I applaud the authors for their efforts and read some of the paper findings with great interest. However, I feel that the paper is premature and a more in depth analysis of the results is needed and in particular a clearer scientific justification for the novel contribution. I also believe that the current scientific methodology is partially floored in the way uncertainty is treated (not) and skill is evaluate. The reason for some of the results are unclear (E.g. bias vs non bias correction).

The methodology is floored as the paper simply ignores any form of uncertainty be it in the hydrological model (e.g. application of Darcys law on a 55km grid??), the observations (discharge measurements have errors between 10-30% and more!), the forcings (partially done) or the post-processing. This is a major floor of the study and
needs to be rectified.

p3473/section 2.1 I cannot see from this section how this model is different in any respects from many of the LSM models quoted in the introduction - please clarify that in the text.

Why is the model run on daily timesteps, when most of the analysis focuses on monthly data?

p3476/3.1 definition of skill. I believe the authors make it too simple for proving that there is 'skill' - skill reference should be adequate (in particular in comparison on a monthly scale and the effort of running such a global model). Monthly discharge can be computed by a simple water balance model without routing (simply dumping the P-E of each catchment to the outlet including the bias correction). Please, see e.g. discussions by Schäfli and Gupta.

P3477/3.2 Your choice of binary scores suffers from the large number of correct rejections - hence I find it very laudable that you publish the contingency tables. However discuss your results with respect to that issue (follow discussion of Finley affair, Murphy 1996 and Stephenson, 200) - you actually note that effect in your results, but given not enough discussion in context.

p3479/3.3 Despite your claims, I cannot see how your evaluation is realistically flood and drought related. Please substantiate your claim that the 5 year return period "provides and acceptable common measure". I have fundamentally no problem with such an evaluation, however, your are completely overstating the significance of your results w.r.t to floods and droughts. Please properly explain how you have derived the return period and also indicate what the results are with other return periods. In addition, I am not sure I understand your numbers, e.g. in a 20 year data set of the Amazon, you seem to have a total of 336 events (table C1). There are only 240 months and you state p3480/L6 "we limit ourselves to monthly..." - please clarify as you presumable used daily data here.
Appendix A. I do not see the evidence for the assumption that daily maxima necessarily transform to monthly maxima. I do not understand your justification for the extrapolation from the Rhine - and I am not sure why it is necessary (rather than just stating this limitation).

Section 4.1: most of your skill improvement comes through the bias correction. At this stage I am unconvinced that there is any value in running a complex hydrological model. Please plot time series for a station where there is no improvement and one where there is a large improvement. Please also plot the climatology you are using. Currently, it is difficult to make heads and tails of these results and in particular understand why the bias correction brings such an improvement.

I liked this section and think a more in depth analysis could actually be very interesting.

Minor points:

I believe that the paper is not set properly into context. There have been multiple publications of the LSMs mentioned in the introduction with respect to their capability in reproducing streamflow. None of these studies is mentioned although they are highly relevant. Just because of different focuses of the two approaches to hydrological modelling does not mean that results with respect to streamflow are different (in particular on a monthly scale). The paper needs to mention these studies and explain the added value to the scientific community (please see a very limited number of references of recent papers below Gong, Papenberger, Yamazki (all this year - amongst many others!) ). In addition, there are multiple papers using meteorological skill scores within a hydrological setting (Laio and Tamea, 2007 amongst many others - also see special issues of HEPEX) thus I cannot see how the contribution is novel in this context. Having said that the way those measures are applied are very interesting and could be the focus of this paper.

Table 1: Where do these basin data come from? Please quote reference.
References

Laio and Tamea 2007, Verification tools for probabilistic forecasts of continuous hydrological variables, HESS

Schäfli, B. and Gupta, H.V., 2007. Do Nash values have value? Hydrological Processes, 21(15), pp. 2075 - 2080


Pappenberger, F., Cloke, H.L., Balsamo, G. Oki, P., Ngo-Duc, T., 2011, Global routing of surface and subsurface runoff produced by the hydrological component of the ECMWF NWP system, International Journal of Climatology , 30(4), 2155-2174


Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 8, 3469, 2011.