Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2019-32-RC2, 2019

Open Access Earth System Science Data Discussions

Interactive comment on "GRUN: An observations-based global gridded runoff dataset from 1902 to 2014" *by* Gionata Ghiggi et al.

Anonymous Referee #2

Received and published: 21 June 2019

The manuscript by Ghiggi et al. presents an extension of the methodology developed by Gudmundsson and Seneviratne (2015,2016) in order to reconstruct monthly runoff rates at global scale at 0.5 deg spatial resolution. Runoff estimates from a large number of small catchments are combined with state-of-the-art atmospheric reanalysis to train a machine learning algorithm (random forests) to predict monthly runoff rates from monthly temperature and precipitation. Runoff estimates are extensively evaluated through cross-validation as well as against independent observed discharge time series and runoff estimates from global hydrological models. The study is technically sound and the manuscript is clearly written. The presented data-set is the first of its kind at global scale.

Main comments:

The authors chose to develop the runoff model from precipitation and temperature, while GS15 found for Europe that this simple model is outperformed by one considering full atmospheric forcing - which is available through GSWP3. This choice must be justified in the methods section. Further, GS15 showed that including land surface parameters (soil texture and slope) does not improve model performance. I can only assume that this is the reason why land surface parameters were not included in this study. This choice needs to be clearly mentioned in the manuscript and it should at least be discussed whether this assumption is also valid on global scale.

GRUN runoff estimates are compared to those from an ensemble of global hydrological models. The ensemble actually consists of global hydrology and water resources models, land surface models and dynamic vegetation models. Neither type was specifically developed or is trained/calibrated to accurately predict runoff at grid cell level - and this is without question where the strength of GRUN lies. This needs to be discussed in a revised manuscript. GHMs have mainly been validated against discharge observations from catchments >10000 km2, acknowledging the uncertainty in runoff estimates at grid cell level originating from model structure and parameters, climate input, and land surface properties. To make for a fair comparison, I challenge the authors to repeat the validation experiment in section 5.3 for the GRDC discharge records. Since the authors use results from the ISIMIP nosoc experiment, catchments heavily impacted by flow regulation and/or water abstraction need to be excluded from this analysis.

The authors argue that GRUN estimates do not include human interference on runoff generation, thus that differences between the GRUN reconstruction and in-situ observations could potentially be used to identify streamflow stations which have a hydrological regime very different from the naturalized flow (P14, L9-10). I strongly disagree given that the screening procedures used are capable of identifying break points in the discharges records caused by the installment of (large) water infrastructure, e.g. dams and diversions, while they likely miss more gradual changes. Consequently, the discharge time series used for training the model are a blend of near natural catchments and those impacted by human activity, and it is unclear to which degree human impact is actually implicitly represented in the reconstruction. This needs to be clearly stated and discussed in the revised manuscript.

The extrapolative power of GRUN is somewhat overstated (in particular P14 L7-16). The developed RF model is (in the best case) capable of predicting short-term changes in monthly runoff/water availability as a function of (changes in) precipitation and temperature with all other boundary conditions constant. Changes in water availability originate from a variety of drivers, e.g. water abstraction/diversion, land use/cover change, reservoir management, and in many cases climate may not be the most important one (see e.g. Haddeland et al., 2014). While I do not question the usability and added value of GRUN in the first place, the authors need to point out these limitations very clearly in a revised manuscript.

The authors evaluate the performance of GRUN in a cross-validation exercise. The corresponding maps displaying six performance metrics show clear regional differences in model performance, while numbers are only provided on global level (Table 1). In order to provide a more detailed picture of model performance for potential users of the data-set, I recommend to report numbers (Q5,Q50,Q95) at sub-global level, at least for the CV-SPACE experiment. This could be climate regions or the SREX regions already used by the authors.

Minor and technical comments:

P4 L17 "Climatological runoff signature" Clarify term or rephrase.

P4, L19-20 Clarify how false zero values were distinguished from e.g. dry out during drought periods

P4, L29 Hydrological signatures are index values to describe hydrological behavior, i.e. there is no single hydrological signature of a discharge time series. Please rephrase.

P5, L4 Correct to "displays"

P5, L5 Rephrase "sudden drops", e.g. step changes in the mean

P5, L1 Correct to "with an area"

P6, L4 "streamflow records heavily impacted by humans"

P6, L25 Correct to "that influence"

P6, L5 Consider rewording to "is assumed to be negligible" since routing is indeed not negligible under certain circumstances as later shown by the authors.

P8, L7 Correct to "Six performance metrics"

P10, L22 Correct to "show higher"

Sect. 6.1 The description of global runoff patterns is somewhat lengthy since it presents rather common knowledge, please shorten.

P12, L25 Which patterns are the authors referring to? Please explain in more detail.

P13, L14-15 Top 5 driest months in terms of runoff anomaly? Please clarify.

P14, L14-15 "to address other new scientific challenges in water cycle research" Either give specific examples or remove.

Fig 2: The resolution of all maps needs to be increased. In addition, I'd recommend to remove the continent outline since it blurs with the observation points.

Fig 3: Consider swapping x and y axis to ease readability and add name of the experiment to the caption. Add results from the CV-SREX experiment to the figure.

Fig 4: The map resolution needs to be improved. Please remove continent outlines to improve readability.

Fig 5b: I recommend to plot all simulated hydrographs in the same color, since the color coding according the NSE value compromises the readability of the plots.

Fig 8: Please increase resolution. The legend lists "LSMs" while the term "GHMs" is used throughout the manuscript. In Fig 8a, consider showing the full range of GRUN ensemble simulations - since the same is done for the GHM ensemble.

Fig. 10a: Please change background color since it is hardly distinguishable from the blues in the color scale.

Tab. 1+2: Please add the 5th and 95th quantile for each performance metric.

References:

Haddeland, Ingjerd, et al. "Global water resources affected by human interventions and climate change." Proceedings of the National Academy of Sciences 111.9 (2014): 3251-3256.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2019-32, 2019.