Peer Review File

Manuscript Title: Environmental performance of blue foods

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referee #1 (Remarks to the Author):

Summary of the key result
- The authors have brought together information about four main environmental stressors attributable to the bulk of aquaculture and fisheries production of blue foods globally. They have analyzed the data in a way that permits examination and ranking of the various production methods for different species groups. Using their methodology, the authors have also indicated the potential influence of various intervention scenarios that are designed to reduce environmental impact.
- To me, the most interesting and potentially useful outcome of this study is the comparison of production methods across many species. The comparison with poultry production provides an interesting perspective.

Originality and significance: if not novel, please include reference
- This study is original and significant in its findings. The scale and scope of the data set the study apart.
- The information and findings will be of interest to a broad spectrum of researchers, including those interested in aquaculture (production, feed production, etc.) and harvesting, and can be helpful in developing research priorities.

Data & methodology: validity of approach, quality of data, quality of presentation
- The authors have applied a sophisticated set of analytical approaches to examine their data. Life Cycle Analysis has been used successfully to look at environmental impacts of a variety of species and production techniques for many years and this has been used extensively here. The quality of the data is high, although there are some gaps, which are mostly acknowledged, e.g., the analysis stops at the farm gate and so does not account for potential environmental cost from transportation or marketing, and there is no consideration of differences in dress out percentages (proportion of production used for human consumption). These considerations may be difficult to quantify and may not greatly affect the research conclusions.
- There is a wide disparity in the number of observations for different species, as indicated in Table S1. Many species are represented by a single farm (n=1), while others are represented by over 400 (shrimp L. vannamei). Most of the species, with a couple of exceptions, that are represented by only a few farms are relatively low in production. While multiple replication is always preferable and excess reliance on information from low n is to be avoided, I believe that in this case, the data are of satisfactory quality, despite the low n in some cases.
- I didn't see consideration of freshwater capture fisheries included in the analyses. The authors should either provide this or provide a comment as to why this is left out (e.g., insufficient or unreliable data).
- I suggest that the y-axes in Tables S2 and S3 be expressed in million tons. E.g., in S2, instead of 2.5e+07, use 20 and indicate in the axis label that these are million tons.
- I think in the equation in line 486, the authors intended to say phosphorus instead of nitrogen.

Appropriate use of statistics and treatment of uncertainties
- I am not an expert in the principal statistical methods used in this study, but from my reading of similar work, it seems that the authors used appropriate statistical methods. They provide clear descriptions of the variability in their data.

Conclusions: robustness, validity, reliability

- The conclusions of these study appear to be robust and valid. The analysis of impact of intervention scenarios provides some insight into the robustness of the conclusions.
Suggested improvements: experiments, data for possible revision
- I suggest that comparative analysis of emerging production technologies, especially recirculating aquaculture systems (RAS) relative to current production methods would provide interesting information. Of course, this is especially challenging as this is a rapidly developing area, especially for salmonids.
References: appropriate credit to previous work?
- References appear to be appropriately cited.
Clarity and context: lucidity of abstract/summary, appropriateness of abstract, introduction and conclusions
- The clarity and context of the text is of the highest quality. The abstract is appropriate and lucid, as are the introduction and conclusion.

Referee #2 (Remarks to the Author):

In the manuscript titled "Environmental performance of blue foods" the authors develop a novel method to investigate the environmental impacts of freshwater and marine food sources. They find that, across five environmental indicators, lower-impact foods such as bivalves and seaweeds in aquaculture systems, and herring and cod in wild-capture fisheries, can have one tenth the impact of their higher-impact counterparts. Similarly, they also show how the impacts of producing a given product can be highly variable. They continue on to show how the impacts of blue foods could be reduced through various strategies, such as improved feed conversion ratios, yields, and feed composition.

The authors further discuss how the five environmental impacts they measured are only a small subset of the impacts that arise from fish production. These include impacts such as antibiotic resistance, invasive species, by-catch, and disease spread. I appreciate that the authors acknowledged that their analysis is not an all-inclusive estimate of the environmental impacts of blue foods, and applaud them for being upfront when recognizing this limitation.

This analysis is needed, will be very useful, and can be seen as an aquatic counterpoint to Poore and Nemecek (2018, Science). Despite the large diversity in blue food production systems, previous life cycle meta-analyses have limited resolution on these food, aggregating all blue foods into between 1-4 categories as seen in Poore and Nemecek, Clark and Tilman (2017), Nijdam (2012), and de Vries and de Boer (2010). Results from the current analysis can therefore be used in a variety of ways, for instance into global food system modelling (and associated scenario analysis), or as an input to create ecolabels for individual food products.

The major limitations of the manuscript is that it does not adequately acknowledge the broader food system and environmental contexts, and that additional information on the methods would be helpful in the main text. More on these major comments, as well as on minor comments and in-line comments are below.

Major comments:
1. Additional nuance of the broader food system and environmental contexts would be helpful throughout the manuscript. A few instances are pointed out below.

L 73: What are the potential implications of large increases in seaweed production? Can these be discussed to some extent, or is it possible to mention that large increases could result in environmental impacts and other shifts we may not be aware of? I'm concerned that as phrased, this paragraph could be interpreted along the lines of "seaweed is low-impact per unit, therefore large increases would also be low-impact". Yet we've seen that large production of low-impact per unit terrestrial based foods (soy, maize, etc) can have high environmental impacts.

L 168: I appreciate the analysis on the energy and GHG efficiency of increasing stock biomass, but I'm also concerned about the knock-on implications of this strategy. For instance, how would stock biomass be increased? Given the general global patterns of most fisheries being fully exploited or over exploited (FAO 2020), increasing stock biomass implies reducing catch rates until stocks have recovered. Does this mean then mean that fish captures would need to decrease in the short and medium time frames? If so, how can this be incentivised, and what implications would this have for food, nutrition, and economic security until stocks have recovered, especially for populations reliant on blue foods for nutrition and income?

L 183-194 (and to a lesser extent the following two paragraphs): This paragraph discusses the potential of blue foods in more sustainable diets. This is great, but it also needs to mention the potential knock-on implications and limitations of transitioning to blue foods. For instance, increasing consumption of capture fisheries and reducing consumption of e.g. poultry or ruminant meat would certainly decrease land use and water use, but how possible is this given the current status of fisheries? If this dietary transition is achieved, what would this mean for the short- and long-term sustainability of fisheries? For substitutions between higher-impact to lower-impact blue foods from fisheries, are lower-impact fisheries underexploited compared to high-impact blue foods? Alternatively, if it is unlikely that demand for blue foods under such a dietary transition could be met from fisheries and would instead need to be met through aquaculture, then this should be mentioned.

2. More information on the methods in the main text would be very helpful.
The paragraph in L35-42 is a great summary of what the authors have done, but I do not know if it will be accessible to readers unfamiliar with life cycle approaches. Some additional information on the methods here (or elsewhere) could help overcome potential communication barriers. For instance, what aspects of production systems are included (e.g. are the impacts of land cover change included?; this and other boundaries are shown in Fig S1, but is not clear from the main text), is there an alternative way to describe "resource input data" that will be more accessible to a general reader?

I recognise the authors are working within stringent word limitations and that most of this is covered in the supplement, but I'm concerned that the current explanation of methods in the main text will prevent a general reader from understanding how the study was conducted.

Minor comments:
1. I recommend also performing the analysis on an edible weight basis. I recognise this will require some assumptions on the proportion of a blue food that is consumed (which often varies across regions, as the authors mentioned in the SI), but this will help put the different blue foods on a more similar scale.

Please explicitly mention the functional unit in the main text before describing the results. This was a bit unclear until reading the caption of Fig 1.

In-line comments:
L21-22: Would it be helpful to reframe this slightly? I think additional specificity here could help communicate the scope of the nutrition issue to less familiar readers. For instance, more than half the adult population is currently malnourished, either being overweight or underweight.

L53-57: Some additional specificity to the land use impacts would be useful, especially if this is used to discuss the biodiversity impacts of different systems. Are there certain systems (pond, etc?), species (shrimp, etc?), or world regions where aquaculture is particularly responsible for conversion of natural habitats? Land use isn't a particularly good proxy for biodiversity impacts because land use does not always imply land conversion, and further, the association between land conversion and biodiversity outcomes is not co-linear and varies by region and type of species. Additional nuance on the systems/species/etc that are responsible for natural habitat conversion would help overcome this limitation. If this isn't possible to address, either through the methods or through published literature, then please reframe this paragraph to avoid the potential interpretation of high land use necessarily implies high biodiversity loss.

Fig 1. Three general thoughts:
1) I enjoy the figure layout, but found some of the colour bands difficult to differentiate. This would likely be even more difficult for people with colourblindness, especially for the reds and greens in panels (c) and (f).
2) Following the above, would a different colour scheme or layout help? Possibly boxplots (although this would lose some of the variance currently shown in the figure), or alternatively adding small bands of black/white to differentiate between the different percentiles of each blue food?
3) Would it be helpful to visualize the impacts for the aquaculture systems on radar or spider plots as a potential SI figure? This would make it possible to compare the performance of each blue food across the five indicators in a single plot, instead of needing to look at multiple panels for a given food.

L 102-110: Would it be worth broadening the discussion to also include non-environmental aspects of sustainability? For instance, culture, social, and economics? I recognise the focus of the paper is on the environmental performance of blue foods, but it's becoming increasingly important to recognise the roles that foods play on the non-environmental aspects of sustainability.

L117-118: If there is space, it would be useful to explicitly mention a few of these non-linear relationships and interactive impacts between different stressors.

Box 1 and Fig 3: Is it possible to account for uncertainty in the risk index of different types of fish?

L152-156: Is yields in this context fish yields in aquaculture systems, yields of crops that are used for fish feed, or both?

L157-167: Three general comments:
1) What is the aggregate benefit of implementing the different interventions simultaneously?
2) Would energy sourcing be a useful lever to look at, at least for the high energy use aquaculture systems?
3) Would it be useful to mention that deforestation free feeds would have other environmental benefits, such as biodiversity?

Fig 4: In panel (a), the colours are very difficult to tell apart, especially the pale yellows and greys. In panel (b), do the lines indicate the change from 0%? Would it be possible to add uncertainty to these estimates given that some of the categories are aggregates of multiple species and/or systems? Alternatively, would it be possible to perform sensitivity on the assumption that e.g. yields go to the upper 20th percentile to add error bars, for instance by evaluating the potential gains at the 10th percentile and the 30th percentile?

Paragraph starting L 202: Can this paragraph be more specific to increase actionability and tangibility? For instance, who can do what, and how can it be done? A short list in this paragraph, supplemented by a longer table in the SI would increase the onus for individual actors (or sets of actors) to shift their behaviour, rather than resulting in the finger-pointing (from individuals, from gov, from business, from producers, etc) that typically results from high-level analysis such as this one.

Methods and SI:

1) General: The approach is sensible and similar to other life cycle analyses. I'm not an expert in Bayesian analysis, so do not feel comfortable commenting on the specifics of the model set up and use of priors.

2) Testing the model: Apologies in advance for my potential misunderstanding of Bayesian stats and Markov Chain Monte Carlo analyses. My understanding is that the framework of the model in this analysis is:

"LCI database + data from paper publications -> MCMC analysis -> estimated results, including extrapolating"

where the role of the MCMC analysis is to create a distribution of estimates and the range of potential error given the data inputs in the LCI database and paper publications.

Is it possible to test the accuracy of your model within this framework, for instance by comparing the estimated results to the results in the paper publications, and/or withholding some of your data when constructing the MCMC model? If it is, this should be done. I realize that this will not be a perfect comparison given the difference between e.g. system boundaries, LCAI vs LCI databases, etc, but this would (a) provide a sense check of how well your model is estimating the impacts of different systems, which would be particularly important for the species with small data samples, and (b) allow you to improve the model by identifying the type of systems where thee model is least accurate.

L 541-549: I recommend adding comparisons with beef and a plant-based food, in addition to the comparison you have with poultry. The comparison with poultry throughout the text and figures is helpful, but it also assumes that most people reading the article or alternatively seeing Fig 1 on e.g. Twitter have general background knowledge in the field. Adding similar lines for beef and a plant-based food would avoid this assumption and increase accessibility to the paper's results.

SI:
Fig S4: The colours are really hard to differentiate.

Fig S12: Please add more information to the figure legend to make the comparison group more apparent.

Fig S14: Similar comment as for Fig 4a – some of these colours are difficult to tell apart.

Referee #3 (Remarks to the Author):

This paper by Gephart et al entitled "Environmental performance of the blue foods" is part of a collection of papers which are the major outputs of the international initiative "Blue Food Assesment" drawing together a diverse set of scholars to identify, review and fill important gaps in our understanding of the role of aquatic foods in global food systems, now and in the future. I strongly support the publishing of this type of meta-research in a journal with a srtong impact like Nature because this paper and its companions are timely. Blue foods are already a principal source of protein for some 3 billion people. In many of the most vulnerable countries, they are also an essential source of long chain polyunsaturated fatty acids and vital micronutrients, such as iron and zinc, and constitute a key preventive measure against premature deaths and stunting. However, Blue Foods tend to be ignored in discussions about how to feed the world's growing population sustainably during the next decades. Their environmental footprints can vary significantly. Some fisheries and farms are responsible for significant greenhouse gas emissions or ecological impacts. But many aquatic foods outcompete terrestrial livestock in terms of environmental impacts. Given their relative health benefits, comparatively small environmental footprint, and the potential for a substantial expansion of global production, foods from ocean and freshwater systems have a key role to play in achieving both a healthy and sustainable food system. Yet many discussions of food neglect blue foods altogether. And when they do include blue foods, they tend to deal with them in sweeping generalizations – typically calling for massive expansion of aquaculture. It is therefore essential to assess the real impacts of the different types of aquaculture and guide its future developments.

This manuscript builds upon a simplified Life Cycle Assessment, developed to allow for broader and more rapid environmental assessments of aquaculture systems, using a harmonized and elegant methodology. This approach provides an overview of the environmental impacts of the most commonly aquaculture species farmed for food provision worldwide. The comparisons between studies on different production systems conducted using different methodologies were harmonized using life cycle inventory (LCI) data, rather than characterised life cycle impact assessment (LCIA) results. This novel and globalized analysis allows to compare very distinct geographical regions and aquaculture systems. The data from the previous inventories (feed use, farm size, etc.) were used to recalculate greenhouse gas emissions, nitrogen emissions, phosphorus emissions, terrestrial land occupation, and consumptive freshwater use for aquaculture systems which are measured as the major environmental stressors. I did not find any weakness in the careful methodology that is provided in the methods section and in the supplementary files. This report is based on data collected from over 1,590 farms and 1,000 unique fishery records around the world for 23 species groups covering over 70% of global blue food production.

The results are quite clear and across all blue foods, farmed bivalves and seaweeds generate the lowest stressors.

Unfortunately, despite they also represent the largest volumes of production for the marine aquaculture, only a few LCIA have been conducted for these two groups except for mussels in comparisons with fish or crustacean farming. I think this represent a weakness in the meta-analysis.
The studied seaweed species are considered by the authors as providing food in comparison with the excluded volume of industrial seaweed cultivation. In fact, Gracilaria (Agarophyton) chilensis is grown in Chile essentially for agar extraction which is used in food texturing but also mainly for microbiological and biotechnological applications. Presently Macrocystis pyrifera is also not considered as a major edible species, nor Laminaria digitata and Saccharina latissima. It is a pity that no LCIA or LCA are available for the major seaweed crops grown in thousands of farms in South Korea, Japan and China (Pyropia spp;, Undaria pinnatifida and S. japonica). Indeed, LCIA were mainly conducted to assess the potential of growing large volumes of seaweeds for producing bioenergy and not for human food. I would suggest to the authors to encourage such studies in a near future to help emerging countries to invest in the potential of cultivating edible seaweeds and promoting alternative diets including seaweeds.
Moreover, the standard stressors used for LCIA do not take into account the ecosystem impacts most notably biodiversity loss. This point is discussed in the manuscript. However, this discussion provides only global impacts that are summarized in Fig.2.
It is clear that the development of the shellfish industry worldwide had a strong impact on the local biodiversity and on the ecosystem functioning. The introduction of the Japanese oyster Crassostrea (Magalenna) gigas led to its escape to establish wild settlement but also to the introduction of dozens of alien species belonging to various groups of marine invertebrates and algae. It also allowed the expansion of epidemic diseases in oyster farms which strongly impacted the producers economically and socially.

