Initial response to RC2

Berit Arheimer (R#2)'s original text in black *with our initial response in* blue*.*

This paper shows recent progress in global river forecasts from the Glofas modelling system. Such data is indeed very useful and appreciated by many users at the global scale, especially by low- and middle-income countries who might not have access to their own river-forecast system. Accordingly, it is very important to evaluate such systems scientifically before launching them operationally.

The paper gives a very good overview of a river-discharge forecast system, which is indeed valuable for the scientific community to learn more about, as such systems are dedicated to national/international institutes with advanced IT infrastructure and operational production.

We thank Berit for her positive words about our manuscript and constructive feedback and suggestions that will help refine our paper.

My main concern with this paper is that I miss a scientific question and the story of what kind of new scientific knowledge we have learnt from using the forecasting system and evaluation method described.

The Glofas model and forecasting system has been described before in the scientific literature and the focus of this paper seem to be that the results are now part of the climate service C3S, but this is hardly a scientific finding. New datasets should rather be published in ESSD, in which Glofas results have already been published. Likewise, the methods used for forecast evaluation are standard and has been published before. For publications in HESS I expect a more scientific analysis of the results and conclusions about new knowledge from the identified scientific achievements with impact on our understanding of Hydrology or Earth Systems. Right now, I have difficulties to find a clear take-home message in the current version of this paper. It is very descriptive and less analytic.

We take on board your point that the take-home messages in the current version could be clearer and will ensure they are sharper in the resubmitted manuscript. The original pre-operational GloFAS (version 1) was indeed described by Alfieri et al. (2013), however we disagree with your point that the scientific details of the fully operational GloFAS (version 2) have already been published in the hydrological scientific literature. Our manuscript is the first time the fully operational real-time forecast configuration has been published. Uniquely, this is the first time the large-sample and long-term reforecast strategy we use for generating and evaluating the forecast skill of GloFAS has ever been published, and we think this provides an important advancement in the area of global hydrological forecasting as well as users of GloFAS forecasts. We do not claim the evaluation method is new, but what is novel is the scale of the evaluation (both in space and across the long forecast range) and how the data and results are delivered to the hydrological community. Additionally, to our knowledge (apologies if incorrect), no other global operational hydrological forecasting system currently provides such a long-term and large-sample set of reforecasts, delivered in a free and open way, with summary evaluation results available on the user interface – we think this is a significant advancement towards transparency and as such, this procedure will be implemented with each new major release of GloFAS; we very much encourage other systems to follow in this direction. Once large sets of ensemble reforecasts are available from multiple forecast systems, even more interesting scientific questions around predictability can be uncovered, and we look forward to future collaborations with your group and WWH in this regard!

Our paper is not simply a data paper so we disagree ESSD is the most appropriate outlet. We want to communicate with the wider hydrological science community our scientific description of the GloFAS configuration, our method for evaluating the skill of GloFAS, and the findings of when and where in the world GloFAS ensemble forecasts are skilful against the two primary benchmark forecasts used in hydrological forecasting. HESS in particular has become a key journal for publishing papers in the area of hydrological forecasting and therefore we hope the editor agrees is a good home.

I therefore suggest to find a scientific angle from current discussions in the research community and tell the story of the results from that perspective.

Interesting scientific questions could for instance be:

- On the method side: How should we evaluate forecasts – _what metrics are there, how do they compare and what does different metrics contribute in understanding/reliability for the user community and research community, respectively?
- Could the metrics presented (and argued for?) in this paper be compared with other metrics, to show their excellence and benefits to users/scientists? (is there a take-home message or guide-lines to the scientific community from using a specific metric/evaluation method compared to another?) What are the options?
- On the understanding of hydrology: what are the attributes for catchments/regions with high or low skills in forecasting? i.e. which processes do we need to learn more about to improve the quality of river-discharge forecasts?
- How does different global river-discharge forecast systems compare to each other? Can we learn from different model setups and elaborations on procedures, process descriptions or geophysical representation?

