Reply to Anonymous Referee #2

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 is specifically validating the quality of three climatic variables coming from different satellites data-streams and models using scientifically proven quality validation methodologies. The three include rainfall, maximum temperature and minimum temperature. Being a research that has been done for the first time that I know of, the paper unravels the different quality of each of these datasets and with evidence provide great knowledge of which is the best among the 6 datasets for each variable. If further validate the same dataset with observed rainfall and satellite from weather stations. Though not conclusive, through this research, one can relate that the CHIRPS dataset is better for rainfall analysis in specific areas which have complex topography with a case study in three East African countries. While the ORH dataset works best for the Tmax/ Tmin variable. The paper specifically highlight the methods used and why and how each one is best.
  East Africa being a complex region of climate analysis. The paper seems to have limited itself to specific sites which might not fully represent the entire region. Despite having fewer observed data the sample areas of interest might limit the imagination of the complexity of the region.

Authors´ response:

- Thanks for pointing out/acknowledging the difficulties in performing this kind of studies. Indeed the observed data is very limited in terms of spatial coverage and length of time period. In addition, getting daily data from the meteorological agencies is not easy, particularly from Kenya and Tanzania due to their data sharing policy. Therefore, for the purpose of validation we used almost all the available stations (210), particularly in Ethiopia, with an elevation, as given in Table 1, from 400-2510 and average elevation of the validation areas from 520-2830 meters. The included station data from Kenya and Tanzania range in altitude from 1097 to 2328 masl. Therefore, based on the stations at different elevations we conclude that our data set (which is the most comprehensive for East Africa to date) reasonably represents the study area even in those parts were no ground observations are available.

- CHIRPS products seemed to work well in some areas while at the same time came in second in other areas. The author should try and indicate by what percentage in all the analysis done was CHIRPS top and if the percentage is worth representing the region as the best dataset.

Authors´ response:

- We concluded CHIRPS to be the best product for rainfall based on the overall analysis (daily-monthly), the multiple statistics used (Table 3 and Figure 4 and 5) and the result of the analysis of rainfall characteristics (Table 6, see below). Table 6 will be added in the revised version page 14 and the text in lines 1-3 will be modified as: *On average, over the 21 validation areas, CHIRPS captures well the number of wet days (99.8 %), average duration of wet (87.5 %) and dry periods (84 %), average* *total rainfall (95.6 %), average amount of wet periods (84.3 %), and average daily rainfall (93 %) (Table 6).*

Specific comments

- In page 9 of the document the author mentions that "The quality of selected stations was checked and extremely high rainfall records during dry seasons were excluded." Through this statement it is not clear what is considered as a dry season and the reason for exclusion of such rainfall dataset remains hanging. Also in consideration of the same, extreme event such as flash floods may be recorded in a single days' rainfall.

Authors´ response:

- During quality control few data were removed with extremely higher rainfall events such as >480 mm/day preceding and following dry days. We have done the quality control with the meteorology-experts in the field (colleagues from National Meteorological Agency, Ethiopia) and these data had been identified as error in inputting the data.

- A few questions to be asked are; Could the x,y decimal places affect the location of a given station ending up reporting a value for a wrong location? For example a station reading of 36.123456, -1.123456 might fall at a different location compared to a reading of 36.123, -1.123. In this reference were the station locations validated?

Authors´ response:

- Good point: It is well conceivable that stations are falsely located in the next grid box of the product especially if you have a product with very high resolution such as CHIRPS and ARC2. For this reason we used all available station information and the extracted data is validated. This is can be a problem if you are comparing station to pixel. But if you are using the station average to area grid cell average the change from 36.123456, -1.123456 to 36.123, -1.123 might not be a problem if they are located inside the validation area as you are taking the average.

- From this paper it is also not clear what the following terms refer to; Wet days, duration of wet days and average amount of wet periods this might be confusing since they all are represented by one unit which is days. For example, when we talk about wet days we say 10 days. If we talk about duration of wet days do we still say 10 days? The same applies to the average amount of wet periods.

