Round #1

Decision

by Ignasi Bartomeus, 2019-02-21 13:10
Manuscript: 10.5281/zenodo.2533857

Methods clarification needed

After carefully reading the manuscript and reading the reviewers comments, I think that analyzing how to improve a common sampling technique used for describing ground arthropod communities can constitute a good contribution to the field. However, I concur with both reviewers that the manuscript needs to be presented in a more clear way, and acknowledge better its limitations. I made a number of wording suggestions in the text to improve clarity, especially about the results presented (see the document here: https://www.dropbox.com/s/3qyv35ljifj1yfc/Manuscript_IB.docx?dl=0). My only main concern is regarding the simulation (see reviewer 2 detailed advise). I am not sure the data at hand allows testing the sampling effort question, but if the authors think so, it should be clearly justified in the paper. Clear recommendations on what pitfall traps are optimal in different conditions would be of great help. I hope our comments are helpful to strengthen the manuscript.

Best, Ignasi Bartomeus

We thank the recommender and the reviewers for their careful revision. Their external perspective helped us to better explain what has been done.

Reviews

Reviewed by Matthias Foellmer, 2019-02-18 21:58

In this study, the authors aim at filling a gap in our understanding of the efficacy and necessary sample size of different pitfall trap types in agricultural fields, for which such tests are lacking, in contrast to more natural habitats. The first part of the manuscript focuses on the analysis and results of the field study component; the second part focuses on a simulation study to evaluate necessary sample sizes of pitfall traps to detect statistically significant effects at alpha = 0.05 in crop fields as a function of proportional cropping system difference, using the information on ground-dwelling spider and carabid communities gained from the field experiment.

In my opinion, there is a lot of valuable data presented, which should provide guidance for future studies. However, the presentation and organization of the manuscript are not clear, making it unnecessarily difficult to understand the paper.

L 39-40: The generalizability of the simulation results is less then indicated (see below), which the abstract will need to reflect.

Here and in the subsequent sentences, I specified that our results apply to arable crops.

L 115-116: rephrase "poorly abundant and diversified arthropod". Maybe "little abundant arthropods with low diversity".

Done.

L 144-163: I strongly suggest providing a diagram illustrating the experimental design. It's hard to keep track of the spatial and temporal aspects.

A new figure (Figure 1) has been added in section 2.2 to illustrate the sampling design.

L 189-190: so eight traps per sampling station? Please clarify.

I am not sure that we mean the same thing by "sampling point". Here, it means the location of each individual trap. At each sampling station, one trap was opened two or three times, as explained just afterwards (last paragraph of section 2.2) and in Figure 1. I explained in the text was I mean by sampling station.

L 202-203: why only five individuals? This seems very low. Please explain.

This small number of individuals was a compromise between the need to achieve rarefaction by levelling down community size across all traps without excluding too many traps for which the number of individuals would be below this threshold. In the case of carabids, this led us to remove 38 data points (with less than five individuals) evenly distributed over the three types of traps (10 for large traps with salt water, 14 for small traps with salt water and 14 for small traps with vinegar).

This paragraph has been added in the text (section 2.3).

L 211: Please specify the removal process

We used a backward selection procedure: at each step, the least significant effect was removed and the new model was assessed again.

However, in order to take into account the comment of another reviewer, all the same terms were finally included in each model (cropping system effect, crop effect and their interaction as covariables, and type of trap as main variable), in this analysis.

L 276 – section 3.1: I'm missing the model output. Please add the table.

The outputs are now provided in Table S1 in which we detail the effects of the types of trap.

L 292 – Figure 1: This looks like boxplots showing raw data, not estimated effects from a model. I suggest providing an effects plot instead.

N. Bartomeus found that this plot was still informative. So, I kept it but the estimated effects are now given in Table S1 in the supplementary material.

L 308 – Figure 2: I think the variables are on the left.

No, the individuals are the traps (each empty circle is a trap) and they are described by the activity-density of each species they contain. The species are therefore the variables.

