Interactive comment on "A pan-African Flood Forecasting System"
by V. Thiemig et al.

We would like to thank the reviewer for giving very useful remarks and advice on our paper. We have taken all of them into account and provided clarifying answers as well as propositions for changes in the text, tables and figures.

Anonymous Referee #3
Received and published: 18 July 2014

The manuscript presents an extension of the European Flood Awareness System to the African continent, using local GIS data sources, ERA interim, GPCP and ECMWF medium range forecasts. The framework is calibrated over 36 gauge stations. The performance of the system is evaluated with respect to the Kling Gupta Efficiency index while the predictability performance is evaluated with respect to the Continuous Rank Probability Skill Score. The paper is well written and organized. The system by itself is an achievement and leverages on previous papers by the same author and co-authors. **It however presents this extension of the system without answering a particular scientific question, or it just remains unclear**. The value of the system might be more emphasized by adding:

i) more technical details on the system with respect to other existing systems and research on large scale flood forecast system (GPCP dataset, calibration of forecasts, downscaling approach, etc),

Table 1 was updated and contains now all HEPS that are operational, pre-operational or under development/evaluation. A detailed review *of all available systems* is, however, beyond the scope of this paper. There are two forthcoming publications which will contain considerable detail and the interested reader is referred to those:

1. Flood Forecasting: a Global Perspective, Thomas Adams and Tom Pagano (Editors), Anticipated Publication date: May 2015
2. HEPEX book

Both are in their final stage and we hope that we can link to them to give the reader the possibility to access more information on this topic.

Additionally, we have dedicated a whole study (DOI: 10.1080/15715124.2011.555082) on the exploration of the current situation of flood forecasting and early warning in Africa. We cited and gave some relevant information to justify our current study in the introduction (see p. 5561, ll. 11-24).

ii) put the reproductability performance (KGE) in context with respect to other existing systems over the region (African Drought Monitoring system, GLOFAS) for example,

We agree that it is interesting to compare the hydrological performance of different hydrological models operating over the same region with each other. To do the suggested, it *is necessary to run the underlying hydrological models (of e.g. the African Drought Monitoring system or GloFAS) over the same time period and evaluate the capability to simulate discharge at the same locations against ground observations, using the same performance indicator (KGE or a comparable one). For GloFAS, we are in the lucky position that it is another of our in-house research projects which grants us easy access to the hydrological modeling environment and we could potentially do such a comparison; but for other hydrological models, particularly the ones not of public domain, it is not easily feasible. Regardless of feasibility issues, we can only see the added value of such a comparison if we would not only compare the ability to reproduce discharge, but also the ability to predict flood events, which would narrow it down to flood forecast systems. More considerations towards a cross-comparison to other flood forecasting systems please see below under iv).*

iii) evaluate/emphasize changes in predictability of the flood across different horizons (short and medium range in particular) in order to clarify the performance of the system at the medium range scale in particular,

We agree with the reviewer, it would give indeed added value to the evaluation of AFFS. To do so, we would need (several) ground measurements for the basins flooding was reported (ideally spread throughout the affected basins). These we could then compare against the AFFS forecasts and analyse if the lead time is a function of e.g. upstream area, average annual discharge, climate zone, topography, etc. . However, we don't have those data at hand, and in most cases we are not sure if they even exist.

To evaluate the flood forecasting performance of AFFS despite the lack of those data, we collected flood-related information, in particular on when, where and with which magnitude a flood event has happened, from several disaster databases (see Section 2.2.1; Figure 2). Those information allowed to verify if the flood events that were observed/reported were also predicted by AFFS. However, those information do unfortunately not have the accuracy and are too sparse to extent the analyses beyond.

It is questionable if the situation of having enough ground measurements in the right place, covering the right time will ever be reached, considering that we are trying to evaluate the flood forecasting performance of AFFS across a whole continent and not for isolated locations. However, in future this analyses may be possible, once the technology of estimating river discharges from remotely sensed data is more advanced and accurate enough.

iv) evaluate with respect to nowcast (NASA flood monitoring) and other medium range forecast system (GloFAS) in order to emphasize the value of a regional system over a global system.

