Earth Syst. Sci. Data Discuss.,
doi:10.5194/essd-2015-38-RC2, 2016

Open Access Earth System Science Data Discussions

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

G.V. Ayzel (Referee)

hydrogo@yandex.ru

Received and published: 9 February 2016

As noted in the first referee comment "The beauty of first publishing the method and results... that the discussion and review comments of GS15 are publicly available" and it's totally true for this article too – following previous article history helps us to understand general concepts and, in addition, each referee can focus on the own key points of presented work without any intersections.

I am really impressed by the described database construction and provided description of the workflow – it is one another significant step to the modern science with shareable data, community models and reproducible research.

For this interesting and useful article and dataset I would recommend such minor revisions related to deeper analysis around a key questions I marked.

General comments: Machine learning model cross-validation in space and time is a valuable tool for accessing performance fluctuations in respect with changing conditions, and for this work it was done well. Nevertheless there are two points I marked as unclear and required more detailed explanation.

1. The most dataset points were derived from the Random Forest model (RFM) implementation to E-OBS forcing data, but only a few points were used to model training and testing – and I guess that performance of RFM roughly decreases for eastern part of Europe (e.g. Russia) because of no observational data were account in the model. It is a theme of a great interest and I recommend describe it more clearly – it may have an impact of the decision to use provided dataset for Eastern Europe domain.

2. This question is mainly related to the GS15 [1] article, but it is the main question for my own interest too. What is the reason of such significant differences between LSM- and RFM-generated monthly runoff? Our recent results (SWAP model runs for ISI-MIP basins) show monthly NSE about 0.8, and I think other models demonstrate the same efficiency. Why should we use E-RUN instead of the averaged LSMs outputs for large-scale case studies?

Specific comments:

Title. Actually it is combined (observational plus RFM-predicted) dataset; Abstract, P1, L1: "hydrological cycle variable" instead of "climate variable"; P1, L13: same comment as previous one; P8, L20: average number is not useful metric, histogram of assigned stations per cell will be suitable; P9, L 29: model efficiency is the same metric with Nash-Sutcliffe-Efficiency; It is necessary to attach figures in high quality to the supplement section.

Data availability assessment: Provided link to the dataset hosted by PANGAEA is not stable, there is the reason to make some reserve open data storage. Reading, testing and visualization was done by the NETCDF4 and Basemap libraries for python (ver.3.5) without any errors.

Please also note the supplement to this comment:
http://www.earth-syst-sci-data-discuss.net/essd-2015-38/essd-2015-38-RC2-supplement.pdf

---

Interactive comment on Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2015-38, 2016.