

Comment on soil-2020-93

Jonathan Sanderman (Referee)

Referee comment on "Quantifying soil carbon in temperate peatlands using a mid-IR soil spectral library" by Anatol Helfenstein et al., SOIL Discuss.,
<https://doi.org/10.5194/soil-2020-93-RC1>, 2021

Synthesis

In this concise paper, the authors have demonstrated that total soil carbon in temperate peatlands, an ecosystem that is underrepresented in the Swiss national database, could be effectively predicted using diffuse reflectance mid-infrared spectroscopy after adding only a very small number of locally representative samples to the larger database. Studies like this that are demonstrating how to best use national spectral libraries to reduce the cost and effort of collecting new soil information are important next steps in the evolution of soil spectroscopy as a routine tool for soil science. While I think this study is a good contribution to the literature on spectroscopy, I think the authors should do more with the data and analyses as I suggest below.

General comments

Why have you chosen to focus on total carbon when the samples have a mix of organic matter and carbonates? There are very few applications of soil information where total carbon is preferred over either total organic carbon or carbonates.

This might be splitting hairs but I don't think the three comparisons in this paper should be presented as three different cases. Rather, the local-only models are being compared to two different ways of using the SSL+spike (either as a global PLSR mode or developing an appropriate subset of the SSL using RS-LOCAL before model building). I bring up this comment because of the way the three modeling approaches are first presented in the abstract and intro. A more detailed explanation of the model choices would likely

eliminate this comment.

Was there a rationale for not also using a memory-based learning approach? If RS-LOCAL can achieve results just as good as or better than MBL, then that is a good argument for the simplicity of RS-LOCAL. Given the diversity in SC content, carbonate content, and soil type in the HAFL dataset, it seems that one subset of the SSL is not going to be as good as subsets specifically built for the individual samples.

The size of the validation sets are changing with increasing number of spike samples. This is setting up a situation where the results are not perfectly comparable in Fig 8. How about restricting your validation set to the 58 samples that are never used in calibration?

I'm struggling with the high bias in the SSL validation results. Could this be due to a limit of 10 components for the PLSR model? With 4000+ samples, it could well be justified to search up to 20 or even 30 components to find the minimum in RMSE. I'm also wondering if you transformed the TC data prior to model fitting (log or square root transform) if that would get rid of some of the curvilinear nature to the fit – Baldock et al. (2013 Soil Research - <https://doi.org/10.1071/SR13077>) found that sq. root transformation really improved model fits. It is very much possible that the bias is real but it would be good to test these two ideas out to see if it helps remove the bias.

I'm not sure section 5.1 is necessary or even really fair to include in this paper. The topic is certainly interesting – can MIR be used to study peat composition? As discussed there is a growing literature doing just this. Where I see a problem is that most of those papers focused exclusively on pure peat soils while the HAFL set of samples compose everything from pure OM to almost pure mineral soil. As your Fig 4 shows, the majority of the variability in the HAFL set comes from this gradient in C content.

Specific comments

Title – Given the focus and brevity of this MS I think a more focused title on applying a national library to peatlands is appropriate. The phrase “efficiently using variation” will not mean much to most readers of Soil.

L8 What do you mean by “organo-mineral diversity”?

L11 “target-feature representations” is jargon that should be avoided in the abstract. Can this be reworded?

L15-19 I found this summary of the results really difficult to digest until I had read the entire paper. Please rephrase and simplify.

L23 "a SSL" – do you mean any SSL or the Swiss SSL here?

L63-64 This sentence was difficult to understand intent. Please rephrase.

L78-82 Can you please spend some time defining in simple terms what memory based learning and transfer learning mean? Few soil scientists will be familiar with these terms.

L110 Change to modelling approaches "for" SC

L115 Shouldn't this be in the methods section?

L125-127 Please rephrase this sentence, it is currently difficult to read.

L130 What method was used for grinding? This would be nice to include given some of the current debate of grinding.

L130 What does it mean to optimize signal-to-noise? Wouldn't we want to maximize S:N?

L138-146 I'd prefer to see preprocessing discussed with the rest of the data analysis

L138 If you tested several pre-processing approaches, please show the results of this testing as supplemental information. It would be interesting to know how much better the SG 1st derivative performed relative to just baseline correction.