Authors´ response:

- Thank you very much for pointing out this issue. The unit for number of wet days (count of wet days in a year) is days/year, for average duration of wet periods (the number consecutive wet days) is days and for average amount of wet periods is mm. We will include the full description of the rainfall characteristics with their units and their need in rainfall modelling with some references (e.g., Jebari et al., 2012) in the methodology part (section 3.2 page 11) of the revised version and the results will be provided in Table 6 as described above (see table below).

- From the paper it is very clear that the author highlights CHIRPS as the best rainfall product while ORH as the best temperature product. CHIRPS comes out better than the rest based on the characteristics described by the author in page 14 but the author has not conclusively stated by how much is CHIRPS better than all this other products if you compare all the statistical analysis done. The Author has only highlighted that

"In general, the observed rainfall characteristics 15 are well captured by CHIRPS compared to CHIRP, ARC2, ORH, RCM, and RCMs."

Authors´ response:

- Thank you very much and we provide a table in page 14 with a summary statistics of the rainfall characteristics as given below (Table 6) in the revised version and the text will be modified as: *In general, the observed rainfall characteristics are well captured by CHIRPS compared to CHIRP, ARC2, ORH, RCM, and RCMs (Table 6).*

- While at the same time pointed out areas that ARC2 has performed better that CHIRPS and CHIRP. Regarding the above, in some instances such as EthioShed4 the CHIRP and CHIRPS have equal R squared while in some areas ARC2 came on top. Through the analysis of all the Sheds analysed what percentage of CHIRPS compared to the rest of the datasets was better.

Authors´ response:

- Yes it is true that both ARC2 and CHIRP have shown higher R squared, considering the biases and errors, in 2 and CHIRPS in 17 of the 21 validation areas. However, in terms of capturing the daily rainfall characteristics (Table 6 ARC2 showed higher deviations compared to CHIRPS. In EthioShed4, CHIRP and CHIRPS have an equal R squared, but in terms of biases CHIRP showed higher biases (data points below and above the regression line) compared to CHIRPS (most of the data points lie in the regression line) as shown in figure 4. We will highlight this more explicitly in the revised version.

- Still in line with that there are some areas where all the R squared were between 0.13 and 0.55, is it possible to elaborate on why such cases occur? Is it the methodology used to model the datasets that limits the correlation with the station data?

Authors´ response:

- The small R squared values were mainly computed for the regional climate models (RCMs). Compared to the satellite based rainfall products (which already include ground observed data), RCMs have a coarse spatial resolution (~ 50 km) and during the downscaling process of the global climate models they include less local information such as topographical features, which makes them weak in synthesising local daily rainfall, particularly in topographically very complex regions.

- Another question of concern is what explains the equal value for CHIRP and CHIRPS as portrayed in EthioShed4?

Authors´ response:

- In EthioShed4 it is true that R square value is the same, but if you see the distribution of the data above and below the regression line there is a difference, which is explained as a bias (over- or underestimation). To compare both products it is also good to see figure 5 (Taylor-diagram), which shows the correlation and standard deviation of each product on monthly time scale. Therefore, in addition to the bias shown in figure 4, figure 5 also shows a deviation between CHIRPS (with slightly better correlation and standard deviation on monthly time scale similar to figure 4) and CHIRP.

- In the introduction the paper highlights CHIRPS as a dataset that has both station and satellite data in it. Might this explain the high correlation?

Authors´ response:

- As we explained in the methodology and discussion part, station data are included in CHIRPS, ARC2 and ORH. But compared to ARC2 and ORH, a larger number of stations are included in CHIRPS. Therefore, on a monthly time scale, the high correlation can be true due to the inclusion of monthly station data in the development of CHIPRS. In the revised version we will include a section in the methodology part to highlight the stations included in CHIRPS (see also replies to reviewer #1).

- Are the same stations in CHIRPS used to validate the CHIRPS product?

Authors´ response:

- Yes – but due to different data processing and inconsistent use of station data in CHIRPS, the data included there are not fully congruent to the station data we used in the correlation. Multiple stations, particularly monthly data, from Ethiopia are included in CHIRPS as shown in table one and discussed in the discussion part. But, not all stations used in this study are included and the stations are not consistently used in the development of CHIRPS due to missing values. For example, in Ethiopia, in Jan/1983 monthly data from 140 stations are included and decreased to 133 in Feb/1983. In addition, in Aug/2005 monthly data from 213 stations are included and decreased to 169 in Dec/2005, which shows the inconsistency in the inclusion of the stations. We will add more information in section 2.2 (data sets) under CHIRPS page 7 line 21 in the revised version.

