Review 2

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.

We thank the reviewer for the time and effort in reading and commenting on our proposed manuscript. We are confident that the issues the reviewer raises can be addressed in a revised manuscript. We understand that certain methods and interpretations were not clearly formulated and will add missing information and rephrase unclear sentences. Additionally, as described in our response above to Reviewer 1, we will report and discuss the effect of spiking more in detail. We respectfully disagree with several repeated main concerns of the reviewer that the pedgogenic heterogeneity of the soils would be a critical problem of our study. Presenting the differences between central African soils to the soils covered by the Sub-Saharan spectral library indicates, on the contrary, the importance of our presented data analysis. Due to the new variability of the soil samples that our central African spectral library adds to the existing continental library, prediction accuracies will be significantly improved for these regions. Our findings and platform also encourages the future addition of new data. Our infrared library therefore helps to more accurately predict central African soils and represents a first step towards filling a critical knowledge gap of this understudied area.

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.

  The reviewer is correct, we present a workflow on how to predict total carbon and total nitrogen of soil samples using our central African SSL together with an existing continental library. We also made all data (spectra, metadata and wet laboratory measurements) and accompanying code openly available on a Github repository. This will not only allow for the reproduction of our analyses but also for new analysis and predictions of soil properties in new studies. Importantly, this will facilitate new soil analyses for this highly understudied area.

As we state in subsection 2.2, L103-106, additional soil properties, which are included in the repository, were analysed. These include pH, texture, total Al, Fe, Ca, Mg, Mn, Na, P, and K. We chose to highlight TC and TN as example properties to demonstrate our predictive models in a concise way. The additional data for the parameters listed above and also the results of the same analyses are available on the GitHub repository. We will modify the abstract and the discussion to re-iterate the availability of these auxiliary data.

- 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;

We agree with the reviewer that these sentences provide limited context for which circumstances the inclusion of chemically associated spectral information was beneficial. As described in our responses to Reviewer 1, the effect of spiking on the prediction accuracy was substantial. We will modify the abstract and body text to maintain consistency of this result throughout.

- 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;

We thank the reviewer for their perspective but respectfully disagree that the results do not look promising. Compared to other large-scale mid-infrared prediction studies (e.g. Dangal et al. (2019), Angelopoulou et al. (2020)) and also to other soil infrared studies, which look at geographical extrapolation strategies (e.g. Padarian et al (2019), Briedis et al. (2020), Gomez et al. (2020)), our results for TC and TN provide a method that yields satisfactory results in a simple and cost-effective manner. In fact, given the variability in soil properties covered by our data the accuracy of prediction exceeded our initial expectation and provides now a tool to further study the role of large scale patterns of soil properties in one of the least studied but fastest changing regions of the world.

Angelopoulou, T., Balafoutis, A., Zalidis, G., Bochtis, D.: From Laboratory to Proximal

Sensing Spectroscopy for Soil Organic Carbon Estimation—A Review. *Sustainability*, *12*, https://doi.org/10.3390/su12020443, 2020.

Briedis, C., Baldock, J., de Moraes Sá, J.C., dos Santos, J.B., Milori, D.M.B.P.: Strategies to improve the prediction of bulk soil and fraction organic carbon in Brazilian samples by using an Australian national mid-infrared spectral library, Geoderma, 373, https://doi.org/10.1016/j.geoderma.2020.114401, 2020.

Dangal, S., Sanderman, J., Wills, S., and Ramirez-Lopez, L.: Accurate and Precise Prediction of Soil Properties from a Large Mid-Infrared Spectral Library, Soil Systems, 3, https://doi.org/10.3390/soilsystems3010011, 2019.

Gomez, C., Chevallier, T., Moulin, P., Bouferra, I., Hmaidi, K., Arrouays, D., Jolivet, C., Barthès, B.G.: Prediction of soil organic and inorganic carbon concentrations in Tunisian samples by mid-infrared reflectance spectroscopy using a French national library, Geoderma, 375, https://doi.org/10.1016/j.geoderma.2020.114469, 2020.

