Interactive comment on "Observational gridded runoff estimates for Europe (E-RUN version 1.0)" *by* Lukas Gudmundsson and Sonia I. Seneviratne

H. Müller Schmied (Referee)

hannes.mueller.schmied@em.uni-frankfurt.de

Received and published: 7 February 2016

The manuscript "Observational gridded runoff estimated for Europe (E-RUN version 1.0) from Lukas Gudmundsson and Sonia I. Seneviratne, submitted to Earth System Science Data, presents a pan-European dataset of monthly runoff as well its description. The authors have further developed previous work (published in HESS) and updated the time series to the end of 2014. Firstly, they selected suitable stations and time series of river discharge from three databases, and trained a machine learning algorithm which is then used to model monthly runoff fields at pan-European scale. Using a cross validation approach and different metrics, the authors tested the approach and concluded that the dataset can be of use e.g. for long term runoff statistics or for analyzing droughts.

The manuscript is well organized and (with a few exceptions) written very clear, especial the different test "batteries" to obtain a homogenized dataset; congratulations! The beauty of first publishing the method and results in HESS (Gudmundsson and Seneviratne 2015 (referred in the following as GS15)) is, that the discussion and review comments of GS15 are publicly available. Major issues are discussed in GS15, so the basic method is accepted and has been development further (e.g. by including Spanish stations). I have only a few issues that should be improved in the final version and would suggest a minor revision.

General comments:

Title: One term which was discussed in GM15 was the question if the dataset can be called "observations", as also a model is included (machine learning). When I read "observational", I assume that this are solely measurements. I therefore suggest to modify the title to "Observation-based gridded runoff estimates for Europe (E-RUN version 1.0)".

I miss a section describing the limitations of the approach. For example, it would be interesting what the approach does with areas which are under heavy human use (e.g. irrigation in Spain) or where reservoirs occur. I expect that the approach has problems in such areas (could be hardly to attribute from climate forcing only) and would like to ask the authors if they could discuss it briefly. In addition, it would be nice to have an interpretation of the jittering time series of Fig. 6. It is hard to see in which season the high / low number of stations are included. I guess that this is due to the assumption at P6, l 24 where summer months e.g. of Spain are flagged as "suspect" but the river might be intermittent for some time in reality. Furthermore, the machine learning algorithm depends (strongly) on the climate forcing used. It would be interesting to assess the effect of the climate forcing to the results of the machine learning algorithm. E.g. Müller Schmied et al., 2016 shows for the global hydrological model WaterGAP that modeled water balance components can differ largely due to the climate forcing. I know that the main task of the manuscript in ESSD is to describe the dataset (and e.g.

not to compare with other datasets) but I think it could be of value to discuss a possible effect in a few sentences. For future work, I would like to encourage the authors to analyze the effect of (different) climate forcing on the model results.

Specific comments

Abstract: Sure, there are different perspectives to which river runoff is belonging. To my understanding, a climate variable would be e.g. precipitation or temperature, but not river runoff. Does it make sense to write climate-related variable instead of climate variable? Further, the reader reads "River runoff is . . . directly linked to the terrestrial water balance" which sounds a bit strange. I would understand river runoff as a component of terrestrial water balance which is directly linked to climate variables – but I think that depends a bit on the perspective. You might consider this point for rewriting if you agree on that.

P1, l23: what is meant by "eventually"? Does it belongs to the "large rivers" or the contribution to discharge (which I would assume)? Please rewrite.

P3, l2: why did you selected only daily observations if you then aggregate it into monthly values?

P3, l12 ff: It can be understood that the Spanish data are only available for 2010 and 2011, but based on Fig 6 I think this is a wrong assumption. You might clarify that in the section for non-Spanish speakers that cannot get the information from the website (which is now showing 2011-2012 in the website title)

P6, l9: the authors stated that they converted discharge into runoff but did not wrote from where they got the basin areas which are needed for such an conversion. I guess it is used from the discharge databases (see metadata in the Supplement) but that has to be stated definitely in the manuscript.

P9, equation 5: Isn't it the well-known Nash-Sutcliffe-Efficiency? Would it be better to reference it as NSE?

Reference list: please check carefully upper-/lowercase in the titles and be consistent.

Fig. 1: the EWA and GRDB stations are very hard to see / distinguish from the GRDB stations, you might use a different color. In general, the figure could be larger.

