

Interactive comment on “Identifying and quantifying geogenic organic carbon in soils – the case of graphite” by Jeroen H. T. Zethof et al.

Anonymous Referee #2

Received and published: 23 July 2019

This is a very well written manuscript that reports a methods comparison for the identification/quantification of graphite in soil. The writing is excellent, and the reporting of results is clear. But the study have a few important shortcomings that I'd like to see addressed before I would deem it acceptable for publication.

1. The manuscript does not provide sufficient context for focusing exclusively on graphite. The reader might interpret the current rationale as a narrow justification for using graphite in the experiment. I understand that graphite may form from metamorphic processes, but is the goal of the study to quantify graphite specifically, or geogenic organic C more broadly? The former seems far too narrow a prospect given the wide range of forms of geogenic C, and the latter is underdeveloped in the study. Without further elaboration, graphite seems a little too specific. Galy, Hemingway and others

C1

are not specific when they refer to “lithospheric” C, Ussiri (cited in the manuscript), Chan (2017, *Thermochimica Acta*) and others have targeted coal, and there is a large and growing literature on pyrogenic C in soils. How would the presence of these affect results? Can graphite be distinguished from other forms of thermally recalcitrant organic C? The authors do well to distinguish between carbonates (which varies widely among dolomite, calcite, etc.) and thermally recalcitrant C, but have not adequately elaborated on graphite versus other forms of geogenic C, let alone pyrogenic C. The distinction between the latter two is obvious using ^{14}C , but the issue here is among geogenic C forms.

2. As a result of the discussion above, and more generally for a method development study, I found the number and range of materials used too small. Only one graphite-containing natural soil and one carbonate-containing natural soil were used for validation. The artificial soil mixtures were made from one OM source far removed from the natural soil (Germany vs. Spain), and “neat” mineral specimens (quartz, muscovite, CaCO_3 (not dolomite or calcite?)). While I understand the desire to create a reductionist, simplified system for initial testing, the result is only nine samples of one mixed-matrix to generate the calibration. This is a highly undersampled relationship. This is critically important because the authors are right to highlight matrix effects, but they do not adequately account for these in the design of the calibration/correlation study. CaCO_3 is not dolomite or calcite, soil minerals are often interstratified, and graphite likely exists in a mineral-associated form fused to the mineral matrix. None of these incipient properties of soil are accounted for in the method development – making the results of limited value. It is a nice proof of principle, but the study needs to go well beyond this given the current state of the literature. I don't see how this study substantially move us further than some of the studies cited within it.

3. The authors identified one of the challenges of thermally distinguishing forms of C as the determination of threshold temperatures. However, the description of how TGA data was processed is inadequate. Phrases like “models were created from the

C2

calibrations sets" and "the best temperature limits for quantification of graphitic C in the calibration sets was determined" are not reproducible. These steps may be the most critical step of the process, but even the most experienced expert in this field would be unable to verify and repeat it. More detail is required here. There are also no details provided on how these data were used in the calibration. The selection of different threshold temperatures somewhat undermines the "smart combustion" approach if it cannot be universally applied. Perhaps there is some elaboration required in the discussion about how much control over threshold temperature there is available with such an instrument, and whether the DIN methods are suitable/adequate standards.

4. I would argue that since this is such a key/core component of the study, that it deserves its own separate subsection within the methods (ie, statistical analyses). I'm not 100% sure that my comments here will be relevant or correct because it was difficult to follow precisely how each of the calibrations were generated. But if I read it correctly, the authors appear to use different methods to generate the calibration curves depending on the quantification method. PLS was used for the FTIR method of quantification, not sure how it was done for what data from TGA, and Pearson correlation for the "smart combustion" data. In all cases, the independent variable should be clearly identified – presumably the "known" quantities of graphite in the mixture (though this should be verified using total C analyses). The dependent variable should also be clearly identified. Lastly, while the calibration/regression lines are shown with error envelopes, none of the data points have errors associated with them. Were the analyses replicated at all? I found no mention of these. Clearly, each of the analytical methods has instrument/analytical error associated with them. How were these accounted for in the study?

4. The methods chosen for comparison were not especially the best available or most appropriate given the current literature. It almost seems as though there was a foregone conclusion that "smart combustion" was going to be the best and the need was to validate these. However, as the authors correctly point out, the use of "smart combus-

C3

tion" doesn't solve the problem of selecting a threshold temperature for distinguishing forms of organic C because it either uses the "wrong" temperature or at a minimum uses the same threshold for all samples. The selection of FTIR was intriguing since the presence of interference bands is well documented. I thought there might be a better spectral method (NIR, MIR, Raman, etc.) that would be better suited to the task. Similarly, the use of TGA has well documented shortcomings in that mass loss reactions are not all attributable to organic C combustion/pyrolysis. In fact, the only method that directly quantified carbon in the current study was the "smart combustion". While the authors discuss the possibility of combining methods, they seem to have missed the opportunity for using EGA during ramped heating - which is essentially what "smart combustion" is.

Interactive comment on SOIL Discuss., <https://doi.org/10.5194/soil-2019-30>, 2019.

C4