L 323: "Over the variability" sounds very odd, please rephrase.

I replaced by "among all the variability". Is it correct now ?

L 2237 – 338: Why for large pitfall traps filled with salt water only?

We did not study all the combinations between modalities of each factor (trap size and type of preservative fluid) because of space constraints in each field. Each experimental field measured 63×65 m and it was not possible to put in it too many traps. It would have led to a too high concentration of the traps, with the risk of interactions between traps. This is why the effect of

the type of preservative fluid was studied on small traps only. The reason for this incomplete factorial design is now justified in section 2.2:

"In order to satisfy the different objectives, to keep a minimal distance of 15 m between each sampling station (location of each individual trap) so that they remain independent and to avoid depopulating the fields with a too high number of traps, we combined an incomplete factorial design with two sampling designs and different trap numbers at three periods in spring (fig. 1 B)".

L 359: I don't think there are supplementary materials.

This section in the end of the manuscript was named "Appendix". I renamed it "Supplementary material".

L 417 – 423: Your test was more limited than this paragraph suggests. I don't think you can generalize to other types of pitfall traps.

Right. The following sentence was added in the paragraph:

"These features [trap size and type of preserving fluid] are the main sources of variation in pitfall trap design, while other characteristics were fixed here (trap and rain guard colour, no funnel), to which our results are subject".

Reviewed by Cécile ALBERT, 2019-01-25 23:19

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?)…

Section 2.4 was partly rewritten. Intermediate sub-titles have been added to make the approach and the structure of the section more obvious.

…especially to have simulated data more variable than observed data.

See answers to you specific questions below.

 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.

The small amplitude of the effects of cropping systems does not mean that the models are not robust. The goodness of fit of the statistical models were initially not provided. R^2 are now shown in Table 1. Marginal coefficients of determination (accounting for crop and cropping system effects) equal 43.5% on average, and the conditional coefficients (accounting also for random effects) equal 68.2% on average. Thus, even if the cropping system effects are small, the statistical models fit quite well the data.

Besides, the most important in the simulations is not really to predict the cropping system effects. We test different values of cropping system effects simply by generatiing different values of activity-density, richness and CWM. The most important here is to predict the effects of the uncontrolled field-year and station effects, so as to generate a variability in the simulated data as can be observed in the fields. Accounting for this variability is essential because it justifies our questions on the sampling effort.

These explanations were added in the end of section 3.2.

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.

Reformulation done.

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.

I tried to add these explanations in the abstract, while remaining concise.

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?).

I clarified this, but in a summary, I cannot add too many details. We compare three "technical elements" (size of pitfall trap, type of preserving fluid and sampling effort) on six response variables (three metrics × two taxa).

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

I added "species or genus" richness, because spider could only be determined at genus level (it is often not possible to determine juvenile spiders at species level).

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

Details were added.

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

Modification done.

L. 49 for both taxa?

Not really, there was a simplification in the text. I now specify that the sampling effort was lower for functional composition than for activity-density in case of spiders and lower than richness in case of carabids.

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

I added "activity-density" and "species richness".

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.

The sentence has been split in two, for clarity.

L 70 tricky formulation

The sentence was simplified.

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

The sentence was reformulated, thank you.

L. 77 their instead of these

Done.

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.

The sentence was rewritten based on your suggestion.

L. 81 Their? Pitfall traps' design?

Precision added.

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

Modification done.

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... '.

Modification done.

L. 97 'on the soil'

Modification done.

L. 98 'potentially reducing'

Modification done.

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

Yes, modification done.

L. 101 'types of preserving fluids'

Modification done.

L. 102 deleted 'preserving fù_:uids'

Modification done.

L. 103 'effective' for what?

I specified "effective to conserve dead individuals".

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

I specified "by humans or wild mammals".

L. 105 'are other possible'

Modification done.

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

Preserving fluids may be attractant or repulsive. Depending on their density and their surface tension, trapped arthropods could also escape from the trap. The capture efficiency includes both aspects. I specified this in the text.