When assessing the environmental benefits of cultivating seaweeds to mitigate eutrophication, and acidification, to prevent oxygen minium and to capture carbon dioxide, it is important to keep in mind that Volatile Short-Live Substances are also an important by-product of the seaweed photosynthesis and metabolism. Their production rate varies by 3-4 orders of magnitude (Leedham et al., 2013 Biogeosci., 10, 3615) depending on algal species, compositions and the metabolic contribution of associated microbiota (King et al., 2001, Mar. Ecol. Prog. S., 224, 69), as well as abiotic and biotic stress factors. Some VSLS like bromoform and iodine oxide contribute to stratospheric ozone loss and increased surface UV radiation (Dessens et al., 2009 Atmos. Sci. Lett., 10, 201; Oman et al., 2016 Geophys. Res. Lett., 43, 9869; Fernandez et al., 2017 Atmos. Chem. Phys., 17, 1673). The impact of VSLS on stratospheric ozone depends strongly on the location of their emission. The tropical areas around coastal Southeast Asia and Australia are of major importance, as they are the main global entrance regions of tropospheric air masses into the stratosphere (Tegtmeier et al., 2012 Atmos. Chem. Phys., 12, 10633) and coincide with the largest macroalgae industrial production worldwide. VSLS emissions from the mid-latitudes and polar regions on the other hand, are less likely to reach the stratosphere due to their fast atmospheric degradation. Their environmental impact needs to be assessed. Budgets for the non-CO2 trace gases from macroalgae habitats, their effects on the atmosphere and development under climate change are lacking in general. As stressed by Popkin in a featured news in Nature (vol. 565, January 17th, 2019),| terrestrial tree forests trees are supposed to slow global warming, but growing evidence suggests they might not always be climate saviours.

The manuscript also concludes that capture fisheries predominantly generate greenhouse gas emissions, with small pelagic fishes generating lower emissions than all fed aquaculture, but flatfish and crustaceans generating the highest. Among farmed finfish and crustaceans, silver and bighead carps have the lowest greenhouse gas, nitrogen, and phosphorus emissions, but highest water use, while farmed salmon and trout use the least land and water. The provided model allows also to identify and quantify improvement opportunities and discuss public and private policy options to realize these improvements. There is no doubt that these conclusions will help policy-makers to select and support research programs and pilot project developments to improve the performances of existing systems and increase sustainability.

Referee #4 (Remarks to the Author):

The paper addresses a highly relevant topic, and given the role that "blue foods" have to play in the coming years and decades, its ambitions are very laudable, i.e. providing a quantitative measure of environmental performances of blue foods (as reflected by title). Conceptually, the provision of estimates from the different taxa groups for selected environmental metrics is indeed relevant and could be useful for further studies in food science. So is the approach to estimate reductions of some environmental stressors, which could bring relevant insights on which levers of fisheries/aquaculture systems to act on.

In spite of these good intentions, I found several critical problems with respect to the documentation of the study (or rather its lack thereof!), the methodological approach used, and the overall content/relevance of the paper. In my opinion, resolving all these issues is a necessary step before the work can be considered again for publication in any journal.

I explain each of them below, followed by additional detailed comments, hoping that it can help the authors rework their study.

1. *Complete lack of transparency in the data behind the estimates*: I am very surprised (not to say "shocked") by the lack of transparency in the study. The paper states early on: "We leverage data from over 1,590 farms and 1,000 unique fishery records around the world for 23 species

groups covering over 70% of global blue food production." (l.38-39). This is the data basis for the entire study, and my first reaction when reading was "Waouh, great!". Unfortunately, my heart sank very rapidly when I realized that this statement is given without any cross-references to where the data are documented. I have been looking into the Methods section, the SI and other papers, incl. by Halpern et al., etc. but to no avail. The Methods section reiterates the statement, and for aquaculture systems, it additionally specifies "Aquaculture inventory data were extracted from a database of aquaculture life cycle assessments initially compiled by1 [Parker et al. 2012] and updated with more recent studies (l.359-360) [….] resulting in a dataset of 112 species-specific LCA models from studies published between 1995 and 2020" (l363-364). As a reader, I would expect as a minimum that these "LCA models" and their references be provided transparently at least in Supplementary Material, with additional key features of the studies (like normally done in most review studies). As a reviewer, access to these information is essential to be able to check the comprehensiveness of the review (for which I have serious concerns; see below comments) and its validity as to the processing of the extracted data (for which I also have concerns; see below too). With the present documentation in the manuscript (and SI), I could not check those aspects, which are key to any peer review process. Note that I give the example of aquaculture above, but this comment applies to all data used in the study. In general, there is no comment as to the origin, quality or validity of the data sources in the paper. Addressing completely these issues of data transparency and quality check is a requirement before the paper can move any further in any scientific journal.

2. *Unsupported claim of data representativeness*: in several places of the manuscript, comprehensiveness of the study is advanced (e.g. "70% of global blue food production", l. 39; "Our results represent the most comprehensive and standardized blue food stressor estimates to date", l.83-84; etc.). The manuscript, incl. Methods section, makes references to Figs. S2 and S3 to back up these statements on representativeness and comprehensiveness. As far as these two figures are presented, they however do not give any indications of either data representativeness nor data comprehensiveness. Figure S3 does not describe the extent of the production originating from the covered farms; this is what would be meaningful to compare to the total national production (right now the plot is useless). Likewise, Fig. S2 shows total aquaculture production on y-axis by number of distinct farms covered in the study on x-axis, but where can we see how many farms in total exist globally (to get an idea of farm coverage)? For example, it seems that a bit less than 50 farms were covered for bivalves, but how many are there in the world? And where are those 50 farms located compared to the global distribution of bivalve farms? And how representative the technologies covered in these 50 farms are of the technology mix for bivalve farming in the world? Based on the graphs displayed, I could go arguing that the 50 farms covered in the study are producing 0.001% or 99.99% of the farmed bivalve production globally, that they are located in a single country or spread in 50 different ones, using the same technology or 50 different types. Figure S2 and S3 would not be able to contradict or clarify any of my statements. This is a major problem as those statements/questions I raise all relate to data representativeness, incl. technological and geographical representativeness (you may also add temporal representativeness, but it is strongly linked to technologies) and also have implications on completeness and uncertainty/variability evaluation. Fully transparent description of these aspects are essential to check and justify the relevance of the study and its results.

3. *Relevance of the paper undermined by lack of depth, lack of contextual positioning and missing discussion on key aspects*: beyond the other problems that I mention above and below (which have implications on the reliability and consistency of the displayed results), I overall find that the result analyses and the discussion points in the paper lack depth and overlook important aspects when addressing environmental performances of blue foods. Here is a summary of some criticisms (not exhaustive):
- There is a lack of contextual perspective in the discussion points. The introduction introduces the challenges that food systems face and the role that blue foods could play, but no link to these aspects is made afterwards. The paper is very much focused on the supply side (and in a narrow perspective and not in a systemic perspective; see below sub-point) and makes very little

connection to the demand side and how blue foods could help address the challenges of food systems while ensuring environmental sustainability (although claiming to do so, e.g. l. 187). A signal that the relationship to a broader context is missing is that there is no positioning at all with the SDG framework, which, given the aims and scope of the study, could be expected (at least in introduction and some discussion aspects).
- The larger importance and role of aquaculture over fisheries (as source of protein) is not mentioned or discussed while the study has a rare advantage to cover both (which could offer opportunities for interesting discussions). In that context, aspects such as the sustainable management of fish stocks, which poses caps/limits to how much fisheries can be used to supply food for human consumption and therefore calls for increasing demand from aquaculture systems, are not mentioned.
- The analysis of the results is strongly biased towards comparisons across species/taxa groups, and there is little discussion on the different environmental performances that can vary considerably across technologies for a same species or within a same taxa group. This is particularly important for aquaculture systems, e.g. pond vs. RAS, etc., where different impact magnitude may be observed, and sometimes leading to trade-offs (when one technology performs better for some environmental problems but not for others). This has been commented in past LCA studies (likely several in the pool of retrieved studies), which could be used as support. In this context, the authors should also discuss the different opportunities to shift from high-impact technologies to low-impact technologies (beyond just fine-tuning some model parameters), although the lack of technology differentiation (e.g. farming systems) in the current study does not make this easy.
- The temporal perspective is also lacking as the study appear quite static and representing past/present state. It does not open up with actual technological outlook (emerging technologies and aquafeed are barely mentioned), as it only addresses impact reduction potentials based on scenarios that are disconnected from reality, e.g. variation of 10% on several parameters --> how realistic? How feasible? What implications on other parameters/system components? This overall dynamic or future-oriented take on the blue foods production sector would be relevant to address in relation to their environmental impacts.
- Conclusions/recommendations from the study (mostly after l. 182), which relate to shifting to low-impact blue foods, should be nuanced and revisited as some of them do not hold. There are some clumsy claims like in l.183, where it is not obvious from the study results how much reduction blue foods can bring; the comparison with chicken (only one displayed) could suggest that the authors recommend to switch most food supply (globally) to bivalves and seaweed... Other claims are mere overstatements. For example: "Our standardized estimates enhance the resolution of the potential role of blue foods within sustainable diets, highlighting opportunities to shift demand from relatively high- to low-stressor blue foods and from terrestrial animal-source foods to comparatively low-stressor blue foods." (l.187-189) This could be misleading to the reader and should be nuanced considering all the limitations of the study. In particular, the authors cannot use the term "sustainable" when only addressing a few environmental problems. Although there is a part of the paper focusing on other stressors, the very truncated coverage of environmental stressors is a major limitation and prevents any such claim about environmental sustainability. Many other impact categories, like toxic releases of chemicals, antimicrobial resistance, seabed damages, fish resource depletion, etc. are highly relevant to consider when addressing "environmental performance of blue foods". Addressing a systemic perspective covering the entire food systems (not just part of the production) is also required in such discussion.

4. *Novelty of the paper in need of justifications*: for aquaculture systems, Bohnes et al. (2019) already conducted a comprehensive review of LCA studies (in "Life cycle assessments of aquaculture systems: a critical review of reported findings with recommendations for policy and system development"; Reviews in Aquaculture (2019) 11, 1061–1079). This study is surprisingly not cited in the current paper, which is a major flaw considering their very close scope. Although the current paper goes beyond aquaculture systems alone and also bring different angles (e.g. testing different reduction potentials), it is important that the authors position their work in

relation to existing literature and argue of the novelty of their work (e.g. estimates of environmental impact intensities are provided in Bohnes et al. (2019), albeit to different extents). In addition, seeing that the authors overlooked such a spot-on literature source, I confess to be much worried about their ability to find all relevant LCA studies relevant to their review scope, and I am quite doubtful of their argued comprehensiveness with regard to the retrieved literature (see next point).

5. *Potential lack of comprehensiveness in the retrieved data*: I may be wrong since the list of considered studies is not documented in the study, but I have suspicions that some were overlooked. The Methods section states that complementary data for aquaculture about "FCR and feed composition data from Monterey Bay Aquarium's Seafood Watch program2 and farm-level data collected for a range of systems in southeast Asia3 [Henriksson et al. 2015] to improve geographic representativeness" (l.368-370). This gives the impression that only ref. 3 (Henriksson et al. 2015) was used to cover Southeast Asia. There are however other sources relevant, e.g. Henriksson et al. 2017 (Indonesia), etc. Several of these sources have already been identified in the (uncited) review paper on LCA of aquaculture systems by Bohnes et al. (2019), which identified 65 LCA studies (covering more than one environmental impact categories) with a total of 179 individual aquaculture systems assessed (+ 38 aquafeed systems). In that paper, it is apparent that more studies than just Henriksson et al. 2015 cover southeast Asia (other regions in the world should also be addressed). Seeing the publication trends in Bohnes et al. (2019), several additional papers must have been published since that review. Again a detailed account of the reviewed studies (and of the method by which they were identified by the authors) is needed to check comprehensiveness in the number of retrieved studies.

6. *Lack of details on raw data consistency and handling*: in addition to providing transparent accounts of which peer-reviewed studies were included to map LCI for fisheries and aquaculture systems, I also require more information on the data from Monterey Bay Aquarium's Seafood Watch program. I briefly checked the Seafood Watch program website, in particular the Methods section described online. Although I find such initiatives to build consolidated databases much needed for environmental sustainability assessment (we need more of such work!), I still miss background details on what studies were included in the database and how the harmonization across studies was performed. I could not find indications of validation/peer review of the database/tool, which, if existed, would anyway have to be recalled succinctly in the present paper since it is an important data source. Harmonising LCI data is not a simple exercise, as the authors of the present study seem to suggest in SI. It usually requires meta-analyses to harmonise methodological choices made in the different LCA studies (see, e.g. entire Special Issue of Journal of Industrial Ecology, https://onlinelibrary.wiley.com/toc/15309290/2012/16/s1, with several papers addressing the methodological challenges of performing meta-analyses and a number of applications for the energy sector). The conduct of such harmonization cannot be documented with just one paragraph in the Methods section and 2 pages of general information in SI. I miss how each study was handled, for example when other LCI modelling framework (system expansion, allocation with other principles) than the one selected by the authors were used: how was the previous allocation undone? How were system boundaries with previously applied system expansion adjusted for the specific credited processes?, etc.

7. *Inappropriate basis of comparison across taxa group*: I would question several claims in the second main part of the manuscript (l.43-96). Roughly summarised, this part takes each environmental stressor and describes comparisons across species to identify low/high environmental performances (e.g. "farmed seaweeds and bivalves generate the lowest emissions, followed by small pelagic capture fisheries, while flatfish and crustacean fisheries produce the highest", l. 46-48). I find such comparisons misleading since the estimated values of the stressors have been based on live weight basis, which is only mentioned in the Supplementary Information. The quantitative evaluation of the stressors is indeed totally disconnected from the demand side and ignores the functionality of the food systems, which could be defined to meet the nutritional needs of the national/global population (you may find interesting discussion on how to define such

functionality in Weidema and Stylianou, 2020). Although live weight has been widely used as basis of comparison in past LCA studies, and I concede there is yet no consensus on how to appropriately define a functional unit for food systems, the use of live weight has been clearly pointed out as an incorrect approach since it completely dismisses the functionality of food systems (see Avadi et al. 2015; Bohnes and Laurent, 2019; Weidema and Stylianou, 2020). The recommendations on satiety/nutrition from Weidema and Stylianou (2020) may be difficult to implement for such large-scale study (although likely the way forward), especially since post-fisheries or post-farming processes are not included, but the use of edible weight should be possible to implement and would definitely be a step forward. Even though cultural differences may mean including some uncertainties in the estimation of edible weight (currently raised as an argument for not going for it in the supplementary information), I would believe it would yield less uncertainty and more comparability across taxa group than done in the current estimates. Although discussion on functionality is outside the scope of such study, 1-2 sentences to clarify this topic are also required in the manuscript (and the Methods section should argue for the approach followed).

8. *Other detailed comments*:
- Title of the paper: it should be nuanced considering the limited coverage of environmental problems (dismissing relevant impacts like antimicrobial resistance, fish resource depletion/overfishing, seabed damages, etc.).
- l. 31: "employing inconsistent methodologies": the methodologies themselves are not necessarily inconsistent (LCA is an ISO-standardized methodology), but the lack of comparability between studies due to different methodological choices is the problem.
- l. 60: "consumptive freshwater use is largely limited to feed production and evaporation15": does this apply to freshwater aquaculture (which constitute the largest proportion of aquaculture)?
- l.72: in terms of impacts, "negative emissions" can yield either positive or negative impacts (if background is a nutrient-rich environment or not). Nuancing may be pointed out (as was done in Halpern et al. for example)
- l. 97-140 (ca. 3rd section of the manuscript): this section could be much sharpened, considering that such topic has been addressed in literature before. It is an essential aspect, which would deserve better introduction, perhaps an overview of relevant stressors to consider when addressing blue foods (beyond those covered), potential occurrences of trade-offs and risk of burden-shifting from one problem to another. A policy angle should also be considered here.
- l.105: antimicrobial resistance could be added as an important problem raised in aquaculture systems. See for example: Reverter et al. 2020 (Nature Communications 11, 1870).
- Box 1 (l. 128 and on): it is important to specify how such graph is to be used (not for showing relationships, but for supporting policy making, e.g. fishing crustaceans in Atl may be less GHG-intensive, but have higher risk for mammals ◊ trade-offs). The authors could also reflect more on the complementarity between impact assessment (e.g. carbon footprint and other stressors assessed) and risk assessment (from the text it seems the two are confused). Note also that if the authors wanted to illustrate trade-offs, there could be easier way at the level of stressors/impacts without involving risk assessment as done here.
- l. 141 "reducing impact": at what levels? globally, while providing food supply? nationally/regionally? or at technology/species level. The study, which does not have a systemic perspective, seems to only address the latter. This should be clarified.
- l. 147-156: the split between 2 paragraphs (in addition to the text that perhaps could be clarified too) is very confusing as they address the same points
- l. 152: to complement earlier major comment: how realistic is it to make variations of 10% for all parameters? Bringing the analysis into an actual context would be relevant instead of keeping the discussion to a statistical analysis. Some attempts are made sometimes, e.g. l.154-156, but it lacks in-depth discussions and connection to actual systems. In that example, the discussion is kept to keeping the same technology, but how about shifting to other technologies that may bring less impacts.
- l.154-156: what about other technology alternatives (instead of fine-tuning/optimising the same technology). Maybe some technologies should simply be abandoned because not much more

optimizable and emerging technologies should relace them (?) As mentioned above, such discussion is needed.
- l. 172-173: “this could create co-benefits for biodiversity impacts (Box 1).” With respect to biodiversity impacts, the coverage in the current study is far from being comprehensive (and certainly not guaranteed though Box 1). Box 1 cannot tell whether or not if low-emission gears used in Central America will increase or decrease the risk; so this is not a good reference. In addition, note that Box 1 actually illustrates and mainly describes trade-offs (increased risk) so no co-benefit (ref. to Box 1 at the end misleading here).
- l.173: “zero emission technologies” --> there will be emissions over the entire life cycle --> nuancing is needed and emphasizing the need for life cycle perspective to avoid burden shifting from fleet emissions to on-land emissions is important to mention here for nuancing. This is part of undertaking a systemic perspective (which was not done in current study).
Figure 4a: I am a bit puzzled by the results for eutrophication: how come that by reducing 10% of crop (feed?) the eutrophication impact increase by 10+%? I may misunderstand the figure (which is difficult to grasp), but I read it correctly, then it is contradicting with the fact that off-farm N and P contributions are negligible in Figure S.4 (assuming the latter is correct; see my comment for Fig. S4 below).
l.184: “most stressors” --> “most considered stressors”. As mentioned above, keep in mind that you only studied a very few environmental problems and other relevant ones are not considered.
- l.185-186: “Capture fisheries vary widely in their GHG emissions but are low impact with respect to the other stressors considered.” --> inappropriate and unsound statement. Taking some indicators as seabed damages, this statement is not true. Again, this discussion part needs to be nuanced considering that the study only has partial coverage of environmental problems. The disregard of this important limitation occurs frequently in the paper and needs to be corrected consistently.
- l.203: “young” ◊ maybe for aquaculture but certainly not for fisheries!
- l. 221-223: role of aquaculture ◊ could have been an entry point for the paper. a pity to wait till the end to get such a contextual statement
- l. 224: word “enable” is definitely an overstatement considering the lack of transparency in the study.

