

Comments on “Which pitfall traps and sampling efforts should be used to evaluate the effects of cropping systems on the taxonomic and functional composition of arthropod communities?” by A. Gardarin and M. Valantin-Morison.

#### General comments

I have read carefully the manuscript. It deals with finding the optimal sampling design to detect differences in arthropod communities in different cropping systems. It explores different dimensions of this sampling design by varying trap size, type of preserving fluid used in traps, and number of traps and seasons used to assess communities for both spiders and beetles. Though I found the idea interesting in general, I have some major concerns regarding the methods and the way things are explained.

Clarity of text. In many places key methodological details are missing, main methods are not explained, rationale for using different methods are not given, many assumptions are not explicitly stated and are hard to follow for the reader. The part on the ‘simulations’ is particularly hard to follow. In some places misunderstanding can come from English formulation issues or too long and tortuous sentences.

Methodological issues. Methods are not well explained in general, but the reader can still follow most of it, excepted the part on the simulations that need to be rewritten (or deleted?), especially to have simulated data more variable than observed data. I am also concerned that the procedure used here might not be appropriate. The general idea is to make a good explanatory statistical model and then use the values behind to generate more data (like a type of bootstrap). The problem is that to do so it seems to be the model should be ‘good’, it should explain a large part of the variance and be robust for predictions as well. In particular, if the effect of cropping practices is very small, there is no way the simulations can do a good job in simulating ‘new data’ that are more realistically different.... So maybe the observation data are not appropriate to make these simulations because there are so small differences from the beginning? If the authors think this is ok, they should explain all that more properly so that the reader can make his/her own opinion on the question.

#### Detailed comments

##### Abstract

L. 23 Seems strange to talk only about innovative techniques. Reformulate: The idea is more being able to see if innovative techniques lead to different communities of arthropods than other type of techniques, being able to compare different cropping techniques.

L. 24 “distinguish between simplified” – I guess what is meant is that in arable crops whatever cropping practices, arthropod communities are simplified in comparison to natural ecosystems. And so sampling techniques that are working in natural ecosystems need to be adjusted to these specific communities. For non-specialists on arthropods in might go a bit fast here.

L. 29 We need to know from the beginning that 4 technical elements are tested: size of pitfall trap, type (?) or presence/absence (?) of preserving fluid used in pitfall traps (again for non-specialist of arthropods, ‘preserving fluid’ comes out of nowhere), sampling effort, and different metrics (not clear how many different metrics are compared?).

L. 36 Replace by: ‘, and a higher species richness and CWM’, richness in itself does not mean anything

L. 40 what is an affordable number? Give a range. What about field-year replicates? We did not know there were such replicates

L. 44 we did not know there were 3 cropping systems. Please give the required elements in the method point (2)

L. 49 for both taxa?

L. 56 why only one of the different community measure (CWM) and not the others?

Introduction

L. 65 formulation. 'Which' refers here to synthetic inputs mostly, this is hard to get that this is rather 'numerous challenges'. Maybe : 'synthetic inputs.

L 70 tricky formulation

L. 72 the logical link is weird. They are used as indicators of anthrop activities because they respond to these activities with communities changes.

L. 77 their instead of these

L 78 Again, I get what is meant but sentence might be clearer. You mean that assessing arthropods communities under different cropping practices can help telling if the practice really is environmental-friendly and preserves soil-functioning and associated services.

L. 81 Their? Pitfall traps' design?

L. 83 the type of preserving fluid used within them -> be precise

L. 86 suggestion: 'riverbeds. However, in these ecosystems, vegetation structure is more complex than in arable fields, leading to more diversified arthropod communities. It is thus unclear to which extent the results... '.

L. 97 'on the soil'

L. 98 'potentially reducing'

L. 99 could? It is an assumption, right?

L. 101 'types of preserving fluids'

L. 102 deleted 'preserving fluids'

L. 103 'effective' for what?

L. 104 who could be intoxicated? Scientists? Animals? Insects?

L. 105 'are other possible'

L. 106 I understand the idea of 'preserving efficiency' for preserving fluid, but not the one of 'capture efficiency', because they attract insects? Because they cannot leave the fluid/trap? Be precise

L. 101-107 ok, so maybe this can be structured with types of fluid, their efficiency regarding different aspects and their drawbacks (toxicity, ect)

L. 112 remove 'ecological studies'

L. 112 'impracticable' why? 20 already impracticable?

L. 112-114 this should be clarified. In some cases it does make sense to determine absolute richness/abundance. What is meant here is that what is ususally done (determining abs values) is not necessary when the goal is to compare how well-being are communities under different cropping systems. Looking at relative values/comparison might require smaller efforts. Right? If yes, please reformulate.

