SOIL Discuss., referee comment RC2
https://doi.org/10.5194/soil-2022-27-RC2, 2022

Comment on soil-2022-27

Stephen Chapman (Referee)

Referee comment on "Improving models to predict holocellulose and Klason lignin contents for peat soil organic matter with mid-infrared spectra" by Henning Teickner and Klaus-Holger Knorr, SOIL Discuss., https://doi.org/10.5194/soil-2022-27-RC2, 2022

General Comments

Several previous studies have used mid infra-red spectra to assess peat organic matter chemistry, particularly holocellulose and lignin content. One in particular had calibrated the spectral data with a set of chemical analyses using various (non-peat) organic materials. This paper aimed to re-examine the calibration using more refined models and gauge the suitability for applying them to peat and peat vegetation. The authors do a thorough job in applying some sophisticated statistical techniques to the data used by Hodgkins et al. (2018) and show how the calibration models can be improved and what the limitations of the current dataset are. From this point of view, the paper is useful and adds to the precision and accuracy of using infra-red spectra to rapidly assess peat chemistry, with possible application to other organic matter types. A weakness is that the authors have restricted themselves to the training chemical data from the Hodgkins et al. paper which are a curious set of paper, wood and plant leaves. In fact, a prior referee of the Hodgkins et al. paper described the set as being “bizarre” and actually it is surprizing that such reasonable results are obtained from such an unlikely dataset. One would normally expect the training set for peat analysis to be based upon samples of actual peat, from as wide a set of sources as possible. Clearly, the authors have limited themselves to what was available from the previous study and have not engaged in any further chemical analysis. To be fair, this weakness is acknowledged (L462) and the need for further representative training data expressed as the next step. Additionally, given the quite detailed statistics, which use a Bayesian approach, some of the methodology might be made clearer to the more general reader as several concepts are not explained and taken for granted. Also, in places, the English is a little unclear or awkwardly expressed; I have tried to indicate some of the more difficult passages below. Finally, there are quite a few errors in the supplementary information.

Specific Comments

Abstract (L1) I feel there is a need to distinguish peat from soil organic matter. True, peat

is a type of soil (a Histosol) and the organic matter of peat is a type of 'soil organic matter'. However, the general expression of 'soil organic matter' is usually associated with mineral soils. Thus, in L3, Hodgkins et al. set out to predict peat holocellulose and Klason lignin contents, not organic matter holocellulose and Klason lignin contents; indeed, they specifically removed the training samples containing silicates.

L13 "mineral-rich samples" – it is unclear whether this refers to peat or to mineral soil. Of course, it could be both. As mentioned later (L96), some peats can become contaminated with mineral material to varying levels, though these are relatively rare occurrences.

L29 "OM derived from plants and SOM" – better to reword as "SOM and OM derived from plants" (OM derived from SOM doesn't make sense).

L37 "few species" – unclear if this is of wood or some other plants.

L46 "a later study" – need to specify which.

L48 Replace "are" with "were".

L49 Replace "are" with "were".

L51 Replace "for example" with ", for example,".

L56 Replace "of" with "in".

L66 Replace "which" with "in which".

L67 I suspect most readers of 'SOIL' will not be so familiar with the term "beta distribution"; some explanation here is warranted.

L77 The meaning of "prediction domain" should be given.

L79 Replace “on” with “to”, replace “e.g.” with “, for example,”.

L80 The meaning of “no problem-specific correlations” is unclear.

L82 Generally, it is not good to start a sentence with “But”; “However,” might be better here.

L83 The “generality” of the training data set is not so obvious – it is quite wood based, since paper itself is derived from wood. Secondly, it could be argued to be a disadvantage since it overlooks any potential diagnostic peculiarities found only in peat.

L97 Replace “predictions” with “prediction”?

L98 Replace “also for” with “that includes”.

L99 Of course, as mentioned above, it would have been better to have a set of genuine mineral soil samples.

L114 “The data are available…” – delete this; it is obvious.

L117 Replace “SOM” with “peat OM”.

L122 “informative priors” – again this term may not be readily understood and should be made clearer. Secondly, what exactly these priors are should be made more specific.

L131 Replace “MCMC” with “Markov Chain Monte Carlo (MCMC)”.

