SOIL Discuss., referee comment RC2
https://doi.org/10.5194/soil-2020-99-RC2, 2021

Comment on soil-2020-99

Anonymous Referee #2

Referee comment on "The central African soil spectral library: a new soil infrared repository and a geographical prediction analysis" by Laura Summerauer et al., SOIL Discuss., https://doi.org/10.5194/soil-2020-99-RC2, 2021

The research aimed to present a mid-infrared soil spectral library (SSL) for central Africa (CSSL) to predict key soil properties, thus allowing (i) for future soil estimates with (ii) a minimal need for expensive and time-consuming soil laboratory analysis. The CSSL contains over 1,800 soils from ten distinct geo-climatic regions (from the Congo Basin and wider African Great Lakes region) for a whole of six hold-out core regions.

The paper is affected by several issues, and therefore I must suggest its rejection.

In the following points, my main concerns:

- General comment: used methods or obtained results do not justify several sentences. In the following points, some example are reported, but many other occurs;

- Abstract: “we present a mid-infrared soil spectral library (SSL) for central Africa (CSSL) that can predict key soil properties”…but after the author state, “We present three levels of geographical extrapolation, deploying Memory-based learning (MBL) to accurately predict carbon (TC) and nitrogen (TN) contents in the selected regions.”. So, you are not presenting a CSSL to predict key soil properties, but “only” some selected soil properties! The authors should be consistent throughout the text.

- Abstract and Discussion: “The Root Mean Square Error of the predictions (RMSEpred) values were between 0.38–0.86 % and 0.04–0.17 % for TC and TN, respectively, when using the AfSIS SSL only to predict the six regions. Prediction accuracy could be improved for four out of six regions when adding central African soils to the AfSIS SSL. This reduction of extrapolation resulted in RMSEpred ranges of 0.41–0.89% for TC and 0.03–0.12% for TN.” Ok, but immediately after I read, “In general, MBL leveraged spectral similarity and thereby predicted the soils in each of the six regions accurately; the effect of avoiding geographical extrapolation and forcing regional samples in the local neighborhood (MBL-spiking) was small)” or, even along the Discussion section (line 309), “We showed that TC and TN in six regions of our CSSL can be accurately predicted”…so, in the same paper, the authors write two opposite things. I agree, according to your results, that the first sentence was more closes to reality than the second one, but this bring to an additional issue, i.e., see point 4;

- Abstract, Discussion, and Conclusions: your results don’t look so “promising” (lines 17, 352) as you state, and some of your results and the following discussion are too much speculative;

- Results and Discussion: authors didn’t explore limits in their proposed method. For instance: issues arising from the use of RMSE to compare predictions among regions with different pedoenvironmental features and, consequently, total C and total N.

- Soil sampling method and approach: soils were sampled according to a prefixed depth technique (Table 1) without considering soil variability in terms of main genetic horizons. So, this means that there is huge variability in processes and, consequently, pedogenetic features. But this problem is not considered as a possible cause of errors in obtained results. This is totally a mistake for this reviewer. Indeed, looking at Table 2, it was clear that a quite high pedovariability exists in investigated soil samples (samples comes from five different RG);

- Whole paper: a group of references should always be avoided. It could be preferred to use a max of 2 refs. after every important statement. Otherwise, it could be quite impossible to verify if reported references was cited in a good way;

- Whole paper: several acronyms appear without any explanation!.

- Whole paper: several typing mistakes occur. Some are reported here (vide infra), but many others occur. Additionally, the correctness of some sentences is questionable;

- Title: too generic and not fully in agreement with obtained results (vide infra). Indeed, I am not sure that you have filled a gap; at least in an accurate way;

- Abstract (line 11): AfSIS!?!

- Introduction (from l. 28-30): “Despite the expected severity of these impacts, our understanding of the effects in the humid tropics are limited by sparse data and uneven distribution of low-latitude research”. Too vague and generic sentences. For instance, such a sentence is not true for many areas of Brazil;

- Introduction (l. 30-31): “which contains the second largest tropical forest ecosystem on Earth and represents a considerable reservoir of soil C (FAO and ITTO, 2011)”. Old reference. Ten years are already gone by. In case of such important statement more recent, an updated information must be reported;

- Introduction (l. 33): “Thus, the projected drastic population growth in the coming decades (Vollset et al., 2020)” a quantification in terms of percentage, or something like this, is always required; otherwise, it is just a vague statement;

- Introduction (l. 35-36): “In the wake of these current and future impacts, more spatially explicit soil information is urgently needed in many research fields.” Again, too vague and generic sentence. Which field of research?;

- Introduction (l. 44): “low cost” always depends on the point of view. What does for the authors “low cost” means? Why not introducing a specific brief paragraph for cost estimation by comparing soil analysis vs. DRIFT spectroscopy;

- Introduction (l. 50-55): too speculative sentences. It seems more an authors’ self-convincement rather than a scientifically based questions;

- Introduction (l. 52-53): sorry, I really don’t know what "positive predictive transfer" means;

- Method (l. 91): WRB, 2006? Really? Are you aware of the 2015 updated version?

