

Interactive comment on “Continental-scale controls on soil organic carbon across sub-Saharan Africa” by Sophie F. von Fromm et al.

Anonymous Referee #2

Received and published: 23 November 2020

General comments:

The manuscript “Continental-scale controls on soil organic carbon across sub-Saharan Africa” describes a continental-scale analysis of associations between soil organic carbon and soil physico-chemical properties across Africa. The manuscript outlines a novel soil dataset collected at the Afsis “sentinel sites”, and then steps through several statistical analyses that tease apart associations between carbon, extractable metals, and soil exchange pools across different domains of climate, soil pH, and soil weathering status. The authors conclude that short-range order (oxalate extractable Al) and to an extent Fe explain much of the variation in carbon stocks in wet/acid soils, whereas exchangeable calcium explains much of the variation in dry/alkaline soils. Soil texture and land use appear largely irrelevant at this scale.

C1

I think this manuscript is excellent and will be a very useful contribution to the study of soil geography. While the primary result has been identified in earlier studies (particularly Rasmussen et al.'s 2018 study), this manuscript applies to a different geographic domain (tropical and subtropical Africa) and with a more systematic data collection effort. It also considers soil weathering status using total elemental inventories and chemical weathering indices, which adds novelty. The results provide clear confirmation of the patterns hinted at in the Rasmussen study, and also point to some new complexities (particularly in relation to Fe). Furthermore, this study applies to data that were collected in a systematic sampling effort—hence these results should be considered more conclusive than those in earlier studies. The manuscript does a good job of balancing different statistical approaches, and stands as an example of how data-driven modelling tools (i.e. random forests) can be used responsibly in a process-oriented way to complement more traditional statistical approaches. While at points the interpretation slides into a more descriptive “data-mining” posture, it is also punctuated with insightful process-based insights. In short: overall this is a strong manuscript!

My main criticisms apply to the way the methods are presented—I think some details are left out or insufficiently documented. I also think that the methods and discussion sections could use more of a “road map” at the start—particularly the discussion, which dives into a description of the correlations between different variables where it could start with some pithy statements summarizing the high level process-based interpretation.

I also would appreciate a bit more discussion of the underlying geographic patterns in the context of African geology (perhaps just a paragraph). I realize that the existing geospatial products don't allow for a thorough quantitative analysis of geologic state factors, but some limited qualitative might be good. More specifically the authors might address how parent material, soil age, and erosion rates vary (or do not vary) across the sampling locations, and how these might exert some influence on the results independent of climate.

C2

Specific comments:

Lines 39-40: The phrase “complex analytical approaches with a large number of parameters” is somewhat opaque. Perhaps substitute something more specific?

Lines 62-63: To be fair here: there is an implicit representation of competition between microbes and minerals in Earth System models via clay content. There are two issues in this case: (1) competition between minerals and microbes is not represented in an explicit, mechanistic way; and (2) clay content doesn't capture the relevant aspects of soil mineralogy or chemistry. I think this manuscript addresses the latter issue more than the former.

Lines 129-131: Was this digestion quantitative? I believe some silicates are resistant to aqua regia. Perhaps clarify whether these should be considered total elemental pools or simply aqua-regia-digestible pools, as this may influence the interpretation of the CIA (though probably not much I imagine).

Line 160: It would be good to include a short overview paragraph at the start of the statistical analysis section explaining the overall strategy. It seems that several approaches were applied to the same data: linear mixed effects models, regression trees, and random forests. I can see how the approaches complement each other (the mixed effects models seem more conservative and permit statistical hypothesis testing while accounting for non-independence of the data, but the CART based approaches can handle non-linearity). This is explained later, but the readers will benefit from a quick signpost at the start. Similarly, the discussion section is hard to follow at the start. I strongly recommend adding a concise paragraph at the beginning of the discussion that identifies the major results. As it stands now the discussion dives right into the details and I can only identify an emergent narrative at the end.

Lines 167-171: I understand that the transformation is necessary for comparing different predictors on the same scale. However, what does the transformation mean with respect to the functional relationships in the data? Are the models linear with respect

C3

to the original scale? I suspect not: a linear model fit to transformed data is not necessarily a linear model with respect to the original data. This is worth noting, even if the analysis stays the way it is.

Line 183: How was the hierarchical clustering done?

Line 204: The spatial partitioning is really laudable. It is surprising how infrequently this is done, and it really should be a community standard. Thank you for being rigorous!

Line 242: Please introduce the marginal/conditional R-squared values before mentioning here. To many readers this distinction might not be obvious.

Figure 2: The univariate linear regression fits in this figure are purely for illustration? Perhaps mention them briefly in the statistical analysis section.

Figure 3 (and throughout): How were confidence intervals obtained? They are reported throughout the paper, but unless I missed something the method used to obtain them is not reported.

Line 289: How was the % variation explained obtained here? Is this an R-squared value for a reduced model? Or is it some sort of variable importance metric? Perhaps something is missing from the methods description?

Line 446: I hope that the data presented in this study are eventually made available in some easy-to-access way. A database of this size and completeness could be extremely valuable to other researchers and would be best archived on some sort of data repository rather than only available on request from the author.

Interactive comment on SOIL Discuss., <https://doi.org/10.5194/soil-2020-69>, 2020.

C4