L139 Is there not a third way? Surely (holocellulose + Klason lignin) ≤ 100%? In fact, it may be < 100% if there are other extracted components not included (I am not sure if this might be the case, since I do not have ready access to the De la Cruz et al. (2016) paper, describing the actual chemical analysis procedure). It is not clear if this constraint has been included within the statistical analysis.

L140 Replace “in” with “within”.

L142 Replace “covering” with “covered”.

LL154-155 “in the form of depth profiles” – this is unclear; are you referring to actual peat depths here? Perhaps needs rewording.

L165 Replace “amont” with “amounts”. The term “regularization” may need some defining. Replace “for” with “of”?

L168 It might have been interesting to compare the approach given in 2. to a PLS (Partial Least Squares) analysis, which has been applied to similar data (see Artz et al.) and is a standard (non-Bayesian) approach for such multivariate data.

Figure 2. The key would be better expressed as “Bayesian models â□i Beta â□i Gaussian”, since both are Bayesian. The sentence “ Dashed grey…” could be omitted; it is pretty obvious.

L276 Define the “fingerprint” region (it could be shown in Figure 1).

Figure 3. Under “Holocellulose (no minerals)”, the blue line at ca. 1590 (negative coefficient > 0.2) is unmarked – should it be?

Table 1. The column “Original model?” is not helpful, more confusing – omit (given in legend, whether .2 or .3).

LL304-305 This is really unsurprising.

LL322-323 Again, this points to the training dataset not being suitable for the task in hand.

L360 “classes” – this is a new term which has not been defined. We are left to guess what these classes are. They are given in Supplementary Figure 11, but at this point we are yet to read this.

L395 “because they interfere less with” – no, the reverse applies: “because they are interfered less by”.

L396 Replace “that” with “why”.

Figure 6 “If the points are…” – I think you can omit this; it is a rather trivial remark and obvious to all.

L403 The ELPD values quoted seem to be the same – is this correct?

L422 Replace “to predict” with “being predicted”.

L423 Replace “on” with “to”.

L427 Replace “on than” with “to than”.

L429 Replace “what” with “as to what”.

L439 Replace “problem is that also” with “problems are that”.

L442 Replace “probably ca” with “can probably”. However, I think “probably” is inaccurate here; I think you say “certainly”.

L444 “calibration transfer” needs definition or more explanation.

L450 I suggest replacing “SOM” with “actual peat samples” – the idea of application to SOM (in general and including mineral soils) comes two sentences later. Hence in L451 I suggest omitting “SOM and”.

L464 “interpreted with caution” – it would be useful to develop this thought more. The

estimates given in Hodgkins et al. have been applied to give far-reaching conclusions regarding peat carbon storage, latitudinal trends and possible responses to climate change. Using the same basis, ideas have been amplified and extended in the Verbeke et al. (2022) paper. Can we still rely on the overall findings from these papers or are the conclusions now in doubt? I suspect that the revised models given here, while improving the accuracy of determinations to some minor degree, will not change the overall pictures presented in these (and perhaps other) derivative papers but some discussion on this would be very helpful and would give a wider context and application.

L466 Delete ", also in practice".

Supplementary Information

A general comment on the information here: apart from the introductory paragraphs to S1 and S2, there is a description of the figure followed by the legend to the figure and often the material is a repeat. Really, it would be better just to have a descriptive legend.

S1 Replace "as baseline" with "as a baseline".

Figure 3. See same comment as for Figure 2 above.

Figure 4. The axis labels for (a) are the wrong way round! For "Sample type" is this the same as "class" in the main text? If so change to "class" (but see on Figure 11 below).

Figure 6. Replace "overestimation" with "underestimation" and "underestimation" with "overestimation".

Figure 8. I would suggest having Training, Peat and Vegetation across the top and peak heights down the side (as laid out in Figure 7). Also (a) and (b) seem to show the same data; can this be right? In the legend it says peat (second row) but it is actually peat in the third row.

Figure 10. Replace "6.31e-16" with "0".

Figure 11. Replace "type 2" with "class 2". It is unclear what these classes are, and where does paper and cardboard fit in? Perhaps indicate in the list of sample types which are

class 1, 2, 3, etc.

Figure 12. The order of regressions mentioned in the description is not the same order as actually depicted. In the legend, replace “binnes” with “binned”, “50” with “20” (I presume), and “line” with “lines”.

Figure 13. The “Dataset” key could be omitted – it is superfluous.