Methods:
- l. 356: ReCiPe mehods “mass balance calculations” --> what is meant? Have the authors altered the ReCiPe characterization factors for the different stressors? Please clarify. Please also specify the versioning of ReCiPe.
- l. 363: what is defined as “LCA models”; is it a study? Is it a compared alternative? You need to clarify!
- Remark: ecoinvent 3.0 is quite old (from 2014/2015) and had some bugs, incl. water flow balance. Any reason why a more updated version was not used? (current version accessible is 3.7)
- l. 419: how did you handle gaps, e.g. species or technologies not covered by emission factors --> hence no impacts? Also in studies reviewed: how to deal with studies only covering part of impact categories, e.g. not providing LCI for all 4 impacts? How did you handle those?
- l. 422-423: what is the coverage of species within each taxa group? Again information of what is actually covered in terms of species, geography and technology, etc. is lacking, rendering any discussion of representativeness impossible to check.

SI file:
- There is a general confusion of concepts between emission and impact in the study in both the manuscript and SI, yielding ambiguity when reading the paper. Table S1 suggests that some impact assessment is performed but the authors often refer to emissions and even units are sometimes incorrect (see Figure Sx, where the units for eutrophying impacts should be expressed in kg-Neq and kg-Peq). Scientific rigor on this aspect should be expected at least in a study addressing environmental performances.
- l.42-43 (and rest of paragraph): more details on system boundaries and scope of the assessment are required. This should describe and argue for what was included/excluded, data quality and

representativeness for specific processes and implications. There is a brief outline of how the "simplified LCA" approach was conducted, with some references to selected papers and LCI databases, but there is no discussion about data representativeness (e.g. how are the data from Agri-footprint representative of all feed production processes in use today? In terms of time, technology and geographical coverage, cf. scope definition of LCA studies).
- l. 51: please specify the correct version of ecoinvent 3. In the manuscript, it is 3.0 (which in my opinion is rather outdated now).
- l.55-61: the use of allocation needs to be defended (relative to the use of system expansion, which according to ISO14044 precedes the use of allocation). An attributional modelling as applied in the analysis-type study by the authors may warrant including system expansion to account for secondary functions that may differ across the different species/taxa groups (e.g. different byproducts) and across different technologies. This may be considered a better reflection of actual market mechanisms. The authors discusses the choice of different allocation methods, which is fine, but before doing so, I encourage them to justify the choice of allocation first.
- l. 77: please indicate ReCiPe 2016 (full name of the methodology)
- Table S1: This table is not informative at all because it does not provide any explicit references for the data described, which is a major flaw (see above comment on lack of transparency). It does not specify how many distinct studies there are either (e.g. the three "1" for aquatic plants: are they the same study or three different studies?; this can be important in terms of methodological consistency across studies). Meta data on time, technology and geographic representativeness associated with each species/taxa group is also absent, albeit essential here. Following up on previous comment: discussion should also be included on taxa groups associated with very few studies (in terms of uncertainties). For example, taxa group with very few studies of different nature (technology/species wise) could explain some of the high variance observed in l. 143-144 (e.g. miscellaneous marine and diadromous fish stressors)?
- Fig. S2: it is difficult to discern what happens in the cluster at the bottom left corner of the graph. Additional information in the caption should help appraise the situation there (may not be relevant once data representativeness is appropriately addressed). In addition, are those data points only the LCA studies on full aquaculture systems, or do they include the feed data points added (which should be separated, in my opinion)? Or is it representing the 1590 farms mentioned in the manuscript? There is a lack of clarifications on what is represented on this graph.
- l. 94: for transparency, I think the background data for all calculated stressor estimates should be disclosed.
- Fig. S4: I may have missed something, but is it not surprising not to have contributions to eutrophication impacts (negligible) from off-farm processes (especially in light of feed production that may consume fertilizers?)? Maybe worth checking, if not done already.
Figure 4 caption: have units on graphs not in captions. Also have the units right (Neq, Peq, etc.)
- Table S3: what is reported? Is it r-square here, or? Please refer to conventional ways of reporting statistical results in scientific publications.
- Captions for Figure S14 and others: complete lack of explanations to understand what the figure shows. Even if some explanations are in the manuscript for some figures, each figure should be self-standing and the reader should not have to piece things together to understand how to read the figure.

Recap of references cited:
Avadi A, et al. (2015) Comparative environmental performance of artisanal and commercial feed use in Peruvian freshwater aquaculture. Aquaculture 435, 52–66.
Bohnes FA, et al. (2019) Life cycle assessments of aquaculture systems: a critical review of reported findings with recommendations for policy and system development. Reviews in Aquaculture 11, 1061–1079.
Bohnes FA & Laurent A (2019) LCA of aquaculture systems: methodological issues and potential improvements. The International Journal of Life Cycle Assessment 24, 324–337.
Henriksson PJG, et al. (2017) Indonesian aquaculture futures – evaluating environmental and socioeconomic potentials and limitations. Journal of Cleaner Production 162, 1482–1490.
Reverter M, et al. (2020). Aquaculture at the crossroads of global warming and antimicrobial

resistance. Nature Communications 11, 1870.
Weidema BP & Stylianou KS (2020) Nutrition in the life cycle assessment of foods—function or impact? The International Journal of Life Cycle Assessment (2020) 25:1210–1216

Referee #5 (Remarks to the Author):

Fisheries and aquaculture are underrepresented in studies on the environmental impacts of foods. This manuscript uses secondary data at the production and input phases to calculate environmental impacts of live weight fisheries and aquaculture products (ie., farm gate /dock sales) using a mass allocation method.

The methods and approach are appropriate and the sample sizes are in line with what's available in the peer-reviewed literature. The main strengths of the paper are the use of life cycle inventory data instead of life cycle impact assessments, which gives this manuscript a distinct advantage over other similar papers. However, this may be lost on readers. It might help to emphasize this point as well as bring in the arguments made by co-authors in Henricksson et al. 2021. GFS https://doi.org/10.1016/j.gfs.2021.100508. The authors also take a global geographic scope including a large proportion of global production, but miss small scale fisheries and aquaculture which is unfortunate but I'm not sure if this is solvable due to data gaps.

Fisheries and aquaculture are included together, however, the treatment of aquaculture in parts of the analysis and supplement is a bit richer. Is this because fisheries already has good synthesis w/ life cycle methods? Or that the bounadry conditions for aquaculture are broader, thus more room for analysis? or something else.

Specific comments

L9 "…but flatfish and crustaceans generating the highest." add greenhouse gas emissions to the end of the sentence?

L77-9 "To compare with terrestrial foods, we estimated stressors for chicken…" I like that the authors selected chickens since they are the most similar in terms of environmental impacts to fish. As the authors note in the intro, granularity is important and something most LCAs miss with blue foods. In that spirit, it would be helpful to note here that the chicken data is from US/Europe and from intensive/vertically integrated farms, which gives readers key information with which to make the chicken/ blue food comparison.

L81-2 "Notably, the groups generating the lowest stressors (e.g., bivalves and small pelagic fishes) also provide the greatest nutritional quality across all forms of aquatic foods2,9." I'm not sure I agree. They are higher in omega-3 and vitamin B12, and zinc for bivalves but similar for most other micronutrients listed in Fig 1 of (Golden et al. in rev).

L125-8 I like the cross-cutting look at marine mammal risk x GHG for fishing gear types. It's important to show where synergies exist. Are decisions about gear ever made with fuel use in mind instead of what gear is most efficient for max sust. yield? I see this comes up again in L171-2 without refs and L215 supported with refs. In L171 I was left wondering how serious this idea was?

L184-5 "…fed aquaculture still outperforms many other animal-source foods." I see that the point of this paper was not to compare aquaculture and animal sourced food, but it might help readers if you specify which ones, for example beef and pork, but not eggs and poultry? Eat Lancet with a

Blue Lens has a nice comparison across species groups, but not with the granularity of this study.

L192 "Further, blue foods provide the highest nutrient richness across multiple micronutrients, vitamins, and fatty acids" as mentioned above, I think this overgeneralizes and oversimplifies the comparison. Should the authors specify what type of fatty acids (polyunsaturated)? If the comparison is between two groups (blue foods vs terrestrial foods) should it be "higher" instead of "highest"?

L144 "High variance in milkfish and miscellaneous marine and diadromous fish stressors points to large potential performance gains per unit…". Was sample size a cause of higher variance?

L152 "…a 10% yield improvement reduces…" I thought you meant yield in processing and then reread the supplement the boundary conditions and scenarios. It might be good to clarify this point.

L206 and L212 could these thoughts be combined? I agree that there is a need for more than just technology (L206) and was wondering when that would be mentioned.

L207 in addition to husbandry more generally, it would be useful to call out disease and mortality specifically as important factors.

Figures.

Fig 1. It may be helpful to define what "credible intervals" are for readers.

Fig 3. The authors switch from median and credibility intervals in Fig 1 to mean and 95% CI in Fig 3. Is this an oversight or is there a reason for switching?

Fig 4. The x-axis is a little bunched. Could the author stretch the plot wider for ease of reading Fig4 A and to distinguish differences in the lollipop plot in Fig 4 B.

Supplement
The fisheries component is not as well characterize in the supplement as the aquaculture data. For example, after reading Fig S1 I was wondering if there was similar table for fisheries?

Fig S4. Blue and green colors are difficult to tell apart

Author Rebuttals to Initial Comments:

We appreciate the time the reviewers took to provide thoughtful and detailed comments to improve our paper. We took the revision opportunity to make several substantial improvements to the paper and its documentation, including:

- Re-estimating the environmental stressors in terms of edible portion, as suggested by two reviewers. We still include the original estimates in terms of live weight in the SI and include both in the SI data file. This will allow our results to be used in a wider variety of contexts as the edible portion connects more directly to diets while the live weight values connect more directly to production. Reassuringly, while the estimate values naturally change depending on which of these is used, the general patterns and conclusions are not sensitive to this decision.
- Adding a new box to provide additional nuance to the shellfish and seaweed conclusions. These groups result in the lowest stressors within our model, but it is important to note some specific consideration related to these two groups. This also provides an opportunity to highlight some methodological decisions that arise in any environmental assessment study.
- Ensuring full study reproducibility. At the time of submission, we did not include all input data in the GitHub repository, though the repository itself was public. We have now made all

input data and code publicly available (https://github.com/jagephart/FishPrint). We also added a workflow figure to the repository so potential users can more easily see how the data files relate to each script and produce each figure in the manuscript. The repository was archived so that this version of the repository has a time stamp and citable DOI (DOI: 10.5281/zenodo.4768324). We also added an additional SI file with the full references of included studies.

- Emphasizing the novelty of the study and methodological approach. Through the reviewer comments, we realized there was some confusion about the novelty of our approach, which uses data underlying published studies and within a database to estimate stressors in a standardized way. This contrasts with studies which compile final stressor estimates from published literature and summarize them without accounting for underlying methodological differences. We added a sentence to the introduction and beginning of the methods to clarify this, though we did not want to add a lot of text about technical methodological points at the beginning of the manuscript, given the broad audience. Critically, we also added a methods flow chart to the SI, which more clearly illustrates the workflow.
- Expanding on broader sustainability and actionability. There are many related issues that one could touch on that are important and relevant to the topic of blue foods in the sustainable diets and how to achieve these. We could not discuss all points at length within the allotted space, but we have added some key text and references to expand our discussion of these issues.

We addressed the specific reviewer comments, point by point, in bold below.

Referee #1 (Remarks to the Author):

Summary of the key result

- The authors have brought together information about four main environmental stressors attributable to the bulk of aquaculture and fisheries production of blue foods globally. They have analyzed the data in a way that permits examination and ranking of the various production methods for different species groups. Using their methodology, the authors have also indicated the potential influence of various intervention scenarios that are designed to reduce environmental impact.

- To me, the most interesting and potentially useful outcome of this study is the comparison of production methods across many species. The comparison with poultry production provides an interesting perspective.

Originality and significance: if not novel, please include reference

- This study is original and significant in its findings. The scale and scope of the data set the study apart.

- The information and findings will be of interest to a broad spectrum of researchers, including those interested in aquaculture (production, feed production, etc.) and harvesting, and can be helpful in developing research priorities.

Data & methodology: validity of approach, quality of data, quality of presentation

- The authors have applied a sophisticated set of analytical approaches to examine their data. Life Cycle Analysis has been used successfully to look at environmental impacts of a variety of species and production techniques for many years and this has been used extensively here. The quality of the data is high, although there are some gaps, which are mostly acknowledged, e.g., the analysis stops at the farm gate and so does not account for potential environmental cost from transportation or marketing, and there is no consideration of differences in dress out percentages (proportion of production used for human consumption). These considerations may be difficult to quantify and may not greatly affect the research conclusions.

- There is a wide disparity in the number of observations for different species, as indicated in Table S1. Many species are represented by a single farm (n=1), while others are represented by over 400 (shrimp L. vannamei). Most of the species, with a couple of exceptions, that are represented by only a few farms are relatively low in production. While multiple replication is always preferable and excess reliance on information from low n is to be avoided, I believe that in this case, the data are of satisfactory quality, despite the low n in some cases.

Response: It is certainly true that some individual species have very low n, which increases the uncertainty in estimates for these groups. A strength of taking a Bayesian approach is that cases with few observations don't strongly influence the estimated parameter distribution, resulting in higher uncertainty in the estimates of the mean. Additional future data for these species would reduce the uncertainty in those estimates. We also accounted for differences in total production of different species within the species groups such that even if we had many observations for a species with little production, it is given less weight in the overall estimate.

The reviewer's point of some species having little representation in the literature is still very important. We expanded our description of the literature review process in the methods and Fig S1 to provide additional detail in the comprehensiveness of the data included. Then, by comparing the species production to the number of observations, we were able to identify groups of the highest priority for future data collection. We updated Figures S3 and S4 and added text to point to these priority species: "Further, farmed carps, bivalves, and seaweeds account for vast amounts of global production yet remain underrepresented in environmental assessments (Fig S3-S4) and therefore should be prioritized for additional research." (Lines 97-99). The underrepresentation of bivalves and edible seaweeds is further emphasized in the new Box 1.

- I didn't see consideration of freshwater capture fisheries included in the analyses. The authors should either provide this or provide a comment as to why this is left out (e.g., insufficient or unreliable data).