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

I adapted the paragraph, but it was already and partly structured by type of fluid, dealing first with ethylene glycol and formalin (pros and cons) and then with salt water and vinegar (pros and possible drawbacks).

L. 112 remove 'ecological studies'

Done.

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

I specified that it is impracticable when it has to be replicated over a network of multiple sites.

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 usually 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.

This sentence was indeed very unclear. You are right, I reformulated.

L. 116 'and poorly diversified'

The sentence was reformulated.

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

I did not want to contradict myself but to put forward that this is problematic. This sentence has been added:" This can be problematic in terms to feasibility."

I also specified that a sampling effort of 20 to 70 traps per site is often impracticable when it has to be replicated over a network of multiple sites to account for landscape effects for instance, in terms working time to settle and harvest the traps and to identify the captured

arthropods.

L. 120 'studies investigating the'

Done.

L. 122 remove 'and so on'

Done.

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 functioning 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

A long sentence summarizing these three points has been added in the end of this paragraph.

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.

The sentence has been reformulated in this way.

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.

This sentence has been added in the end of the second paragraph, where the question of the metrics was already alluded:

"The most commonly utilized metrics to assess carabid and spider assemblages are activity-density (as a measure of abundance), taxonomic richness (as a measure of taxonomic diversity) and measures of functional composition, such as community weighted means of traits (Eyre 2012 ; Martin et al 2019)."

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.

The second sentence of this paragraph was modified in such a way.

M&M

L. 146 'where' instead of at which

Done.

L. 146 why are they called 'innovative'?

I specified the cropping systems are innovative for their objectives and their combination of techniques.

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

For clarity, this sentence has been moved to the end of the section (2.1).

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

The sentence has been rewritten: "Each cropping system is replicated three times, but the replicates do not always contain the same crop simultaneously (they are however cultivated

with the same crop sequence)".

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?

I have modified the sentence to explain why is meant by "high-environmental": "Pesticide applications were allowed but the treatment frequency index and ploughing frequency were halved in comparison with conventional cropping systems".

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

Yes. I added in the text the amounts of mineral nitrogen used in each system because they were strongly reduced, even if they were not prohibited in any system.

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

Yes, I added in the text the amount of mineral nitrogen used in this L-GHG system. It would be too simplistic to describe the cropping systems by yes/no in terms of tillage, pesticide and fertilization use. I preferred to specify the amount or the frequency of their use.

L. 178 'flushed'?

Correction done.

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

Each year, we placed eight large traps in five fields (one field per crop×cropping system combination). A new figure (Fig. 1) has been added to explain the sampling design. Is it more understandable now?

L. 184 remove 'the same'

Modification done.

L. 185 '100%' meaning?

I removed "100%", which was not useful. Only vinegar was used in this kind of trap.

L. 186 is there a reason there are 8 of one sort and 4 of the other?

Using eight traps for each treatment (24 in total per field) would have led to a too high concentration of the traps, with the risk of interactions between traps. We therefore chose to compare the effect of preserving fluid with four traps in each treatment only. I tried to explain this briefly in the text by adding this in section 2.2:

"In order to satisfy the different objectives, to keep a minimal distance of 15 m between each sampling station (location of each individual trap) so that they remain independent and to avoid depopulating the fields with a too high number of traps, we combined an incomplete factorial design with two sampling designs and different trap numbers at three periods in spring (fig. 1 B)".

Is there a reason the larger are put for a longer period of time? Please be precise.

We chose to compare the effect of type of preserving fluid during the months where the arthropod activity-density was the highest. This explanation has been added in section 2.2.

So a total of 80 traps?

Yes. This precision has been added in the text. A new figure (Figure 1) has been added to clarify the sampling design.

L. 201 Type of

Modification done.

L. 205 why 5?

This small number of individuals was a compromise between the need to achieve rarefaction by levelling down community size across all traps without excluding too many traps for which the number of individuals would be below this threshold. This sentence has been added in the text.

