We thank the reviewers for their comments. They have been of much help for us in revising the manuscript. Below we have copied their comments in *italics* *text.* We give our response in non-italics text.

A small remark. The page and line numbers given in the specific comments do not match with the page numbers in the PDF document. (Probably because of some conversion done by the journal?). This has sometimes made it difficult to determine exactly which passage the comment refers to. We give therefore in our response the page and line number in the PDF document we thought reviewer comments refer to. We hope that we have correctly addressed the specific comments of the reviewers.

**Referee: 1**

**Comments to the Author**

**General comments**

­

*In this paper authors put to the fore various problems associated with the use of trait databases that are routinely addressed by those using such or such database but maybe considered as grey literature. In that respect providing a global picture may be useful. It is quite strikening that invertebrate ecologists tend to design their own specific methodology. This is true with biomonitoring with every biologists designing an index or a method for its own place thinking that what was existing was not enough appropriate and deserving improvement. We have quite a similar thing with trait databases. From an initial idea, that was connected with theory testing (habitat templet) promoting fuzzy coding for traits to incorporate trait plasticity (but also lack of knowledge) of species in a specific French floodplain, trait database evolved towards compiling all the existing knowledge of biological and ecological traits of European genera. In parallel in the US another trait database was developed with another point of view and it is a bit a pity that this database at this time did not think about harmonisation with the European one. And then different countries aimed at providing trait databases for other part of the world where freshwater ecology was especially active. One useful aspect of the present paper is Table S1, which describes the meaning behind each trait designed by authors. In my opinion, this table should not appear as supplementary.*

The table is quite large and spans multiple pages. Therefore, we opted to put a reduced version of the table in the paper. Table 1 shows now as an example two traits (predator and swimming) and differences in ther definitions across trait databases. Table S1 was left in the SI and shows definition differences for all traits we used.

*However one aspect lacking in my opinion in this table or elsewhere in the paper is the rationale for using such and such traits. A major interest of the use of trait is hypothesis testing and this has been put to the fore in many papers (see e.g. Statzner & Bêche (2010) 10.1111/j.1365-2427.2009.02369.x). What hypothesis can be made in terms of grouping features or traits (to take the authors terminology) if we are in the direction of intercontinental comparison?What do we expect from trait response? I think some preliminary thinking would be welcome. By the way, Verbek et al. (2008)* [*https://doi.org/10.1111/j.1365-2427.2008.02035.x*](https://doi.org/10.1111/j.1365-2427.2008.02035.x) *elaborated a trait data base for ponds using a different approach as those used in the proposed databases, which is much more in connexion with life-history strategy theory.*

We wanted to show differences in trait definitions and how harmonisation affects trait-environment relationships. Thus, we needed to constrain our selection of traits to those that occur in all used trait databases. We tried to make this point stronger in the method section were we introduced the traits. However, the trait databases we used were different in the traits they present and the coverage of trait information (e.g. no ecological traits in the New Zealand database, almost no information on dispersal). Also, we added a sentence in the method section when we introduce the traits that some of the traits potentially help in coping with disturbances and that they relate to ecosystem functions.

*In addition, looking at the bibliographical records of the present paper, it seems that the trait approach started after 2000 with the new century thus ignoring pioneering work of the Rhone team that appear in Freshwater Biology in 1994. Why? In this FWB issue, one can find one of the first collective trait database at the species level. So I suggest to put more on the foundation of trait database to add some theoretical to a paper that is rather pragmatic and technical.*

*A second aspect that could be addressed by authors is functional diversity. In the paper, they deal with trait composition. I may be wrong but it seems that the harmonisation may decrease the variety of trait categories found within each database (e.g. from seven size categories towards 3). This should have impact on functional diversity. At least, one need to be convinced that it is not true. This could be easily tested on Szocs et al. (2014) study.*

Yes, harmonisation decreases the number of traits for some grouping features (or to use another terminology, the number of trait categories within one trait).

We calculated functional richness (FRic), functional evenness (FEve), and functional diversity (FDiv) for the Szöcs et al., 2014 dataset and the aggregated and harmonised trait datasets. Invertebrates analysed in Szöcs et al., 2014 were sampled within three years. We accounted for this fact by regressing each functional diversity (FD) metric with year and extracting the residuals. The residuals of the Original study were then compared with the residuals for the aggregated and harmonised datasets through regression and inspection of the R^2 for each FD metric. We have added the analysis on FD in the effects of harmonisation and trait aggregation on inferences regarding trait-environment relationships.

