Comment on egusphere-2022-488
Damien Beillouin

Community comment on "Quality Assessment of Meta-Analyses on Soil Organic Carbon"
by Julia Fohrafellner et al., EGUsphere,
https://doi.org/10.5194/egusphere-2022-488-CC2, 2022

General comments:

This paper investigates the quality of meta-analyses published on the effect of agricultural practices on soil carbon in Europe. This is an important and timely topic as the number of meta-analyses is increasing strongly. It is therefore timely to reach out this research community to raise awareness of these issues.

The paper is overall well structured, and well referenced, with an introduction clearly defining the terms and objectives. For the material and method, the authors have adapted quality analysis grids from meta-analyses used in others scientific fields. The authors repeatedly emphasise the specificities of soil science work that would require specific meta-analysis methods or analysis grids. I was not convinced by these arguments, at least, as currently formulated. I also have some questions about some of the 17 criteria used to judge the quality of meta-analyses. The authors identified 31 published meta-analyses dealing with agricultural practices on soil carbon in Europe (on cropland). Recent work has identified more than 100 meta-analyses on these issues worldwide. Given that most of the meta-analyses are global (across continents), this low number identified by the authors raises questions for me.

Detailed comments:

Introduction:

L 48: “Particularly, the use of meta-analysis as a tool to investigate the effects of agricultural management practices on relevant response variables, such as yield or soil physical or chemical parameters, is becoming increasingly prominent (Valkama et al., 2019, 2015).”

- Are the two references particular example of meta-analyses on this subject? If so, how do you choose these references? (the first meta-analyses in the academic field?).
- Or do these references analyze the interest and use of meta-analyses in agricultural field?
“Because of their close relationship, many applications of meta-analyses in ecology are also transferable to the field of agriculture and soil science.”

- What do you call “applications of meta-analyses”. Does this refer to the method itself? Or the use of the results of meta-analyses?
- Doesn’t this contradict your arguments about the specificity of soil carbon meta-analyses?

“Provide clarification by synthesizing conflicting evidence from primary studies.”

- Mention that this clarification is provided by the increase statistical power/increase precision of the mean estimates of the effect?

“Nevertheless, research on agriculture and soil encounters issues, which are often specific to these fields. Firstly, changes in soil are often slower than other physiological and biogeochemical changes; e.g., changes within plant tissue. Therefore, long time experiments are needed to detect treatment effects on soil parameters or soil health indicators, like soil organic carbon (SOC)”

- Doesn’t this need of long term experiment also apply to fields in ecology (e.g. effect of disturbances on the composition/recovery of populations,...) and medicine (e.g. effect of environmental conditions/substances on the prevalence of some cancers,...), or other scientific fields?

“Therefore, it is crucial to define not only the treatment but also the control of the experiments precisely to allow computation of heterogeneity”

- I totally agree, but I don’t think this difficulty is specific to agronomy studies. In fact, all effect-sizes are metrics quantifying the relationship between two entities (generally the control and the treatment).

“A good example is bulk density, which can be measured in a field experiment or estimated using pedotransfer functions in order to compute SOC stocks from concentrations”

- An other example could be the soil depth, sometimes very different across studies, and not always precised.

“However, they are formulated rather generally”

- Could you precise what generally formulated means? Philibert and Beillouin papers presented 8 and 20 quality criteria, respectively (defined in table 1 and Figure 6 of their article respectively).
L 93. “but do mainly focus on systematic reviews and maps and contain elements not necessary in meta-analysis (e.g. registration, gathering a maximum of available relevant literature or performing critical appraisal)”

- Could you justify that registration, gathering a maximum of available relevant literature or performing critical appraisal are not required for meta-analyses concerning soil organic carbon.
- For example, the same research question can be addressed in parallel by several research teams, duplicating the necessary research efforts and investments on the same topic (could also occur in agronomic/soil science academic field) -> registration could potentially allow avoiding these problems.
- + in the discussion you mention “The publication of protocols prior to a meta-analysis would benefit the method by allowing constructive criticism and suggestions for improvement by the scientific community (Moher et al., 2015; Brandt et al., 2013).” Is this sentence coherent with the one L 93?
- gathering a maximum of available relevant literature is generally recommended (e.g. by the search of several bibliographic databases). Incomplete base of scientific paper in meta-analyses could lead to wrong results.

Looking at the criteria #1 in Table 1, you give a higher score to meta-analyses using >4 databases. This contradict thus your above-mentioned sentence?

- The quality of the primary research is very variable. Weighting the evidence according to the (estimated) quality of each paper could avoid to give a large importance to the papers with the lowest quality.

L 137. “Moreover, the interest in SOC sequestration and subsequent increase in related publications raises the question whether there are meta-analyses synthesising this knowledge”

à A (partial) answer could be found in “A global overview of studies about land management, land-use change, and climate change effects on soil organic carbon” Published in 2021. The authors search exhaustively the literature to identify meta-analyses on soil organic carbon. They identify 192 meta-analyses, and then characterized the interventions, outcomes, temporal dynamics of publication and spatial distribution of the primary studies synthetized in these meta-analyses. They also analyzed Also some quality criteria of the analyses (based on indicators found in Phillibert et al., and Beillouin et al. ). See for example Fig 5 of their publication and paragraph 3.3.

L 148. “This study aims to quantitatively analyze 31 meta-analyses, “

- Does these 31 meta-analyses represent a subset of all meta-analyses published on Soil organic carbon focusing on cropland, and European region? How did the authors define that a meta-analysis focused on Europe when most published meta-analyses use global data? did they exclude all meta-analyses that included data outside Europe? Or on the contrary, did they include all meta-analyses with at least one primary study in Europe?