Padarian, J., Minasny, B., McBratney, A.B.: Transfer learning to localise a continental soil vis-NIR calibration model, Geoderma, 340, 279-288, https://doi.org/10.1016/j.geoderma.2019.01.009, 2019.

- 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.

We thank the reviewer for this comment and agree that it can make sense to use the $RMSE_{pred}$ to compare between regions but strictly together with the range of the measured attribute since data distributions are different. Nevertheless, the $RMSE_{pred}$ is an appropriate error metric to compare the predictive capacity across the of the three modeling strategies, as assessed by individual regions (e.g., Table 3). As answered above in response to Reviewer 1, we will modify the manuscript to better reflect limitations of obtained accuracies. We will also discuss geophysical and environmental variability between the regions more in depth.

- 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);

We respectfully disagree with the reviewer’s opinion. We agree that using samples that were sampled per horizon would have been an advantage for using the data for pedogenetic interpretation later on. However, such data is rare at continental scales. We would also like to have a complete chemical and pedological (soil forming factors) characterisation of the collected, analysed and modeled soils, but this is a cumbersome endeavour to explore in full detail (XRD, geomorphology, land use (history), etc.). This is simply not feasible for the size and extent of our soil collection and thus deemed beyond the purview of this study. Given that the depth increments for samples included here did in most cases not exceed 10 cm increments we believe that our predictions can still yield considerable depth explicit information. Samples were taken in a way that a large variety of mineral and organic mixtures are covered. Soil spectroscopy can naturally deal with such soil complexity. In terms of the methodology used here, since our data covers a significant variability of soil conditions, our library can be used for samples taken with fixed depth increments or sampled pedogenetically following horizon boundaries. An additional advantage of depth explicit sampling is the fact that for example TC and TN stocks can now be accounted for by various volumes of soil. One of the nice features of infrared reflectance spectroscopy is that it generates signals arising from absorption features of chemical bonds that are distinctive of functional groups and the organic or mineral compounds that contain them. Spectra offer an integrative fingerprint to comprehend major chemical complexity and selected physical properties in soils in combination with statistical modeling. One of the key assumptions (and generally the foundation of predictive capacity) is that chemical relatedness is sufficiently reflected in the spectra. In the case of memory-based learning with a nearest neighbour (distance) approach, chemical relatedness and thus the pedogenic resemblance is even enforced in the modeling process via a nearest neighbor approach. Variability in soil processes and soil dynamics are undoubtedly the latent driving forces behind the chemical composition of the measured soils.However, we specifically highlight that the predictive errors must be directly related to the representativeness in terms of chemical composition of soils and number of samples that were available in respective modeling strategies and regions (see Table 3). Furthermore, information on soil transforming factors such as land use, parent material, and other environmental conditions, which affect the biogeochemical attributes of soils, was already included in the submitted version of the manuscript (see e.g. Table 1 and Table 2).

- 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;

While we appreciate the reviewer’s perspective, we tried to limit chains of citation where possible. We were careful in our selection of references and are confident that each reference we cited for a given statement is suitable. Since infrared spectroscopy is at the boundary of disciplines; it involves interdisciplinary methods that were developed in different fields, e.g. statistics, statistical learning, general soil science, chemistry, physics, chemometrics, pedometrics, electrical engineering (signal processing). In these situations, it is necessary to cite often a series of papers that describe complementary parts of the overall approach and method.

We also checked the SOIL guidelines and there is no limit with regards to the number of references that can be cited together for supporting our statements.

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

We will thoroughly check to make sure acronyms are all defined in the revised version.

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

We thank the reviewer for the comment. The mentioned typos will be corrected and the manuscript will be carefully reviewed for spelling and grammar by a native English speaker.

