Interactive comment on "A European map of groundwater pH and calcium" *by* Michal Hájek et al.

Anonymous Referee #1

Received and published: 8 January 2021

General comments

The authors use a suite of statistical model and machine learning approach to develop a European-wide database of groundwater pH and calcium composition from a large-sample of on-site measurements across Europe. These groundwater parameters are major predictors of ecological status and biodiversity of groundwater bodies. This new database provides some improvements over existing similar database and constitutes an important contribution to the scientific community as well as present and future sustenance of groundwater dependent ecosystems.

I believe that the manuscript is suitable for publication. But the authors could provide clarifications on the specific comments raised below:

Specific comments for authors

1. You have performed quality control to filter unfit datasets before the geographical modelling and numerical analysis, yet modeling exercise such as this are associated with prediction uncertainties. Can you provide an estimate of uncertainties associated with the estimated GW-pH and GW-Ca produced by the model?

3. Is it possible to improve the performance of the random forest model? Which additional predicting variable(s) (even if such information is scarce) could be added to improve the results?

3. What are the criteria for selecting the three predictor variables (soil pH, lithology, and precipitation) used as proxies for predicting groundwater pH and Ca concentration?

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-319, 2020.