*Finally, I understand that authors wanted to use the same approach as Szocs et al. (2014) to analyse the effect of their harmonisation but RDA assumes a linear trait response to environment. Is this true? In addition the common application of community weighted trait matrix is now highly criticised (see Peres-Neto et al. 2017, 10.1111/ecog.02302). Other tools exist such as fourth corner that would appropriately help to make the demonstration that trait harmonisation has similar performance as the initial trait data.*

We re-analysed the data of Szöcs et al., 2014 (original dataset) and the aggregated and harmonised trait datasets with the suggested fourth corner approach to analyse which traits respond to conductivity. The term fourth corner has been used in the ecological literature to describe the problem of detecting associations between species traits and environmental factors and many approaches have been developed in the past to solve this problem (Niku et al., 2021). We assume the reviewer refers to the classical fourth corner method developed by Legendre et al., 1997.

We applied two types of analysis: first fourth corner was used with all traits from the original study and the corresponding traits in the aggregated and harmonised trait datasets. Second, we selected the traits that have been identified by Szöcs et al., 2014 to increase (shredder, life cycle duration > 1 year, gills, bi/multivoltine, ovoviviparity) or decrease (univoltine, eggs in clutches/cemented or fixed, life cycle duration <= 1 year) with salinity. These selected traits have also reacted to salinity in other studies.

For some sites, environmental and abundance data have been acquired multiple times per year. Szöcs et al., 2014 used the full data and did not consider repeated sampling. For the sites with multiple samplings, we selected only those sampling events with the highest macroinvertebrate abundances per year and accordingly only used the conductivity data for these sampling events. Fourth corner analysis was carried out with the fourthcorner() function from the ade4 package

Using fourth corner analysis with the full trait data yielded no statistical significant relationships for any traits across all datasets (Results attached). Fourth corner with the selected traits revealed significant relationship to conductivity for the traits shredder and ovovivparity for the original dataset. In three out of the six aggregated and harmonised datasets (*direct\_mean, weighted, harmonised;not aggregated*) four to five significant relationships (traits shredder, univoltinism, bi/multivoltine, ovoviviparity, and aquatic oviposition; Results attached) were detected according to the fourth corner method. In the remaining three datasets (*stepwise\_median*, *stepwise\_mean*, and *direct\_median*) fourth corner detected significant relationships for the traits univoltinism and bi/multivoltinism. Results were only slightly different when not considering multiple samplings (original dataset: one significant trait-conductivity relationship, others 3 to 4; Results attached).

The fourth corner analysis detected fewer trait-conductivity relationships than the RDA approach. Interestingly, the fewest significant relationships have been found for the original trait dataset. Furthermore, our analysis suggests that the fourth corner approach depends on the number of traits used. Recently, Miller et al., 2019 compared the fourth corner approach (different versions termed weighted correlation metrics), community-weighted mean regression, and model-based approaches to the fourth corner problem with regard to Type 1 Error inflation and statistical power to detect trait-environment associations with real datasets and simulation. They concluded that the fourth corner approach has lower power than model-based approaches and that power can also be low when the number of species exceeds the number of sites (Miller et al., 2019).

Given that we used the RDA approach to enable comparability with the original study and the mentioned power problems of the fourth corner analysis we propose to keep the RDA approach, but to acknowledge its limitations and those of the CWM. We added….

If the reviewer’s wish, we could of course re-do the analysis with a model-based technique.

**Specific comments**

*p.6 li 26-27. These bibliographical examples are far from covering the extent of the use of invertebrate traits in freshwater ecology*

(Probably refers to Page 5 li 26-27 in the PDF).

We added a part to the sentence mentioning the use of invertebrates traits in biomonitoring and functional ecological studies.

*p.6 li 32-35 You probably could add here Statzner et al. 2007 DOI 10.1007/s10531-007-9150-1 who claimed the need for a European collective trait database. Also Gayraud et al. (2003)* [*https://doi.org/10.1046/j.1365-2427.2003.01139.x*](https://doi.org/10.1046/j.1365-2427.2003.01139.x) *provided trait data for large river invertebrates species (truly the trait data are not public)*

(Probably refers to Page 5 li 32-25 in the PDF).

We thank the reviewer and added the references.