We agree. As AFFS is a continental medium-range flood forecast system, we think that a cross-comparison to other flood forecasting systems would be very valuable from the point of view to understand the strengths and limitations of various flood forecasting systems, and with that, foster future enhancements in this field. For that reason we are currently working on a cross-comparison between AFFS and another flood forecasting system, GloFAS. However, in order to draw a valid comparison the "base-line" conditions have to be equal.

That means, the systems have to be run over the same time period and the same spatial domain, and evaluated the same flood events. As GloFAS didn't exist back in 2003, but is pre-operational since late 2011, we are working on a cross-comparison based on the year 2012 (and maybe also 2013). We intend to publish the results in a peer-reviewed journal.

A cross-comparison between AFFS and flood monitoring system, however, is according to our opinion not feasible as the main target of a flood monitoring system is not to detect floods and not to forecast them (which they are not capable of as they are not designed for that purpose). Hence, comparing a flood monitoring system to a flood forecast system would result into a not valid comparison as the flood monitoring system would be unreasonably "punished". Therefore we are determined to avoid such comparisons.

We modified the discussion:

Additionally, a cross-comparison study of AFFS with other (global) forecasting ~~or nowcasting~~ systems covering the African continent (such as e.g. GloFAS) is necessary to gain a deeper understanding on the particular strengths and limitations of AFFS~~, as well as to examine issues such as whether there is a necessity for a hydrological model, or the detail of output products required to be useful for the end-users~~. Note, in order to draw a valid comparison the general set-up of the comparison i.e. systems have to be equal, meaning that the systems under comparison have to be run over the same time period and the same spatial domain, and evaluated the same flood events. As GloFAS did not exist back in 2003, a cross-comparison within this study was not feasible, but will be focus of future research.

Comments:

- No evaluation of the modeling performance (Kling Gupta Efficiency or NSE) with respect to other hydrologic set ups over Africa: Princeton's drought monitoring, NASA global nowcasting and GLOFAS systems for example.

Please see reply to main comment ii) above.

- No evaluation with respect to global nowcasting POD and FAR numbers from the NASA system or GLOFAS for example. It would give some context on the value and performance of the system. Analysis in the day 1-3 range are relatively short term and could be compared to a nowcast system. Might be good to emphasize the capability of such system at short term and medium range in order to communicate the value of the system across multiple horizons and facilitate understanding of results by users who may be using multiple systems. And what is the added value of a continental flood forecast system like EFAS/AFFS with respect to GLOFAS?

We think the added value of a continental system over a global one is:

a) ***Ownership*** *– EFAS has been taken over by the European Commission to run specifically for Europe where specific interest and endusers exist* *b) **Data** – coming with the ownership is also access to data. We have a lot more data from the European partners than you will be able to assemble for a global system*

*c) **Mandate** – EFAS has a mandate to act for Europe also because of the European civil protection. It is a European system run by Europeans and for Europeans.*

Data and mandate are linked to the Ownership obviously. The idea would be that an AFFS would be run by an African organisation for Africa in a similar way.

Related to the comparison question, please see reply to main comment iv) above.

- No reference to other existing global flood forecast system or research with African basin application like Wu et al., Schumann et al.

Please see reply to main comment i) above. Table 1 was being updated.

As already mentioned under i) we have dedicated a whole study (DOI: 10.1080/15715124.2011.555082) on the exploration of the current situation of flood forecasting and early warning in Africa. We cited and gave some relevant information to justify our current study in the introduction (see p. 5561, ll. 11-24). Our intention was not to repeat too many information that was already published, but rather link to it. If the reviewer, however, is convinced that the current study would gain from a more in-depth evaluation on that topic we would include it. But, please keep in mind that the research by Wu et al., which is indeed sound and interesting, is focused on flood monitoring/detection and not on forecasting.

