Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-59-AC1, 2020

Geoscientific Model Development Discussions

Interactive comment on "Using SHAP to interpret XGBoost predictions of grassland degradation in Xilingol, China" *by* Batunacun et al.

Batunacun et al.

906301021@qq.com

Received and published: 26 October 2020

1, Thank you for your interesting submission. This paper presents an interesting suite of tools for investigating an important topic of research. I have only some minor comments to make before publication:
Response: We appreciate your suggestions for our manuscript and we consider all your comments as very useful. We have addressed each of your comments below.
2, As with any ML interpretation, one wonders how more generally useful this is to other regions. Have you considered whether testing against historical datasets is worthwhile?
In other words, have you tried to apply this method to re-analyses previously studied grassland degradation? If not, are there similar areas of focus this might work on?

Response: Thank you for your useful question.
We have not yet been able to test this method in another region or to historical datasets.
But the method in this study has certainly the potential for transferability for two reasons:
First of all, we have used this method on another topic and dataset in the same region, namely for studying land degradation. Please see Land-use change and land degradation on the Mongolian Plateau from 1975 to 2015ïijŇ a case study from Xilingol, China. Land Degradation Development 29: 1595–1606. DOI: 10.1002/ldr.2948. XGBoost and SHAP presented an excellent performance as well.
Secondly, the datasets in this study (land use and driver data) are public and available and could be replaced by other datasets. For more information please see the data description in the manuscript, line 171-173.
Based on this, we believe XGBoost and SHAP provide large potential to be applied to other datasets, regions and topics as well.
Actually, the datasets in this manuscript are historical data and we predict the dynamic grassland degradation (newly added grassland degradation, NGD) between 2000-2015 in Xilingol based on historical data from 1975 to 2015. Then we use the historical data from 2000-2015 to test the predicted NGD between 2000-2015. The results indicated that it is worthwhile testing against the historical data.
More detailed information about the data in this manuscript please check line 143-170.
In addition, for a ML model,
3, Please can you comment on the computer hardware required for training?
Response: Done.
I comment the computer hardware in "Code and data availability" section, please check line 582-594. The used XGBoost algorithm including the SHAP library runs well on a modern (Intel or AMD) PC (4 cores or more, 16 GB RAM). The training and the simulation were made on Linux as operating system but should work also under Windows.

4, Please include a Zenodo, or other archive, snapshot of the data used in this study. Thanks!
Response: Thank you for your careful comments, I have published the python code at GitHub and Zenodo. The data also has been described clearly in GitHub README.md. The results in this paper could be reproduced by using data in GitHub.
Please check the following link of the python script:
Link: https://zenodo.org/record/3937226.Xw2M6egzZPY
DOI: 10.5281/zenodo.3937226
I have given the specification in the manuscript, please check the attached file.
Please check line 284-285, 560-563

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-59, 2020.