- 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;

The results clearly show that the presented soil spectral library drastically reduces the need of novel chemical measurements because the new library adds complementary information which improves the trade-off between the amount of classical re-analysis to be done in the lab and estimation accuracy. This is an important step forward in order to enable researchers from developing countries with limited funds to gain data on soil properties without the need of extensive chemical analyses (something that was not possible for tropical Africa before). For many soil parameters Infrared spectroscopy can reach similar accuracies together with traditional laboratory reference measurements. Every method has flaws and errors occur also in wet chemistry analyses (e.g. preparation). If there is considerable uncontrollable variation in the chemical measurements, spectroscopy-based approaches excel at reducing the bias in the measure of interest. The estimation accuracies obtained in the regions using the relevant spectral data and libraries were very close to typically reported accuracy limits for total carbon, for example (for references and more details see comment above).

- Abstract (line 11): AfSIS!?!

Thanks for spotting this acronym standing for Africa Soil Information Service. We will replace the acronym with the full title in the revised manuscript.

- 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;

We agree this sentence was perhaps too vague, however, it is true that there is a general tendency of sparse soil data availability in the humid tropics. We will rephrase the sentence to say more explicitly that there is in particular a lack of soil data for the humid tropics of central Africa.

- 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;

We will replace the reference from FAO and ITTO (2011) by a more recent one.

- 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;

We agree with the reviewer that a quantification is useful and will change the sentence as following:
“*Human populations in Uganda, Rwanda and the DRC are projected to more than double in the coming 80 years (Vollset et al. 2020). Such dramatic growth will likely contribute to further agricultural conversion.* “

- 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?;

We thank the reviewer for this comment. Soil data applies to multiple disciplines in environmental science, ranging from agricultural to soil, biogeochemistry and climate sciences.

- 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;

We thank the reviewer for this comment. With costs we mean the monetary expenses for soil laboratory analyses. In our opinion, this sentence already explains why these costs are low: fast, simple handling, less work, minimal chemical consumables. This further allows high repeatability and coverage of spatial soil heterogeneity, which we will add to the sentence.

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

We are not fully sure at what the reviewer is getting at. The paragraphs elaborates on the benefits of soil spectroscopy including defined, targeted workflows. References are given. No scientific questions were raised.

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

Thanks for this hint, we will repeat the answer to the exactly same question reviewer 1 posed above:

With “positive predictive transfer” we describe the information transferred from a large infrared library for a new calibration of a local set as described by Padrian et al. (2019). The calibration of a new local set using a large-scale spectral library can be complex in soil science, especially when the local set covers a different geographical domain than the library. Soil spectral libraries become particularly useful when a large amount of their relevant information can be extracted in a way that it improves prediction accuracy (positive transfer) and minimizes the number of additional costly local reference measurements for quantifying soil properties in the local set (accuracy-cost trade-off). To avoid technical jargon we will rephrase the paragraph L48-L55 and move it to L60, where it fits better into the context:

“One of the main aims of establishing large-scale SSLs is to minimize the need for future wet chemical analyses (e.g., Nocita et al., 2014; Stevens et al., 2013; Shi et al., 2014; Viscarra Rossel et al., 2016). However, these libraries often span vast geographical areas that include different soil types and climate zones, which comprise complex soil organic C forms and mineral compositions. Due to this heterogeneity, predictions rendered by global linear regression models are often unfeasible for new local soil property assessments at a regional, field or plot-scale, especially when the new set covers another geographical domain than the library. Pandiran et al. (2019) could considerably improve prediction accuracies for a new local set when using a compositionally related subset from a large-scale SSL together with a small number of local reference analyses. The cost-accuracy trade-off can be met when the accuracy of the library-based prediction is similar to the one made when applying a local but more costly calibration strategy. Several data-driven methods have proven to be successful to overcome this issue, for example RS-LOCAL (Lobsey et al., 2017) and memory-based learning (a.k.a local learning e.g. Ramirez-Lopez et al., 2013; Shenk et al., 1997; Naes 1990). In addition, other promising approaches have also been proposed, although they require more research (e.g. deep learning (Ng et al. 2019), fuzzy rule-based systems (Tsakiridis et al. 2019)).”

