Dear Prof. Geoff Pegram,

We would like to thank you for your time and effort that you put into the review of this manuscript! In the following a point to point response to each of your comments.

Sincerely yours,

The authors

126: The sites in Fig 1e do not match with Figs 2 and 4. In Fig 2, station 32 is mis-labelled.

Clarification:

Lines 122 to 128 describe the discharge observations and links to Figure 1e to see the ground observational network that we used.

Figure 2 is mentioned in the preceding section (ll. 114-120). This section describes the sources that were tapped onto to compile a list of flood events that were reported for Africa in 2003 in various disaster databases. This database is visualized in Figure 2 and is used as reference to verify AFFS forecasting performance.

Figure 4 is first time mentioned in the methodology part under section “flood events” (ll. 331-2). It is stated that information regarding AFFS forecasted flood events including location and timing was compiled into another database, which is shown in Figure 4.

Hence, all three Figure show different information: Figure 1e the observational network; Figure 2 the list of reported flood events and Figure 4 the list of forecasted flood events using AFFS.

Changes to the manuscript:

- We ask the typesetting team to not shrink Figure 1 to such a small size as we provide it in A4 size. This figure contains many important information related to the study area and it is important that the reader can actually see the information of the maps and also read the legends.
- Changed ll.125-6 to: The resulting ground observation network comprises 36 discharge measuring stations holding daily observations between 2003 and 2008 (Fig. 1e).
- For Figure 4, we decided to move ll.329-332, which holds the first description) it to the results part under the section “flood events” as it is an actual result and should not be mentioned earlier. We hope that this will also prevent any further confusions.

220: it is not clear what data were used for calibration - were these actual riverflow records from gauging stations? Were these the set mentioned in line 125 ff from "ground observation network comprises 36 discharge measuring stations holding observations between 2003 and 2008 (Fig. 1e)."?

Clarification:

Yes, LISFLOOD was calibrated using daily discharge records at the 36 locations shown in Figure 1e. The calibration was done between 2003 and 2008.

Changes to the manuscript:

- Ll. 220-222 were changed to: For the pan-African set-up, LISFLOOD was calibrated at the 36 locations shown in Figure 1e (black dots) using daily discharge records, over a time period of five years (2004–2008; 2003 used as warm-up). Those 36 sub-catchments correspond to 11 hydrological basins.

255-270: the use of KSE' with apostrophe rather than prime as defined in eq (1a) is inconsistent, here and elsewhere - 336ff.

Clarification:

The KGE with apostrophe (=KGE') denotes the modified version of the Kling-Gupta Efficiency. Throughout the manuscript the denotation KGE' has been used consistently used, also in Eq 1a and in l. 336ff. The only time KGE was mentioned in the text without apostrophe is in ll.270f "The benefits of using KGE' over KGE or Nash–Sutcliff Efficiency are discussed by Gupta et al. (2009) …", where reference is made to the original version of the Kling-Gupta Efficiency.

Changes to the manuscript:

- Ll.270f were changed to: The benefits of using the modified version of the Kling-Gupta Efficiency (KGE') over the original one (KGE) or Nash–Sutcliff Efficiency are discussed by Gupta et al. (2009)
- Figure 6 was updated and includes not the apostrophe

283: the reader has to wait until line 329 to realise that the meaning of Pobs was "Information regarding observed flood events" , so please fix

Clarification:

P^{obs} does not mean "Information regarding observed flood events". It refers to the cumulative distribution function of the observation. In ll 281-283 it is mentioned: "… the CRPS (Continuous Rank Probability Score), […] compares the cumulative distribution function of a probabilistic forecast (PhydEPS), to the observation (Pobs), …"

Changes to the manuscript:

- ll 281-283 were changed to: "… the CRPS (Continuous Rank Probability Score), […] compares the cumulative distribution function of a probabilistic forecast (PhydEPS), to the cumulative distribution function of the observation (Pobs), …"

325: add " see equations 4 to 6"

Clarification:

Ll 325ff show the equations POD, CSI and FAR and also numbers them as 4, 5 and 6 respectively. I don't know exactly where we should include "see equation 4 to 6" at that point, therefore I would suggest to do the following change.

Changes to the manuscript:

- ll 324ff was changed to: They are calculated as follows: [then the 3 equations follow]. All are expressed as percentages. The optimum value for POD and CSI is at 100%; whereas it is 0% for FAR.

332: again the mismatch between Figures 1e and 4 - this makes the argument difficult to follow.

Clarification:

This is a misunderstanding by the reviewer of what we are showing in Figure 1e and Figure 4; which has already been addressed under his comment related to l.126. We hope that this issue is solved by that explanation, plus the changes.

347: the poor validation of the Zambezi appears to be principally because it was during a short period of relatively low flows, not the full range

Clarification:

Yes, we agree that this is also a contributing factor.

