Response to Anonymous Referee #1

In this paper, the authors reported a newly developed river discharge dataset at the global scale using a meteorological reanalysis dataset and evaluated its performance. Since this river discharge dataset is very promising in terms of high resolution (0.1 degree) and feasibility for real-time update, it will benefit potential users among hydrology and related-field communities. The contents of this paper are also suited for this journal. This paper is well organized. But there is some room for improvement before publication. In particular, since this paper targets a new release of river discharge data and is intended to be published in the journal specialized for scientific data, methods and processes used in producing the dataset should be solidly and clearly written.

Thank you for your positive comments and constructive feedback. We believe your clarifications have sharpened the manuscript and made it clearer for the reader. Our responses to your comments are provided below in blue together with your original review in black.

MAJOR COMMENTS

Figure 1: Since the spatial resolution of LISFLOOD (0.1deg) is finer than that of ERA5 (0.28125deg), I guess a kind of downscaling techniques was used to produce the LISFLOOD dataset. However, there is no information (except "been resampled" in L138) on this process in this paper. How did the authors produce runoff data at a finer resolution in this paper? Did the authors weight the ERA5 runoff value (by something) during the "resampling"? Did the authors consider terrain effects within an ERA5 cell in allocating surface/subsurface runoff to multiple 0.1deg land cells? Please provide the procedure in detail.

This question was also asked by Anonymous Referee #2. In order to be consistent with the operational GloFAS procedure, the runoff fields from ERA5 were downscaled using the simple nearest neighbour method from the native resolution to the 0.1° LISFLOOD grid. The task was done using the open source 'pyg2p' module with interpolation 'grib_nearest' option in Python (https://pypi.org/project/pyg2p/). No weights were applied to runoff values and terrain effects within the ERA5 cell were not considered. We will add the additional detail on how the downscaling was done in the resubmitted manuscript.

Sect. 2.2 and Figure 1: The authors describe surface and subsurface runoff data originally generated from the HTESSEL land model. I think the runoff scheme directly affects the river discharge data, but less information about it is provided. To which depth of soil layer did the authors consider as the subsurface runoff? Regarding the description in L125-127, how much delays were considered before the subsurface water returns back to the river channel in the LISTFLOOD ground water module? Does it depend on the soil properties?

The HTESSEL land surface model is used to calculate the water balance at the land surface. Excess precipitation and snowmelt are partitioned as surface runoff or infiltrated into a four-layer soil column (7 cm depth for top layer and then 21, 72, and 189 cm) at each ERA5 grid cell, before draining from the bottom of the soil column as sub-surface runoff. Water moving (vertically) through the soil column does depend on soil properties. There are six soil texture classes in HTESSEL (e.g. coarse or very fine) that determine hydraulic properties. Therefore, the soil properties will determine the amount of time it takes for water to exit at the bottom of the soil column as sub-surface runoff. Further information on the details is provided in Balsamo et al. (2009).

As mentioned in L125-127, the HTESSEL sub-surface runoff is used as input to the LISFLOOD groundwater module, which consists of two parallel linear reservoirs (upper zone for quick and lower zone for slower groundwater flow) that store and subsequently transport water to the river channel with a time delay. In Hirpa et al. (2018), the upper zone time constant was given a default value of 10 days with a lower (upper) bound of 3 days (40 days) during calibration. The upper zone time constant has a default value of 200 days with a lower (upper) bound of 40 days (500 days) during calibration.

L132-135: The authors describe flow alteration by lakes and reservoirs, but readers cannot figure out how much the flow is altered by them. Did the authors use a kind of algorithms of flow alteration or dam manipulation? The authors also discuss the limitation of this dataset as "While GloFAS-ERA5 reanalysis does represent major dams and reservoirs on the modelled river network, it does so in a simplified way and does not include operational operating schedules for individual structures. (L298-299)" in a later section, but due to the lack of description on dam operation schemes employed in this paper, it is very difficult to have a clear image on that. What does "a simplified way" mean? In addition, how the authors treat river water withdrawal from rivers for human activities (agriculture, industrial, etc.) in this dataset? Please provide information about it in detail.

Reservoir outflow is calculated with a set of simplified rules depending on their filling level, and balances water recharge if storage is below normal or release if above normal. There is a minimum outflow to ensure the downstream river does not dry up, and a non-damaging release so the reservoir does not reach full capacity. Simplified reservoir operating parameters were used based on expert opinion (outlined in Zajac et al., 2017) given lack of availability of global operational release records.

As mentioned in our reply to Anonymous Reviewer #2, we propose to add a new table (Table A below) to accompany Fig. 1 in the resubmitted manuscript that makes it clearer for the reader to find the open access publications outlining the full methodological details of the key components of GloFAS-ERA5:

Table A: Scientific papers and model documentation for the key components in the production of GloFAS-ERA5 v2.1 river discharge reanalysis dataset.