In conclusion to the specific comments.

- The paper is very clear on how the validation is done. However, more can be done to ensure that these products are regarded as the best products as indicated by the author.

Authors´ response:

- Thank you very much for the comment and we will consider your comments in the revised version, e.g. by adding a summary statistics table (e.g., Table 6) as mentioned above.

- The paper currently is validating the products for areas with low observed dataset. Perhaps, the author can use historical analysis as a means of validation too. Also, an elaborate point of validation would be to highlight how the non-blended datasets such as CHIRP is performing compared to observed station data in regions that have well established network of weather stations such as the developed countries. Then further, validating the CHIRP against the CHIRPS. This will basically ensure less redundancy.

Authors´ response:

- Thank you very much and we elaborate on CHIRP with ground observation as compared to CHIRPS in more detail in the revised version as recommended. Concerning the use of historic data: we did that to the extent possible, limited by the lengths of available time series, but in general using 30-year datasets. However, it would go beyond the scope of this paper to extend the analysis to further (developed) countries and regions. Adding even more data, figures and respective discussions to the paper does not seem feasible to us. We definitely agree that it would be worthwhile to use the same approach of validating climate data products in other countries and regions. It is well conceivable that obtained results in terms of best products might look different, which could be due to various factors: higher spatial resolution, better general data quality, higher homogeneity of the region in terms of topography etc. To conclude: we feel that extensions to further regions and ultimately to the global scale requires (a) separate study/studies.

Technical comments

- In page 7, the Dekadal should come after pentadal since the former represents 10 days and the later represents 5 days.

Authors´ response:

- Thank you very much and we will fix this in the revised version.

- In page 13, 17 is numerical while three and one are text – you might want to use either for all.

Authors´ response:

- It is very common to convert numbers <10 to text, but not larger numbers. This will be checked with editorial policies of HESS.

- In Page 20, it is indicated that "The products are available with higher spatial and temporal resolution and for longer periods." – doesn't longer periods mean the same as temporal resolution?

Authors´ response:

- No, the terms have different meanings. Temporal resolution is used to indicate the time scale such as daily, dekadal and monthly; longer periods refers to the length of the time period/series such 30 or >30 years.

Table:

Table 6: Summary of daily rainfall characteristics retrieved from multiple rainfall products and averaged over the validation areas of Ethiopia, Kenya and Tanzania. The value which comes closest to the observed value is highlighted in bold and values in brackets give the standard deviation.

Rainfall characteristics	Obs.	CHIRP	ARC2	CHIRPS	ORH	HadGEM2	MPI	GFDL	RCMs
Number of wet days (days/year)	189.58	351.06	162.98	**189.26**	192.14	205.08	243.55	210.42	299.37
Average duration of wet periods (days)	5.86 (11.4)	167.96 (171.8)	4.80 (7.6)	**5.13** (8.1)	3.02 (4.0)	11.70 (26.2)	12.17 (28.7)	9.37 (18.3)	21.36 (48.7)
Total amount of precipitation (mm/year)	953.63	**980.24**	671.62	912	1027.02	841.73	1055.7	1253.38	1068.6
Average amount of wet periods (mm)	30.2 (84.4)	498.43 (562.6)	20.56 (48.6)	**25.46** (63.4)	15.64 (34.6)	50.12 (165.3)	55.45 (183.8)	59.64 (166.7)	78.88 (245.1)
Average duration of dry periods (days)	5.37 (9.5)	1.53 (0.82)	6.01 (10.8)	4.5 (6.3)	2.55 (3.2)	6.91 (10.8)	**5.67** (8.3)	6.55 (10.1)	3.55 (4.3)
Average daily precipitation (mm/day)	5.28	2.78	4.16	4.88	**5.4**	3.84	4.19	5.69	3.48

Reference:

Jebari, S., Berndtsson, R., and Bahri, A., (2012). Soil erosion estimation based on rainfall disaggregation. J. Hydrology. 436-437; 102–110.