villarini, G., Wood, E. F., Chaney, N. W., Sheffield, J., Villarini, G. and Wood, E. F.: 20 Development of a High-Resolution Gridded Daily Meteorological Dataset over Sub-Saharan Africa: Spatial Analysis of Trends in Climate Extremes, Httpdxdoiorg101175JCLI--13-004231 [online] Available from: http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-13-00423.1 (Accessed 30 November 2016), 2014.

Dembélé, M. and Zwart, S. J.: Evaluation and comparison of satellite-based rainfall products in Burkina Faso, West Africa, Int. J. Remote Sens., 37(17), 3995–4014, doi:10.1080/01431161.2016.1207258, 2016.

Duan, Z., Liu, J., Tuo, Y., Chiogna, G. and Disse, M.: Evaluation of eight high spatial resolution gridded precipitation products in Adige Basin (Italy) at multiple temporal and spatial scales, Sci. Total Environ., 573, 10 1536–1553, doi:10.1016/j.scitotenv.2016.08.213, 2016.