GloFAS-ERA5 component	Description	Reference
ERA5	Global reanalysis dataset using ECMWF Integrated Forecast System (IFS) model cycle 41r2 from 1979 to present	Hersbach et al. (2020)
ERA5 runoff	Surface and sub-surface runoff within ERA5 generated using the HTESSEL land surface model	Balsamo et al. (2009)
LISFLOOD river discharge	River discharge generated using LISFLOOD hydrological and channel routing model to route runoff into and through the river network and provide groundwater storage. LISFLOOD includes lake, reservoir and human water use routines	Burek et al. (2013)
Lakes and reservoirs used	Incorporated 463 lakes and 667 reservoirs	Zajac et al. (2017)

in GloFAS	into the GloFAS river network	
Calibration of LISFLOOD used in GloFAS	LISFLOOD was calibrated against daily river discharge from 1287 observation stations worldwide	Hirpa et al. (2018)

Sect. 4.3: The authors provide monthly performance of this dataset. Such information is very useful, however, it is very difficult to interpret this seasonality, because the results are (probably) a mixture of contributions from both the northern and southern hemispheres. Have the authors made similar analysis for each hemisphere? The authors state "Attribution of such biases in the GloFAS-ERA5 reanalysis is outside the scope of this data paper (L293)", however, practical information on the seasonal performance of this dataset will be very beneficial for potential data users. In my view, the authors should add and show, at least, whether a larger bias ratio observed in the months of November to March than the other months (Fig 7c) is attributable to winter discharge from the northern hemisphere or summer discharge from the southern hemisphere (or a mixture of them; or from some specific regions).

This is a very good suggestion. We have conducted your proposed analysis (Fig. A) and found that while the overall GloFAS-ERA5 monthly performance in each hemisphere does not change substantially from the global analysis (Fig. 7 in the paper), there are some differences worth reporting. We therefore propose to add Fig. A to the resubmitted manuscript as a new Fig. 8 together with the following additional paragraph in Sect. 4.3:

"Results are grouped into northern (n=1268 stations) and southern (n=533 stations) hemispheres in Fig. 8. The overall GloFAS-ERA5 monthly performance in each hemisphere does not change substantially from the global analysis (Fig. 7). Nevertheless, there are some differences. The KGESS and bias ratio from the northern hemisphere (Fig. 8a and c, respectively) tend to follow the global analysis most strongly (i.e. Fig. 7a and c, respectively), which is not surprising given 70 % of all stations are located in the northern hemisphere. However, a higher proportion of southern hemisphere stations show large positive biases from April to June, compared to November to March in the northern hemisphere. The largest proportion of stations with negative KGESS in the southern hemisphere are found from August to October (Fig. 8a). These months correspond with lower southern hemisphere correlation (Fig. 8b) and a higher proportion of stations with large positive variability ratios (i.e. GloFAS-ERA5 has higher variability than observed river discharge)."