Fig. 3: In case of fraction of suspect days is 0.0 – is the color white or also a bit reddish? If reddish, I would re-color it to white as otherwise, it is hard to interpret (for me). Furthermore, the figure caption is not very self describing.

Fig. 4: If the aim is to be able to distinguish only blue from red, than it is ok. For me it is very hard to distinguish the bluish and reddish colors. So I would suggest to use more different colors if it is the aim to distinguish within the categories.

Fig. 6: in caption "temperature" should be lower case. What is meant by "first/last available observation at each station" in case of > 1 station within a 0.5° grid cell?

Fig. 7: Please add to figure caption "monthly" (30% missing monthly values are selected) to be concise.

Fig. 8: Figure labels are not fully self describing. Please write out "cv" either in the figure or in caption. Is there a special reason for the different colors of boxplots? For me, it is a bit hard to see the median in the cv-time box-plot (black on dark-blue color), and in general, I do not see a reason to have the boxplot colored at all. Please use a light color (or white) for both boxplots instead.

Assessment of the data itself (downloaded from Pangaea website at 2016-02-07)

Citation at pangaea should be updated. Download runs without any problem.

First visualization using panoply. Time selector is located as mean of the month (which is, sometimes e.g. 16.1. 00:00 or 14.2. 12:00). Values of data are in mm/day, but calculated/displayed are monthly means. It might be better for interpretation of the data (esp. for someone who did not read the manuscript carefully), if monthly sums are provided (mm/month) and time iterator are consequently monthly time steps. For

example, time unit is given as "days since. . ." but providing 769 time steps. This is a bit unclear, as there are 769 monthly time steps. I suggest to modify the timestep from day to month.

I investigated the 30 years between 1979 and 2008 (I selected time period using Climate Data Operators), calculated monthly sums (ignoring leap years) and visualized long term average runoff in mm/yr. During that, I recognized that values are not available for all grid cells and each month (compare the Figure 1 here against e.g. your Fig. 9). For example, the last years of Iceland are missing, which I guess is due to the missing climate observation data. There are also other grid cells affected (e.g. Turkey). I suggest to describe in the manuscript that data are discontinuous for some regions/grid cells.

I could not resist to compare the long term average runoff with those from WaterGAP 2.2 (STANDARD, see Müller Schmied et al., 2014). Figure 1 shows (for all grid cells that contain values in every month of E-RUN) the long term average of E-RUN, WaterGAP 2.2 as well as the absolute differences. I was so interested, as WaterGAP is calibrated against the long term average river discharge (from GRDC database) for many basins in Europe, and that could be some kind of "fair" comparison. This is not completely the case, as station correction factor CFS in WaterGAP (see Müller Schmied et al., 2014) is applied only at the grid cell where the calibration station is located (and cannot be meaningful back-transferred to the basin itself). The comparison is also not "fair" as climate forcing differs (E-OBS for E-RUN, WFDEI for WaterGAP). Anyhow, the brad pattern is comparable, and in many regions, differences are within +-100 mm/yr.

References

Gudmundsson, L. and Seneviratne, S. I.: Towards observation-based gridded runoff estimates for Europe, Hydrol. Earth Syst. Sci., 19, 2859-2879, doi:10.5194/hess-19-2859-2015, 2015.

Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., Oki, T., Port-

mann, F. T., Reinecke, R., Riedel, C., Song, Q., Zhang, J., and Döll, P.: Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use, Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2015-527, in review, 2016.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18, 3511-3538, doi:10.5194/hess-18-3511-2014, 2014.
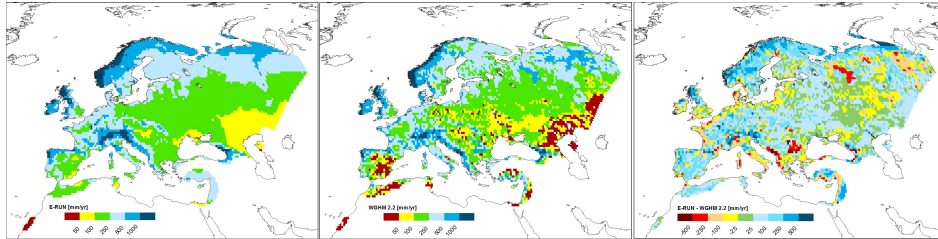
Fig. 1. Long term average cell runoff for the years 1979-2008 for E-RUN 1.0 (left), same for WaterGAP 2.2 (STANDARD, center), and differences (E-RUN – WaterGAP, right); all values in mm/yr.