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

We will update the reference to the newer version, thank you!

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

We agree that we missed to add this important information and will therefore change it.

See our suggested changes under Review 1, *L145-161*

- 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?;

  Highly weathered tropical soils are often acidic (pH < 6) and don’t contain any inorganic carbon and therefore assumptions might be made that total carbon would correspond to organic carbon.

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

  We will add the required information (see reviewer 1, L99)

- 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?;

  This will be changed accordingly (see reviewer 1, L113, L125)

- 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;

  Some cluster areas were excluded because they did not have enough samples to provide reliable results (< 80 samples per region). We agree that this is not clearly presented. For a new version of the manuscript we will these regions from this table. A new table with all the regions will be presented in a supplementary table in the appendix.

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

  The transformation from reflectance into absorbance is not arbitrary. Instead it is based on the Lambert-Beer’s law (please see https://en.wikipedia.org/wiki/Beer%E2%80%93Lambert_law) which dictates that the concentration of the components in a matrix influence the way in which that matrix absorbs radiation. Although this law does not 100 % apply for opaque materials, it serves as the fundamental theoretical basis for quantitative analysis in vibrational infrared spectroscopy and it is the underlying reason why scientists use the calculated absorbance as the starting point for the numerical analysis of their spectra. This is evidenced by countless studies (e.g. Baes and Bloom, 1990; Baharom et al., 2015; Barthès, et al., 2020; Gogé, et al., 2014; Minasny et al., 2013; Peng et. al, 2013). Therefore, since conversion from reflectance into absorbance is considered as elemental in vibrational spectroscopy, we do not see the need to provide detailed justification and references to support this procedure. However, if the reviewer has a particular reference in mind, we would be happy to consider it for citation in our manuscript.

Baes, A. U., & Bloom, P. R.:. Fulvic acid ultraviolet-visible spectra: Influence of solvent and pH, Soil Science Society of America Journal, 54, 1248-1254, https://doi.org/10.2136/sssaj1990.03615995005400050008x, 1990.

Baharom, S. N. A., Shibusawa, S., Kodaira, M., & Kanda, R.: Multiple-depth mapping of soil properties using a visible and near infrared real-time soil sensor for a paddy field, Engineering in Agriculture, Environment and Food, 8, 13-17, https://doi.org/10.1016/j.eaef.2015.01.002., 2015.

Barthès, B. G., Kouakoua, E., Coll, P., Clairotte, M., Moulin, P., Saby, N. P., ... & Chevallier, T.: Improvement in spectral library-based quantification of soil properties using representative spiking and local calibration–The case of soil inorganic carbon prediction by mid-infrared spectroscopy, Geoderma, 369, https://doi.org/10.1016/j.geoderma.2020.114272, 2020.

Gogé, F., Gomez, C., Jolivet, C., & Joffre, R.: Which strategy is best to predict soil properties of a local site from a national Vis–NIR database?, Geoderma, 213, 1-9, https://doi.org/10.1016/j.geoderma.2013.07.016, 2014.

Minasny, B., McBratney, A. B., Stockmann, U., & Hong, S. Y.: Cubist, a Regression Rule Approach for use in Calibration of NIR Spectra, Picking Up Good Vib, 630, 2013.

Peng, Y., Knadel, M., Gislum, R., Deng, F., Norgaard, T., de Jonge, L. W., ... & Greve, M. H.: Predicting soil organic carbon at field scale using a national soil spectral library, Journal of Near Infrared Spectroscopy, 21, 213-222, 2013.