Response: This is a gap in our analysis, which we attempted to address but were unable to find sufficient data. One study did find very low greenhouse gas emissions from fisheries in Swedish lakes (0.77 kg CO_2-eq per kg edible portion) (Hornborg and Främberg 2020), but we were hesitant to point to this as it is not likely representative of the vast majority of inland fisheries. We were also able to find two FAO publications that touch on this topic (Welcomme 2011; Ainsworth et al. 2021), but they only state "Carbon emissions from capture fisheries was taken as the average carbon emissions from several freshwater fish species. Carbon emissions from capture fisheries stem from gear construction, fuel use from boats and refrigeration." We emailed the contacts to see if it would be possible to access the underlying data but have not heard back at this point. We also contacted another researcher at FAO who is working on this topic and he indicated that the work is still in progress and that few

countries have provided high quality data. These efforts suggest that there is not sufficient data at this point to include an estimate for inland fisheries or to properly represent many marine small-scale fisheries. We added text to ensure that this gap is more clearly communicated: “Our capture fishery data primarily represents commercial marine fisheries25. However, subsistence marine and inland catches often utilize non-motorized or no vessels which likely generate few emissions, but there is insufficient data on fuel use across the diversity of small-scale fishing methods to reliably estimate emissions.” (Lines 93-96).

References:

Ainsworth, R., Cowx, I.G. and Funge-Smith, S.J., 2021. A review of major river basins and large lakes relevant to inland fisheries.

Hornborg, S., & Främberg, A. (2020). Carp (Cyprinidae) fisheries in Swedish lakes: a combined environmental assessment approach to evaluate data-limited freshwater fish resources as food. Environmental management, 65(2), 232-242.

Welcomme, R., 2011. Review of the state of the world fishery resources: Inland fisheries. FAO Fisheries and Aquaculture Circular, (C942), p.I.

- I suggest that the y-axes in Tables S2 and S3 be expressed in million tons. E.g., in S2, instead of 2.5e+07, use 20 and indicate in the axis label that these are million tons.

Response: We agree this looks better. We updated these two figures and have improved their interpretability by changing the units as suggested and using a log scale so we can add labels to the points. Additionally, we scaled the points by the number of studies so that this context is clearer and to better identify priority areas for future data collection.

- I think in the equation in line 486, the authors intended to say phosphorus instead of nitrogen.

Response: That is correct. We made this edit. Thank you for catching this!

Appropriate use of statistics and treatment of uncertainties

- I am not an expert in the principal statistical methods used in this study, but from my reading of similar work, it seems that the authors used appropriate statistical methods. They provide clear descriptions of the variability in their data.

Conclusions: robustness, validity, reliability

- The conclusions of these study appear to be robust and valid. The analysis of impact of intervention scenarios provides some insight into the robustness of the conclusions.

Suggested improvements: experiments, data for possible revision

- I suggest that comparative analysis of emerging production technologies, especially recirculating aquaculture systems (RAS) relative to current production methods would provide interesting information. Of course, this is especially challenging as this is a rapidly developing area, especially for salmonids.

Response: We appreciate this comment and the importance of system type for the resulting environmental stressors. At a coarse level, the stressor distributions for each system type points to some potential trade-offs that have also been identified in other work (e.g., some RAS systems with high GHG emissions, but with low land and water compared to some pond systems; Fig S8). We decided to focus our study on the taxa group level since many food systems studies, policy decisions and consumer guidance are at this resolution, without information about production systems or other farming practices that are important for explaining the full range of estimates within a taxa group. While we collected system type information during our data compilation process, there are an insufficient number of system type-taxa group observations to produce reliable estimates for comparison in this study. We do see the taxa group level estimates as an important benchmark against which future technological improvements can be compared through future work and new data collection. We added text to emphasize the importance of system type and role of future technological advancement to underscore this as a key area for future work: "Our estimates represent a snapshot of current production, but future work on emerging production technologies, feed innovations and growing sub-sectors is important for tracking changes against these benchmarks." (Lines 99-100). We also added text in the "Reducing blue food environmental impacts" section to point to the importance of production system: "While some variability within a taxa group is due to differences in on-farm practices, production technology is an important factor across stressors49 as variability in stressors for a given species reared in different farming systems can be considerable (e.g., 50)." (Lines 187-190).

References: appropriate credit to previous work?

- References appear to be appropriately cited.

Clarity and context: lucidity of abstract/summary, appropriateness of abstract, introduction and conclusions

- The clarity and context of the text is of the highest quality. The abstract is appropriate and lucid, as are the introduction and conclusion.

Response: We would like to thank the reviewer for taking the time to review our work and for their positive review and helpful comments.

Referee #2 (Remarks to the Author):

In the manuscript titled “Environmental performance of blue foods” the authors develop a novel method to investigate the environmental impacts of freshwater and marine food sources. They find that, across five environmental indicators, lower-impact foods such as bivalves and seaweeds in aquaculture systems, and herring and cod in wild-capture fisheries, can have one tenth the impact of their higher-impact counterparts. Similarly, they also show how the impacts of producing a given product can be highly variable. They continue on to show how the impacts of blue foods could be reduced through various strategies, such as improved feed conversion ratios, yields, and feed composition.

The authors further discuss how the five environmental impacts they measured are only a small subset of the impacts that arise from fish production. These include impacts such as antibiotic resistance, invasive species, by-catch, and disease spread. I appreciate that the authors acknowledged that their analysis is not an all-inclusive estimate of the environmental impacts of blue foods, and applaud them for being upfront when recognizing this limitation.

This analysis is needed, will be very useful, and can be seen as an aquatic counterpoint to Poore and Nemecek (2018, Science). Despite the large diversity in blue food production systems, previous life cycle meta-analyses have limited resolution on these food, aggregating all blue foods into between 1-4 categories as seen in Poore and Nemecek, Clark and Tilman (2017), Nijdam (2012), and de Vries and de Boer (2010). Results from the current analysis can therefore be used in a variety of ways, for instance into global food system modelling (and associated scenario analysis), or as an input to create ecolabels for individual food products.

The major limitations of the manuscript is that it does not adequately acknowledge the broader food system and environmental contexts, and that additional information on the methods would be helpful in the main text. More on these major comments, as well as on minor comments and in-line comments are below.

Response: Thank you so much for taking the time to review our manuscript and for your positive feedback on the value of our work. We agree that there are aspects of the broader food system and environmental contexts that we were not initially able to incorporate due to space constraints, but we have worked to include greater acknowledgement of these issues throughout, especially relating to the specific points you raised. We respond to your comments below and appreciate the suggestions as we feel they have improved the manuscript.

Major comments:

1. Additional nuance of the broader food system and environmental contexts would be helpful throughout the manuscript. A few instances are pointed out below.

L 73: What are the potential implications of large increases in seaweed production? Can these be discussed to some extent, or is it possible to mention that large increases could result in environmental impacts and other shifts we may not be aware of? I'm concerned that as phrased, this paragraph could be interpreted along the lines of "seaweed is low-impact per unit, therefore large increases would also be low-impact". Yet we've seen that large production of low-impact per unit terrestrial based foods (soy, maize, etc) can have high environmental impacts.

Response: We agree that this is an important point, as no food production is free of environmental impacts and the total impact can be substantial even if the impact is low on a unit of product basis. Since life cycle assessment is best positioned to provide product-level estimates, rather than total impacts, we do not want to direct focus too much on the total impacts. In order to make sure this point comes through though, we added a brief sentence in the new Box 1, along with other nuances related to seaweeds and bivalves. This new box also points toward impacts associated with seaweed production that fall outside the system boundaries of our study:

"Box 1: Bivalves and seaweeds perform favourably across the estimated stressors, but also highlight key assumptions and nuances. First, bivalve estimates change by nearly five-fold when expressed in terms of edible portion (Fig 1) compared to live weight (Fig S10) due to the shell weight. Second, some processes falling outside our system boundaries represent a potentially large fraction of life cycle emissions for these groups, even if still small in absolute value in some cases. For seaweeds, a large proportion of GHG emissions can occur at the drying stage[26] while for bivalves, CO_2 emissions during shell formation[27] and high emissions associated with live product form transport[28] can be important. Third, impacts on biogeochemical cycling and habitats are highly context dependent. For example, the systems represented here extract nitrogen and phosphorus, which could be problematic in nutrient poor environments. Additionally, stratospheric ozone effects from volatile short-lived substances depends on the location and varies widely across species[29,30]. Fourth, sustainable diet recommendations based on these or similar results must account for differences in nutrition content and bioavailability, a particularly important consideration for seaweeds[31]. Finally, these systems are underrepresented in the literature, particularly for edible seaweeds (Fig S3). As recommendations point toward the potential of these groups, it is important to increase data on these systems, deepen understanding of the above nuances,

and be mindful of the total impacts associated with large-scale production on coastal habitats." (Lines 112-129)

We also added text to the discussion focused on several challenges to scaling up: "Major challenges remain for shifting demand, as well as meeting increased demand. While improved management offers potential opportunities for expanding some production from low stressor capture fisheries, uncertainty remains around the extent and feasibility of rebuilding many fisheries48. Additional research is needed to understand the total environmental impacts of large-scale expansion of low per unit stressor foods, especially for system-specific impacts (Box 1). Increasing production also requires creating appropriate incentives and reducing barriers for producers. Historical food system transitions required public investment technologies that could be scaled-up by the private sector and public policy leadership57." (Lines 246-253).

L 168: I appreciate the analysis on the energy and GHG efficiency of increasing stock biomass, but I'm also concerned about the knock-on implications of this strategy. For instance, how would stock biomass be increased? Given the general global patterns of most fisheries being fully exploited or over exploited (FAO 2020), increasing stock biomass implies reducing catch rates until stocks have recovered. Does this mean then mean that fish captures would need to decrease in the short and medium time frames? If so, how can this be incentivised, and what implications would this have for food, nutrition, and economic security until stocks have recovered, especially for populations reliant on blue foods for nutrition and income?

Response: It is true that rehabilitation of stocks will probably need reduction in fishing pressure and that this will have short term consequences for the many stakeholders involved (or possibly long term, depending on the characteristics of the fishery). Significant debate also remains about the feasibility and timeframe for increasing fishery stocks, which will influence the scope of potential consequences. As a result, we certainly agree that there are numerous potential unintended social and economic consequences of this. We would argue that this is likely true of other interventions as well (e.g., incentivizing specifc farming methods or species). We would also suggest that the likelihood of different consequences being incurred by producers would depend at least in part on the governance context and policy approach to implementing the intervention itself. While we cannot fully cover the scope of potential unintended consequences or necessary considerations, we added text to point to the need for context-specific evaluation of potential social and economic trade-offs associated with environmentally-oriented blue food interventions: "Further, without careful consideration for local contexts and inclusion of relevant stakeholders, environmentally focused interventions can generate

social and economic trade-offs that undermine broader sustainability goals." (Lines 287-289).

L 183-194 (and to a lesser extent the following two paragraphs): This paragraph discusses the potential of blue foods in more sustainable diets. This is great, but it also needs to mention the potential knock-on implications and limitations of transitioning to blue foods. For instance, increasing consumption of capture fisheries and reducing consumption of e.g. poultry or ruminant meat would certainly decrease land use and water use, but how possible is this given the current status of fisheries? If this dietary transition is achieved, what would this mean for the short- and long-term sustainability of fisheries? For substitutions between higher-impact to lower-impact blue foods from fisheries, are lower-impact fisheries underexploited compared to high-impact blue foods? Alternatively, if it is unlikely that demand for blue foods under such a dietary transition could be met from fisheries and would instead need to be met through aquaculture, then this should be mentioned.

Response: These are valuable points and we certainly agree with the complexity around these issues. While we have limited space to cover the full range of food-system wide implications, we added text to acknowledge the issues of diet substitution and to specifically address the issue of production potential and sourcing limitations for blue foods: "Major challenges remain for shifting demand, as well as meeting increased demand. While improved management offers potential opportunities for expanding some production from low stressor capture fisheries, uncertainty remains around the extent and feasibility of rebuilding many fisheries48. Additional research is needed to understand the total environmental impacts of large-scale expansion of low per unit stressor foods, especially for system-specific impacts (Box 1). Increasing production also requires creating appropriate incentives and reducing barriers for producers. Historical food system transitions required public investment technologies that could be scaled-up by the private sector and public policy leadership57." (Lines 246-253).

This comment also prompted us to add the point that even without expanding capture fishery production, these estimates help us value the "replacement cost" of avoiding fishery declines: "Capture fisheries vary widely in their GHG emissions but are low impact with respect to the other stressors considered. This underscores the value of sustainably managing wild fisheries to avoid the environmental replacement cost that would be incurred under fish catch declines24." (Lines 233-236).

Finally, we added text to the introduction to position the paper as serving as a benchmark for current production within an evolving sector: "It is therefore critical to examine the environmental performance across the diversity of blue foods in a robust, methodologically consistent manner to serve as a benchmark

within a rapidly evolving sector as blue food demand increases21, production shifts toward aquaculture and production technologies advance." (Lines 34-37).

2. More information on the methods in the main text would be very helpful.

The paragraph in L35-42 is a great summary of what the authors have done, but I do not know if it will be accessible to readers unfamiliar with life cycle approaches. Some additional information on the methods here (or elsewhere) could help overcome potential communication barriers. For instance, what aspects of production systems are included (e.g. are the impacts of land cover change included?; this and other boundaries are shown in Fig S1, but is not clear from the main text), is there an alternative way to describe "resource input data" that will be more accessible to a general reader?

I recognise the authors are working within stringent word limitations and that most of this is covered in the supplement, but I'm concerned that the current explanation of methods in the main text will prevent a general reader from understanding how the study was conducted.

Response: Thank you for highlighting this as we certainly want the value of our methodological approach to be communicated clearly to an audience beyond the life cycle community. In an effort to improve this without overwhelming the introduction with technical details, we 1) harmonized the language around the methods throughout the manuscript; 2) revised the methods summary in the introduction to read "We identify a set of key life cycle inventory data (i.e., material and energy input and farm-level performance data) from published studies and datasets to which a harmonized methodology is applied." (Lines 40-42); 3) added additional text to summarize the data inputs in the Methods section: "These data, mainly from published seafood LCA studies and complemented by additional data, together with background data from the ecoinvent v3.6 LCI database, were used to aggregate environmental stressors, considering lifecycle fuel use, electricity generation, agricultural field emissions, land-use change for Brazilian soybeans, among other things (see Fig S1)" (Lines 459-463) and a new methods flow chart to the SI (Fig S1).

Minor comments:

1. I recommend also performing the analysis on an edible weight basis. I recognise this will require some assumptions on the proportion of a blue food that is consumed (which often varies across regions, as the authors mentioned in the SI), but this will help put the different blue foods on a more similar scale.

Please explicitly mention the functional unit in the main text before describing the results. This was a bit unclear until reading the caption of Fig 1.

Response: We now express the results in terms of edible portion in the main text and indicate this functional unit in the caption of Figure 1. We moved the estimates in terms of live weight to the SI and include both sets of results in the stressor estimates file. We added an additional explanation of the functional unit issue in the SI: "Stressor estimates can be expressed in terms of different functional units such as the live weight at the farm gate or the edible portion. Stressors estimates expressed in terms of live weight equivalent enable the data to be easily connected to production data, which is also typically expressed in terms of live weight, while the edible portion is more relevant to blue food demand and the role of blue foods in diets. As a result, we calculate the stressors in terms of both edible portion (Fig 1) and live weight equivalent (Fig S10). It is encouraging that the general patterns and findings of this paper are largely unaffected by the choice of functional unit. The largest change is for the bivalves, so we added text to illustrate the issue of functional units in the new Box 1 (Lines 112-128). We also added "results expressed in terms of edible weight" in the first paragraph of the "Blue food environmental stressors" section (Line 57).

In-line comments:

L21-22: Would it be helpful to reframe this slightly? I think additional specificity here could help communicate the scope of the nutrition issue to less familiar readers. For instance, more than half the adult population is currently malnourished, either being overweight or underweight.

Response: We revised this sentence to add specificity. It now reads: "Yet, it still fails to meet global nutrition needs7, with 820 million people lacking sufficient food8 and with one in three people globally overweight or obese9." (Lines 21-23).

L53-57: Some additional specificity to the land use impacts would be useful, especially if this is used to discuss the biodiversity impacts of different systems. Are there certain systems (pond, etc?), species (shrimp, etc?), or world regions where aquaculture is particularly responsible for conversion of natural habitats? Land use isn't a particularly good proxy for biodiversity impacts because land use does not always imply land conversion, and further, the association between land conversion and biodiversity outcomes is not co-linear and varies by region and type of species. Additional nuance on the systems/species/etc that are responsible for natural habitat conversion would help overcome this limitation. If this isn't possible to address, either through the methods or through published literature, then please reframe this paragraph to avoid the potential interpretation of high land use necessarily implies high biodiversity loss.

