

Dear Editor,

As requested we added the “persistence” as benchmark during the evaluation phase. In the following the changes that were done in respect to the last manuscript version.

Thank you and best regards,

The authors

- 1) Textual modifications to embed the requested change of adding the “persistence” as benchmark.

In this study, the CRPSS was calculated for each lead time at the 36 key locations all over Africa. **Two different benchmarks were considered; once the seasonal mean (here: moving average considering 30 days before and after the respective observation) and once the persistence (here: the last observation is kept constant over the forecasting range; Bauer-Gottwein et al, 2015).** Hence, the here computed CRPSS evaluates the advantage of using the flood forecast calculated by AFFS in comparison to using the seasonal mean or persistence as alternative approximation to a forecast.

An average CRPSS was computed for all stations **and also** for different geographical locations (Eastern, Southern and Western Africa) **and for different precipitation classes (< 600, 600–1000, >1000).**

In Fig. 7 the **two different sets of** CRPSS are plotted over the 10 days lead time. **Both sets show the same general tendency: 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. **~~The decomposition of the CRPSS for different ranges of average annual precipitation amounts indicates that the predictability of streamflow is generally slightly lower in arid areas (average amount of annual precipitation <600 mm).~~**

Figure 8 compares the forecast to **the different benchmarks (seasonal mean (left) and persistence (right),** see Sect. 2.4.2.1) and indicates the number of days the forecast is skilful – this is also called the limit of predictability. A few stations indicate that a skilful forecast can be achieved up to Day 10, and that at some stations no skilful predictions have been made for this year in comparison to the **seasonal mean or the last observation respectively.**

Both, the positive CRPSS and the limit of predictability > 0 show that hydrological calculations based on AFFS are on average more skilful than using the last observation (=persistence) or the seasonal mean as forecast. This is in particular true for the first two days, in which the CRPSS is

remarkably high (persistence as benchmark); but also for the remaining forecasting range in which the CRPSS varies between 0.4—0.5 (seasonal mean as benchmark).

- 2) Figure 7 and 8 were updated. Now including two different benchmarks.
- 3) Further changes:
 - a. Numeration of figures was corrected
 - b. Section numbering was corrected
 - c. Links to different sections were corrected