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

Done.

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

These specifications have been added in the end on the introduction. I also defined activity-density in the first paragraph of section 2.3 and the other measures in the subsequent paragraphs.

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

In section 2.2, the crop×cropping system interaction was only taken into account as a covariable in each model, I do not think it is a problem. In section 2.3, in which we evaluate the significance of the crop, the cropping system and their interaction, we indeed must perform type III sum of squares in the ANOVA to account for the fact that we have an incomplete factorial design. This was initially not specified in the text, I added it.

L. 215-216 why theses analyses?

We wanted to explore a bit more the effect of type of traps, beyond the six response variables mentioned previously. The aim was to check if the types of traps led to differences in the recorded community composition. Additional justifications were added here (last paragraph of section 2.3) and this was stated in the introduction (aim to compare the effects of type of traps on community composition).

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

Rare species were only removed in the multivariate analysis of community composition described in this paragraph. The analysis of richness is described in the previous paragraph and it was performed on the whole dataset.

I do not understand the second question. Rare species strongly influence the construction of the axes of the factorial plan resulting in a bad representation of the individuals and of the variables.

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

The multivariate analysis provides a visual representation of the communities captured in each kind of trap. The PERMANOVA realizes a statistical test and make it possible to test if the type trap has a significant effect on the composition of the community. The sentence was slightly reformulated to make it clearer.

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

The metrics (activity-density, richness and CWM of body size) obtained with different sampling efforts are compared with GLMM models. The effect of sampling effort (i.e. number of traps) is assessed at $P<0.05$, as explained later in the text.

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

I reformulated to indicate that we simulated a range of cropping system effects.

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

The first sentence was rephrased in this way: "We first developed statistical models relating the activity-density, richness and CWM body size of carabids and spiders to the kind of cropping system".

In the second sentence, "simulate" was replaced by "predict".

I hope this will be clearer now.

L. 233 the most numerous ones

Yes, I added in the text that this was the type of trap for which we had the largest dataset.

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

In the first analysis, in which we compared the effect of the type of trap, we already pooled the data of May and June, as explained in section 2.3. Here, we added data for the additional month April. This should be clearer with the new Figure 1.

L. 239 not clear what the station effect is

This effect is the same as in section 2.3: "A random station effect, nested within the field×year effect, was used to take into account the location of the pitfall trap within the field."

Since this analysis in section 2.4 is the same as in section 2.3, and as suggested by N. Bartomeus, this paragraph was removed and we only refer to the previous analysis.

L. 246 why 2-40?

I added the following explanation in the text:" This sampling effort encompasses and extends beyond the most common practices in ecological and agronomic studies (e.g. Woodcock 2005; Eyre et al. 2016; Engel et al. 2017)".

Another practical reason is that the number of traps could not be indefinitely increased due to constraints of time for calculations (R is not very efficient for this). This number of traps, combined with one or three field-years replicates, was sufficient to compare these different strategies of sampling effort on our response variables.

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

You are right, the sentence was imprecise. The reference system was simulated once and comparisons were carried out on pairs of cropping systems. The text has been amended accordingly.

L. 251 'for each of the'

Modification done.

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

The following explanation was added in the text:

"The range of positive effects was consistent with our observations and with the literature (Djoudi et al. 2018; Henneron et al. 2015; Rusch et al. 2014), while negative effects were tested for exploration purposes".

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

The following explanation was added in the text:

"Since cropping system experiments generally provide two to four field replicates (e.g. Hossard et al. 2014; Meyer et al. 2019), we simulated here one, three and six field-year replicates. To focus on the effect of number of field-year replicates, only three observed crop×cropping system combinations were studied here".

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

In each tested situation (i.e. each combination of cropping system effect in a given crop, of sampling effort, for each response variable and each taxonomic group), we compared the metrics of arthropod communities between the simulated system and the reference one. We evaluated how far the metrics revealed a significant difference between the two systems. I tried to explain this better in the text.

