I have read with interest the draft untitled "Identifying and quantifying geogenic organic carbon in soils – the case of graphte". Overall, I have found that the draft is very clear. To be published in SOIL, I consider that the authors should provide a proper description of the soils they used in the study. I have also several (rather minor) concerns and questions that have to be answered before I can recommend the publication of this draft in SOIL.

Dear Reviewer,

First of all, we want to thank you for your time reviewing our article, your kind comments and your effort helping us improving the work. We will address your concerns and answer your questions, point by point.

- *I consider that the authors should provide a proper description of the soils they used in the study*

Thank you for addressing this important point. We indeed forgot the inclusion of a proper soil description. Hereby the revised part of the materials and method section:

"Top soil and fresh rock samples from a nearby outcrop, were taken from a field site in Rambla Honda, Sierra de los Filabres (37°07'43" N, 2°22'30" W / Southern Spain). The area is located in the Nevado-Filabride complex and contains Devonian-Carboniferous slaty micaschist with graphite and garnets crossed by abundant quartz veins (Puigdefábregas et al., 1996). Carbonates found in the soil sample (0.18 % C) originated from pedogenesis and dust deposition as the parent rock does not contain carbonates. Soil material was taken from the topsoil (0-5 cm, without sieving crust) under the grass tussock *Macrochloa tenacissima* to ensure a substantial amount of OC was present. The soil itself could be classified as Skeletic Leptosol (loamic colluvic) following the World reference base for soil resources (2014). Additional soil material was collected from a field near the town of Alboloduy (37°04'09" N, 2°36'43" W), hereafter referred to as AB soil, with similar vegetation and (climatic) conditions. The lithology consists of feldspathic mica schist (IGME, 1979), but without natural graphite and with a relative high CaCO3 content (1.87 % C). The AB soil was also sampled from the topsoil, without sieving crust, under the grass tussock *Macrochloa tenacissima*. According the World reference base for soil resources (2014) the soil could be classified as Skeletic Leptosol (loamic). The soil samples were dried at 40°C and sieved to ≤ 2 mm." Original p. 4, line 27 – p. 5, line 2

1) *Graphitic C can be found in rocks. Depending on P and T conditions experienced by the sediments, we do not get necessarily pure graphite. I am therefore wondering if the graphite standard material (Merck) is similar to the graphite found in the soils developed in micaschists. I would appreciate to see Raman signatures of the graphites used in the study. We can also imagine that some graphitic C with lots of defaults may evolved before the final oxidation step. In this case, such graphitic C would not be recovered in the ROC fraction. Can the authors discuss or rule out this hypothesis?*

It is right that pressure and temperature conditions are important factors determining the degree of graphitization. From literature we have an estimation what the conditions have been for the sampled rock, but the degree of graphitization of the standard is not provided by the manufacturer. We arranged Raman spectra (see Figure 1), which we will include in the revised version to shed light on potential differences between the standard and natural graphite as found in the rock and soil sample. As both the graphitic schist and the standard show a highly similar pattern, it can be assumed that they have a similar state of graphitization. The D1 (~1350 cm^{-1}), G (~1580 cm^{-1}) and D' (~1620 cm^{-1}) peaks indicated in Figure 1, can all be attributed to graphitic C, whereby the ratio between the D1 and the sum of all three peaks is a clear indication for the degree of graphitization (Beysacc et al., 2003; Ferrari, 2007). For the Standard (Merck) the ratio is 0.20, while for the graphitic schist we

obtained 0.34, which are both indicating well organized carbon (< 0.5, Beysacc et al., 2003). No peaks are observed around the 1200 and 1500 cm⁻¹ bands, including the soil sample, which would have indicated the presence of pyrogenic / black carbon components (Sadezky et al., 2005, Schmidt et al. 2002). The G-peak for the soil sample (calibration 1) coincides with the D' peak, as could be expected in a sample with different carbonaceous substances, but can also indicate defects in the crystalline graphite structure formed by weathering.