- 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;

This is information given from the AfSIS spectral library, which was previously measured using the standard operation procedure of the Soil-Plant Spectral Diagnostics Laboratory of the World Agroforestry Center. We found it important and therefore added it to the manuscript. The aggregation of 32-co-added internal measurements into one final spectrum per measured replicate in different wells is a strategy proposed by the OPUS BRUKER software (Bruker Optics GmbH, Germany), which is common on different IR spectrometers. Previous internal tests in our lab confirmed that there was no added benefit doing more than four measurements on replicates in different wells, evaluated on the modeled outcome, which is the proper way of testing a measurement protocol. For example, Peng et al. (2014) report that no further prediction improvements were found by increasing replicates beyond 3 replicates, and some even show deleterious effects at excessive number of replicates likely due to higher chances causing excessive scattering. Our samples were finely powdered and have a relatively low spectral variability, and the scattering effects were alleviated by thoroughly testing single preprocessing methods and combinations thereof.

Peng, Y., Knadel, M., Gislum, R., Schelde, K., Thomsen, A., Greve, M.H.: Quantification of SOC and Clay Content Using Visible Near-Infrared Reflectance–Mid-Infrared Reflectance Spectroscopy With Jack-Knifing Partial Least Squares Regression, Soil Science, 179, 325-332, https://doi.org/10.1097/SS.0000000000000074, 2014.

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

We appreciate this perspective, however, we were trying to adhere to the classical stylistic guidelines of SOIL in which results are presented in a “pure” form divorced from discussion and interpretation. We furthermore disagree with the opinion of the reviewer that the results were detached from the context. We clearly document that the central African MIR SSL adds complementary soil information with regard to what is already available in the library of the Africa Soil Information Service. The way we developed the estimation scenarios reflects one of the key practical issues that motivates doing spectral research, namely the fact that we use an existing library and predict understudied regions with it and therefore minimizing additional costs for new soil wet chemistry analyses. These analyses were done with statistically sound methods. The results section follows these strategies and presents our finding in a clear structure. We provide insights into patterns we found, what worked and what not, and above all, we round up our findings with a recipe.

- 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;

Actually, we explain the use of multivariate methods before section 3.1 The first reference to a multivariate approach (within our manuscript) is given in section 2.4 (Spectral resampling and pre-processing) of the materials and methods. Sections 2.5 (Modeling and prediction data) and 2.6 (Predictive modeling) also explain the use of multivariate methods.

Concerning the use of principal component analysis, unfortunately the reviewer does not provide any information, clue or references to scientific literature supporting the claims about “PFA” being “better” for “variability interpretation” than PCA. We assume that with “PFA” the reviewer refers to Principal Factor Analysis (as she/he does not provide the name of the method in full). Unfortunately, we did not find scientific references reporting the convenience of using PFA over PCA in the soil spectroscopy literature. Although we cannot claim what method is best (PFA or PCA) for infrared spectroscopy data (and it is not at all the purpose of our paper), we do know that PCA is a well suited method for the purpose of data visualization (which our only aim for using it). Whether PFA would add some benefit for our data visualization is then debatable.

Please also note that we do not use PCA as data pretreatment, therefore we do not explain PCA as such. Finally we used PCA, as it is the standard method for latent variable extraction and exploration in chemometrics (please see Cordella et al., 2012) and its use can be considered as standard in soil spectroscopy for exploratory analysis and visualization (e.g. Stenberg et al., 2010; Viscarra Rossel and Chen, 2011; Nocita et l., 2013; Sanderman et al., 2020).

We will be very grateful to the reviewer if he/she could share with us scientific literature about PFA in spectroscopy that we could use to consider the use of this method.

Finally, we agree with the reviewer that we could provide more details to the reader on “the implementation” of PCA and will add this information accordingly.

Cordella, C. B.: PCA: the basic building block of chemometrics, Analytical chemistry, 47, http://dx.doi.org/10.5772/51429, 2012.

Nocita, M., Stevens, A., Noon, C., & van Wesemael, B.: Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy, Geoderma, 199, 37-42, https://doi.org/10.1016/j.geoderma.2012.07.020, 2013.

Sanderman, J., Savage, K., & Dangal, S. R.: Mid-infrared spectroscopy for prediction of soil health indicators in the United States, Soil Science Society of America Journal, 84, 251-261, https://doi.org/10.1002/saj2.20009, 2020.