- The system is mentioned to include the effect of reservoirs but the reference to Haddeland et al. is a summary of multiple large scale reservoir models using monthly generic operating rules. Which model was used and how appropriate is it for flood forecasting at a daily time scale? How sensitive were the results to the regulation from the reservoirs and reservoir modeling uncertainties – would you have gotten the same results without the reservoirs?

The reservoir module in LISFLOOD is developed within the JRC. Water reservoirs are simulated as points, with a given water storage potential in Mm3, and outflow operation rules. The outflow behavior of a reservoir is described by a number of parameters. First, each reservoir has a total storage capacity S [m^3]. The relative filling of a reservoir, F, is a fraction between 0 and 1. There are three 'special' filling levels. First, each reservoir has a 'dead storage' fraction, since reservoirs never empty completely. The corresponding filling fraction is the 'conservative storage limit', L_c. For safety reasons a reservoir is never filled to the full storage capacity. The 'flood storage limit' (L_f) represents this maximum allowed storage fraction. The buffering capacity of a reservoir is the storage available between the 'flood storage limit' and the 'normal storage limit' (L_n). Three additional parameters define the way the outflow of a reservoir is regulated. For e.g. ecological reasons each reservoir has a 'minimum outflow' (O_{min}, [m^3/s]). For high discharge situations, the 'non-damaging outflow' (O_{nd}, [m^3/s]) is the maximum possible outflow that will not cause problems downstream. The 'normal outflow' (O_{norm}, [m^3/s]) is valid once the reservoir reaches its 'normal storage' filling level. Depending on the relative filling of the reservoir, outflow (O_{res}, [m^3/s]) is calculated as:

$$O_{res} = min\left(O_{min}, \frac{1}{\Delta t} F \cdot S\right) \qquad F \leq 2L_c$$

$$O_{res} = O_{min} + (O_{norm} - O_{min}) \frac{(F - 2L_c)}{(L_n - 2L_c)} \qquad L_n \geq F > 2L_c$$

$$O_{res} = O_{norm} + \frac{(F - L_n)}{(L_f - L_n)} \cdot max\{(I_{res} - O_{norm}), (O_{nd} - O_{norm})\} \qquad L_f \geq F > L_n$$

$$O_{res} = max\left(\frac{(F - L_f)}{\Delta t} S, O_{nd}\right) \qquad F > L_f$$

with:

S: Reservoir storage capacity [m^3]
F: Reservoir fill (fraction, 1 at total storage capacity) [-]
L_c: Conservative storage limit [-]
L_n: Normal storage limit [-]
L_f: Flood storage limit [-]
O_{min}: Minimum outflow [m^3/s]
O_{norm}: Normal outflow [m^3/s]
O_{nd}: Non-damaging outflow [m^3/s]
I_{res}: Reservoir inflow [m^3/s]

Defining these parameters peak flows can be levelled off and distributed on a longer timescale for a certain extent. As we don't have information on the reservoir management operation rules the normal storage volume, the flood storage volume, the normal outflow and the non-damaging outflow parameters are calibrated to match the discharge in the nearest downstream gauge station. Ofcourse, in reality humans are in control, who may take different decisions for reservoir outflow.

Results from a sensitivity study over Southern Africa showed that the calibration parameters related to infiltration and groundwater processes are much more important compared to the reservoir parameters. Therefore, simulations with or without the reservoir routine should not affect our results.

Because the reservoir routine is not important for the results we only limit the manuscript by adding two references who describes the method:

(1) The revised LISFLOOD manual: Burek, P., Van Der Knijff, J., and De Roo, A.: LISFLOOD, distributed water balance and flood 30 simulation model – Revised user manual 2013 (Ispra: Institute for Environment and Sustainability), 2013.
(2) And a reference which described the method (although very brief) and is public available: De Roo, A., Odijk, M., Schmuck, G., Koster, E., Lucieer, A. (2001), Assessing The Effects Of Land Use Changes On Floods In The Meuse And Oder Catchment. Physics and Chemistry of the Earth, Part B, Vol. 26, No. 7/8, 593-599.