We will include the next paragraph to the method section together with the Raman spectra as in Figure 1:

"Raman spectra were made of the soil of calibration set 1, the standard (Merck) and graphitic schist, using a Thermo Scientific DXR Smart Raman Spektrometer, with 532 nm laser and a power output of 9 mW. Before the measurement, sample was pressed in aluminium cups. Peaks obtained were integrated using Lorentzian profiles fitting in Origin 2019.

In Figure 1 it can be seen that the spectra of the graphite standard (black) was highly similar to the graphitic schist (red). The D1 (~1350 cm⁻¹), G (~1580 cm⁻¹) and D' (~1620 cm⁻¹) peaks indicated in Figure 1, can all be attributed to graphitic C, whereby the ratio between the D1 and the sum of all three peaks is a clear indication for the degree of graphitization (Beysacc et al., 2003; Ferrari, 2007). For the graphite standard (Merck) the ratio is 0.20, while for the graphitic schist we obtained 0.34, which are both indicating well organized carbon (< 0.5, Beysacc et al., 2003). No peaks were observed around the 1200 and 1500 cm⁻¹ bands, including the soil sample (blue, Fig. 1), which would have indicated the presence of pyrogenic / black carbon components (Sadezky et al., 2005; Schmidt et al., 2002)"
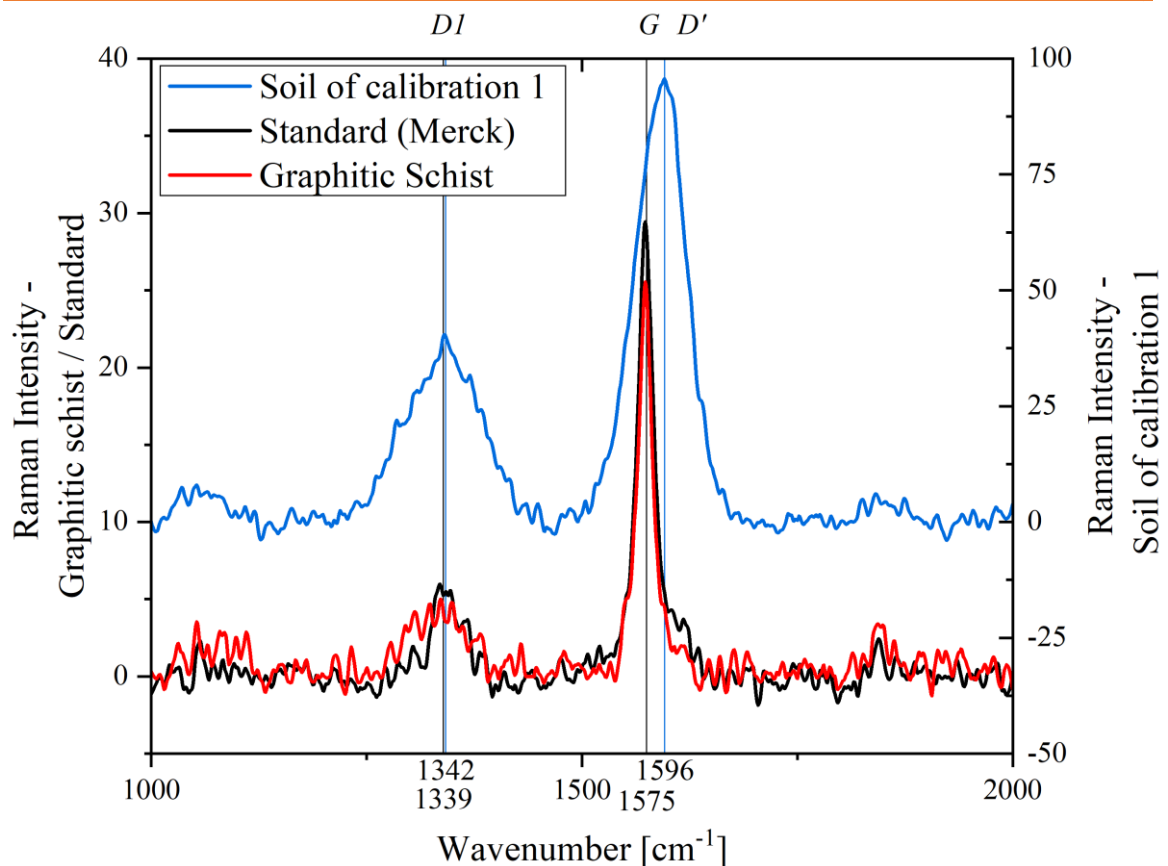
Figure 1. Raman spectra of the graphite standard (Black), graphitic schist (red) and soil of calibration set 1 (i.e. natural graphite containing soil, Blue). Vertical lines indicate the peaks for amorphous carbon (1342/1339 cm⁻¹) and peaks for graphitic carbon (1575 cm⁻¹ standard/schist and 1596 cm⁻¹ for soil of calibration set 1). Indicated are the D1 band (1350 cm⁻¹), caused by plane defects and heteroatoms in the carbon structure, G (1580 cm⁻¹), crystalline carbon i.e. pure graphite, and D' band (1620 cm⁻¹), caused by disordered graphitic lattices.