L. 116 'and poorly diversified'

L. 118 This contradicts what comes above where it is said that large sampling efforts are not feasible (though we do not know why)

L. 120 'studies investigating the'

L. 122 remove 'and so on'

L. 110-128 in general this section contains many different ideas and it is not 100% clear -> 1) we want to compare practices so maybe smaller sampling efforts are ok?, 2) Communities are poor and little abundant so this may still require a high sampling effort, 3) different metrics can be calculated abundance, richness but also trait-based metrics that better relate to fuctionning and these might require also different levels of sampling effort? Right? So overall you could say that there is no idea of the level of sampling effort needed because it depends on three different elements, the 1, 2, 3 I just give above

L. 129 sentence could be more general -> main goal is to determine how sampling should be conducted in order to characterise differences among communities under different cropping systems, this include comparing pitfall traps characteristics, the associated preservative fluid, and the sample effort.

L. 140 -> the three metrics? Please add above something clear on the different taxa and metric studies, preferentially before line 110. There should be an explanation of the metrics saying that different metrics can characterize communities, some being more related to functioning.

L. 141 we are missing a sentence saying what is done here -> large field sampling campaign (multi-year?) among crop systems to assess the optimal sampling design.

M&M

L. 146 'where' instead of at which

L. 146 why are they called 'innovative'?

L. 148 'each'? Which ones? How many different types?

L. 149 unclear, there is a rotation every year in each field?

L. 152 Given ploughing and pesticide are allowed, it is unclear why it is 'high-environmental'? Explain, and more environmental than conventional: less inputs? Different inputs?

L. 159 allowed. Is it allowed in the PHEP too?

L. 163 and are pesticide and fertilization allowed?

It would help if all the differences are clear: tillage yes/no, pesticide yes/no, fertilization yes/no

L. 178 'flushed'?

L. 182 8 large traps? Where, how? Why 8? 5 different treatments... so what 8 by treatment?

L. 184 remove 'the same'

L. 185 '100%' meaning?

L. 186 is there a reason there are 8 of one sort and 4 of the other? Is there a reason the larger are put for a longer period of time? Please be precise.

So a total of 80 traps?

L. 201 Type of

L. 205 why 5?

L. 205-206 please reformulate for a better English phrasing

L. 207-208 Yes, 6 response variables, that is what we want to know in the intro, together with a definition (what is activity-density?) and with an explanation of why these measures matter

L. 210 is it ok to put an interaction effect given that the design is not fully-crossed?

L. 215-216 why these analyses?

L. 218 for richness estimates is it ok to remove rare species? How about the sometimes important effects of rare species?

L. 221 why this analysis and explain what it is in one sentence.

L. 226 what is the threshold used to say communities are similar or different?

L. 229 'by varying', what do you mean? Analysing how much they differ from?

L. 231/ 242 'modelled', 'simulate' what do you mean?

L. 233 the most numerous ones

L. 234 why is it pooled over the year now and it was not for the first analysis?

L. 239 not clear what the station effect is

L. 246 why 2-40?

L. 249 pairs of cropping systems? I do not understand that. They could be all compared to the reference without looking at pairs?

L. 251 'for each of the'

L. 252 why beyond the observed values? It is wanted or a result? Is that not an issue if simulated values are too different from observed ones? It means differences among cropping systems might be overestimated with the simulations? This needs to be explained

L. 255 why 6 years, why 3 cropping types, please explain, the reader does not know

L. 262 models of what explained by what? Community metric as a function of sampling effort?

L. 265 I thought trap type was standardized to larger traps at this point

L. 265 what situation? The reader get lost at that point

L. 266 how is the difference among ref and other crop system measured?

## Results

L. 278 replace by 'average'

L.280-285 Are all the differences significant? Tuckey tests announced on line 294 are not described in the M&M are they?

L. 308 variables or species on the right panel?

L. 314-326 If I understand well, differences among crop systems are already hardly different, even with the entire set of observed data? Does it mean the dataset is too small, or that there are no differences? More environmental friendly designs are not so much better in termes of communities? Or the LGHG is already good in preserving biodiversity?

Fig. 3 is hard to understand. Sampling effort should be the x-axis, why are there some positive and negative difference? Why are the difference smaller with larger sampling efforts? How can it be that what is estimated with 8/16 communities (arrows) are better than what is estimated with 40 samples?

L. 357 Definitely I think I did not get what has been done here. Why are the levels of differences 'chosen' (ex. 30%)?

## Discussion

Line 420 'type of preserving'

L. 475, so then it seems inappropriate to do so

L. 479-516 Depends on the robustness of the corresponding methods

L. 492 Why unrealistic? Here for sure you already had a design with replicated trap size, fluid, ect. If the idea is to compare one crop type to another one with only one size and fluid type, then 40 might not be unrealistic? Or does It mean that differences are too small here?