Changes to the manuscript:

- ll. 345ff were changed to: The difference in KGE' between the calibration and the validation period is largest in the Zambezi catchment. Possible reasons can be seen in Fig. 6b. At this particular location there is a lack of data in the calibration period. On top of that, the flow during the few years for which data were available was relatively low in comparison to the one in the validation period, hence the calibration did not cover the full range of flow conditions which surely contributes to a suboptimal calibration.

357: "Olifants river" is not mentioned in Figures 1, 2 and 4

Clarification:

In l.357 there is the reference made to Figure 6c. In that figure 6c it is seen that the station we are looking at in the Olifants river is called “Loskop North”, which is also mentioned in Figure 1e.

They are not mentioned in Figure 2 and 4 as those figures shows different things as already explained above and modifications were done to prevent any future confusions.

367: Figures 7a and 7b have been switched

Clarification:

The reviewer is absolutely right. We switched the plots because like that they were better in line with the order that they are mentioned in the manuscript. We also created the plot newly, but the wrong (not updated one) was included in the manuscript. We rectified that now.

375: Figure 8 is difficult to read for 2 reasons, the image and text are very small and the colour palette does not grade from light to dark nor does it have a comfortable colour gradation, so the message is lost. Rather plot the two sets of predictabilities as histograms versus sizes of named catchment areas?

Clarification:

We fully agree with the reviewer the plots have been printed too small and like that it is very difficult to see the differences. It is not like that if it was printed in the size we provided the figure, which is then very clear.

Changes to the manuscript:

- request to the typesetting team to not shrink Figure 8 (and also Figure 10) to a size that it is very difficult to impossible to see the information displayed. Thank you.

387: ENS is not in the glossary - does it mean Ensembles?

Clarification:

Here: ENS = meteorological ensemble predictions

Changes to the manuscript:

- l.387 was changed to: “… the decrease in forecasting performance cannot be affiliated to possible inaccuracies of the ~~ENS~~ meteorological ensemble predictions only, but there must be other additional influencing factors.”

413: POD etc - it would help the reader if these were qualified here

Changes to the manuscript:

- l. 413 was changed to: “… resulting into a general Probability of Detection (POD) of 69 %, a False Alarm Rate (FAR) of 29% and a Critical Success Index (CSI) of 54 %.

418: 10 km^2 should be 10 km^3

Corrected.

434: This is the first mention of the Sabi other than in the introduction - what makes it special? - also Fig 1 is very difficult to read - enlarge?

Clarification:

As mentioned in the introduction, in addition to the evaluation of the hindcast, we wanted to illustrate the flood forecasting performance of AFFS on a concrete example to show the reader to potential outcome/ products. We applied all the methods described in Section 2.3 to arrive to this forecast. As it is in an ungauged basin, none of the statistics can be calculated on this forecast, but it can be crossed compared with reports found in disaster databases. Doing so, it shows that AFFS is capable to predict flood events even in ungauged river basins, which emphasises at the same time its value.

We agree that Fig 1 is too small. This is not a mistake by the authors as the figure is originally in A4, but is printed in the .tex file too small. We assure to pay attention that the figure will be printed in a good size once the official typesetting is done.

525: the Glossary (not labelled as such) splits the last sentence of the acknowledgements.

That is also a typesetting error, and we hope that it will be rectified.

Dear referee,
We thank you kindly for your suggestions which will surely add value to the current manuscript and also thank you for pinpointing to a mistake that has unfortunately happened. Please find below our response including the adjustment on the manuscript to each point raised.
Sincerely Yours,
The authors

---

1. **The forecasting system does not use assimilation of in-situ or remotely sensed hydrological observations (river discharge, soil moisture, storage change etc.) to increase sharpness and reliability of predictions. I think somewhere in the discussion, this opportunity should be mentioned and discussed as one way to strengthen the system in the future.**

We totally agree with the reviewer on this point and therefore suggest the following addition to the manuscript [discussion part (page 15, ll496):
This shows that the system works well with a minimum number of ground observations, while at the same time, it indicates a good potential for further improvements once more observational records become available. In this context remote sensing might become a valuable alternative source of observed land surface hydrological fluxes. Surface-water related signals are translated into estimates of e.g. streamflow, soil moisture, land-surface temperature, surface water height and inundation extend. The research on the potential benefit of assimilating those estimates into hydrological applications is increasing steeply and have shown already now promising results (Andreadis et al., 2007; Glen el al., 2014; Hirpa et al., 2013,2014; Munier et al., 2015; Pedinotti et al., 2014; Revilla-Romero et al., 2014; Tarpanelli et al., 2013; Wanders et al., 2014).
As assimilating those data might improve the forecast ability of AFFS by e.g. improving estimates of the initial conditions or the timing of the flood peak, it should be focus of future research.
Furthermore […]