L150 What was the TC method for NABO samples?

L163 Was tuning based on RMSE or R2?

L167 Why did you limit PLSR to only 10 components especially when using the entire SSL? This might be a reason for the high bias in the results.

L168 If you ran cubist and don't show the results, then why even mention it here? Given this is a really short paper, I'd like to see the Cubist results included.

L180 Why did you choose to stop at a 50/50 cal/val split? Many studies often use 70/30 and 80/20 splits?

L180 Why was KS used for sample selection and not conditioned Latin hypercube sampling or other techniques? I don't think there is anything wrong with KS but some justification would be good.

L181-183 Are these sentences included to justify not having locals models of less than 15 samples as done with the SSL and RS-LOCAL? If yes, it may be easier just to state that you did not build models of less than 15 samples.

L188 For the SSL and RS-LOCAL models, why didn't you try a no-spike case study? It would be really interesting to know how the SSL stands up with no new local information.

L195 Can you better explain how RS-LOCAL searches for subset K and how you ended up with K being the same for all levels of m ?

L212 A short section describing how you evaluated the results of the different models would be helpful.

Fig 2 I found this figure to be very helpful, thanks for including. I suggest either changing the color or using hash marks for NABO as you cannot see that part of the bar in black and white.

L218 Picking only one region (2000 cm^{-1}) to highlight here seems arbitrary. There are other regions with clear patterns with SC.

Fig 3 & 4 The color ramp is not colorblind friendly and does not reproduce in black and white. There are lots of great colorblind friendly options available at <https://colorbrewer2.org/> and other websites.

L229 Can you add a measure of skewedness to Table 1?

L231 Please tell us how many samples in the Swiss SSL are organic soils. To me there seems to be fairly good coverage of the HAFL data.

L233 There are quantitative ways of assessing differences in PC space such as calculating centroids and hulls. You can also calculate a resemble matrix and then apply multivariate ANOVA analysis to be truly quantitative here.

Fig 5 I suggest using very light symbols for BDM samples not used in validation. Also move these symbols to the back; they currently cover most of the other data making this figure difficult to read.

L240 I'd like to see the RS-LOCAL tuning results presented in the supplement. I'm curious as to how important each parameter was.

Fig 6 & 7 are not very informative given how you are trying to discuss results across increasing number of spike samples. Fig 8 is really the only figure that really matters for interpreting the results presented in section 4.3. I would suggest either moving both figures to supplement or just keep Fig 6 as an example of the different model fits at one level of spiking.

Fig 8 Bias is also in units of %. You might want to mention in the caption that RMSE axis is log.

L282-283 Are you referring to overall correlation or to specific bands with this sentence? It is unclear.

L300 Are you saying these are transmission measurements instead of diffuse reflectance? I'm not familiar with all of these studies but I know the Matamala et al 2019 study used diffuse reflectance. A really interesting study using MIR spectroscopy on peat soils to look at decomposition state by correlating MIR data with NMR data was done by Hodgkins et al (2018 Nature Communications - <https://www.nature.com/articles/s41467-018-06050-2>).

L309 Do you think the strong bias in the spiked SSL models is a common feature or is it specific to your application of an SSL to peatlands? Or is it an artifact of using PLSR

instead of Cubist or other ML models.

L313 You say that you cannot infer if lower SC range is better predicted than the higher SC range but you have the data to do exactly such a test. You certainly have the space to divide out the validation results into low v. high C ranges to better see where the bias and lack of fit is creeping in.

L325-326 This sentence makes it sound like you did exactly what I just asked for in the last comment but did not present the findings. This sentence cannot be in the discussion if you do not present the data elsewhere.

L330-331 In general be careful about making statements about being the first to do something. There are very few truly original ideas in science. In this case, the USDA NRCS KSSL MIR spectral library contains 2000 organic soil samples – see papers by Wijewardane et al. 2018 and Dangal et al. 2019.

L346 – Ramirez-Lopez et al (2019 EJSS - <https://doi.org/10.1111/ejss.12752>) incorporated FTIR prediction uncertainty into digital soil mapping.