Some graphitic C, especially with lots of defects or impurities in its mineral structure, might indeed evolve before the final oxidation step of the smart combustion method, resulting in an underestimation of the graphitic C content of the soil as it is not taken into the ROC fraction. We also hypothesized in the discussion that radicals, released from other minerals by temperatures of 700°C and higher, might induce graphite evolution under anoxic conditions (page 11, Line 28-33). According to the measurements with the smart combustion method about 6% of the total C was lost in samples of quartz + graphite standard (i.e. Calibration set 2), while the Graphitic schist lost 2% of the total C, although the graphite standard had a higher structural organization (lower ratio, as discussed above). We will extend line 28-33 (page 11) of the discussion to increase clarity about the loss of graphite before the final oxidation phase and include a Figure (Figure 8c below) visualizing this loss as follow:

"As shown by Hayhurst and Parmar (1998), very small impurities in the graphite can cause a small part of the graphite to pyrolyse during anoxic conditions at higher temperatures. Graphitic C of lesser graphitization, might therefore result in a larger loss of graphitic C during pyrolysis and a greater underestimation of the graphitic C content. Taking a closer look at the measurements of the artificial soil, reveals that a small part of the graphite started to oxidize under anoxic conditions (Fig. 8). The measurement of graphite in quartz, as in calibration set 2, showed that about 6% of the total carbon was lost during the pyrolysis phase, while for the graphitic schist this loss was 2% (Fig. 8c), resulting in an underestimation of the graphitic C content. Bews et al. (2001) suggested that at temperatures higher than 700 °C, radicals like HO_2 and OH might act as reactant with the pure C. Furthermore, in the method comparison study for recovering different types of black C, Roth et al. (2012) suggested a (relatively) strong catalytic effect of oxides on black C oxidation, which was most predominant in soils." p.11 line 28-33
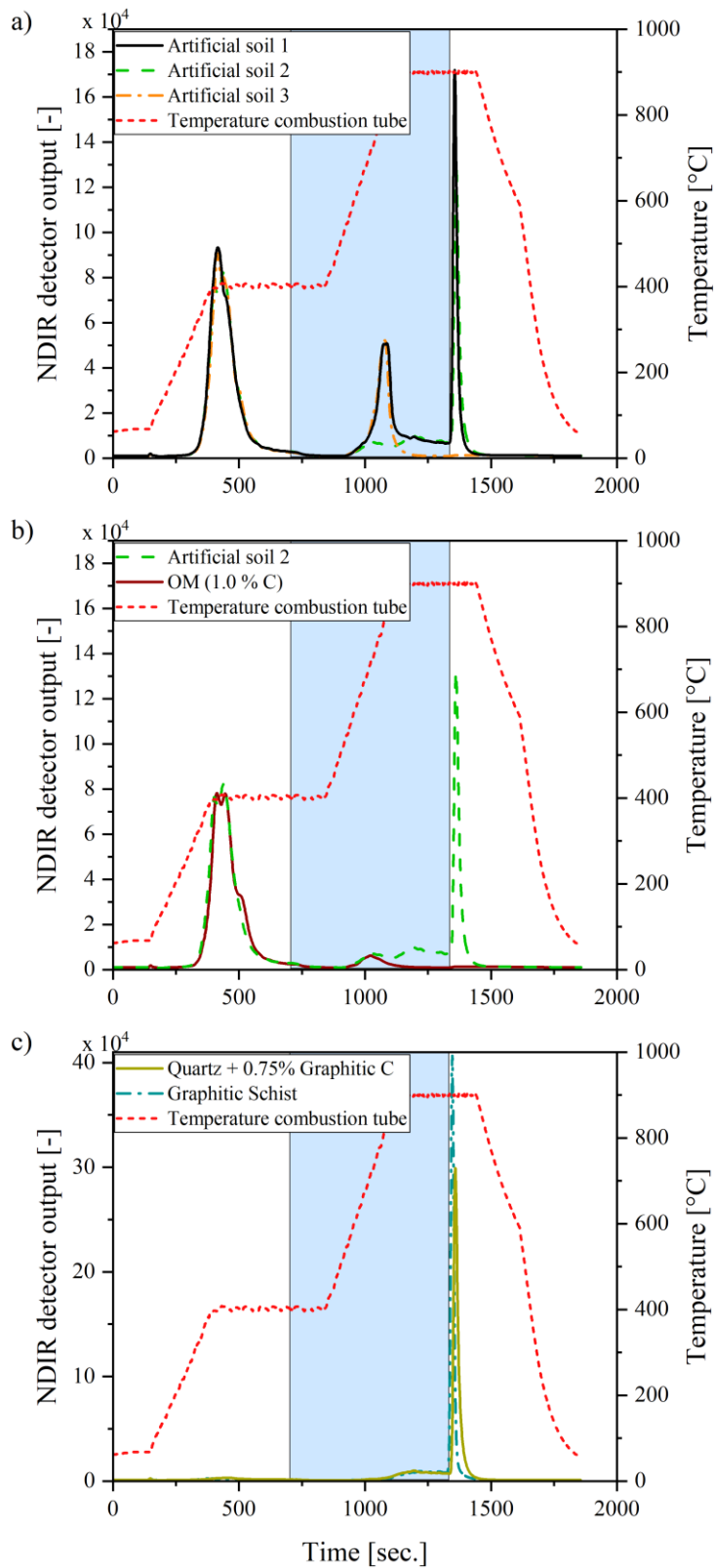