Thank you for this valuable list of avenues for further research. Most of these would be very interesting standalone studies in their own right. We agree with your bullet point 3 that the results of forecast skill could be expanded in the current version of the manuscript to provide more insight into the attributes of catchments/regions with high or low skill. Great suggestion and we will add this analysis to the resubmitted version of the manuscript.

Please, find some detailed comments on current manuscript below. Apologies for mentioning my own work, but I am very eager to start comparing model results at the global scale soon. 😊

Yes indeed, we are already collaborating on a comparison between GloFAS and World-Wide HYPE in terms of hydrological simulation performance, but the work is not yet completed. We also agree that the prospect of being able to compare multiple operational global forecasting systems (and not only GloFAS and WWH, but others as well) in terms of ensemble forecast skill would provide extremely valuable information to the scientific and forecast user community and are too are very eager to participate further.

Introduction

Line 31: Reference Blöschl, et al. 2019 does not evaluate risks or hazards.

We will modify this sentence in the resubmitted manuscript.

Line 37: also note the global and continental scale forecasting based on sharing the world-wide HYPE model:

> *Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C. M., Hasan, A., and Pineda, L., 2020. Global catchment modelling using World-Wide HYPE (WWH), open data and stepwise parameter estimation, Hydrol. Earth Syst. Sci. 24, 535–559, https://doi.org/10.5194/hess-24- 535-2020*

Thank you for the suggested paper. We will add it to this section in the resubmitted manuscript.

Line 60-70: In fact, global river forecasts and reforecasts are also available at https://hypeweb.smhi.se/ where the user can subscribe to seasonal forecasts with monthly data. In addition, 1-10 days forecasts with

thresholds based on return periods of high flows can be ordered at https://hypeweb.smhi.se/water-services/data-delivery-services/

Thank you for this information. GloFAS has been providing an on-demand tailored user data service free of charge since it went pre-operational in 2011, but the present manuscript outlines a step-change in the service of large scale GloFAS data. What would be fantastic for the community is to be able to access sets of reforecasts generated from many hydrological forecast centres in a standardised format and central data service portal. This has been common practice in the weather and climate fields for years and has really facilitated advancement in forecasting science.

Section 2. even though Glofas has been evaluated against observed river flow in previous publications, it would be helpful to include such information about model performance vs absolute values also. For instance, Fig 4 could also include colors of KGE performance (modelled values vs observed values) in the circles showing location of gauges. This would make this figure much more informative and help the reader a lot to judge model performance. Please, check the color coding in Arheimer et al., so the overall pattern of model performance could be compared. Please, also mention median KGE at global scale (no you only say that it was skillful, which is very vague).

Given the full evaluation of the hydrological model performance of GloFAS version 2.1 is published in another paper (Harrigan et al., 2020) we only summarised the results briefly here and pointed the reader to the original paper, where the detailed statistics (such as median KGE at the global scale, which is 0.31) and indeed all raw statistics for each of the 1801 stations were provided within the Supplementary Information to allow for comparison. However, we agree that having the hydrological model performance results summarised in Figure 4 would to aid the reader to interpret the forecast results in the present manuscript. We believe it is most helpful to do so using the modified KGE as a skill score (KGESS) against a mean flow benchmark, following Knoben et al. (2019) as done in Harrigan et al. (2020), and will add it this information to the current manuscript.

Section 3: please start with some sentences summarizing the evaluation concept – _e.g. that you use scores with met. model vs observed met. model (_“a perfect weather model”) and correlation with observations. It would also be interesting for many users to actually see some scores to absolute values as well – or at least to discuss the difficulties here.

We mention explicitly that the forecast skill using the CRPSS (i.e. Sect. 3.3) is “verified against GloFAS-ERA5 river discharge reanalysis used as proxy observations [or ‘perfect model’] (following Alfieri et al., 2014)” in L254-235. We have made it clear that both the benchmark forecasts and the verifying observations are based on the river discharge reanalysis rather than in situ station observations. This approach is common practice in forecast evaluation (e.g. as in Pechlivanidis et al. (2020) mentioned below), but for the benefit of a broader audience we will outline the justification of this approach in Sect. 3 of the resubmitted manuscript - i.e. that it allows for the prediction range to be determined regardless of systematic hydrological model error and that forecast skill can be calculated for any location in the world.

