SOIL Discuss.,
https://doi.org/10.5194/soil-2018-28-AC2, 2018
© Author(s) 2018. This work is distributed under
the Creative Commons Attribution 4.0 License.

Interactive comment on "Using deep learning for Digital Soil Mapping" *by* José Padarian et al.

José Padarian et al.

jose.padarian@sydney.edu.au

Received and published: 30 November 2018

Thanks again for your comments.

- Also the test dataset is only 49 sampling points versus the 1744 (with augmented data). Was the split repeated?

The 49 points were selected on the original data, before data augmentation, in order to have the same test for both cases. The split was not repeated because the idea is to exclude the test set from the training. In a repeated sample, some of the test data would be present in the training or validation sets, defeating the purpose of having a "completely independent" test set.

- About test set selection

As we mentioned in the manuscript (line 1, p8), we performed a random selection.

Since the distribution of the OC is not normal, the distribution of the test set is not exactly the same (see attached figure) and that explains the lower error.

We acknowledge that it would be possible to sample the original data in a different way (sampling by quantiles), and that would increase the error shown in Figure 5 (only in the test set). Most certainly, the error wouldn't be significantly different than the validation error, since the test set would have a distribution just like the rest of the data, which error is accounted during the validation.

We propose adding to the revised version a paragraph explaining this and the attached figure.

Interactive comment on SOIL Discuss., https://doi.org/10.5194/soil-2018-28, 2018.

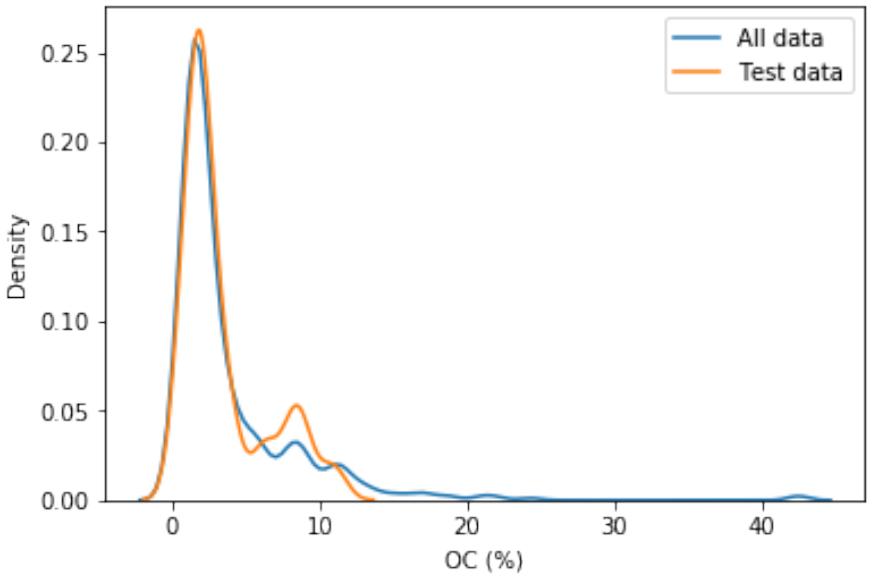

Fig. 1. Distributions