Figure 8. Examples of smart combustion measurements of the artificial soils (a), the fresh OM component (b), the graphite standard and graphitic schist (c). The blue area delineates the part where O_2 is substituted for N_2 and the temperature program is displayed by the red dashed line. Note that artificial soil 2 (green) is without $CaCO_3$ and artifical soil 3 (orange) is without graphite.

2) *2◦ Pyrogenic C (pyOC) can also resist to high T under anoxic conditions. In this case, some pyOC may be recovered in the ROC fraction. What would happen if the studied soil contains both graphite and pyOC? It may have been interesting to add charcoalin the tested mixtures. If the presence of pyOC is a limit to the method, it should be discussed.*

We understand from your question that we did not discuss sufficiently the potential interference of thermally resistant OM (like pyrogenic C / black C). Therefore, we will elaborate more on this topic in the discussion section, as showed below. Including other forms of thermally resistant C fractions and examine how we could distinguish between them would be a very interesting topic for further investigation. Nonetheless we can expect, with the current settings, that pyrogenic C might end up in the ROC fraction of the smart combustion method.

"When the sample contain other forms of thermally resistant OM or even black carbon, which are not pyrolyzed during the anoxic phase, this C component is likely to end up in the graphitic C fraction with the smart combustion method. Especially as most temperature boundaries are empirically derived (Pallasser et al., 2013; Ussiri et al., 2014), a pre-test with continues heating under oxic conditions, is therefore recommended to get an idea which/how many substances are present in the sample. According the Raman spectra (Fig. 1), no indications were found for the presence of black C in the soil and rock samples, as it should have created peaks / increased Raman intensity around the 1200 and 1500 cm^{-1} bands (Sadezky et al., 2005, Schmidt et al. 2002). Further studies should focus on temperature boundaries of different substances in relation to their properties and see how for instance graphitic C can be distinguished from other thermally stable C components." After original p.12, line 11

