The statement in the methods (lines 109-110) is quite brief in regards to how data was extracted from the meta-analyses. If I am following correctly it appears the authors visually from graphs or quantitatively from tables if effect sizes were significantly above zero, below zero or were not different from zero, and then counted the n=value from these studies. It is not entirely clear to me how effect size is counted. Is this one per study (i.e. one effect size per MA?) or the total in the studies that were combined, and therefore 1 effect size = 1 field site embedded within one meta-analysis? I believe it is the latter. I encourage the authors to clarify.

The number reported at line 108 is the total number of overall (average) effects sizes (i.e. the number of overall effects sizes for all combinations of management practices and target variables reported among all MA’s included in our analysis. We will clarify this.

I also encourage the authors to consider if the language "increase, decrease, no effect" is the same as "positive, neutral, negative". The latter have a connotation that is assumed by the reader, while the former are (in my opinion) more descriptive of the actual effects. For example, a decrease in runoff, while "negative" in nature, could be perceived as a positive effect of a management practice.

Yes, we did consider this aspect. The direction of the effect sizes reported in the paper are purely statistical and therefore have no connotation of value. We decided to report the effects in a statistical sense, because in some instances it’s not clear whether an effect would be beneficial or detrimental.

It was noted briefly in the caption to figure 5 (line 163) that the effects are statistical, but we will also make this clear in the text in the revised version of the paper.

We will also add the following information more clearly on how the graph was constructed:
To build Figure 5, we used a qualitative approach. From the figure reported in the meta-analyses, the overall effect size was extracted qualitatively. This means that if the average log response ratio and the entirely of 95% confidence interval was larger than 0 (so a response ratio of 1, 0 in log scale), it was considered as positive. The word "positive" here refers to the sign of the overall response ratio on a log scale. If some of the confidence interval overlapped 0, a "neutral" effect was chosen. If all confidence intervals were smaller than 0, the overall effect was considered as "negative". For all overall effects retrieved, the number of individual effect sizes used to compute the overall effect was also collected. For a given pair of practice and soil parameters, a pie chart representing the overall effects as a proportion of the number of individual effect sizes they summarised, is represented. There are situations where several meta-analyses can have information on the same pair, which can lead to pie charts with multiple colours. For instance, if a MA1 has a positive overall effect from 25 individual effect sizes and MA2 has a negative overall effect from 75 individual effect sizes, the pie chart will be one quarter orange (positive) and three quarters blue (negative). Note that that these proportions are qualitative and not quantitative. Indeed, we do not have the information on whether the individual effect size is positive, negative or neutral, but we only interpret the overall effect reported in the meta-analysis. The same figure with individual effect size would give other proportions (=quantitative approach). However, the trends between the qualitative and the quantitative approach will be similar. Several meta-analysis can also contain information about the same primary studies. This is shown in the redundancy analysis later.

Although I found the manuscript generally easy to read and follow, I believe that the short intro and methods section are incongruent with the very long results & discussion. I encourage the authors to review and look for opportunities to be more concise. For example the discussion of tillage is quite lengthy, and although it is interesting to cover some elements of yield impacts, tradeoffs, etc. I am not sure that lengthy discussion as to the paper. And further, although the introduction discusses/focuses on conservation agriculture, the practices outlined in the paper are far broader in scope. I encourage the authors to consider a slight reframing of why specific practices were selected to focus on in this analysis. I do appreciate that the selection was broad, as it allows for a quick visualization/comparison of what properties and practices are more widely studies, but again think this decision needs more clarification.

Yes, we agree. We did mention the terms climate-smart and regenerative agriculture in the introduction at lines 40-43. In the revised version, we will make it clearer that our synthesis is not exclusively focused on conservation agriculture.

I disagree with the authors final assessment in line 534 that continuous living cover reductions in SWS/recharge will outweigh increases due to carbon sequestration. While there may be concern of water limitations in drier climates, I do not believe the evidence presented supports this perception as there are few studies that look directly at this effect (judging by the gaps in evaporation and water content associated with the cropping systems practices). The benefits discussed in terms of soil physical improvements, plus benefits from carbon, with the uncertainties of rainfall variability and increases in precipitation in many places, even semiarid regions, make this assertion one that deserves more attention.

Yes, we agree. We will modify this statement to say that although there is some cause for concern, there is currently little empirical evidence that would allow us to draw any firm conclusions, and that more research is therefore needed.
There is a small typo on line 112 - believe the first word in the sentence should be "We..."

Thanks. We will correct this