Response: We agree that the land-biodiversity relationship is very context dependent, and we added text to this paragraph to make this point: "Land use, especially conversion of natural areas, results in a range of context-dependent biodiversity impacts and GHG emissions23 and creates potential trade-offs with alternate uses, including production of other foods." (Lines 59-61). In this new text, we also added the point that land use can serve as an indicator for potential trade-offs with alternate uses so that the focus does not seem so focused on biodiversity. We do not however expand on this point further in this section since we cover the context-dependent nature of biodiversity impacts in more detail in the section that follows. We did add to the later text to include the point that the impacts on biodiversity can be limited in certain contexts: "Local context, such as ecosystem function, carrying/assimilating capacity, and species composition influence how stressors translate into environmental impacts39,40. Notably, land use impacts on biodiversity depend on the land use history and ecological context41. While all land used for food cultivation represents habitats converted at one point, avoiding additional agricultural expansion is important for preventing further habitat loss42. This is also true for on-farm land use by aquaculture, where conversion of ecologically valuable ecosystems, such as mangrove forests23 that serve as critical carbon sinks43 and nursery habitats, can generate severe impacts." (Lines 147-154).

The text in this section already included a point about the nonlinear relationship between stressors and impacts, but we also added an example: "Individual stressors may also have nonlinear relationships with impacts or act interactively44,45, such as climate change impacts compounding land use patterns that limit climate refuges or migration options46 or resulting in more frequent disease outbreaks, that increase antibiotic use and risk of antibiotic resistance." (Lines 155-158).

Fig 1. Three general thoughts:

1) I enjoy the figure layout, but found some of the colour bands difficult to differentiate. This would likely be even more difficult for people with colourblindness, especially for the reds and greens in panels (c) and (f).

Response: Since this paper is part of the broader Blue Food Assessment, we have been asked to use a color palette selected by a graphical design firm. As a result, we are limited to a fixed set of colors. We did shift the color palette to darker colors in Figure 1 to help distinguish the colors better. Fortunately, the differences between colors across panels is less important than the shading, as the color functionally duplicates the information about which stressor is depicted.

2) Following the above, would a different colour scheme or layout help? Possibly boxplots (although this would lose some of the variance currently shown in the figure), or alternatively adding small bands of black/white to differentiate between the different percentiles of each blue food?

Response: We agree that the shades are difficult to differentiate for some of the colors. In order to improve this, in the revised version of Figure 1, we now scaled the central portion of the distribution to a darker shade for the lighter colors. This creates a wider scale to differentiate among the bands.

3) Would it be helpful to visualize the impacts for the aquaculture systems on radar or spider plots as a potential SI figure? This would make it possible to compare the performance of each blue food across the five indicators in a single plot, instead of needing to look at multiple panels for a given food.

Response: We really like this suggestion as it helps illustrate the relative rankings and trade-offs in a compact way. We added a supplementary figure (Figure S5) with this visualization. We opted to retain the original structure of Figure 1 in the main text since it allows us to include the uncertainty in the estimates and to include capture fisheries (which would not be as compelling as a spider plot with only one spoke having non-zero values).

L 102-110: Would it be worth broadening the discussion to also include non-environmental aspects of sustainability? For instance, culture, social, and economics? I recognise the focus of the paper is on the environmental performance of blue foods, but it's becoming increasingly important to recognise the roles that foods play on the non-environmental aspects of sustainability.

Response: We fully agree on the importance of these other dimensions of sustainability. Given this paper's focus on the environmental aspects, we unfortunately cannot allocate the space this topic deserves within the word restrictions. We added a sentence to point toward such considerations in the context of the interventions: "Further, without careful consideration for local contexts and inclusion of relevant stakeholders, environmentally focused interventions can generate social and economic trade-offs that undermine broader sustainability goals." (Lines 287-289). We would also like to note that this paper is part of a collection of papers in the Blue Food Assessment and that other papers within this collection tackle this issue more directly, specifically ones focused on justice in blue food systems, on small-scale producers, and on patterns of demand.

L117-118: If there is space, it would be useful to explicitly mention a few of these non-linear relationships and interactive impacts between different stressors.

Response: In addition to the examples within the references provided, we added an example in the text of the interaction of climate change impacts and land use: "Individual stressors may also have nonlinear relationships with impacts or act interactively44,45, such as climate change impacts compounding land use patterns that limit climate refuges or migration options46 or resulting in more frequent disease outbreaks, that increase antibiotic use and risk of antibiotic resistance." (Lines 155-158).

Box 1 and Fig 3: Is it possible to account for uncertainty in the risk index of different types of fish?

Response: Unfortunately, this information was not included in the studies from which we drew the risk assessments. As a result, we cannot include uncertainty in the risk dimension, but still see the relationship as valuable since the average risk value suggests synergies in some cases, but trade-offs in others.

L152-156: Is yields in this context fish yields in aquaculture systems, yields of crops that are used for fish feed, or both?

Response: This is just referring to the aquaculture yield. To clarify this, we revised the sentence to read: "Holding all else constant, a 10% fish production yield improvement reduces land and water use for freshwater pond systems by 1–9% (Fig 4a)."

L157-167: Three general comments:

1) What is the aggregate benefit of implementing the different interventions simultaneously?

Response: This is a great question given that real world improvements to food system sustainability will require multiple simultaneous interventions. We aimed to ground our scenarios in the range of improvements observed in previous studies, in the range of plausible values based on our compiled data or on frequently proposed interventions. Unfortunately, this limits what types of interventions we have evidence for and we fear that including multiple scenarios in and additive way could severely over- or under-represent the aggregate benefit (depending on whether we double count an effect or possibly

don't count a multiplicative interaction among interventions). Some of these interventions likely have limited plausible interactions though and we think this would be a great area for future work. We now point to this in the levers section: "While we looked at individual interventions, improvements will likely occur through a suit of interventions and the synergistic or antagonistic interactions of interventions represents an important area for future work." (Lines 268-271).

2) Would energy sourcing be a useful lever to look at, at least for the high energy use aquaculture systems?

Response: While we did not look at specific energy sourcing changes (either shifts among sources for the energy grid or shifts between, for example, diesel and petrol), the zero-emission electricity scenario represents one extreme shift in energy sourcing. This scenario is only presented in the SI (Fig S15) since it only has an impact on the GHG estimates. The largest change in this scenario is for shrimp, followed by miscellaneous diadromous fish and trout. In an effort not to expand the number of scenarios much further, we must exclude numerous potentially interesting scenarios. However, we have improved our model documentation and study reproducibility (https://github.com/jagephart/FishPrint) so that future work can leverage these models to explore additional scenarios.

3) Would it be useful to mention that deforestation free feeds would have other environmental benefits, such as biodiversity?

Response: Yes, we agree and think it is a great addition. We now include the following text: "Comparing changes in feed sourcing, we found switching to deforestation-free soy and crops reduced GHG emissions by 6–54% (Fig 4b). This could create a co-benefit of also reducing biodiversity impacts." (Lines 203-205).

Fig 4: In panel (a), the colours are very difficult to tell apart, especially the pale yellows and greys. In panel (b), do the lines indicate the change from 0%? Would it be possible to add uncertainty to these estimates given that some of the categories are aggregates of multiple species and/or systems? Alternatively, would it be possible to perform sensitivity on the assumption that e.g. yields go to the upper 20th percentile to add error bars, for instance by evaluating the potential gains at the 10th percentile and the 30th percentile?

Response: We made a number of improvements to Figure 4 to improve its readability: 1) we changed the color from the yellow-blue color scheme to red-blue (still drawing upon the colors from the Blue Foods Assessment color palette though); 2) we shortened the labels in b and aligned the stressor categories in parts a and b; 3) we reduced the margin sizes and added arrows for any change greater than 50% to make the other lines more visible. The change is relative to the current value, where 0% would be no change. To clarify this, we edited the figure caption to read: "Change (%) in each stressor under four scenarios (defined in Table S8) relative to the current estimate."

Unfortunately, there is not a particularly clear way to add in uncertainty to this figure that we could come up with. While we appreciate the suggestion for the sensitivity analysis for the scenarios that leverage the stressor distributions, there is nothing analogous for many of the other scenarios, like a substitution of FMFO with fishery by-products. While we attempted this, it seemed to make the figure more confusing by implying less uncertainty in some scenarios than others.

Paragraph starting L 202: Can this paragraph be more specific to increase actionability and tangibility? For instance, who can do what, and how can it be done? A short list in this paragraph, supplemented by a longer table in the SI would increase the onus for individual actors (or sets of actors) to shift their behaviour, rather than resulting in the finger-pointing (from individuals, from gov, from business, from producers, etc) that typically results from high-level analysis such as this one.

Response: We appreciate the point the reviewer is making here and brought in some more specific examples of actors and related initiatives into the text: "A range of actors will be important for stimulating a shift to more sustainable production methods and, for instance, nation states, civil society and the private sector all have important roles to play. Private sector pre-competitive collaborations, e.g. SeaBOS63 and the Global Salmon Initiative can help stimulate production improvements at scale. Likewise, government-led initiatives helping small-holders improve their farming practices through e.g., access to high quality feeds, seed and broodstock, are crucial for closing the aquaculture performance gap64–66." (Lines 275-279). The literature related to actors and policy levers is too large for us to adequately summarize in an SI table and would be deserving of its own targeted effort. We did however add a reference to a recent paper that reviews historical food system transformations, which covers some of these issues: "Historical food system transitions required public investment technologies that could be scaled-up by the private sector and public policy leadership57." (Lines 252-253).

Methods and SI:

1) General: The approach is sensible and similar to other life cycle analyses. I'm not an expert in Bayesian analysis, so do not feel comfortable commenting on the specifics of the model set up and use of priors.

2) Testing the model: Apologies in advance for my potential misunderstanding of Bayesian stats and Markov Chain Monte Carlo analyses. My understanding is that the framework of the model in this analysis is:

"LCI database + data from paper publications -> MCMC analysis -> estimated results, including extrapolating"

where the role of the MCMC analysis is to create a distribution of estimates and the range of potential error given the data inputs in the LCI database and paper publications.

Is it possible to test the accuracy of your model within this framework, for instance by comparing the estimated results to the results in the paper publications, and/or withholding some of your data when constructing the MCMC model? If it is, this should be done. I realize that this will not be a perfect comparison given the difference between e.g. system boundaries, LCAI vs LCI databases, etc, but this would (a) provide a sense check of how well your model is estimating the impacts of different systems, which would be particularly important for the species with small data samples, and (b) allow you to improve the model by identifying the type of systems where thee model is least accurate.

Response: The estimation process described is not quite what was done. We used report and expert knowledge on values for parameters (e.g., FCRs) as priors, or used a weakly informative prior when no information was available and updated the estimates for FCRs with LCI data extracted from the reviewed studies within a Bayesian framework. This data was brought together with other data sources, such as the energy source mix and country-specific evaporative loss rate. The posterior distributions for the parameters were used to calculate the stressors as "generated quantities." We aimed to make this process clearer through the addition of Figure S1. A major benefit of this approach is that small sample sizes result in estimates of the mean with greater uncertainty, leading to more transparency in our confidence in the means presented based on the available evidence. This process means that for each species, the model parameters represent the statistical best estimate of the mean based on all available data points. A comparison between our estimates and other literature value estimates is then entirely based on differences in the models themselves. As pointed out, each study varies in its system boundaries, allocation method, which farms are represented, etc., which will shift their estimates. There is not a succinct way to present such comparisons within the

paper, and in a sense, it undermines the point that the estimates from these different models are not in fact comparable.

Nevertheless, to provide some ballpark comparisons, our median estimate of the mean for GHG (live weight equivalent) ranges from about 0.6 tonnes CO_2-eq for bivalves to 7 tonnes CO_2-eq for miscellaneous diadromous fishes and the GHG emissions from Bohnes et al. (2019) presents a median around 2.7 tonnes CO_2-eq (live weight equivalent) across all studies. The other metrics in Bohnes et al. (2019) are not in units that can be directly compared to our units. Poore and Nemecek (2018) reports values for farmed crustaceans and farmed fish, but they are presented in terms of 100 g protein and apply economic allocation. That paper uses 230 g protein per kg of fish (edible portion) and 150 g protein per kg of head free shrimp (for crustaceans) and it appears that their estimates are then roughly 23000 kg CO_2 per tonne edible for crustaceans and 7500 kg CO_2 per tonne edible for fish. Our shrimp values for economic allocation and edible weight are a little less than 10000 kg CO_2, with the difference likely being related to pond emissions and possibly including things other than shrimp in crustaceans, as our average value for fish is very close to 7500 kg CO_2, depending on how the different fish taxa are weighted (most medians are between 5000 and 1000 kg CO_2). Our GHG values also compare reasonably well to MacLeod et al. (2020), which presents the results as kg CO_2-eq/kg carcass weight (note that there are also some differences in system boundary and other methodological choices). They estimate bivalve emissions of 3-4 compared to our 1.4 [0-4 95% credible interval], catfish as 3-6 compared to our 7.2 [6.3-8.2 95% credible interval], cyprinids as 2-5 compared to our 3.6-7 [2-5.5 for silver/bighead and 5.9-8.2 for misc carps 95% credible intervals], misc marine fish as 6-18 compared to our 11.5 [9.2-13.9 95% credible interval], salmonids as 3-6 compared to our ~4 [4.2-6.3 for salmon and 4.1-6.7 for trout 95% credible intervals], 6-14 for shrimps and prawns compared to our 9.1 [8.1-10.0 95% credible interval], and 3-6 for tilapia compared to our 10.7 [9.7-11.8 95% credible interval].

L 541-549: I recommend adding comparisons with beef and a plant-based food, in addition to the comparison you have with poultry. The comparison with poultry throughout the text and figures is helpful, but it also assumes that most people reading the article or alternatively seeing Fig 1 on e.g. Twitter have general background knowledge in the field. Adding similar lines for beef and a plant-based food would avoid this assumption and increase accessibility to the paper's results.

Response: While we see the value of adding additional foods to provide the kinds of comparisons you suggest, it would open the paper to a much wider scope that would require substantial time that is not available during the allowed revision window. Consequently, we do not believe we would be able to do a proper job reviewing and modeling stressors for these other foods.

Further, adding additional terrestrial foods would likely merit more discussion than we have room to dedicate to the topic.

SI:

Fig S4: The colours are really hard to differentiate.

Response: We updated the colors in Figure S4 to improve the contrast between the two colors. We also changed the order of the bars to align with Figure 1 and removed the duplicated y-axis labels to utilize more of the space for the data. We similarly revised Fig S11 to align the colors for on- and off-farm sources of stressors across the two figures.

Fig S12: Please add more information to the figure legend to make the comparison group more apparent.

Response: We expanded the figure legend to include the same information as Figure 4, including clarifying that the change is relative to the current estimate and referencing the table where the scenarios are defined: "Fig S15 | Additional aquaculture scenario results. Change (%) in stressor values three additional scenarios (using mass allocation; defined in Table S5) relative to the current estimate."

Fig S14: Similar comment as for Fig 4a – some of these colours are difficult to tell apart.

Response: We have updated the SI figures (Fig S15-S18) to align with the changes in Fig 4 (describe above), including changing the colors of Fig S14.

Referee #3 (Remarks to the Author):

This paper by Gephart et al entitled “Environmental performance of the blue foods” is part of a collection of papers which are the major outputs of the international initiative “Blue Food Assesment” drawing together a diverse set of scholars to identify, review and fill important gaps in our understanding of the role of aquatic foods in global food systems, now and in the future. I strongly support the publishing of this type of meta-research in a journal with a strong impact like Nature because this paper and its companions are timely. Blue foods are already a principal source of protein for some 3 billion people. In many of the most vulnerable countries, they are also an essential source of long chain polyunsaturated fatty acids and vital micronutrients, such as iron and zinc, and constitute a key preventive measure against premature deaths and stunting. However, Blue Foods tend to be ignored in discussions about how to feed the world's growing population sustainably during the next decades. Their environmental footprints can vary significantly. Some fisheries and farms are responsible for significant greenhouse gas emissions or ecological impacts. But many aquatic foods outcompete terrestrial livestock in terms of environmental impacts. Given their relative health benefits, comparatively small environmental footprint, and the potential for a substantial expansion of global production, foods from ocean and freshwater systems have a key role to play in achieving both a healthy and sustainable food system. Yet many discussions of food neglect blue foods altogether. And when they do include blue foods, they tend to deal with them in sweeping generalizations – typically calling for massive expansion of aquaculture. It is therefore essential to assess the real impacts of the different types of aquaculture and guide its future developments.

This manuscript builds upon a simplified Life Cycle Assessment, developed to allow for broader and more rapid environmental assessments of aquaculture systems, using a harmonized and elegant methodology. This approach provides an overview of the environmental impacts of the most commonly aquaculture species farmed for food provision worldwide. The comparisons between studies on different production systems conducted using different methodologies were harmonized using life cycle inventory (LCI) data, rather than characterised life cycle impact assessment (LCIA) results. This novel and globalized analysis allows to compare very distinct geographical regions and aquaculture systems. The data from the previous inventories (feed use, farm size, etc.) were used to recalculate greenhouse gas emissions, nitrogen emissions, phosphorus emissions, terrestrial land occupation, and consumptive freshwater use for aquaculture systems which are measured as the major environmental stressors. I did

not find any weakness in the careful methodology that is provided in the methods section and in the supplementary files. This report is based on data collected from over 1,590 farms and 1,000 unique fishery records around the world for 23 species groups covering over 70% of global blue food production.