3) *10 samples for calibrating a model is definitely a too low number of samples. It would have been highly surprising to get nice results with such a low number of samples. It may have been interesting to use all the samples to design calibration models. We can't exclude that with a nice sample set containing 500 samples with known graphite concentrations, a convincing IR-based model can be designed.*

We agree that more samples could potentially improve the model. On the other hand, it is frequently shown in the literature that the performance of IR spectroscopic models for predicting soil properties increases with sample set homogeneity (e.g., Grinand et al., 2012), i.e., calibration and validation become more precisely when focussing on samples from similar or identical sites and soil matrixes. Here, we like to point out that two representative matrix substances were used as calibration samples: quartz sand and the soil of interest (soil 1), which was later used for validation. The R^2 and RMSEP of the calibrations were quite sufficient (R^2=0.96 and 0.99; RMSEP=0.24 and 0.10). The samples with the unknown graphite concentrations were of exactly the same matrix (Quartz, soil 1). So, the models we used were very specific in addition to the high R^2 and low RMSEP. Since further samples are not available, we calculated a model including both calibration data sets, soil 1 and quartz. This PLSR model used 3 components and an R^2 of 0.96 and an RMSEP of 0.24 (Fig. 2). These values were at the same level as found for the single models (see above). Nevertheless, all models substantially overestimated the graphite content. Therefore, we do not think that the use of more samples of different origin would improve the prediction / validation. Against the backdrop of the literature (specific graphite absorption bands that have been reported in the literature are only valid for oxidized graphite), the failed predictions of the graphite contents were plausible. We modified the discussion correspondingly, however, - if you agree - we do not intend to add the Figure (here shown as Fig. 2) to the text:

 "The calibration between infrared spectra and graphite contents of the calibration sets yielded promising results (Figs. 1a and 1b) and could also be used for a cross-validation (Fig. 5). Although the same substrate materials and similar contents of graphitic C were used in the validation, the
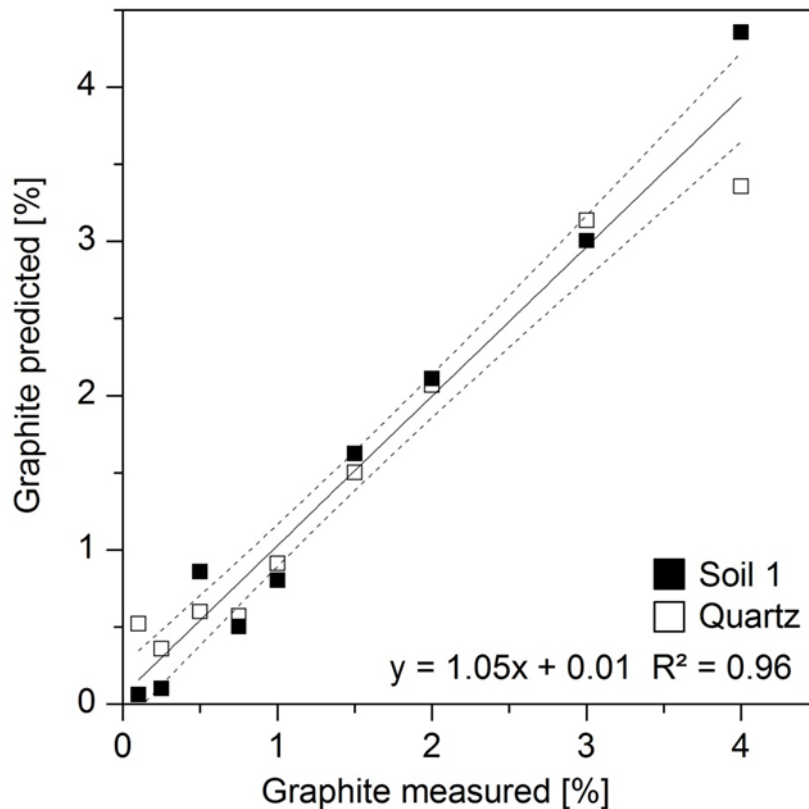
Original line p.9, 17-20

Figure 2. Prediction plot of the PLSR model using a joint dataset of soil 1 and quartz.

