EGUsphere, referee comment RC1
https://doi.org/10.5194/egusphere-2022-488-RC1, 2022

Comment on egusphere-2022-488

John Koestel (Referee)

Referee 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-RC1, 2022

I am thanking the authors of egusphere-2022-488 for this exceptionally well-developed manuscript, in which they investigate the quality of available peer-reviewed meta-analyses on soil organic carbon. I only have some one major point that I want the authors to address. In other words, I recommend a publication after minor revisions.

There is mismatch between the literal meaning of the term "meta-analysis" and the meaning it has for the authors of this manuscript. The literal meaning of "meta-analysis" is something like "transcending analysis" or "analysis of "several" analyses". As such, the term does not imply whether this transcending analysis is qualitative or quantitative or which statistical methods need to be used for a meta-analysis. Some examples of the definition of meta-analysis in dictionaries are

"a research method that combines the results of several related studies to produce better results" (Cambridge Dictionary)

"Meta-analysis is a statistical process that combines the data of multiple studies to find common results and to identify overall trends." (Dictionary.com)

"a quantitative statistical analysis of several separate but similar experiments or studies in order to test the pooled data for statistical significance" (Merriam-Webster)

Also, the definition given by the authors (L36-41) does not specify what kind of statistical methods need to be used in a "proper" meta-analysis. I very much agree with the authors that there are more and less suitable methods. However, I am unconvinced that the three cut-off criteria (effect size, standard deviation, studies weighted by 1/variance) defined by the authors are suited to distinguish between correctly and wrongly executed meta-

analyses. In some fields, the available data in the literature does not allow for investigating effect sizes, since the studies were not carried out using a design that allows classical statistical testing. An example for such a field are soil physics. Here, respective measurements are so time-consuming and the number of "moderators" so large that studies have been carried out with the goal of process understanding. Still, statistical analyses (e.g. correlation analyses or machine learning approaches) can be applied and still deliver useful insight. Likewise, there are cases were the correct standard deviation of repeated measurements are not given by the original publications (e.g. in cases were the data is log-normally distributed). Using the number of replicates (or a thereof derived measure) as a weight instead of 1/variance maybe the only possible way to weight the data in a more reasonable fashion. Please also consider that the variance may not only reflect measurement error, but also express small-scale heterogeneity in soil properties. In the case of organic carbon measurements, pooled satellite samples can be measured and the problem be circumvented. In the case of tension-disk infiltrometer measurements, each point in the field needs to be measured individually. Following a randomized block design would often be too expensive. Therefore, replicated measurements in publications are often only pseudo replicates. Provided there are sufficient (pseudo)replicate measurements, the geometric mean hydraulic conductivity at a plot with large small scale heterogeneity will reflect the hydraulic conductivity similarly well as the one in a field with small small-scale heterogeneity. Only its variance will be much higher. Using 1/variance as a weight will then up-weight locations with small soil heterogeneity not the once with the better measurement precision.

I am not per se opposed to the idea that a proper meta-analysis must include effect sizes calculated from measurements weighted by 1/variance. If the majority of the meta-analysis community defines "meta-analysis" in this fashion, this may be the way to go. It is however very probable that the authors of the here reviewed "meta-analyses" had not been aware of this very strict and specific definition. I therefore urge the authors to better explain why a meta-analysis must include effect sizes calculated as outlined above and suggest than a term for studies that also analyze the results of several source publications using statistical methods, albeit not the ones required by an orthodox meta-analysis. Alternatively, I suggest the term "meta-analysis" as a broader term and the term "orthodox meta-analysis" for conducting a meta-analysis in the sense of Koricheva and Gurevitch (2013).

A minor thing: the plural of meta-analysis should be meta-analyses. At some places in the manuscript it is correctly spelled, at others it is not.