In the manuscript we added on page 5567 line18 the reference where the reservoir module is described:

"The model structure was extended to also account for large reservoirs (Burek et al., 2013; De Roo et al., 2001) as well as ……

- The atmospheric forcing is at 71km spatial resolution but LISFLOOD runs at 0.1 degree grid cell. What is the downscaling approach? Also, the atmospheric forcing is

mentioned to be adjusted with respect to GPCP precipitation. What motivated the choice of that specific dataset and which calibration approach was selected?

To drive the LISFLOOD model for the pan-African domain, the ERA-Interim meteorological forcing with the GPCP-rescaled precipitation was used. The original Gaussian grid is about 71km. The current pan-African setup of LISFLOOD uses a 0.1^0 degree grid, which means that all the datasets were resampled to 0.1^0 degree horizontal resolution.

The ERA-Interim + GPCP has a few advantages compares to for e.g., remotely sensed precipitation:

(1) As stated in the manuscript, Balsamo et al. (2010) reported on systematic biases in the ERA-Interim precipitation data. Therefore, the ERA-Interim precipitation is corrected using the GPCP dataset and the method "calibrates" the monthly precipitation amount. Details of the rescaling method can be found in Balsamo et al. (2010). Hence, not us corrected the ERA-Interim, but the ECMWF

(2) The big advantage of the ERA-Interim GPCP corrected dataset is that it has a large time window. As stated in the manuscript (page 5570 line 10) this long time series is used to calculate the hydrological thresholds (2 and 10 year return periods).

(3) Global coverage. Therefore this dataset is suitable to set up a global forecasting system like GloFas. In a next study, we would like to perform a comparison between AFFS and GloFas. AFFS is forced with the ERA-Interim GPCP corrected dataset to be already consistent with the GloFas system which also will be calibrated with this dataset.

(4) The choice for the ERA-Interim over another SRFE (such as TRMM) was that the NWP model behind the ERA-Interim and ECMWF-ENS is quite similar. As we are using the first for calculating the initial conditions and the second for running the hydrological forecasts we anticipate the possibility of "jumps" between initial conditions and forecast that might happen when using two different data sets.

- The flow forecasting seems to go through some post processing, which one? Or is it normalization only for the analysis

The flood forecasts are based on threshold exceedances see "identification of flood events" (p.5566, point 4).
In order to be able to work with the contingency table we had to proceed in a "Boolean kind of manner" differentiating between "flood" and "no flood". The probabilistic approach, indicating the likelihood of a flood event can be derived from the threshold exceedance maps that show the number of ensemble members (hydrological forecasts) exceeding a particular threshold (such as 2- or 10-year return period; see e.g. Figure 9).

- given that the motivation is focused on the medium range scale, the title might need to reflect that as well " A pan-African medium range ensemble flood forecast system" for example

Thank you. We initially liked our short and catchy title, but the reviewer is right and hence we followed his/her suggestion and changed the title to "A pan-African medium-range ensemble flood forecast system".

- P5566:"more than 10 clustered river pixels are affected." How does that affect basin of smaller size or with complex topology? How does it differ from other system and how sensitive is it to those criteria?

Thank you to the reviewer for pinpointing at the number 10, which is a typo and should be 40. This alone does probably not explain the issue; therefore the following: AFFS was designed as a medium-range probabilistic flood forecasting system. As such it is not targeting to predict small-scale flood events such as flash floods. To introduce the criteria of a minimum amount of pixels that have to be affected reduces the noise in the small upstream catchments and increases the precision of the forecasts. The decision for 40 clustered pixels was done based on the experience gained from the European system EFAS.

Of course this number could be changed, but the analysis could be cumbersome considering the large amount of maps to analyse (1000 maps daily ← 50 members, 10 day forecast, 2- and 10-year threshold). As the 40 clustered pixels worked well for us, we did not experiment further, but of couse this could be done in future.

Figure 7: is it below or over the 10 day lead time?

We changed it to make it clearer into:

Continuous Rank Probability Skill Score over the range of the 10 day lead time.