4) *The authors hypothesized that ROC content would match graphite content. It is not too far but not perfect. Why don't the authors try to design a model based on ROC results as they did with IR and TGA results?*

Thank you for the suggestion to create a model for correcting the ROC value. We considered this as well, but, as can be seen in supplementary Figure 4 of the manuscript, the total carbon measured with the smart combustion method (i.e. the Soli-TOC device) is the same as what is obtained by the more accurate elemental analyser. Therefore, the slight underestimated graphitic C content results from the differentiation between carbon fractions by the temperature-oxidation program and not as a results of the direct output of the sensor. Furthermore, correcting the ROC/graphitic C value using the calibration set would not improve the graphitic C estimation as a model build with calibration set 1 would result in a slight overestimation of calibration set 2 (notice slope > 1.00) and vice versa (see Figures 3 and 4, here below). This difference in underestimation of the ROC fraction was attributed to impurities in the graphite and/or presence of radicals, as discussed in the second paragraph of section 4.3. Especially the presence in radicals will differ from sample to sample, as it depends on the matrix composition, which would give the same matrix issue in the model creation

as with the FTIR and TGA methods. Therefore, we suggest keeping the ROC content as it is derived by the device and clarify this decision in the method section by adding:

"The Soli-TOC device directly converts the NDIR signal in the C content of the different components, as calibrated with CaCO3. Creating an additional model to correct the C output, is introducing an additional error in the measurements. Therefore, we analysed the direct C output, as measured in the ROC fraction. Triplicate measurements were averaged, whereby the average coefficient of variation between replicates was 2.7%, and a Pearson correlation test was performed between the obtained ROC data and calibration sets to analyse how well the graphite content was measured." Original p.7, line 5-7
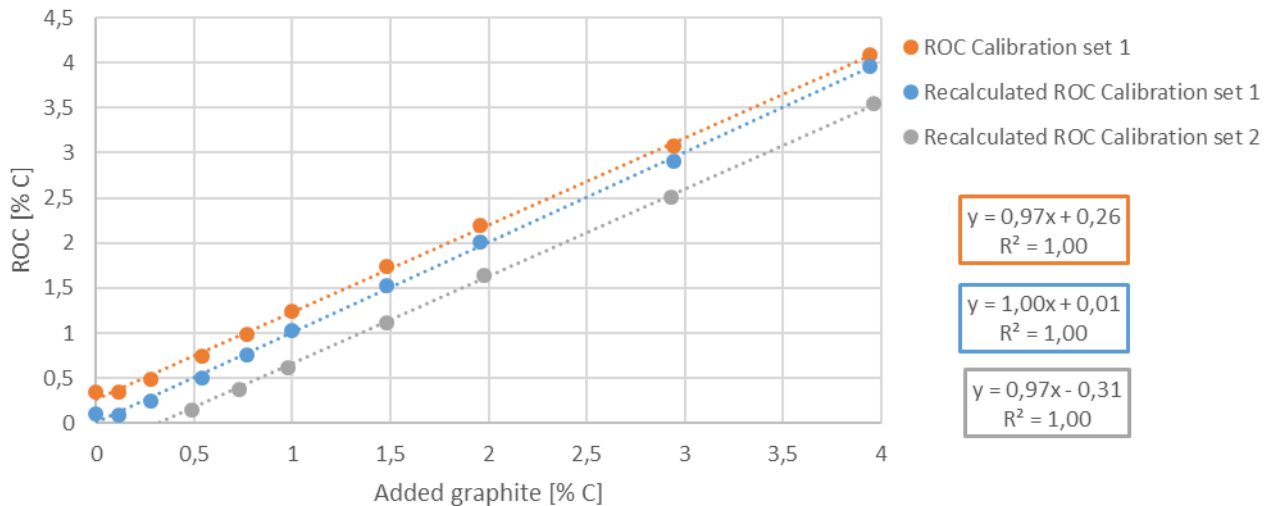