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

This was an error, sorry. It was a trap location random effect. This has been corrected in the text.

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

Here, I referred to the "situation" defined at the beginning of the paragraph. Each situation is a combination of one response variable, one taxonomic group, one crop type, a number of traps per field, a number of field-year replicates and the effect value of the simulated cropping system. The order of sentences was modified in this paragraph, I hope it will be clearer now.

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

The difference was assessed with generalized linear mixed models, as explained three sentences before in the same paragraph. Some sentences were reformulated and their order was modified.

Results

L. 278 replace by 'average'

Modification done.

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

Right. I added a sentence to announce it in section 2.3.

L. 308 variables or species on the right panel?

Both. The individuals are the trap in the left (each empty circle is a trap). They are described by the activity-density of each species they contain. The species are therefore the variables. I tried to improve the legend to explain it.

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 terms of communities? Or the LGHG is already good in preserving biodiversity?

I share most of your hypotheses and they may all be more or less true:

1. Some cropping systems (No-Pest vs L-GHG) are very contrasted and I would not say that there is no difference. But it is true that the reference system (PHEP) is already more environmental friendly than conventional cropping systems, which minors the potential benefits of the other two systems.

2. Our sampling design was initially not aimed at capturing such differences in cropping system effects (these effects are only analyzed secondarily here), but at comparing different methods and sampling efforts. To be able to better distinguish the cropping systems, more years of sampling would have been necessary.

3. In addition, the experiment has been set up in an intensive cropping area, with a low diversity potential at landscape level, which may limit the recolonization of some fields by the biodiversity.

Fig. 3 is hard to understand. Sampling effort should be the x-axis

I guess you expected the sampling effort to be on the x-axis since we varied the sampling efforts in our simulations. We also varied the different metrics of arthropod communities. Since the objective was to determine the level of sampling effort to distinguish two cropping systems, we reported on the y-axis the minimum sampling effort (over all the tested ones), as a function of the difference between the compared cropping systems.

why are there some positive and negative difference?

We compare arthropod metrics between two cropping systems: the metrics in the productive system is taken as a fixed reference and we simulate a range of metrics values below or above the reference. We report of the x-axis the difference between these two values. In section 2.4, we wrote: "We varied these metrics from almost -100% to +150% relative to the reference cropping system".

Why are the difference smaller with larger sampling efforts?

The figure should not be interpreted in that way. When the difference is small between the two cropping systems we compare, then we need a large sampling effort.

How can it be that what is estimated with 8/16 communities (arrows) are better than what is estimated with 40 samples?

It is not better but worse. The arrows indicate the observed difference of metrics between the studied cropping systems and the reference one (productive system PHEP). The graphs indicate that, with one field year replicate (blue curve), a sampling effort of 30 to 40 traps (or more) would have been necessary for several metrics to be able to find a significant difference between the two systems. This is far more than our realized experimental effort and it partly explains why we often did not find significant differences between the metrics in our observations.

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

I guess that you did not understand that we varied simultaneously (1) the numbers of pitfall traps and field-year replicates and (2) the difference of community metrics between two compared cropping systems. For each level of simulated difference between the two metrics (on the x-axis), we then computed the minimal sampling effort (response variable on the y-axis) that was required to find significant difference between them. This is why the level of difference is "chosen". Does this answer to your question well?

These explanations were added in section 2.4.1 to make the general approach clearer.

Discussion

Line 420 'type of preserving'

Done.

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

In this part of the discussion, we explained that (1) the random trap location and field-year effects accounted for a significant part of the explained variance in the statistical models and that (2) the dataset generated took into account this important variability.

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

I hope that the previous responses already answer to this comment.

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?

You are right, the term "unrealistic" was inappropriate and subjective. The feasibility completely depends on the number of crops and cropping systems to be compared. I replaced it by "very high" and I also specified that it was a larger sampling effort than we had in our design.