

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2
<https://doi.org/10.5194/hess-2021-481-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on hess-2021-481

Kolbjorn Engeland (Referee)

Referee comment on "Parsimonious statistical learning models for low-flow estimation" by Johannes Laimighofer et al., Hydrol. Earth Syst. Sci. Discuss.,
<https://doi.org/10.5194/hess-2021-481-RC2>, 2021

The manuscript is well written, comprehensive and add new knowledge to the hydrological community on the performance of machine learning approaches to our modelling and prediction challenges. All the new models are well compared to existing models applied to the same dataset as well as relevant results from the scientific literature.

I have some minor comments to the manuscript.

Introduction:

From the conclusions, it seems like one more research questions could be added: what are the most important predictor variables for low flows in Austria.

Figure 2:

Units could be added to the x-axis

Section 2.2:

It would be good to have some clarifications.

- Did you use 'residual catchments' or the total upstream catchment area for each gauging station when calculating low flows and catchment characteristics?
- Why choose the grid point nearest the gauging station. Would not catchment center of gravity be a better choice?
- When using the 5x5 minutes gridded dataset, how did you account for altitude dependency of temperature (and possibly precipitation)?

Line 140:

The GAM approach has previously been used for prediction of low flows in Ouarda et al (2018)

3..1.3:

The use of the numbering of each step in sPLS is confusing.

3.2:

Could be useful to repeat the complete name of RFE in a similar way that it is done for all the regression approaches

I would be happy if you add one sentence on how nested CV is used. Is it correct that all data are used on both loops of the CV, the only difference is that you make different splits for each of the loops?

Figure 4:

It is difficult to see the colours of the green and red vertical lines. It might help if the colours are a bit brighter

Lines 435-440:

I think that also data quality might be an important factor when winter low flows are calculated from data. My experience from Norwegian data is that water levels often increase due to freezing in the river. The standard rating curve can therefore not be used to estimate discharge. Instead, discharge is estimated by gap-filling approaches like interpolation, use of donor stations or simple hydrological or statistical models.

Conclusions:

In addition to the importance of different predictor variables, is it possible to summarize how they influence low flows? Do low flows increase or decrease with these predictors?

Reference

Ouarda, Taha B. M. J.; Charron, Christian; Hundecha, Yeshewatesfa; St-Hilaire, André; Chebana, Fateh (2018). Introduction of the GAM model for regional low-flow frequency analysis at ungauged basins and comparison with commonly used approaches. *Environmental Modelling & Software*, vol. 109. p. 256-271.