Figure 3. Correction of the ROC values by using a model based on calibration set 1 (graphitic soil with added graphite). In orange the original measured graphitic C content (ROC) of calibration set 1 is plotted against the added graphite. As can be seen by the linear trend line, the graphitic C content is originally overestimated. A simple correction model was created, resulting in an exact estimation (Blue), but the same model resulted in an underestimation of graphitic C in calibration set 2 (Grey, Quartz with added graphite).
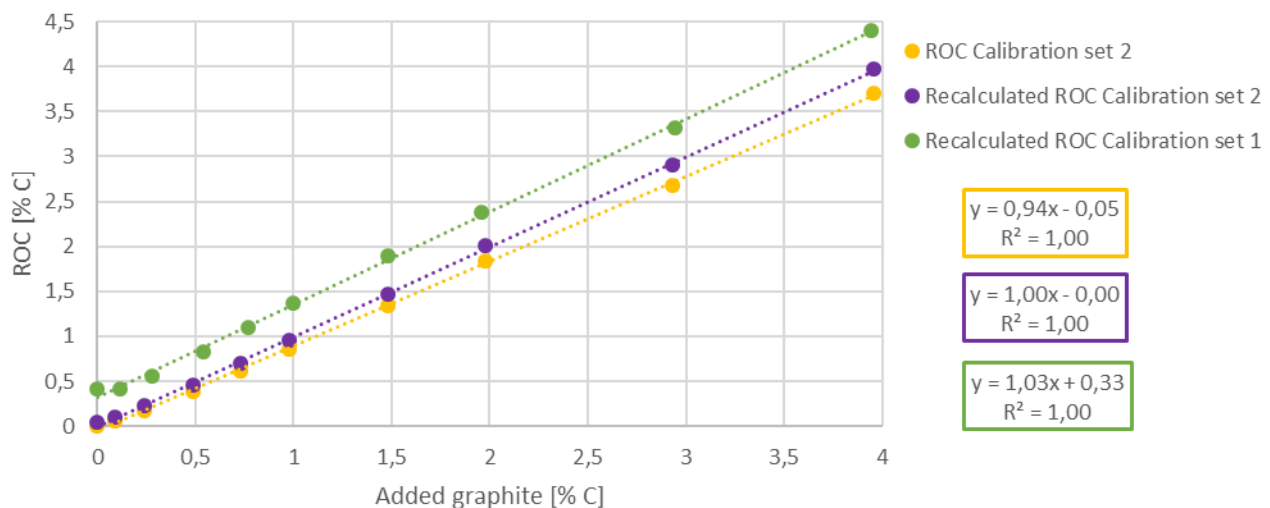

Figure 4. Correction of the ROC values by using a model based on calibration set 2 (Quartz with added graphite). In yellow the original measured graphitic C content (ROC) of calibration set 2 is plotted against the added graphite. As can be seen by the linear trend line, the graphitic C content is originally slightly underestimated. A simple correction model was created, resulting in an exact estimation (purple), but the same model resulted in an overestimation of graphitic C in calibration set 1 (Green, Quartz with added graphite).

5) *I do not understand the Figure 5. I suggest improving the explanations on this Figure or removing it*