- Method (general comment): What about the way you selected “latent variables” for the global calibration you did for optimizing spectral pretreatment?;

- Method: “Note, even if the proportion of samples with inorganic carbon was very low (5%), the term TC will be used in the study.” As usual! Why do you need to specify such an obvious aspect?;

- Method: I think that the way you pretreated your soil samples should be specified;

- Method: “A gold standard was used as a background material for all measured soils” which kind of “standard”? It was a reference soil certified material? Why not including such important information?;

- Method (Table 2): For this reviewer, it was not so clear if you used all the reported nr. of soil samples. It would help if you were more clear from this point of view;

- Method: “Reflectance was transformed into absorbance (1/reflectance) before further processing and subsequent modeling.” No reference!;

- Method: “Four replicates per sample were measured and an average of 32-co-added scans were used for each sample” why? Four replicates are enough for you? If yes, you need to explain the reasons from a statistical representative viewpoint;

- Results (general comment): very aseptic. It looks like a technical report totally detached from the context;

- Results (paragraph 3.1 and Fig. 3): I discover for the first time that the authors applied a multivariate approach too. In particular, they used a PCA. Unfortunately, they didn't explain to us anything about how it was implemented. This is really unusual for this reviewer. Indeed, when a multivariate tool is used, data-pretreatment represent a pivotal matter, but the authors didn't explain anything about this. Additionally, several authors, statisticians included, clearly demonstrated that PFA was better for variability interpretation in a soil dataset with soil data;

- Line 268: soils rather than "sols";

- Results (lines 267-268): "This was expected as the principal component analysis indicates that the sols of these regions might not be properly represented by the AfSIS library." Where? I don't see such an information from PCA;

- Results (lines 276-279): I do not fully agree with the suggested reasons for the total C and N predictions underestimation trend in the six investigated regions. Indeed, several outliers occur in your dataset. This was typically due to an underestimation in investigated pedovariability (vide supra);

- "Results" and "Discussion" (general comment): both these parts are full of "could", "may", "might", etc. I understand that caution is always required in a scientific text, but some more certainties should be given. So, I wonder: are the authors sure enough of the applied method and the validity of the obtained results or not? As a reviewer, the text has several methodological drawbacks, which bring me to hypothesize that all these doubts could be the demonstration of a low statistical robustness of obtained results;

- Discussion (line 309): "We showed that TC and TN in six regions of our CSSL can be accurately predicted". Honestly, I am not agreed. In previous pages and Tables, total C and N prediction can be rarely defined as "accurate";

- Discussion (line 309): "The advantage of using MBL is that it finds spectrally similar observations for every new observation to fit specific models". This is an obvious observation that can be written for every prediction "model";

- Discussion (line 312): ")"…?;

- Discussion (general comment): extremely redundant with the "Results" section. A combination of the "Results and Discussion" section it would have improved the paper in terms of overall quality, clarity, and readability;

- Discussion (general comment): readability is made really low due to the presence of too many acronyms. I understand that several acronyms characterize the whole paper, but some strategies would have improved readability (for instance, avoiding its use while preferring a "recall" of their original meaning);

- Discussion (line 319-323): another obvious observation that strongly affect your paper in terms of novelty;

- Discussion (line 324-326): "We conclude that the particularly high soil diversity in these two regions in terms of soil biogeochemical properties introduces additional complexity in the soil spectral prediction workflow" this is the point! Even if, in my opinion, it would be better to use "soil bio-physical-chemical features" rather than "soil biogeochemical properties". However, this clearly confirm all my previous doubts, and I am astonished that the authors recognized such a big issue only at the end of their paper without additional insights about this;

- Discussion (line 324-326): "Regions that occupied the same score space of the first two principal components as the corresponding other regions and the AfSIS SSL (Figure 3) showed only a minimal effect from spiking (Figure 1)" where I can see such an outcome? It is not contained in Fig. 3 and 1 for sure;

- Discussion (l. 348-250): "Even though spiking is described as particularly effective in improving performance of small sized models (Guerrero et al., 2010), spiking, in our study, did not have as strong of an effect as reported by earlier studies (e.g., Guerrero et al., 2014; Seidel et al., 2019; Barthès et al., 2020; Wetterlind and Stenberg, 2010)…and the reason is!?!;

- Discussion (l. 353-354): "The addition of geographically proximal regions to the large-scale library, which are included in our CSSL, improved prediction accuracy significantly". Sorry but once again, I disagree with the authors. From your reported results, it seems that accuracy improved but not in a so highly significant degree;

- References: Total nr. of references: 77…too much for an original article; Total nr. of references before 2011 > 20; Self-citations > 10