SOILD

Interactive
comment

Interactive comment on "No Silver Bullet for Digital Soil Mapping: Country-specific Soil Organic Carbon Estimates across Latin America" *by* Mario Guevara et al.

Mario Guevara et al.

mguevara@udel.edu

Received and published: 10 April 2018

Responses to referee 1

The authors appreciate the overall support and constructive feedback provided by the reviewer, which will improve the clarity of the paper. We recognize that improvement is needed for the methodological sections. Please note that our conclusion about no best method has been recognized in other studies (Ho & Pepyne, 2002; Qiao et al., 2015) but it is still an ongoing discussion. We argue that there are no best methods for statistical modeling in a quantifiable basis, and we argue that the no-free lunch statistical theorem should apply also for digital soil mapping.

Printer-friendly version

Discussion paper

Specific response to reviewer

1. Weighted overall measure of mapping accuracy maybe more informative than stats per country?

We have calculated an overall fit using each one of the methods for all Latin America and the overall fit is ∼0.41 of explained variance. That said, our overall objective was to evaluate the performance of each algorithm for each country regardless of size and information (i.e., data points). We recognize that with time the WoSIS database will grow as there are multiple sources of information within countries. Our framework assumes that we cannot conclude that RK and RF are the "best methods" if we cannot prove that in fact they are always the "best methods" for all countries (i.e., small and large countries). Although RF and RK were the 'best methods' in terms of absolute accuracy (see Taylor diagram in Figure 2 that for Mexico and Brazil (as in most of the countries), all methods (points) overlap each other and fall in the same range of error, variance and correlation. We recognize that the methods section can improve the explanation of the methods used for model evaluation and how to interpret them (e.g., Talyor diagrams, see (Taylor, 2001; Carslaw & Ropkins, 2012). Please note that an overall measure of mapping accuracy for the 5 methods across all the datasets was performed for bulk density, coarse fragments and SOC (density and mass). Figure 1 shows that all models have relatively similar prediction capacity (except for bulk density where RK generate the worse predictions), supporting the "no free lunch" theorem. We will work in a revised version to improve clarity in the methods and discussion section to make this argument clear.

2. Building models with <30 training points can lead to artifacts

We agree that this recommendation is common for geostatistics. That said, this recommendation does not necessarily apply for machine learning. Based on the quantile theory (e.g., Meinshausen, 2006) and examples for regression and classification (e.g., Zhu, 1997; Pearson et al., 2007) we argue that having just a few points (i.e., ∼10) it is

possible to achieve reliable predictions based on a machine learning framework. We recognize that with time the WoSIS database will grow as there are multiple sources of information within countries. Consequently, our approach can be replicated and tested in the future. We strongly believe that this information represents an important baseline for Latin American countries that have limited information and training for digital soil mapping. Thus, our efforts are empowering capacity building across Latin America.

3. Is local better than global? Is it justified to stitch maps produced by countries vs using global models? (note that reviewer skipped #3 so number in answers are shifted from the reviewer comments).

We fully agree that more information and discussion about global/Latin America vs country-specific analyzes is an ongoing discussion for digital soil mapping. We ran several models for all Central America (i.e., "global" approach) and found very low accuracy and high bias ($\sim$2% of explained variance), a "global" Latin American fit improved explained variance up to $\sim$41%. We can include these results and discussion in the revised version. Please note that Figure 1 shows the comparison of the five methods applied to SOC related properties considering all available data for Latin America (i.e., "global" approach). We will make this clearer in the methods section and discussion in a revised version as suggested by the reviewer.

4. Evaluation of the accuracy of predictions should be based on e.g. 5-fold CV with re-fitting

We appreciate the comment from the reviewer and we will improve clarity of the methods in a revised version. We argue that a fair comparison (either five- or ten-fold) means that the same method was applied to all datasets. Consequently, we selected 10-fold based on previous studies (Borra & Di Ciaccio, 2010). We clarify that the model evaluation is based on interpretation of Taylor Diagrams, which are common practice for model comparison and evaluation for climate research. We will make this clearer in a revised version. We have seen that the lower density of points the lower variance

is achieved by increasing the number of folds. Previous studies have shown that under specific cases a 10-fold validation slightly decreases the variance (Markatou et al., 2005). However, a visual inspection and descriptive statistics of the training dataset is always recommended as good practice (Esbensen & Geladi, 2010) but it is not practical with very large datasets. We argue that the differences in accuracy associated with the number of folds are significantly lower than the differences between models (which use the data under different assumptions); thus, we believe that our approach is robust.

5. Github repository does not contain all points and grids you have downloaded from ISRIC WoSIS and worldgrids.org.

Our code is designed to process the original WoSIS dataset, thus we do not provide a "dataset" as the information is available in WoSIS.

Other detailed comments in the manuscript and questions are available here.

The authors will properly address the detailed comments. Note from the detailed comments that there is still large uncertainty of >5x5 km pixels soil carbon estimates (Tifati et al. 2017), which is the main reason why multi-scale mapping exercises including coarse grained maps are continuously required to better understand the sources, the spatial controls and error propagation pathways of carbon cycle-related models. The authors appreciate your constructive review. The authors believe that properly addressing your concerns along the document will benefit the value of the paper.

References

Borra S, Di Ciaccio A (2010) Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. Computational Statistics & Data Analysis, 54, 2976–2989.

Carslaw DC, Ropkins K (2012) openair âĂŤ An R package for air quality data analysis. Environmental Modelling & Software, 27–28, 52–61.

Esbensen KH, Geladi P (2010) Principles of Proper Validation: use and abuse of re-sampling for validation. Journal of Chemometrics, 24, 168–187.

Ho YC, Pepyne DL (2002) Simple Explanation of the No-Free-Lunch Theorem and Its Implications. Journal of Optimization Theory and Applications, 115, 549–570.

Markatou M, Tian H, Biswas S, Hripcsak G (2005) Analysis of Variance of Cross-Validation Estimators of the Generalization Error. Journal of Machine Learning Research, 6, 1127–1168.

Meinshausen N (2006) Quantile Regression Forests. J. Mach. Learn. Res., 7, 983–999.

Pearson RG, Raxworthy CJ, Nakamura M, Townsend Peterson A (2007) ORIGINAL ARTICLE: Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. Journal of Biogeography, 34, 102–117.

Qiao H, Soberón J, Peterson AT (2015) No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation (ed Kriticos D). Methods in Ecology and Evolution, 6, 1126–1136.

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research: Atmospheres, 106, 7183–7192.

Tifafi M, Guenet B, Hatté C Large Differences in Global and Regional Total Soil Carbon Stock Estimates Based on SoilGrids, HWSD, and NCSCD: Intercomparison and Evaluation Based on Field Data From USA, England, Wales, and France. Global Biogeochemical Cycles, 2017GB005678.

Zhu A-X (1997) A similarity model for representing soil spatial information. Geoderma, 77, 217–242.

Please also note the supplement to this comment:

https://www.soil-discuss.net/soil-2017-40/soil-2017-40-AC1-supplement.pdf