Stenberg, B., Rossel, R. A. V., Mouazen, A. M., & Wetterlind, J.: Visible and near infrared spectroscopy in soil science, Advances in agronomy, 107, 163-215. https://doi.org/10.1016/S0065-2113(10)07005-7, 2010.

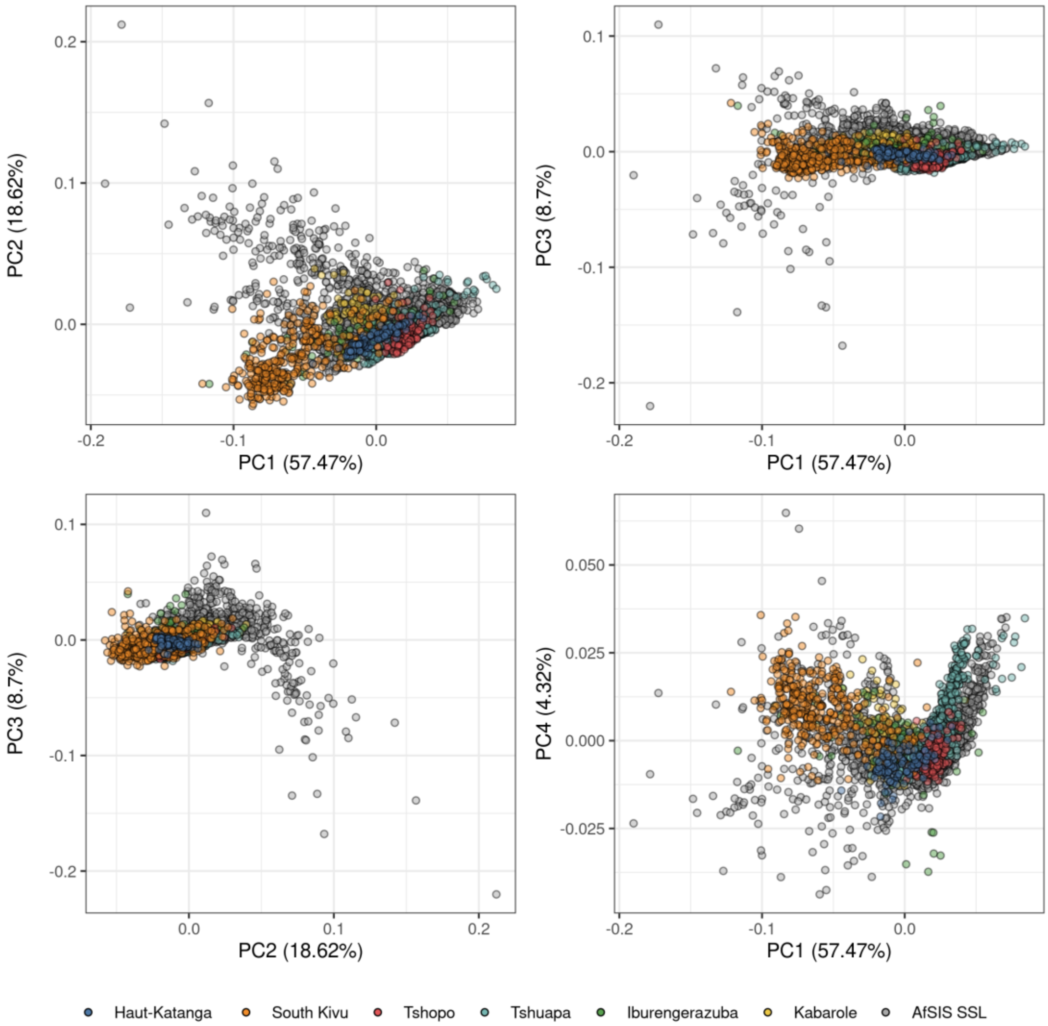
Viscarra Rossel, R., & Chen, C.: Digitally mapping the information content of visible–near infrared spectra of surficial Australian soils, Remote Sensing of Environment, 115, 1443-1455, https://doi.org/10.1016/j.rse.2011.02.004, 2011.