*p.8. Well this is the mathematical definition of fuzzy coding. But the terminology "fuzzy coding traits" which was first described by Chevenet et al. (1994) DOI:10.1007/s10531-009-9688-1 is based on trait affinities that are then converted to proportions. It could be fair to add the reference. Note that the idea itself of making profiles was already in a paper by Bournaud et al. (1992) doi.org/10.1002/rrr.3450070205 (but the authors did use the term fuzzy coding).*

(Probably refers to Page 7 li 12 - … in the PDF)

We added the reference to Chevenet et al. (1994)

*p.8. li 40- see Dolédec et al. (2000) 10.1127/archiv-hydrobiol/148/2000/25 and even more interestingly Gayraud et al (2003)* [*https://doi.org/10.1046/j.1365-2427.2003.01139.x*](https://doi.org/10.1046/j.1365-2427.2003.01139.x)

(Probably refers to Page 7, li. 40 -... in the PDF)

? Include in discussion

*p.24 li 11-12. This looks like a sentence for discussion not for results?*

*(*Probably refers to Page23*, li. 11-12* in the PDF*)*

“Overall, harmonised trait data and all aggregation methods produced similar distributions of RDA species scores compared with the original analysis, and they resulted in similar, but fewer, traits that distinguished upstream and downstream sites than the original analysis”

*p.24. Species scores for the first RDA axis ? what about other axes ?*

*(*Probably refers to Page23*)*

The RDA was carried out with one onstrained variable (conductivity). Hence, the first RDA axis corresponds to the only constrained axis (i.e. the conductivity gradient). All other axes are actually PCA axes (unconstrained axis) and represent the variance not explained by conductivtiy. Thus, they were not of interest for our analysis.

*p.28. This trait and many others was coded in the initial trait database proposed in a Freshwater Biology issue in 1994 Ecology of the Upper Rhône River: A Test of Habitat Templet Theories. Volume 31, 304 pp. ISSN 0046-5070 by Statzner et al. Surprinsingly, this trait was not included in the Tachet database.*

Probably refers to the trait piercer and this paper:

https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1365-2427.1994.tb01756.x

**Referee: 2**

**Comments to the Author**

*Review of the submission FWB-P-Mar-21-0143 entitled “Tackling discrepancies in freshwater invertebrate trait databases: Harmonising across continents and aggregating taxonomic resolution”. Authors compared databases of invertebrate functional traits compiled for different regions (Europe, North America, Australia and New Zealand) to identify differences in grouping features / traits, taxonomic resolution and trait codification. Also, authors explored different methods to aggregate trait data codified at genus level into family level data.*

*I congratulate the authors for this manuscript, which is absolutely timely and necessary. Building a global dataset of invertebrates is still challenging, but studies like this one will pave the road towards this goal. I appreciate that authors share the data and code necessary to reproduce their results.*

Thank you

*The manuscript is very well written and structured. Methods are sound and robust (but see my comment about using RDA to illustrate your example). I will just provide some comments that, from my point of view, can contribute to improve some aspects of the paper: 1) I suggest revising invertebrate trait definition and try to provide a broader perspective that links not only to environmental responses but also with ecosystem functioning. 2) The way in which authors illustrate the example to test the effects of the different type of trait aggregations is a bit hard to follow. I think that using RDA complicates a bit the whole thing and I suggest a potential alternative. 3) The fact that the authors are addressing so many different aspects (discrepancies in trait coverage, taxonomic resolution, trait codification, etc. + aggregation effects + illustration using the salinity study) makes the paper a bit dense, especially in the Methods section. Authors should try to better balance content for these goals and to re-assess if they could move some paragraphs to supplementary material to facilitate reading.*

We moved the paragraph on the “effect of phylogeny and trait variability on aggregation outcomes” to the supplementary materials. The simulation resulted largely in similar results for the different trait aggregations and also showed expected results (range of trait affinities increased with trait variability). We mention the simulation briefly in the “Comparison of family-level aggregated traits with family-level assigned traits” section in the methods part of the paper.