Section 4: the Glofas results could be compared with results from another model, using the same metrics across Europe, presented by:

> *Pechlivanidis, I. G., Crochemore, L., Rosberg, J., & Bosshard, T. (2020). What are the key drivers controlling the quality of seasonal streamflow forecasts? Water Resources Research, 56, e2019WR026987. https://doi.org/10.1029/2019WR026987*

One of the key benefits of providing the GloFAS data openly and free of charge on the Copernicus Climate Data Store (CDS) is that it now facilitates further scientific evaluation and inter-comparisons of similar forecasting systems offering their data in the same way. We however think a more appropriate comparison

of GloFAS forecasts with Pechlivanidis et al. (2020) is to undertake it with the GloFAS-Seasonal system (Emerton et a., 2018), which is forced by the SEAS5 climate output. All GloFAS-Seasonal data including a comprehensive set of reforecasts are now available through the CDS: https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas-seasonal-reforecast?tab=overview; for a higher resolution hydrological seasonal forecasting system at European scale, the EFAS-Seasonal complete dataset is also available from the CDS (https://cds.climate.copernicus.eu/cdsapp#!/dataset/efas-seasonal-reforecast?tab=overview).

To further explore and evaluate the added value of the Glofas system, it could also be compared to warning issued by National forecast services for specific regions or countries, or to soft information from new items reporting floods, to check if the alerts actually captured something real.

Evaluation of flood events against a wider set of observations is a very good idea and something that will be expanded in future assessments, but is outside the scope of this first evaluation to determine overall ensemble forecast skill against the two key scientific benchmark forecasts: persistence and climatology.

Line 265: Attribution is also discussed in the above-mentioned paper. It is another interesting scientific analysis, which deserves much more attention – _also in this global study of model performance. Such an analysis would make this paper much more scientifically interesting.

This point is related to the one you bring up in the above bullet list of interesting further scientific questions. We agree and will explore the attributes of catchments/regions with high or low skill and present the results of this analysis within Sect. 4 of the resubmitted manuscript.

I am looking forward to read a new more elaborated version of this paper, with a scientific discussion linked to the methodological description.

We appreciate your time to review our manuscript and thank you again for your constructive feedback and many ideas for further research and collaborations!

Kind regards,
Shaun Harrigan on behalf of all co-authors

References

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: GloFAS – global ensemble streamflow forecasting and flood early warning, Hydrol. Earth Syst. Sci., 17, 1161–1175, https://doi.org/10.5194/hess-17-1161-2013, 2013.

Alfieri, L., Pappenberger, F., Wetterhall, F., Haiden, T., Richardson, D. and Salamon, P.: Evaluation of ensemble streamflow predictions in Europe, J. Hydrol., 517, 913–922, doi:10.1016/j.jhydrol.2014.06.035, 2014.

Emerton, R., Zsoter, E., Arnal, L., Cloke, H. L., Muraro, D., Prudhomme, C., Stephens, E. M., Salamon, P. and Pappenberger, F.: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0, Geoscientific Model Development, 11(8), 3327–3346, https://doi.org/10.5194/gmd-11-3327-2018, 2018.

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H. and Pappenberger, F.: GloFAS-ERA5 operational global river discharge reanalysis 1979–present, Earth Syst. Sci. Data, 12(3), 2043–2060, doi:https://doi.org/10.5194/essd-12-2043-2020, 2020.

Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling– Gupta efficiency scores, Hydrol. Earth Syst. Sci., 23, 4323–4331, https://doi.org/10.5194/hess-23-4323-2019, 2019.

Pechlivanidis, I. G., Crochemore, L., Rosberg, J., and Bosshard, T.: What are the key drivers controlling the quality of seasonal streamflow forecasts? Water Resources Research, 56, e2019WR026987. https://doi.org/10.1029/2019WR026987, 2020.