- Line 268: soils rather than “sols”;

We thank the reviewer for spotting this typo.

- 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;

Figure 3 shows the coverage of different PC spaces of the certain regions compared to the AfSIS SSL, which is coloured in black. The first three components explain more than 70 % of the variance in the spectra and therefore showing these three components is adequate to analyse differences between regions. Moreover, the distances in a score space provide a useful tool to analyze similarities/dissimilarities (see review 1 and comments/answers above). We agree the graph can be presented in a simpler and clearer way and will change to a PC1-PC2 and PC1-PC3 plot, as suggested by reviewer 1. Moreover we will add more information on how we performed the principal component analysis in the methods section (see above).



- 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);

We thank the reviewer for this comment but we respectfully disagree. Of course, there is a high pedogenic variability between the soils, however, using the similarity based approach of memory-based learning we overcome this issue. Please find a detailed answer to a similar comment above. In these four lines 266-279 which the reviewer points out, we do not discuss outliers: there was a general trend of underestimation of the predictions (Haut-Katanga, South Kivu, Tshopo, Tshuapa for TC) and (Haut-Katanga, South Kivu, Tshopo, Tshuapa and Iburengerazuba for TN) for all predicted spectra (downwards shift from the 1:1 line in Figure 4). This overestimation was less pronounced in strategy 2 and strategy 3. Outliers, i.e. soil samples with large distances to the continental AfSIS SSL and therefore different in their chemico-physical properties, were removed from these analyses. These samples cannot be accurately predicted by the library and need therefore to be traditionally analysed. We will emphasize this more in depth in the revised manuscript.

- “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;

The reviewer is correct, using these words too often leaves the impression of uncertainty. That was not our intention and we will change this accordingly. However, we are confident of the correctness and robustness of our methods and 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”;

We kindly disagree with the reviewer. For this large scale continental study, these results are accurate with reasonably low prediction errors, especially when comparing them to studies covering similar large geographical areas (see comment and references above).

- 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”;

The reviewer might have misunderstood the methods of our modeling approach. General predictive models are trained with all available calibration data and the new observations are predicted by this “global” model, regardless of the similarity to the observations in the calibration set. As we described in the introduction, line 65, with memory-based learning, a predictive model is trained specifically for the prediction set using a subset of samples in a library based on their similarity/dissimilarity. Therefore we don't see the problem with this sentence.

- Discussion (line 312): ")"…?;

Thank you very much for spotting this typo.

- 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;

We thank the reviewer for this suggestion, however, we followed the guidelines of SOIL. Please see above.

- 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);

We agree with the reviewer that in general, too many acronyms make it hard to follow a text. Nevertheless, we do not think we used too many acronyms in this manuscript. We abbreviated the two spectral libraries (CSSL and AfSIS SSL), the modeling method (MBL, PLS, WA-PLS), statistics ($RPIQ_{pred}$, $RMSE_{pred}$, PCA), the soil properties (TC, TN), a long country name (DRC), spectroscopic specific terms (IR, FT-IR, MIR), which are all very common in soil infrared spectroscopy publications. We will verify the EGU style guide and contact the editor to discuss whether we should add a short overview of the abbreviations at the beginning of the manuscript.

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

We thank the reviewer for this comment, but again, we strongly disagree. The contrary is the case: exactly with these lines as the reviewer points out, we highlight the novelty and importance of our research and results. We establish a soil spectral library with soil samples from the humid central African tropics including forest soils with high organic carbon contents. This area has not been covered by the previously established continental AfSIS infrared library yet (Figure A1) and is still highly understudied. With our proposed infrared library we bring a new soil variability and improve predictions for soil TC and TN (as well as many other soil parameters) for central African regions (for more details please see comments above on a similar question). However, we will add Figure A1 to the main text and rephrase these sentences to make this more clear.