*Hope that the following comments are helpful for the authors. Apologies in advance if I misunderstood anything.*

*Invertebrate trait definition and related concepts*

*Considering that the “invertebrate traits” term is central to the paper and that authors are trying to harmonize terms and data, I find fundamental to provide an explicit definition of what authors consider by invertebrate traits.*

*Unfortunately, most studies dealing with invertebrate traits do not provide such a definition and this may cause some confusion. While the term “functional trait” is more commonly used in terrestrial ecology (Dı́az et al., 2001; Violle et al., 2007; Luck et al., 2012; Moretti et al., 2017), terms such as “biological trait” , “species trait” or “ecological trait” (Townsend & Hildrew, 1994; Usseglio-Polatera et al., 2000) are usual in freshwater invertebrate ecology*. ­*Functional trait definitions focus on individual characteristics that are influence organism performance (McGill et al., 2006; Violle et al., 2007), that determine responses to environmental change or effects on ecosystem functioning (Dı́az et al., 2001; Reiss et al., 2009).*

*In addition, the first lines of the Introduction (P6, L8-18) advantages of the invertebrate (functional) trait approached focused in its better capacity to predict responses to environmental change. However, (functional) traits are also useful to predict changes in ecosystem functioning and trophic dynamics (Gutiérrez-Cánovas et al., 2021) - see Díaz et al. (2001) definition*.

- Added this in the introduction

Other comments for the Introduction:

Previous works have advocated for a harmonisation of invertebrate trait databases and should be acknowledged in the introduction (Maasri, 2019).

Please, provide some example of studies showing limitations of the invertebrate trait approach (e.g. Hamilton et al., 2020)

Methodology to evaluate “Effects of harmonisation and trait aggregation on inferences regarding trait-environment relationships”

I guess that the justification of using the RDA method based on CMW traits is that authors wanted to keep the comparability against the original study. However, I found that the approach used here produced results and plots that are not very intuitive and may hamper the illustration exercise.

Alternatively, authors can compare salinity effects on CWM under different aggregation scenarios. Model coefficients (effect sizes) or measures of goodness-of-fit can be used to compare the performance of the different aggregations.

I understand that this is just an illustration of a potential application and by no means I don’t want to suggest unjustified efforts. I’m just asking for an improvement in the presentation of these results to enhance understanding.

Outlook

Authors mentioned several times that the word 'terminology' in the outlook to show that this study have set a starting point for the development of a standardized terminology. However, I don't feel that this is the strongest contribution of the paper (at least, I don't feel this should be placed here) and I suggest moderating the mention of implications regarding terminology (P32, 15-30), which are not really the focus of this study. This is not a big deal, but I think it's worth trying to produce a more focused Outlook section.

Specific comments:

P15, L40. Not sure if authors used a mixture of harmonised and non-harmonised grouping features. Please, clarify.

P7, L20: “Ecological property” is a very broad term. Please, provide a brief definition or replace by a more specific term.

P7, L31: There is something odd here: “When taxonomic composition or data in on trait database are at a more precise taxonomic level than other trait data…”. Please, revise and correct if necessary (e.g. “When taxonomic composition or data in a trait database are at a more precise taxonomic level than other trait data…”).

P10, L40-43: I wonder why authors did not combined (summed) trait affinities. If we assigned the highest affinity, we can underrepresent trait affinity. E.g. trait 1: affinity, 20%, trait 2: affinity 10%; combined trait 1+2: affinity should be 30%, not 20% (same logic if we use integer numbers). Please, clarify.

P30, L16-19: This sentence is difficult to follow. I suggest reworking to enhance understanding (try a more direct narrative). “Aggregation of trait information at the species or genus level to a single estimate at the family level implies a precision that is not necessarily present, especially not for traits with high variability or if trait information at species level is missing.”

Tano Gutiérrez-Cánovas

References

Dı́az S., Cabido M., Díaz S. & Cabido M. (2001). Vive la différence: Plant functional diversity matters to ecosystem processes. Trends in Ecology and Evolution 16, 646–655. <https://doi.org/10.1016/S0169-5347(01)02283-2>

Gutiérrez-Cánovas C., Worthington T.A., Jâms I.B., Noble D.G., Perkins D.M., Vaughan I.P., et al. (2021). Populations of high-value predators reflect the traits of their prey. Ecography, 1–13. <https://doi.org/10.1111/ecog.05438>

Hamilton A.T., Schäfer R.B., Pyne M.I., Chessman B., Kakouei K., Boersma K.S., et al. (2020). Limitations of trait-based approaches for stressor assessment: The case of freshwater invertebrates and climate drivers. Global Change Biology 26, 364–379. <https://doi.org/10.1111/gcb.14846>