L 166. “The 17 quality criteria were structured according to three groups”

- How did you choose which criteria among the numerous references used to keep as
relevant in your analyses and the ones redundant or not useful?

Table 1. Quality criteria 6.

- Could you elaborate on why you consider log ratios to be 'better' effect sizes than hedge's g? Numerous advantages of hedge's g also exist (see for example Borenstein's book). Numerous meta-analyses in ecology use this specific metric.
- What do you call "non standard metrics"? Depending on the specific problem analyzed in the meta-analysis, specific type of effect-size could be used (proportion, correlations, ....). These metrics could be the most suitable for some particular meta-analyses, and thus and they should not receive a low-quality rating for this criterion, I think.

Table 1 Criteria 7. Standard deviation extracted >> from each study =2.

- some result papers do not present a dispersion indicator (e.g. SD, CIs,..). Are you suggesting that these papers should be excluded from the database (for the meta-analysis to get 2 points for this criterion)? Moreover, techniques for imputing missing variances are available, and could present an interesting alternative when some papers do not present a dispersion indicator.

Table 1 Criteria 8.

- Can you explain how a meta-analysis weights only part of the studies by their variance?
- Furthermore, how do you consider studies that weight the effect sizes by the sample size (number of data)? See for example doi:10.1177/0013164409344534

Table 1 Criteria 11.

- Is using Excel for a meta-analysis intrinsically suboptimal? Maybe, it exists some macros or other means to apply all the specifics of meta-analysis models (random forest, proper calculation of tau, ...)? In my opinion, this criterion is not precise enough, even confusing. Perhaps you are only focusing on the type of model used in the meta-analyses? (Or maybe analysing if the method is transparent and reproducible? in which case it is useful to have the type of software used and the codes)

Table 1 Criteria 12.

- How do you consider here the three levels random-effect meta-analytical models that allow for effect-size dependence within a study?

Table 1 Criteria 14.

- How do you consider meta-analyses that do test for publication bias and find one: do
they get a lower score because the results need to be interpreted carefully, or do they get a higher score because they tested for publication bias?

L 206. "The results were compared with the meta-analyses identified by Bolinder et al. (2020), who synthesized meta-analyses studying the effects of several management practices on SOC changes in agroecosystems"

- See also maybe:

L 224. “The aim was to assess how many meta-analyses were conducted on a certain management practice and whether their quality was sufficient to stop the production of new meta-analyses on the respective practice”

- Quality is only one of the criteria for judging the value of continuing the synthesis efforts. A low precision of the effect estimated globally, or new sub-group analyses as a function of co-variate or geographical region, or in combination of other factors, can in my opinion justify the interest of continuing the synthesis efforts in a field (or on a particular practice), even if meta-analyses already exist.

**Results:**

L 254. “Scores also experienced a rise (15-year period) and related with the publication year ($y = -1889.8980+0.9437^x; R^2 = 0.39$)”

- Is the slope significant? Do you also include "pseudo-meta-analyses" in this analysis?
- It is interesting to note that similar trends have been observed in others papers, see for example El-Rabbany et al 2017, Jamshidi et al 2018, Beillouin, 2019 (Figure S3).

L 279. “The "Meta-analysis" group consisted of nine quality criteria, “

- the term meta-analysis may not be the most appropriate to deal with a sub-part of the criteria (all criteria are linked to meta-analyses). Replace by statistical analysis? or ..?

L 288. “Nevertheless, we urge authors to extract SDs for each study and further weight them by the inverse of variance in order to conduct a high-quality meta-analysis”

- This information is not always presented in experimental studies. (see my comment above)
“Only about 25% of meta-analyses had no problems with non-independence of effect size, while the rest extracted several effect sizes per study”

- This is not always a problem, depending on the statistics made (see for example three levels meta-analyses, meta-analyses with variance-co-variance matrix – Lajeunesse, 2011, …).

Figure 6.

- The "real" meta-analyses (according to your criteria), therefore have lower quality than the average of the others, is that right? Is your criterion relevant, then?

Discussion:

“...A quality criterium, which is of special significance to the soil and agricultural field, is the inclusion of grey literature”

- Again, I am not totally convinced that this issue is specific to the soil and agricultural field.

L 451. “Therefore, common- and random-effects models are not useable, leading to difficulties in assessing heterogeneity (Gurevitch et al., 2018).”

- I am not sure to understand this sentence? You suggest that when not weighted at all, we can neither use common-effect models nor random-effects models?
- What are the different type of function used to weight studies?

L 461. “Effect sizes might show a certain amount of variability that cannot be explained by sampling errors alone,”

- Depend on the assumption of the model used. Fixed-effect models consider that the measured effect-size differ to its ‘true effect size’ (and those of all others studies) only because of sampling errors.

“...Therefore, we suggest that the topic is well covered for the moment and no further global meta-analysis is needed until there is a substantial number of new results”

- Not agree. Some interesting new meta-analysis results could be produced, e.g. with new method for analyzing the data or new questions. For ex. https://doi.org/10.1038/s41558-021-01075-w,
L 577. “Quality assessment of meta-analyses, especially in the complex agricultural set up, are highly warranted to harness the power of meta-analyses”

- Does “power” refer to statistical power? Or do you mean potential biased results?

L 578. “We demonstrate that meta-analyses in soil and agricultural research encounter specific issues, which differ to other fields like medicine, environment or ecology”

- Not totally convinced by this part.