- 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;

  We agree with the reviewer, that we should also include physical properties to the sentences and will change it as following:

  “We conclude that the particularly high soil diversity in these two regions in terms of soil biogeochemical and soil physical properties introduces additional complexity in the soil spectral prediction workflow.”

  However, we kindly disagree with the reviewer about seeing an issue behind this sentence. As already answered in the comment above, this argument points out the importance of our study and our data we contribute to the scientific community. The complexity and differing chemical, biological and physical properties in soils from the Congo Basin will improve future soil analyses for these particular regions and bring new variability to already existing soil spectral libraries (see comments/answers above on a similar question). The positive impact of spiking (reducing $RMSE_{pred}$, increasing $RPIQ_{pred}$ values) underlines this argument. These regions have not been covered by the existing continental library, moreover they can also not be represented by the soils of the other central African regions. Adding the region specific soil properties by spiking (Table 4, Figure 5) has shown to be effective and will also be effective and be improved by the future addition of new data. We acknowledge that this has not been discussed enough in the discussion and will add this accordingly.

- 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;

  We assume the reviewer addresses the lines 340-341 with this comment (instead of the indicated lines). Figure 3 presents the first three components of a principal component analysis of the pre-processed MIR spectra, which cover together more than 70 % of the variance. Therefore, we argue that the 3D visualization of these score spaces is a first indication of differences, in case of large (e.g. mahalanobis) distances. South Kivu (orange) clearly covers a large area differing from the AfSIS SSL and also from the other regions. Also Iburengerazuba tends to spread in the same direction. In our opinion, these larger distances can be used to discuss the performances of the strategies. Spiking had a positive effect on all regions (Figure 5, will be corrected), which can be explained by the addition of closer and more similar samples to the prediction models. We agree that the 3D plot is not appropriate and will change it toPC1-PC2 and PC1-PC3 plots, as suggested by the reviewer 1 (see above).

- 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!?!;

We fully agree with the reviewer that the effect of spiking has to be discussed more in depth. The effect was actually pronounced for all regions. Spiking reduced the $RMSE_{pred}$ for all regions for TC and TN and increased $RPIQ_{pred}$ values. The positive effect of spiking is due to the addition of local samples to the models and therefore adding information of the target region. We will add more explanations to the results and discussion section.

- 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;

We thank the reviewer for this comment. We understand that this sentence is not clear and we will rephrase it as following:

“Six central African regions were predicted for soil TC and TN with sufficient accuracy using the large-scale AfSIS soil spectral library only. The general positive effect of adding geographically closer samples to the AfSIS SSL (strategy 2) underlines the usability of spectral libraries for new regions. The generally positive effect of strategy 3, spiking of all regional predictions for TC and TN with samples from the target area, encourages the future amendment of currently existing libraries to improve prediction accuracy. “

We respectfully disagree with the reviewer, that prediction accuracy did not improve between the three strategies. For a study in this scale the prediction errors were on one hand more than sufficient for most scientific and applied uses and on the other hand, they were considerably improved at least for strategy 3 compared to strategy 1. The accuracy gain is of course relative and there are different requirements on accuracy depending on the interests and the possibilities to invest in more expensive laboratory wet chemistry analyses. We will describe and discuss these trade-offs in more detail to emphasize this change.

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

We thank the reviewer for checking our references attententively. We use citations to confirm our statements where required. We will carefully go through all of them and re-evaluate them to see if we can reduce it to a smaller number. Indeed, there is a problematic tendency in modern scientific writing to only cite the most recent references that often make claims that were established much earlier by original studies. We therefore kindly disagree with the reviewer that the older references would be problematic. Moreover, we only added self-citations that were absolutely necessary. The presented library stems from both soil archives and data collected within different projects, universities and institutes. Most of the sample sets have already been published (Table 1), therefore we find it crucial to cite the original studies. Without these collaborative research and data collection efforts, we could not have created this library.