Luck G.W., Lavorel S., Mcintyre S. & Lumb K. (2012). Improving the application of vertebrate trait-based frameworks to the study of ecosystem services. Journal of Animal Ecology 81, 1065–1076. <https://doi.org/10.1111/j.1365-2656.2012.01974.x>

Maasri A. (2019). A global and unified trait database for aquatic macroinvertebrates: the missing piece in a global approach. Frontiers in Environmental Science 7, 1–3. <https://doi.org/10.3389/fenvs.2019.00065>

McGill B.J., Enquist B.J., Weiher E. & Westoby M. (2006). Rebuilding community ecology from functional traits. Trends in Ecology and Evolution 21, 178–185. <https://doi.org/10.1016/j.tree.2006.02.002>

Moretti M., Dias A.T.C., de Bello F., Altermatt F., Chown S.L., Azcárate F.M., et al. (2017). Handbook of protocols for standardized measurement of terrestrial invertebrate functional traits. Functional Ecology 31, 558–567. <https://doi.org/10.1111/1365-2435.12776>

Reiss J., Bridle J.R., Montoya J.M. & Woodward G. (2009). Emerging horizons in biodiversity and ecosystem functioning research. Trends in Ecology and Evolution 24, 505–514

Townsend C.R. & Hildrew A.G. (1994). Species traits in relation to a habitat templet for river systems. Freshwater Biology 31, 265–275. <https://doi.org/10.1111/j.1365-2427.1994.tb01740.x>

Usseglio-Polatera P., Bournaud M., Richoux P. & Tachet H. (2000). Biological and ecological traits of benthic freshwater macroinvertebrates: Relationships and definition of groups with similar traits. Freshwater Biology 43, 175–205. <https://doi.org/10.1046/j.1365-2427.2000.00535.x>

Violle C., Navas M.-L., Vile D., Kazakou E., Fortunel C., Hummel I., et al. (2007). Let the concept of trait be functional! Oikos 116, 882–892. <https://doi.org/10.1111/j.2007.0030-1299.15559.x>

**Referee: 3**

**Comments to the Author**

This is an interesting paper that presents a detailed approach for aggregating and harmonizing

invertebrate trait information from trait databases of different regions. This is largely a methodological paper and the authors have done an excellent job to detail the process for future reference and reproducibility. I particularly like the sensitivity analysis ad reanalysis of the harmonized/aggerated data to test the effects of salinization. The paper is well written, but the methods are understandably long- I am wondering how these can be repackaged so that a reader is not lost in the details. I also found some typos and instances of repetition that can be checked and rectified during revision. I only have minor specific comments that I have captured below.

Specific comments

Abstract: The abstract is rather too long and is difficult to capture the problem and key findings and recommendations of the study. They need to be revised, shortened and sharpened for clarity.

Pg 8, Ln 31: Correct ‘data in on trait’ for clarity

Pg 16, Ln 26-38: What is the difference between these two paragraphs? They seem to have the same information.

Figure 3: The colours used here are difficult to differentiate and need to be improved.

Pg 24 Ln 12: By saying that ‘may be useful’ shows that you have very low confidence in your harmonized and aggregated trait data. This is really where the whole study on harmonization rests, and you should be more clear on the reliability of the trait data you have created. More discussion on what could be the reasons for this low confidence, or performance in evaluating the effects of salinization is needed.

Legendre, P., Galzin, R., & Harmelin-Vivien, M. L. (1997). Relating Behavior to Habitat: Solutions to the Fourth-corner Problem. *Ecology*, *78*(2), 547–562. https://doi.org/10.2307/2266029

Miller, J. E. D., Damschen, E. I., & Ives, A. R. (2019). Functional traits and community composition: A comparison among community-weighted means, weighted correlations, and multilevel models. *Methods in Ecology and Evolution*, *10*(3), 415–425. https://doi.org/10.1111/2041-210X.13119

Niku, J., Hui, F. K. C., Taskinen, S., & Warton, D. I. (2021). Analyzing environmental‐trait interactions in ecological communities with fourth‐corner latent variable models. *Environmetrics*. https://doi.org/10.1002/env.2683

Szöcs, E., Coring, E., Bäthe, J., & Schäfer, R. B. (2014). Effects of anthropogenic salinization on biological traits and community composition of stream macroinvertebrates. *Science of The Total Environment*, *468–469*, 943–949. https://doi.org/10.1016/j.scitotenv.2013.08.058