- Andreadis, K.M., Clark, E.A., Lettenmaier, D.P., Alsdorf, D.E., 2007. Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. Geophys. Res. Lett. 34, L10403. doi:10.1029/2007GL029721
- Gleason, C.J., Smith, L.C., Lee, J., 2014. Retrieval of river discharge solely from satellite imagery and at-many-stations hydraulic geometry: Sensitivity to river form and optimization parameters. Water Resour. Res. n/a–n/a. doi:10.1002/2014WR016109
- Hirpa, F.A., Hopson, T.M., De Groeve, T., Brakenridge, G.R., Gebremichael, M., Restrepo, P.J., 2013. Upstream satellite remote sensing for river discharge forecasting: Application to major rivers in South Asia. Remote Sens. Environ. 131, 140–151. doi:10.1016/j.rse.2012.11.013
- Hirpa, F.A., Gebremichael, M., Hopson, T.M., Wojick, R., Lee, H., 2014. Assimilation of Satellite Soil Moisture Retrievals into a Hydrologic Model for Improving River Discharge, in: Lakshmi, V., Alsdorf, D., Anderson, rtha, Biancamaria, S., Cosh, M., Entin, J., Huffman, G., Kustas, W., Oevelen, P. van, Painter, T., Parajka, J., Rodell, tthew, Rüdiger, C. (Eds.), Remote Sensing of the Terrestrial Water Cycle. John Wiley & Sons, Inc, pp. 319–329.
- Munier, S., Polebistki, A., Brown, C., Belaud, G., Lettenmaier, D.P., 2015. SWOT data assimilation for operational reservoir management on the upper Niger River Basin.

- Pedinotti, V., Boone, A., Ricci, S., Biancamaria, S., Mognard, N., 2014. Assimilation of satellite data to optimize large-scale hydrological model parameters: a case study for the SWOT mission. Hydrol Earth Syst Sci 18, 4485–4507. doi:10.5194/hess-18-4485-2014
- Revilla-Romero, B., Thielen, J., Salamon, P., De Groeve, T., Brakenridge, G.R., 2014. Evaluation of the satellite-based Global Flood Detection System for measuring river discharge: influence of local factors. Hydrol Earth Syst Sci 18, 4467–4484. doi:10.5194/hess-18-4467-2014
- Tarpanelli, A., Brocca, L., Melone, F., Moramarco, T., 2013. Hydraulic modelling calibration in small rivers by using coarse resolution synthetic aperture radar imagery. Hydrol. Process. 27, 1321–1330. doi:10.1002/hyp.9550
- Wanders, N., Karssenberg, D., de Roo, A., de Jong, S.M., Bierkens, M.F.P., 2014. The suitability of remotely sensed soil moisture for improving operational flood forecasting. Hydrol Earth Syst Sci 18, 2343–2357. doi:10.5194/hess-18-2343-2014

2. **I believe it would strengthen the analysis and the paper if the “seasonal mean” benchmark was replaced by a “climatology” benchmark (i.e. historic mean for each day of the year). A 60-day moving average implies quite a lot of smoothing in these highly seasonal river systems.**

We thank the reviewer for this important point. The ESP approach is already nearly identical to this request. The ESP uses observations from the same (and the 10 following days) calendar day from the last 20 years as forecasts (thus representing the calendar day mean as the ensemble of the last 20 years observations). The ESP is harder to beat than a mean for the same day as it also represents the dynamic of the forecasted days. The information requested can be already seen via the comparison between ESP and AFFS forecasts, which is discussed in the text. This computations would not add any further information and findings.

3. **I am having problems with figure 7, esp. with the curves showing performance relative to the persistence benchmark. Persistence is a perfect forecast for a forecasting horizon (fh) of 0 and then quickly degrades with increasing fh. I would therefore expect that CRPSS using persistence as benchmark starts relatively low (must be negative for fh=0) and then increases. The steep descent in CRPSS with increasing fh shown in fig 7b does not look reasonable to me. In essence, one would expect that persistence is harder to beat for short fh than for long fh. I know that the persistence benchmark has been added in the first revision of the manuscript and would suggest that the authors check up on this. I also noticed that the labeling of fig 7 subplots has been mixed up in the caption.**

The reviewer is perfectly right on this point. Reason for this “strange behavior of the graph” is that the “1-“ had unfortunately been forgotten at the beginning of the calculation. We are very grateful to the reviewer for pinpoint to this mistake that is now corrected.

In Fig. 7 the two different sets of CRPSS are plotted over the 10 days lead time. ~~Both sets show the same general tendency:~~ Comparing against the seasonal mean as benchmark the average CRPSS is decreasing as the lead time advances, meaning that the error increases, i.e. AFFS’ skill to forecast streamflow decreases (Fig. 7a). This is also confirmed by the number of stations with positive CRPSS, which continuously decreases over the 10 days lead time (Fig. 7b). Decomposing the CRPSS for different regions in Africa shows ~~(here: seasonal mean as benchmark)~~ that only a small number of stations in Eastern Africa (20 %) have skilful streamflow predictions, while in Western Africa the majority of stations

(70–90 %) show skilful streamflow predictions. Comparing against the last observation as benchmark, shows however that the skill of the AFFS’ prediction is increasing steeply at day 2 and remains high till the end of the forecast (day 10). This shows that after 2 days lead time the forecasts based on AFFS are much more skillful than presuming that the last observation will remain stagnate throughout the forecasting period.