The results are quite clear and across all blue foods, farmed bivalves and seaweeds generate the lowest stressors.

Unfortunately, despite they also represent the largest volumes of production for the marine aquaculture, only a few LCIA have been conducted for these two groups except for mussels in comparisons with fish or crustacean farming. I think this represent a weakness in the meta-analysis.

The studied seaweed species are considered by the authors as providing food in comparison with the excluded volume of industrial seaweed cultivation. In fact, Gracilaria (Agarophyton) chilensis is grown in Chile essentially for agar extraction which is used in food texturing but also mainly for microbiological and biotechnological applications. Presently Macrocystis pyrifera is also not considered as a major edible species, nor Laminaria digitata and Saccharina latissima. It is a pity that no LCIA or LCA are available for the major seaweed crops grown in thousands of farms in South Korea, Japan and China (Pyropia spp;, Undaria pinnatifida and S. japonica). Indeed, LCIA were mainly conducted to assess the potential of growing large volumes of seaweeds for producing bioenergy and not for human food. I would suggest to the authors to encourage such studies in a near future to help emerging countries to invest in the potential of cultivating edible seaweeds and promoting alternative diets including seaweeds.

Response: We conducted a targeted search for aquatic plant studies to improve the representativeness of edible aquatic plants, but as you point out, there are unfortunately few studies on this topic. We revisited the list of studies and were able to add additional data points from Marín et al. (2019) for Gracilaria and Porphyra seaweeds grown in China. These values are now represented in our results. Additionally, we would like to note that Saccharina latissima is the major species consumed in Europe (Thomas et al. 2021). Nevertheless, the available data falls short of representing the global production of seaweeds, and especially of edible seaweeds. We agree with your suggestion and added text to encourage such studies: "Further, farmed carps, bivalves, and seaweeds account for vast amounts of global production yet remain underrepresented in environmental assessments (Fig S3-S4) and therefore should be prioritized for additional research. Our estimates represent a snapshot of current production, but future work on emerging production technologies, feed innovations and growing sub-sectors is important for tracking changes against these benchmarks." (Lines 97-101).

References:

Marín, T., Wu, J., Wu, X., Ying, Z., Lu, Q., Hong, Y., Wang, X. and Yang, W., 2019. Resource use in mariculture: a case study in Southeastern China. Sustainability, 11(5), p.1396.

Thomas, J.B., Sodré Ribeiro, M., Potting, J., Cervin, G., Nylund, G.M., Olsson, J., Albers, E., Undeland, I., Pavia, H. and Gröndahl, F., 2021. A comparative

environmental life cycle assessment of hatchery, cultivation, and preservation of the kelp Saccharina latissima. ICES Journal of Marine Science, 78(1), pp.451-467.

Moreover, the standard stressors used for LCIA do not take into account the ecosystem impacts most notably biodiversity loss. This point is discussed in the manuscript. However, this discussion provides only global impacts that are summarized in Fig.2.

It is clear that the development of the shellfish industry worldwide had a strong impact on the local biodiversity and on the ecosystem functioning. The introduction of the Japanese oyster Crassostrea (Magalenna) gigas led to its escape to establish wild settlement but also to the introduction of dozens of alien species belonging to various groups of marine invertebrates and algae. It also allowed the expansion of epidemic diseases in oyster farms which strongly impacted the producers economically and socially.

Response: Yes, it is certainly true that biodiversity loss, invasive species, and disease are all categories that we are unable to quantify due to their local context and poor coverage in LCA. In fact, Troell et al. (2017) found that "Climate change, habitat change, pollution, and overexploitation are impact categories commonly entailed within the LCA framework, addressing all major drivers, except invasive species, for biodiversity loss, identified by the Millennium Ecosystem Assessment (2005)." We do not have sufficient space to incorporate examples of each kind of impact, but we have added some detail on local context to the section on these other impacts.

We also added text to the discussion on the total impacts of large-scale expansion of the low stressor blue foods: "Major challenges remain for shifting demand, as well as meeting increased demand. While improved management offers potential opportunities for expanding some production from low stressor capture fisheries, uncertainty remains around the extent and feasibility of rebuilding many fisheries48. Additional research is needed to understand the total environmental impacts of large-scale expansion of low per unit stressor foods, especially for system-specific impacts (Box 1)." (Lines 246-252).

References:

Troell, M., N. Kautsky, M. Beveridge, P. Henriksson, J. Primavera, P. Rönnbäck, C. Folke, M. Jonell (2017) Aquaculture, Reference Module in Life Sciences, Elsevier, https://doi.org/10.1016/B978-0-12-809633-8.02007-0.

When assessing the environmental benefits of cultivating seaweeds to mitigate eutrophication, and acidification, to prevent oxygen minimum and to capture carbon dioxide, it is important to keep in mind that Volatile Short-Live Substances are also an important by-product of the seaweed photosynthesis and metabolism. Their production rate varies by 3-4 orders of magnitude (Leedham et al., 2013 Biogeosci., 10, 3615) depending on algal species, compositions and the metabolic contribution of associated microbiota (King et al., 2001, Mar. Ecol. Prog. S., 224, 69), as well as abiotic and biotic stress factors. Some VSLS like bromoform and iodine oxide contribute to stratospheric ozone loss and increased surface UV radiation (Dessens et al., 2009 Atmos. Sci. Lett., 10, 201; Oman et al., 2016 Geophys. Res. Lett., 43, 9869; Fernandez et al., 2017 Atmos. Chem. Phys., 17, 1673). The impact of VSLS on stratospheric ozone depends strongly on the location of their emission. The tropical areas

around coastal Southeast Asia and Australia are of major importance, as they are the main global entrance regions of tropospheric air masses into the stratosphere (Tegtmeier et al., 2012 Atmos. Chem. Phys., 12, 10633) and coincide with the largest macroalgae industrial production worldwide. VSLS emissions from the mid-latitudes and polar regions on the other hand, are less likely to reach the stratosphere due to their fast atmospheric degradation. Their environmental impact needs to be assessed. Budgets for the non-CO_2 trace gases from macroalgae habitats, their effects on the atmosphere and development under climate change are lacking in general. As stressed by Popkin in a featured news in Nature (vol. 565, January 17th, 2019),| terrestrial tree forests trees are supposed to slow global warming, but growing evidence suggests they might not always be climate saviours.

Response: We thank the reviewer for this mechanistically sound point. While the invoked mechanism exists, and is important, as the reviewer suggests, it is well outside of the scope of this paper. We do not claim that our paper addresses every environmental aspect of aquaculture. Rather, conforming with most all literature on environmental impacts of food and agriculture, we focus on a non-exhaustive list of quantified measures (those which are most often used to compare foods across the food system). Nevertheless, to acknowledge some of the nuances of seaweed production, we added a new box (Box 1) to underscore that some of the impacts specific to certain systems are not included within our analysis and are context dependent. We included a point within this box related to volatile short-lived substances, as suggested: "Third, impacts on biogeochemical cycling and habitats are highly context dependent. For example, the systems represented here extract nitrogen and phosphorus, which could be problematic in nutrient poor environments. Additionally, stratospheric ozone effects from volatile short-lived substances depends on the location and varies widely across species29,30." (Box 1).

The manuscript also concludes that capture fisheries predominantly generate greenhouse gas emissions, with small pelagic fishes generating lower emissions than all fed aquaculture, but flatfish and crustaceans generating the highest. Among farmed finfish and crustaceans, silver and bighead carps have the lowest greenhouse gas, nitrogen, and phosphorus emissions, but highest water use, while farmed salmon and trout use the least land and water. The provided model allows also to identify and quantify improvement opportunities and discuss public and private policy options to realize these improvements. There is no doubt that these conclusions will help policy-makers to select and support research programs and pilot project developments to improve the performances of existing systems and increase sustainability.

Thank you very much for taking the time to review our work and for the constructive comments, which have improved the text.

Referee #4 (Remarks to the Author):

The paper addresses a highly relevant topic, and given the role that "blue foods" have to play in the coming years and decades, its ambitions are very laudable, i.e. providing a quantitative measure of environmental performances of blue foods (as reflected by title). Conceptually, the provision of estimates from the different taxa groups for selected environmental metrics is indeed relevant and could be useful for further studies in food science. So is the approach to estimate reductions of some environmental stressors, which could bring relevant insights on which levers of fisheries/aquaculture systems to act on.

In spite of these good intentions, I found several critical problems with respect to the documentation of the study (or rather its lack thereof!), the methodological approach used, and the overall content/relevance of the paper. In my opinion, resolving all these issues is a necessary step before the work can be considered again for publication in any journal.

I explain each of them below, followed by additional detailed comments, hoping that it can help the authors rework their study.

Response: We thank the reviewer for their recognition of the value of the goal of our study and for taking the time to carefully review our work. We have substantially expanded our study documentation in the SI, along with creating a reproducible workflow that integrates the data and the code and is publicly available on Github (https://github.com/jagephart/FishPrint) as well as archived on Zenodo (DOI: 10.5281/zenodo.4768324). A flow chart detailing the data and code is also included on Github to facilitate review and use of the code. We believe there may have been some misunderstanding with our methodological approach, which we have worked to clarify in the responses below, text, and through the addition of a methods flow chart (Fig S1). While we are very constrained by space for additional text, we have also integrated text to incorporate the feedback related to the context. We expand on these points in the responses to the specific comments below.

1. *Complete lack of transparency in the data behind the estimates*: I am very surprised (not to say "shocked") by the lack of transparency in the study. The paper states early on: "We leverage data from over 1,590 farms and 1,000 unique fishery records around the world for 23 species groups covering over 70% of global blue food production." (l.38-39). This is the data basis for the entire study, and my first reaction when reading was "Waouh, great!". Unfortunately, my heart sank very rapidly when I realized that this statement is given without any cross-references to where the data are documented. I have been looking into the Methods section, the SI and other papers, incl. by Halpern et al., etc. but to no avail. The Methods section reiterates the statement, and for aquaculture systems, it additionally

specifies “Aquaculture inventory data were extracted from a database of aquaculture life cycle assessments initially compiled by1 [Parker et al. 2012] and updated with more recent studies (l.359-360) [....] resulting in a dataset of 112 species-specific LCA models from studies published between 1995 and 2020” (l363-364). As a reader, I would expect as a minimum that these “LCA models” and their references be provided transparently at least in Supplementary Material, with additional key features of the studies (like normally done in most review studies). As a reviewer, access to these information is essential to be able to check the comprehensiveness of the review (for which I have serious concerns; see below comments) and its validity as to the processing of the extracted data (for which I also have concerns; see below too). With the present documentation in the manuscript (and SI), I could not check those aspects, which are key to any peer review process. Note that I give the example of aquaculture above, but this comment applies to all data used in the study. In general, there is no comment as to the origin, quality or validity of the data sources in the paper. Addressing completely these issues of data transparency and quality check is a requirement before the paper can move any further in any scientific journal.

Response: Following the *Nature* publication guidelines, we agreed at the submission stage to publish the underlying data to the study, including all reference information for underlying studies upon acceptance of the manuscript. However, to assist in the review process, we have included a folder of all underlying data within our GitHub repository. This in turn has been linked with a citable DOI with a time stamp of the revision point through Zenodo, further increasing the findability and accessibility of the analysis (DOI: 10.5281/zenodo.4768324). This data accompanies the code, which is publicly available on Github (https://github.com/jagephart/FishPrint). The Github repository now includes a diagram mapping the relationship between datafile and scripts. This data consists of all data compiled for this study. We do not re-publish raw data for data existing in other databases, but we cite these already public databases (e.g., FAO, NOAA) and provide the code we used to download and/or process this data. We have also added additional detail on the inclusion criteria to the SI (Fig S1) and posted a data source sheet for all studies considered with reasons for any studies excluded in a supplementary file. We address the points related to the representativeness of the data below and further detail the inclusion criteria point in response to a later point.

2. *Unsupported claim of data representativeness*: in several places of the manuscript, comprehensiveness of the study is advanced (e.g. “70% of global blue food production”, l. 39; “Our results represent the most comprehensive and standardized blue food stressor estimates to date”, l.83-84; etc.). The manuscript, incl. Methods section, makes references to Figs. S2 and S3 to back up these statements on representativeness and comprehensiveness. As far as these two figures are presented, they however do not give any indications of either data representativeness nor data comprehensiveness. Figure S3 does not describe the extent of the production originating from the covered farms; this is what would be meaningful to compare to the total national production (right now the plot is useless). Likewise, Fig. S2 shows total aquaculture production on y-axis by number of distinct farms covered in the

study on x-axis, but where can we see how many farms in total exist globally (to get an idea of farm coverage)? For example, it seems that a bit less than 50 farms were covered for bivalves, but how many are there in the world? And where are those 50 farms located compared to the global distribution of bivalve farms? And how representative the technologies covered in these 50 farms are of the technology mix for bivalve farming in the world? Based on the graphs displayed, I could go arguing that the 50 farms covered in the study are producing 0.001% or 99.99% of the farmed bivalve production globally, that they are located in a single country or spread in 50 different ones, using the same technology or 50 different types. Figure S2 and S3 would not be able to contradict or clarify any of my statements. This is a major problem as those statements/questions I raise all relate to data representativeness, incl. technological and geographical representativeness (you may also add temporal representativeness, but it is strongly linked to technologies) and also have implications on completeness and uncertainty/variability evaluation. Fully transparent description of these aspects are essential to check and justify the relevance of the study and its results.

Response: The claim we make in the introduction is that the taxa groups analyzed cover over 70% of global blue food production. We worked to ensure these are as representative of global production as possible by comparing the geographic distribution of the studies to production and weighting species-level estimates by production values within the taxa groups. Additionally, to ensure that the studies included represent commercial scale production, we exclude all hypothetical and experimental studies from the analysis. Since there is no global data available that we are aware of on production methods, number of farms or production by farm to which we can compare the data, we cannot include such a claim. Instead, our representativeness claim is with respect to the taxa group and geographic coverage (Fig S3-S4). We now point this limitation out explicitly in the methods: "Since global data on production method or intensity is not available, we cannot ensure that the studies are representative of these attributes. Further, while we have both small- and large-scale aquaculture producers represented in the data, we do not have complete data on the size of all farms or the size distribution of farms globally, so we cannot compare the representativeness with respect to producer size. However, all studies included do represent commercial-scale operations." (Lines 542-548). Additionally, we have revised the SI figures to communicate the data coverage more clearly with respect to species and geography (Fig S3 and S4).

We would also like to emphasize that underrepresented groups identified here are important areas for additional work, as pointed out in the revised text: "Further, farmed carps, bivalves, and seaweeds account for vast amounts of global production yet remain underrepresented in environmental assessments (Fig S3-S4) and therefore should be prioritized for additional research." (Lines 97-99). As such, the estimates aim to represent a current snapshot of the stressors by taxa group and serve as a benchmark for the evolving sector.

3. *Relevance of the paper undermined by lack of depth, lack of contextual positioning and missing discussion on key aspects*: beyond the other problems that I mention above and below (which have implications on the reliability and consistency of the displayed results), I overall find that the result analyses and the discussion points in the paper lack depth and overlook important aspects when addressing environmental performances of blue foods. Here is a summary of some criticisms (not exhaustive):

- There is a lack of contextual perspective in the discussion points. The introduction introduces the challenges that food systems face and the role that blue foods could play, but no link to these aspects is made afterwards. The paper is very much focused on the supply side (and in a narrow perspective and not in a systemic perspective; see below sub-point) and makes very little connection to the demand side and how blue foods could help address the challenges of food systems while ensuring environmental sustainability (although claiming to do so, e.g. l. 187). A signal that the relationship to a broader context is missing is that there is no positioning at all with the SDG framework, which, given the aims and scope of the study, could be expected (at least in introduction and some discussion aspects).

Response: We expanded the contextual motivation of this paper and now explicitly mention the three SDGs most closely aligned with the work: "As a critical source of nutrition8,10 generating relatively low average environmental pressures1,2,11,12, blue foods present an opportunity to improve nutrition with lower environmental burdens, in line with the Sustainable Development Goals to improve nutrition (Goal 2), ensure sustainable consumption and production (Goal 12), and sustainably use marine resources (Goal 14)." (Lines 23-27). As one of the core papers in the Blue Foods Assessment, which is deeply involved in the UN Food Systems Summit, this work is also well aligned with the action tracks for the summit, an exciting opportunity to raise the profile of aquatic foods within global food systems dialogues.

