

Interactive comment on “Game theory interpretation of digital soil mapping convolutional neural networks” by José Padarian et al.

Anonymous Referee #1

Received and published: 24 April 2020

General comments

This paper introduces the use of the SHAP (Shapley additive explanations) coming from the Game Theory as a mean to assess the relative implications of covariates in a complex model from machine learning. The methodology is tested on a CNN (Convolutional Neural network) developed to model SOC content at 5 depth intervals in Chile.

The purpose of this study is significant giving the difficulty to interpret complex models developed on machine learning techniques, especially in a spatial context. The expression ‘black boxes’ describes those models pretty well. Here, the authors provide a solution to decrypt those black boxes.

C1

The example provided here illustrates that SHAP could be promising allies in the future modelling works of many soil scientists, especially in a spatial context. Indeed, applying SHAP allowed the authors to confirm that the CNN model captured clear relationships (along with environmental thresholds) existing between SOC and different covariates, and this at the local and the global levels. SHAP also gave information of the relative implications of the covariates in the CNN model for different contexts.

This paper is of good scientific quality, well organized and written. The reading is fluid. I think this paper has its in SOIL after adding few complementary information.

Specific comments

It would be really appreciated having a bit more details in part 3.1, especially about the model functioning and the choice of the training and test sets.

In addition, it could be nice having few information about the validation of the model, its goodness-of-fit, somewhere in the text before presenting the SHAP results, please. Maybe a simple point plot representing the predicted vs observed data based on the test set for at least the 0-5cm depth interval?

I just wonder why you did not directly input land cover/land use data in the model giving the important influence of this parameter on SOC content?

Technical corrections

p.2, l.1: ‘researchers are able to reduce...’

p.3, l.25: The reference needs to be reformatted I think.

Interactive comment on SOIL Discuss., <https://doi.org/10.5194/soil-2020-17>, 2020.

C2