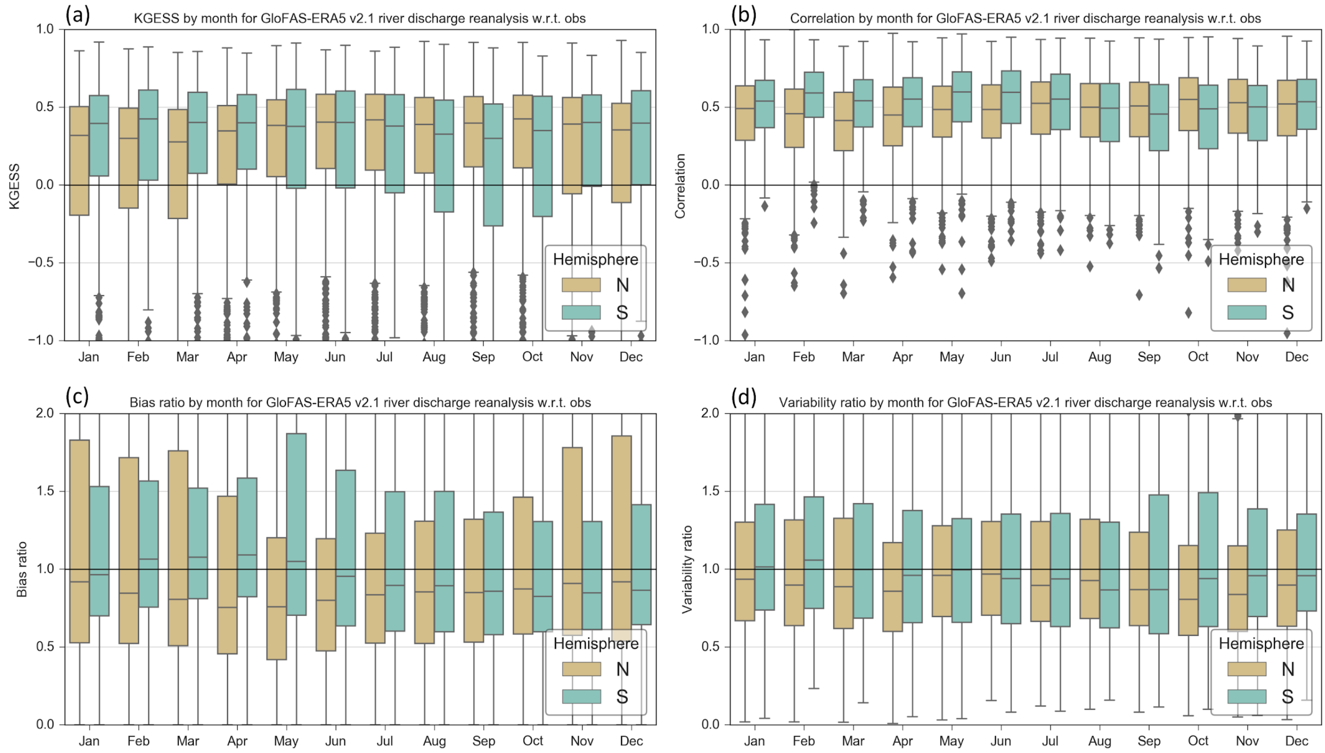

Figure A: Performance metrics for each month by hemisphere. Modified Kling-Gupta Efficiency Skill Score (KGESS) (a) with decomposition of KGE' into Pearson correlation (b), bias ratio (c), and variability ratio (d). Brown (green) boxes represent the IQR and horizontal grey line the median for the northern (southern) hemisphere. Whiskers extend to the most extreme data point, unless the data point is more than 1.5 times the IQR from the box and is instead represented as an outlier (grey diamond).

MINOR COMMENTS L139:
Is a one-year spin-up enough for this simulation? Probably this depends on the groundwater module or dam operation schemes (the information is not clearly written in the current manuscript, though) used in this model.

The initialisation routine within LISFLOOD has been designed so that one year is sufficient to estimate the initial state of all the state variables. For state variables that are fast responding and can reach equilibrium quickly (e.g. storage in the upper groundwater zone), a value of 0 can be used at start of the run and one year is more than long enough. However, for other state variables that are slowly responding it is correct that one year can be too short in some model calibration routines, especially for catchments with large groundwater stores. In order to avoid the need for very long spin up periods, LISFLOOD calculates a "steady-state" storage amount for the lower groundwater zone during a long-term pre-run, and thus reduces the lower zone's spin up time (Burek et al., 2013). We will add a sentence on this to the resubmitted manuscript.

L191: The authors used "1801 catchments" here, but this expression might be confusing if there are multiple gauge stations in a large river system. I think dividing this sentence into two parts (and used "1801 stations" in the former one) will be clearer for understanding.

We will change this to "1801 stations" in the resubmitted manuscript.

Sect. 4.2: The authors discuss the results by using both the bias ratio (beta) and PBIAS, but this might be confusing. For example, "-9%" in L241 is PBIAS, due to its negative value, I guess.

The bias term in the decomposition of the KGE' is the bias ratio β but can easily be converted to the more widely used percent bias (PBIAS) by (β - 1) x 100 (as shown in L205-206 in the submitted manuscript). PBIAS was also used in Lin et al. (2019) when considering what we mean by a "very good" bias error in global hydrological modelling (i.e. ±20 %), mentioned on L235-236. Our original intention was to report the bias in the text in the more widely used PBIAS form. However, your point is valid that it might actually be more confusing to the reader. We will therefore report bias errors as bias ratios and variability errors as variability ratios as per Equations 2 and 3 in the resubmitted manuscript.

We really appreciate your time and insight in reviewing our manuscript,

Kind regards,
Shaun (on behalf of all co-authors)

References

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M. and Betts, A. K.: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System, J. Hydrometeor., 10(3), 623–643, doi:10.1175/2008JHM1068.1, 2009.

Burek, P., van der Knijff, J. M. and de Roo, A. P. J. D.: LISFLOOD - Distributed Water Balance and Flood Simulation Model - Revised User Manual, Publications Office of the European Union, doi: 10.2788/24719, 2013.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G. D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P. de, Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J.-N.: The ERA5 Global Reanalysis, Quarterly Journal of the Royal Meteorological Society, doi:10.1002/qj.3803, 2020.

Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E. and Dadson, S. J.: Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data, J. Hydrol., 566, 595–606, doi:10.1016/j.jhydrol.2018.09.052, 2018.

Zajac, Z., Revilla-Romero, B., Salamon, P., Burek, P., Hirpa, F. A. and Beck, H.: The impact of lake and reservoir parameterization on global streamflow simulation, J. Hydrol., 548, 552–568, doi:10.1016/j.jhydrol.2017.03.022, 2017.