We included some additional demand related motivation in the introduction but have reserved most of the detail for the final section of the paper given the word constraints: "It is therefore critical to examine the environmental performance across the diversity of blue foods in a robust, methodologically consistent manner to serve as a benchmark within a rapidly evolving sector as blue food demand increases21, production shifts toward aquaculture and production technologies advance." (Lines 34-37). We have expanded the text in the final section of the paper related to interventions and the demand side in response to later comments.

- The larger importance and role of aquaculture over fisheries (as source of protein) is not mentioned or discussed while the study has a rare advantage to cover both (which could offer opportunities for interesting discussions). In that context, aspects such as the sustainable management of fish stocks, which poses caps/limits to how much fisheries can be used to supply food for human consumption and therefore calls for increasing demand from aquaculture systems, are not mentioned.

Response: We do not have sufficient space to go into much detail on the nutritional contribution of aquaculture and fisheries nor the full scope of issues related to stock management. However, the original text points to the nutritional importance in the opening paragraph (not in terms of protein as the primary contribution is through micronutrients and essential fatty acids). We did add text to indicate the likely shift toward aquaculture as one element of the paper's motivation: "It is therefore critical to examine the environmental performance across the diversity of blue foods in a robust, methodologically consistent manner to serve as a benchmark within a rapidly evolving sector as blue food demand increases21, production shifts toward aquaculture and production technologies advance." (Lines 34-37).

We also expanded on the management of capture fisheries in two places in the final section of the paper. First, we now point out that the capture fishery estimates can be used to understand the replacement cost of sustainably managing capture fisheries: "Capture fisheries vary widely in their GHG emissions but are low impact with respect to the other stressors considered. This underscores the value of sustainably managing wild fisheries to avoid the environmental replacement cost that would be incurred under fish catch declines24." (Lines 233-236). Second, we raise this point in the new text discussing the limitations of scaling up some of the interventions: "While improved management offers potential opportunities for expanding some production from low stressor capture fisheries, uncertainty remains around the extent and feasibility of rebuilding many fisheries48." (Lines 247-249).

- The analysis of the results is strongly biased towards comparisons across species/taxa groups, and there is little discussion on the different environmental performances that can vary considerably across technologies for a same species or within a same taxa group. This is particularly important for aquaculture systems, e.g. pond vs. RAS, etc., where different impact magnitude may be observed, and sometimes leading to trade-offs (when one technology performs better for some environmental problems but not for others). This has been commented in past LCA studies (likely several in the pool of retrieved studies), which could be used as support. In this context, the authors should also discuss the different opportunities to shift from high-impact technologies to low-impact technologies (beyond just fine-tuning some model parameters), although the lack of technology differentiation (e.g. farming systems) in the current study does not make this easy.

Response: As the reviewer points out, many past studies have focused on comparing different production technologies. This is an important body of work for guiding the sector and is likely best suited for full life cycle

assessments focused on the direct comparison of the potential systems. We chose to focus on taxa groups because it is the level at which national production data is available, and the level at which consumer-facing products are available. The lack of a best estimate for these major food-systems stressors at the taxa group level therefore represents an important information gap. To make these estimates as representative as possible, we excluded all experimental production systems and included species weightings within the taxa groups so that the estimates best reflect current production.

Nevertheless, we did leverage information about the production technologies in our data collection and are able to speak to these to some extent (Fig S8 and S9). We added text that more clearly points to these figures to emphasize the importance of production technology and intensity. In addition, we added a reference that provides a comparison between RAS and net-pen/ponds on species level: "While some variability within a taxa group is due to differences in on-farm practices, production technology is an important factor across stressors49 as variability in stressors for a given species reared in different farming systems can be considerable (e.g., 50)." (Lines 187-190).

- The temporal perspective is also lacking as the study appear quite static and representing past/present state. It does not open up with actual technological outlook (emerging technologies and aquafeed are barely mentioned), as it only addresses impact reduction potentials based on scenarios that are disconnected from reality, e.g. variation of 10% on several parameters --> how realistic? How feasible? What implications on other parameters/system components? This overall dynamic or future-oriented take on the blue foods production sector would be relevant to address in relation to their environmental impacts.

Response: Our goal was to provide the best current estimates of stressors for blue food species groups and use this model to understand potential opportunities to improve the environmental performance. The 10% change referred to looks at the influence of each parameter as a lever for reducing the stressors, holding all others constant, and in many ways also serves to explore uncertainties. We complement this though by a running range of scenarios in Fig 4b and provide referenced examples of the magnitudes of an intervention or the contexts in which the intervention has been suggested in Table S8. For example, moving each observation down to the 20th percentile for the parameter within its group incorporates an aspect of feasibility as there are currently systems producing the same species at that level of efficiency. Other scenarios represent suggested interventions that could be on the horizon, such as replacing FMFO with fish by-products. We cannot however estimate the change for an emerging technology since, by its very nature, there is not data

available at a relevant scale that could be used to model this change. We did however add text to point to emerging technologies as a priority area for future work: "Further, farmed carps, bivalves, and seaweeds account for vast amounts of global production yet remain underrepresented in environmental assessments (Fig S3-S4) and therefore should be prioritized for additional research. Our estimates represent a snapshot of current production, but future work on emerging production technologies, feed innovations and growing sub-sectors is important for tracking changes against these benchmarks." (Lines 97-101).

- Conclusions/recommendations from the study (mostly after l. 182), which relate to shifting to low-impact blue foods, should be nuanced and revisited as some of them do not hold. There are some clumsy claims like in l.183, where it is not obvious from the study results how much reduction blue foods can bring; the comparison with chicken (only one displayed) could suggest that the authors recommend to switch most food supply (globally) to bivalves and seaweed... Other claims are mere overstatements. For example: "Our standardized estimates enhance the resolution of the potential role of blue foods within sustainable diets, highlighting opportunities to shift demand from relatively high- to low-stressor blue foods and from terrestrial animal-source foods to comparatively low-stressor blue foods." (l.187-189) This could be misleading to the reader and should be nuanced considering all the limitations of the study. In particular, the authors cannot use the term "sustainable" when only addressing a few environmental problems. Although there is a part of the paper focusing on other stressors, the very truncated coverage of environmental stressors is a major limitation and prevents any such claim about environmental sustainability. Many other impact categories, like toxic releases of chemicals, antimicrobial resistance, seabed damages, fish resource depletion, etc. are highly relevant to consider when addressing "environmental performance of blue foods". Addressing a systemic perspective covering the entire food systems (not just part of the production) is also required in such discussion.

Response: We maintain that the current phrasing in relation to sustainable diets is accurate. By providing standardized estimates for the five stressors presented, which are the main five used for terrestrial food production, the analysis does enhance the resolution of the role of aquatic foods in sustainable diets. Enhancing the resolution does not imply that every aspect of sustainability is included. Including every sustainability aspect of every food is not feasible and would require not only standardizing many of the categories you mention above for aquatic foods, but also many kinds of impacts that are specific to terrestrial production systems. Throughout the text, we added "considered" to stressors to clarify when we are referring only to quantified stressors. We also revised the text to be more explicit about the comparison to terrestrial animal source foods: "Blue foods already have great potential for reducing food system environmental stressors. Unfed aquaculture results in negligible values for most considered stressors, and many fed aquaculture groups outperform industrial chicken, the most efficient major terrestrial animal-source food." (Lines 230-233). Finally, we expanded/clarified nuance in

the results and other environmental impacts throughout the manuscript (as detailed in response to other reviewer comments). This includes the text in the new Box 1, which details some of the important considerations for bivalves and seaweeds.

4. *Novelty of the paper in need of justifications*: for aquaculture systems, Bohnes et al. (2019) already conducted a comprehensive review of LCA studies (in "Life cycle assessments of aquaculture systems: a critical review of reported findings with recommendations for policy and system development"; Reviews in Aquaculture (2019) 11, 1061–1079). This study is surprisingly not cited in the current paper, which is a major flaw considering their very close scope. Although the current paper goes beyond aquaculture systems alone and also bring different angles (e.g. testing different reduction potentials), it is important that the authors position their work in relation to existing literature and argue of the novelty of their work (e.g. estimates of environmental impact intensities are provided in Bohnes et al. (2019), albeit to different extents). In addition, seeing that the authors overlooked such a spot-on literature source, I confess to be much worried about their ability to find all relevant LCA studies relevant to their review scope, and I am quite doubtful of their argued comprehensiveness with regard to the retrieved literature (see next point).

Response: The authors are aware of Bohnes et al. (2018) and Bohnes et al. (2019), and these happen to be among the many relevant studies we were not able to cite in our manuscript due to the *Nature* journal limit of 50 references. We now cite this paper to support the importance of different production technologies: "While some variability within a taxa group is due to differences in on-farm practices, production technology is an important factor across stressors49 as variability in stressors for a given species reared in different farming systems can be considerable (e.g., 50)." (Lines 187-190).

However, we disagree that it has significant overlap with the presented study given the fundamentally different methodological approach. The presented study is not a review of reported impact intensities as is provided in Bohnes et al. (2019) and other LCA reviews, but rather a re-estimation of impact intensities based on underlying inventory data from LCAs and other studies. In doing so, we standardize the stressor estimates so they are comparable. This is not the case in Bohnes et al. (2019), which states "It is worth noting that no further harmonization of the results was performed (e.g. consideration and/or alignment of system boundaries or LCI modelling framework), and the remaining differences must therefore be regarded as a source of uncertainty." We would argue this is an important distinction and novel advancement of our work and suspect the reviewer would agree based on comment 6. Further, our analysis excludes many of the experimental and hypothetical production systems included in Bohnes et al. (2019) and weights the group estimates

according to global production, which we view as an important advancement for benchmarking current production. Notably, this differs from the goal of Bohnes et al. (2019), which aimed to look at trends within the literature and to compare across production systems. We clarified this point in the introduction to emphasize the novelty of the paper (Lines 40-42), though we keep this brief as the main text since we are space limited and this paper is aimed at a broad audience. Finally, we emphasize this point further in the methods and through the addition of a methods flow chart (Fig S1). We respond to the point about the studies included below.

5. *Potential lack of comprehensiveness in the retrieved data*: I may be wrong since the list of considered studies is not documented in the study, but I have suspicions that some were overlooked. The Methods section states that complementary data for aquaculture about "FCR and feed composition data from Monterey Bay Aquarium's Seafood Watch program2 and farm-level data collected for a range of systems in southeast Asia3 [Henriksson et al. 2015] to improve geographic representativeness" (l.368-370). This gives the impression that only ref. 3 (Henriksson et al. 2015) was used to cover Southeast Asia. There are however other sources relevant, e.g. Henriksson et al. 2017 (Indonesia), etc. Several of these sources have already been identified in the (uncited) review paper on LCA of aquaculture systems by Bohnes et al. (2019), which identified 65 LCA studies (covering more than one environmental impact categories) with a total of 179 individual aquaculture systems assessed (+ 38 aquafeed systems). In that paper, it is apparent that more studies than just Henriksson et al. 2015 cover southeast Asia (other regions in the world should also be addressed). Seeing the publication trends in Bohnes et al. (2019), several additional papers must have been published since that review. Again a detailed account of the reviewed studies (and of the method by which they were identified by the authors) is needed to check comprehensiveness in the number of retrieved studies.

Response: We added an SI csv file with full reference information for all studies considered, along with a column indicating why a given study was excluded. Shortened references are provided at the observation level with the data inputs for the study.

Our literature review process built off the reference list from Parker 2012, which established a database of LCA studies and LCI data and has been updated since initial publication. We then supplemented this set of studies with the data from the Seafood Watch Database and by comparing to the reference list compiled for a systematic scoping review performed by author Gephart as part of a grant submission and the references in Poore and Nemecek (2018). We then conducted targeted literature searches for species and geographies that were poorly represented in the data. Finally, we brought in data from three unpublished studies conducted by coauthors on species/regions of importance

for global aquaculture production. During the revision, we cross-referenced our list of studies with those in Bohnes et al. (2019). This added five farming (i.e., not feed production) studies, though none met our inclusion criteria. Prior to excluding any studies, this process resulted in 84 studies/datasets considered for inclusion. The fact that few to no additional studies meeting our inclusion criteria were added upon considering additional review reference lists gives us confidence that we have very good coverage of available studies fitting our study parameters.

The inclusion criteria are described in the methods as follows: "Studies were systematically filtered to exclude those assessing experimental or hypothetical systems, agri-aquaculture systems, polyculture systems producing species in more than one taxa group, and those for which no inventory data were reported, resulting in a dataset of 61 studies and datasets published between 1995 and 2020." (Lines 483-486). To ensure that this process is clearer, we added a methods flow chart to the SI with these details (Fig S1).

6. *Lack of details on raw data consistency and handling*: in addition to providing transparent accounts of which peer-reviewed studies were included to map LCI for fisheries and aquaculture systems, I also require more information on the data from Monterey Bay Aquarium's Seafood Watch program. I briefly checked the Seafood Watch program website, in particular the Methods section described online. Although I find such initiatives to build consolidated databases much needed for environmental sustainability assessment (we need more of such work!), I still miss background details on what studies were included in the database and how the harmonization across studies was performed. I could not find indications of validation/peer review of the database/tool, which, if existed, would anyway have to be recalled succinctly in the present paper since it is an important data source. Harmonising LCI data is not a simple exercise, as the authors of the present study seem to suggest in SI. It usually requires meta-analyses to harmonise methodological choices made in the different LCA studies (see, e.g. entire Special Issue of Journal of Industrial Ecology, https://onlinelibrary.wiley.com/toc/15309290/2012/16/s1, with several papers addressing the methodological challenges of performing meta-analyses and a number of applications for the energy sector). The conduct of such harmonization cannot be documented with just one paragraph in the Methods section and 2 pages of general information in SI. I miss how each study was handled, for example when other LCI modelling framework (system expansion, allocation with other principles) than the one selected by the authors were used: how was the previous allocation undone? How were system boundaries with previously applied system expansion adjusted for the specific credited processes?, etc.

Response: The challenges highlighted by the reviewer are the major motivation for our work. The issues highlighted by the reviewer are confounding factors for past review studies. Our argument is that practitioner choices are too

discrete to be harmonized at the level of impact assessment results, which is why we only use inventory parameters which are readily identified and extracted and less subject to the methodological choices to which the reviewer refers (feed conversion ratios, general feed compositions, energy inputs to farm operations, pond area and yields, etc.) to inform our model. By doing this we circumvent issues such as system expansion, allocation, etc. predefined by individual studies. Moreover, we argue that the nature of LCA does not allow for meta-analysis in its true form, as sample sizes and deviations rarely are reported (e.g. see Henriksson et al. 2021). Given the reviewer's feedback, alongside other reviewer comments, we have expanded the articulation of these advancements in the revised manuscript so that these advantages of our work are clearer to all readers. We need to strike a balance in the introduction for this novelty to come through without becoming overly technical for the intended broad audience. Therefore, we still include only a brief summary of the methodological advancement, which we revised to read: "Here, we provide standardized estimates of greenhouse gas (GHG) emissions, consumptive freshwater use (water use), terrestrial land occupation (land use), and nitrogen (N) and phosphorus (P) emissions for blue foods. We identify a set of key life cycle inventory data (i.e., material and energy input and farm-level performance data) from published studies and datasets to which a harmonized methodology is applied." (Lines 38-42). We then expand on this in the Methods and SI, particularly with the addition of Fig S1, which provides a flow chart for the overall analysis. A more technical flow chart of how the code works is also included with the code on Github (https://github.com/jagephart/FishPrint).

References:

Henriksson, P.J., Cucurachi, S., Guinée, J.B., Heijungs, R., Troell, M. and Ziegler, F., 2021. A rapid review of meta-analyses and systematic reviews of environmental footprints of food commodities and diets. Global Food Security, 28, p.100508.

7. *Inappropriate basis of comparison across taxa group*: I would question several claims in the second main part of the manuscript (l.43-96). Roughly summarised, this part takes each environmental stressor and describes comparisons across species to identify low/high environmental performances (e.g. "farmed seaweeds and bivalves generate the lowest emissions, followed by small pelagic capture fisheries, while flatfish and crustacean fisheries produce the highest", l. 46-48). I find such comparisons misleading since the estimated values of the stressors have been based on live weight basis, which is only mentioned in the Supplementary Information. The quantitative evaluation of the stressors is indeed totally disconnected from the demand side and ignores the functionality of the food systems, which could be defined to meet the nutritional needs of the national/global population (you may find interesting discussion on how to define such functionality in Weidema and Stylianou, 2020). Although live weight has been widely used as basis of comparison in past LCA studies, and I concede there is yet no consensus on how to appropriately define a functional unit for food systems, the use of live weight has been clearly pointed out as an incorrect

approach since it completely dismisses the functionality of food systems (see Avadi et al. 2015; Bohnes and Laurent, 2019; Weidema and Stylianou, 2020). The recommendations on satiety/nutrition from Weidema and Stylianou (2020) may be difficult to implement for such large-scale study (although likely the way forward), especially since post-fisheries or post-farming processes are not included, but the use of edible weight should be possible to implement and would definitely be a step forward. Even though cultural differences may mean including some uncertainties in the estimation of edible weight (currently raised as an argument for not going for it in the supplementary information), I would believe it would yield less uncertainty and more comparability across taxa group than done in the current estimates. Although discussion on functionality is outside the scope of such study, 1-2 sentences to clarify this topic are also required in the manuscript (and the Methods section should argue for the approach followed).