We hope that Figure 5 in the manuscript, which summarizes the results of the three tested methods, becomes clearer with this extended Figure description:
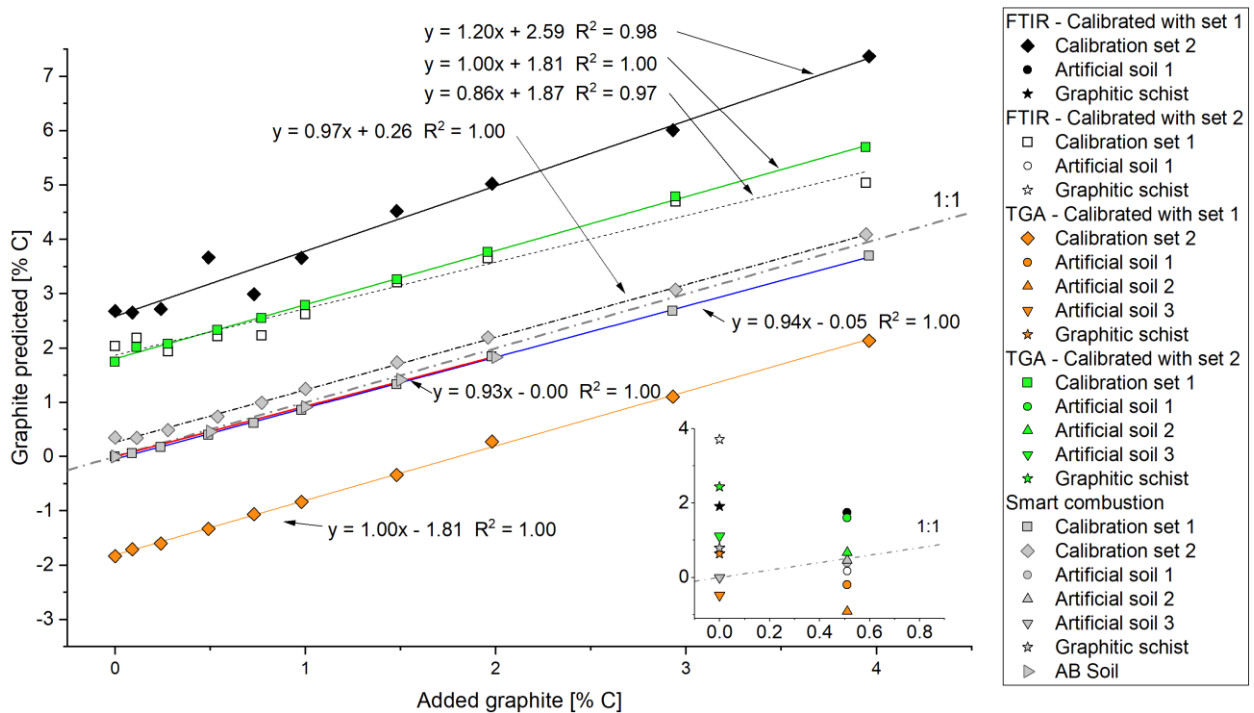
"Figure 5: Overview of the predicted amount of graphite in the calibration sets (squares/diamonds), artificial soil (inset, circles/triangle), graphitic schist (inset, stars) and AB soil (right pointing triangle) as measured with the different methods. Black symbols: graphite prediction by FTIR, model from calibration set 1; White: FTIR, calibration set 2; Orange: graphite prediction by TGA, model from calibration set 1; Green: graphite prediction by TGA, model from calibration set 2; Grey: graphite prediction by smart combustion. Exact data is given in Table 2."

References

Beyssac, O., Goffé, B., Petitet, J. P., Froigneux, E., Moreau, M., Rouzaud, J. N. 2003. On the characterization of disordered and heterogeneous carbonaceous materials by Raman spectroscopy. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 59(10), 2267-2276.

Ferrari, A.C. 2007. Raman spectroscopy of graphene and graphite: Disorder, electron–phononcoupling, doping and nonadiabatic effects. Solid State Communications 143, 47–57.

Grinand, C. ,Barthès, B.G., Brunet, D. , Kouakoua, E. , Arrouays, D., Jolivet, C., Cariac, G., Bernoux, M. 2012. Prediction of soil organic and inorganic carbon contentsat a national scale (France) using mid-infrared reflectancespectroscopy (MIRS). Eur. J. Soil Sci. 63, 141 - 151.

Jorio, A. and A.G.S. Filho 2016. Raman Studies of Carbon Nanostructures. Annual Review of Materials Research 46, 357-382.

Sadezky, A., Muckenhuber, H., Grothe, H., Niessner, R., Pöschl, U. 2005. Raman microspectroscopy of soot and related carbonaceous materials: spectral analysis and structural information. Carbon, 43(8), 1731-1742.

Schmidt, M. W., Skjemstad, J. O., Jäger, C. 2002. Carbon isotope geochemistry and nanomorphology of soil black carbon: Black chernozemic soils in central Europe originate from ancient biomass burning. Global Biogeochemical Cycles, 16(4), 70-1.

Sparkes, R., Hovius, N., Galy, A., Kumar, R.V., Liu, J.T. 2013. Automated analysis of carbon in powdered geological and environmental samples by Raman Spectroscopy. Applied Spectroscopy 67, 779-788.

World Reference Base for Soil Resources 2014. IUSS Working Group WRB. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106. FAO, Rome.