

Geosci. Model Dev. Discuss., chief editor comment CEC2  
<https://doi.org/10.5194/gmd-2022-44-CEC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Reply on AC1

Juan Antonio Añel

---

Chief editor comment on "A machine learning methodology for the generation of a parameterization of the hydroxyl radical" by Daniel C. Anderson et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2022-44-CEC2>, 2022

---

Dear authors,

The size of the parameterizations is not a problem. First, you must not use GitHub. GitHub is not a suitable repository for scientific publication. GitHub itself instructs authors to use other long-term archival and publishing alternatives, such as Zenodo. So first, please, remove any mention of GitHub from your manuscript. Any limitation in size is imposed by the suitable repository you decide to use. You have already used Zenodo, and the limitation of Zenodo is 50 GB per dataset, and you can upload multiple datasets. Given that a full year of parameterizations takes approx. 12 GB, as you report, is entirely feasible. However, I understand that although the ideal would be to upload and store everything, maybe the parameterizations for all the years don't need to be uploaded to assure minimum replicability of your work. In this case, I request you a minimum dataset that lets to obtain the figures included in the manuscript (output files produced with the simulations with the parameterizations, but only the variables plotted). It would be desirable to have the parameterization files for the 1980-2005 or 1980-2019 periods. Please, let me know if this is feasible for you.

We do not ask for the data so that the other party can apply them to another model, but to comply with our obligation of assuring that your research is replicable by others, as the scientific method requests. What is necessary is precisely the same model, inputs, etc., that you have used to produce your results. In this way, whether the files and information are helpful or not for others or the fact that they are misused is irrelevant to the goals pursued by the code and data policy.

Finally, reviewing your manuscript again, I have noted that you have not included the version number of the XGBoost and Python that you have used. Please, reply with it and include it in any future reviewed version of your manuscript.

Regards,

Juan A. Añel

Geosci. Model Dev. Executive Editor