Response: We debated the functional unit to present extensively as the team agrees that both are relevant in different contexts. For the purposes of connecting to production data, the live weight estimates are most relevant. Of course, when looking at the demand side, the edible portion is more relevant. Converting to edible portion introduces additional issues related to differences in processing efficiency and definitions of edible across geographies and cultures. To strike a better balance, we now present the results both as live weight equivalents in the supplementary information (Fig S10) and in terms of edible portion in the main text (Fig 1). We also added new text to the SI regarding the functional unit: "Stressor estimates can be expressed in terms of different functional units such as live weight or edible portion at the farm gate. Stressors estimates expressed in terms of live weight equivalent enable the data to be easily connected to production data, which is also typically expressed in terms of live weight, while the edible portion is more relevant to blue food demand and the role of blue foods in diets. As a result, we calculate the stressors in terms of both edible portion (Fig 1) and live weight equivalent (Fig S10). Edible portions as a fraction of the live weight can vary substantially due to differences in anatomy (Table S1). Further, edible yields are also influenced by cultural preferences and processing efficiencies. Fish heads, for example, are a delicacy throughout most of Asia, while they are discarded in Europe and North America. Other edible parts, such as swim bladders from Pangasius catfish, can yield much higher values than the fillets21. While these result in higher utilization, in other cases, inefficiencies in processing can result in a lower percentage of the live weight mass being utilized. Other byproducts are also often utilized for other processes, such as raw materials for fishmeal and fish oil production, but details on these utilization rates are incomplete. Moreover, these efficiencies vary across space and time. To standardize the definition of edible at the global level, we use the muscle fraction to represent the edible portion. Note that in both cases plants are expressed in terms of dry weight due to the fact that water loss varies substantially after harvest up to the farm gate, even pre-processing. In our analysis, all impacts reside in the edible portion. As utilization approaches 100%, the stressor estimates would approach the estimates expressed in terms of live weight. Even though the two allocation

methods result in different stressor estimates, the relative performance of the blue food groups is robust to the choice of functional unit." (SI Lines 121-141).

We also include a table with all edible portion values applied (Table S1).

As now pointed out in the revised SI text above, while the magnitude of the stressors of course change when expressed as live weight versus edible portion, in most cases the relative rankings are very similar for either functional unit. As would be expected, the largest changes are: bivalves and lobsters for wild capture GHGs and farmed tilapia moved up approximately two positions in most stressors. For the farmed bivalves, they generally just swapped positions with the farmed seaweeds. There were a couple other position swaps, but these were for groups that were nearly indistinguishable to start with. We point to this issue more specifically in the new Box 1 text: "First, bivalve estimates change by nearly five-fold when expressed in terms of edible portion (Fig 1) compared to live weight (Fig S10) due to the shell weight." (Lines 113-115).

Finally, we also want to highlight that this is one of several papers from the Blue Food Assessment and given that we cannot address all topics, we want to highlight that two parallel papers are looking at demand patterns (Naylor et al. *in revision*) and another paper is putting Blue Food in the context of nutrition (Golden et al. *in revision*). There are many important and interesting questions that lie at the intersection of these papers which cannot be covered within a single paper.

8. *Other detailed comments*:

- Title of the paper: it should be nuanced considering the limited coverage of environmental problems (dismissing relevant impacts like antimicrobial resistance, fish resource depletion/overfishing, seabed damages, etc.).

Response: We revisited the title and ultimately decided that we feel the title is accurate and substantially more succinct than the alternatives we came up with. While the quantitative portion of the analysis is limited to the five main stressors, we do dedicate a section of the paper to other environmental impacts that we deemed impossible to quantify at present due to data shortages and

missing cause-effect pathways (e.g., all shortlisted impacts are strongly influenced by ecological factors that are poorly understood and extremely hard to normalize in a meaningful way at a global scale). We also identify this as an important area for further development in the text: "Capturing the full suite of environmental impacts will require more systematic data collection and methodological advancements. This is crucial for informing policy decisions and realizing the potential contributions of blue foods to sustainable diets while avoiding undesirable trade-offs." (Lines 159-162).

- l. 31: "employing inconsistent methodologies": the methodologies themselves are not necessarily inconsistent (LCA is an ISO-standardized methodology), but the lack of comparability between studies due to different methodological choices is the problem.

Response: This sentence was intended to point to the lack of comparability due to differences in methodological choices. We revised this sentence to now read: "Finally, estimates combining results of published life cycle assessments undertaken for different purposes and consequently employing incompatible methodologies19,20, cannot be compared reliably." (Lines 32-34).

While we find the revised phrasing an improvement, we would like to note that previous work has pointed to inconsistencies as an issue for LCA reviews:

"Peer reviewers of scientific articles should also be aware of these critical reviews and of the methodological issues indicated therein to prevent studies with insufficient documentation and/or inconsistencies—as some identified in the current review—from being published. Such practice should eventually contribute to bring more consistency and reliability in LCA studies to support decision- and policy-making processes in fields as important and relevant as the aquaculture sector." -Bohnes et al. (2018)

- l. 60: "consumptive freshwater use is largely limited to feed production and evaporation15": does this apply to freshwater aquaculture (which constitute the largest proportion of aquaculture)?

Response: Evaporation in this sentence refers to evaporation from freshwater surfaces on farms. To clarify this, we modified the sentence to read: "Although blue foods are produced in water, consumptive freshwater use is largely limited to feed production and on-farm evaporative losses for freshwater production16, with feeds accounting for >80% for all groups apart from carps and catfish (Fig S6)." (Lines 67-70).

- l.72: in terms of impacts, “negative emissions” can yield either positive or negative impacts (if background is a nutrient-rich environment or not). Nuancing may be pointed out (as was done in Halpern et al. for example)

Response: We agree and added text to the new Box 1 to draw attention to this point: “Third, impacts on biogeochemical cycling and habitats are highly context dependent. For example, the systems represented here extract nitrogen and phosphorus, which could be problematic in nutrient poor environments.” (Lines 119-121).

- l. 97-140 (ca. 3rd section of the manuscript): this section could be much sharpened, considering that such topic has been addressed in literature before. It is an essential aspect, which would deserve better introduction, perhaps an overview of relevant stressors to consider when addressing blue foods (beyond those covered), potential occurrences of trade-offs and risk of burden-shifting from one problem to another. A policy angle should also be considered here.

Response: The objective of Fig 2 is to point to the other relevant stressors for aquaculture and capture fisheries, pointing out that the quantitative analysis in the preceding section only includes five of these. One kind of trade-off is highlighted in Box 2. We also expanded our discussion of the importance of local context and interacting/nonlinear effects in this section. We leave the policy angle to the final section of the paper (changes we have outlined elsewhere in our response).

- l.105: antimicrobial resistance could be added as an important problem raised in aquaculture systems. See for example: Reverter et al. 2020 (Nature Communications 11, 1870).

Response: We agree that this is an important issue, which is why antibiotic resistance is highlighted in Fig 2 and application of antibiotics is pointed to as an additional stressor in the text (Line 140). To further this point, we now also include the climate change-disease-antibiotic use/antibiotic resistance risk as an example of an interaction among stressors: “Individual stressors may also have nonlinear relationships with impacts or act interactively44,45, such as climate change impacts compounding land use patterns that limit climate refuges or migration options46 or resulting in more frequent disease outbreaks, that increase antibiotic use and risk of antibiotic resistance.” (Lines 155-158).

- Box 1 (l. 128 and on): it is important to specify how such graph is to be used (not for showing relationships, but for supporting policy making, e.g. fishing crustaceans in Atl may be less GHG-intensive, but have higher risk for mammals trade-offs). The authors could also reflect more on the complementarity between impact assessment (e.g. carbon footprint and other stressors assessed) and risk assessment (from the text it seems the two are confused). Note also that if the authors wanted to illustrate trade-offs, there could be easier way at the level of stressors/impacts without involving risk assessment as done here.

Response: We use Box 2 (previously Box 1) to accomplish two goals. First, to illustrate one way that biodiversity risk can be integrated with LCA results. Such an approach builds off previous work on the integration of risk assessment and LCA. While there are fundamental constraints to integrating the two, the integration leveraging multiple approaches is valuable for identifying trade-offs (Guinée et al. 2017). Second, we aim to show that the stressors are not suitable surrogate variables for the other impacts, as there are trade-offs in some cases but not in others.

While trade-offs could have been presented in another way, we wanted to utilize the complementarity of LCAs and ERAs – that together may be utilized as a comprehensive method. LCA covers emissions and resource use of a fishery (quantitatively), whereas ERA aims to take a holistic view of potential ecological risks (qualitative or semi-quantitative). This also supports the local context point in the text (Line 154-155). Together, these points do not make any specific policy claims, but rather highlight the need for integrated management considerations across multiple environmental dimensions.

We focus on the capture fisheries-biodiversity relationship within the main text since this is an important aspect not covered in the five quantitative stressors. However, for an additional representation of the trade-offs and synergies among the aquaculture stressors, we added a spider diagram to the SI (Fig S5).

References:

Guinée, J.B., Heijungs, R., Vijver, M.G. and Peijnenburg, W.J., 2017. Setting the stage for debating the roles of risk assessment and life-cycle assessment of engineered nanomaterials. Nature nanotechnology, 12(8), p.727.

- l. 141 “reducing impact”: at what levels? globally, while providing food supply? nationally/regionally? or at technology/species level. The study, which does not have a systemic perspective, seems to only address the latter. This should be clarified.

Response: We believe that this subheading is sufficiently clear (particularly with the 40 character limit) in the context of the text that precedes as well as the text of the section itself, all of which is focused on the taxa group level and estimates on a per tonne edible portion basis. However, we did add text to the final section on the issues with scaling up interventions.

- l. 147-156: the split between 2 paragraphs (in addition to the text that perhaps could be clarified too) is very confusing as they address the same points

Response: The first paragraph referenced discusses FCR as a lever while the second paragraph discusses yield (in terms of fish production per unit area) as a lever, so we are not sure what is meant by them addressing the same points.

- l. 152: to complement earlier major comment: how realistic is it to make variations of 10% for all parameters? Bringing the analysis into an actual context would be relevant instead of keeping the discussion to a statistical analysis. Some attempts are made sometimes, e.g. l.154-156, but it lacks in-depth discussions and connection to actual systems. In that example, the discussion is kept to keeping the same technology, but how about shifting to other technologies that may bring less impacts.

Response: We covered these points in the earlier response but reiterate here that part a and b of Fig 4 accomplish two separate goals as a primarily serves as a sensitivity analysis to show how strong of an effect each parameter has on the outcome variable while part b is really focused on more realistic potential interventions. The realism of these scenarios is discussed in the earlier response.

- l.154-156: what about other technology alternatives (instead of fine-tuning/optimising the same technology). Maybe some technologies should simply be abandoned because not much more optimizable and emerging technologies should relace them (?) As mentioned above, such discussion is needed.

Response: Our analysis represents a snapshot of current production and our scenarios are based on values represented within the data itself as being plausible as well as literature values and substitutions that have been proposed. We do not feel comfortable speculating about the effects of technology for which we have no data but do now point to emerging technologies as an important area of future research: "Our estimates represent a snapshot of current production, but future work on emerging production technologies, feed

innovations and growing sub-sectors is important for tracking changes against these benchmarks." (Lines 99-101).

- l. 172-173: "this could create co-benefits for biodiversity impacts (Box 1)." With respect to biodiversity impacts, the coverage in the current study is far from being comprehensive (and certainly not guaranteed though Box 1). Box 1 cannot tell whether or not if low-emission gears used in Central America will increase or decrease the risk; so this is not a good reference. In addition, note that Box 1 actually illustrates and mainly describes trade-offs (increased risk) so no co-benefit (ref. to Box 1 at the end misleading here).

Response: We agree that the coverage of biodiversity impacts is not comprehensive and that the example in the box does not replace data on biodiversity impacts in all the LCA studies we have used. The box is only intended to show how these tradeoffs and co-benefits can play out in one specific case to inspire thinking about this in a wider context. The full sentence referred to is "In some cases, this could create co-benefits for biodiversity impacts (Box 1)" while the box (now Box 2) states "Integrating GHG emissions with marine mammal risk assessments reveals that some low-GHG emission gears are associated with higher marine mammal risks (e.g., gillnets and entangling nets; Fig 3), while bottom trawls show the opposite." (Lines 169-172). This supports the point that in some cases there will be co-benefits. In the main text, we further stressed that there are both cases of synergies and trade-offs ("For example, stressor estimates can be combined with ecological risk assessment to reveal instances of both trade-offs and synergies (Box 2)"), which underscores the need for contextual information related to biodiversity impacts that we further emphasize in the revised Box 2 text.

- l.173: "zero emission technologies" --> there will be emissions over the entire life cycle --> nuancing is needed and emphasizing the need for life cycle perspective to avoid burden shifting from fleet emissions to on-land emissions is important to mention here for nuancing. This is part of undertaking a systemic perspective (which was not done in current study).

Response: We revised this to "low emission technologies."

Figure 4a: I am a bit puzzled by the results for eutrophication: how come that by reducing 10% of crop (feed?) the eutrophication impact increase by 10+%? I may misunderstand the figure (which is difficult to grasp), but I read it correctly, then it is contradicting with the fact that off-farm N and P contributions are negligible in Figure S.4 (assuming the latter is correct; see my comment for Fig. S4 below).

Response: Since N and P emissions are dominated by the on-farm portion, the observed change relates to the increase in N and P inputs from the other feed components. A decrease in each feed component is related to a proportional increase in the other feed components as the feed proportions must sum to one. We added a sentence to the methods to ensure this point is clear: "We first tested the influence of model parameters on estimated stressors by reducing each parameter by 10%, holding all other parameters constant, and computing the difference between the baseline and perturbed stressor. In the case of feed composition, when one feed component is decreased, the others are increased proportionally such that the proportions still sum to one." (Lines 679-682). This is not intended to be a perfect representation of substitution of feed ingredients, as these groups serve different roles in the diet, but rather a test of the model sensitivity to the feed composition to demonstrate the strength of various lever options.

l.184: "most stressors" --> "most considered stressors". As mentioned above, keep in mind that you only studied a very few environmental problems and other relevant ones are not considered.

Response: We made this edit.

- l.185-186: "Capture fisheries vary widely in their GHG emissions but are low impact with respect to the other stressors considered." --> inappropriate and unsound statement. Taking some indicators as seabed damages, this statement is not true. Again, this discussion part needs to be nuanced considering that the study only has partial coverage of environmental problems. The disregard of this important limitation occurs frequently in the paper and needs to be corrected consistently.

Response: The sentence highlighted already specifically says "with respect to the other stressors considered" which we believe sufficiently points to the fact that this statement is only in reference to the limited number of environmental stressors considered within the study. Further, we dedicate the "From stressors to ecosystem impacts" section of the paper and Fig 2 to pointing out the many stressors that are relevant to fisheries and aquaculture but not within the scope. During the revision, we checked again to ensure consistency in our presentation of this.

- l.203: "young" maybe for aquaculture but certainly not for fisheries!

Response: We added "in aquaculture" to clarify this. The sentence now reads: "As a young and rapidly growing sector, there are many promising technological innovations in aquaculture (e.g., recirculating aquaculture systems, offshore farming and novel feeds)." (Lines 260-262)

- l. 221-223: role of aquaculture could have been an entry point for the paper. a pity to wait till the end to get such a contextual statement

Response: We added a related framing point to the introduction: "It is therefore critical to examine the environmental performance across the diversity of blue foods in a robust, methodologically consistent manner to serve as a benchmark within a rapidly evolving sector as blue food demand increases21, production shifts toward aquaculture and production technologies advance." (Lines 34-37).

- l. 224: word "enable" is definitely an overstatement considering the lack of transparency in the study.

Response: We disagree with this statement given that all data and code are included with this paper. At the manuscript submission stage, we agreed to follow Nature's guidelines to make all data available upon publication. All model code and input data to reproduce all analyses and figures in the study is already available on Github (https://github.com/jagephart/FishPrint). Within this revision, we have also committed improvements to the code documentation to facilitate future use, including a data/code flow chart and a more descriptive readme file mapping the data files to each script and output.

Methods:

- l. 356: ReCiPe mehods "mass balance calculations" --> what is meant? Have the authors altered the ReCiPe characterization factors for the different stressors? Please clarify. Please also specify the versioning of ReCiPe.

Response: We deleted "mass balance calculations" as it is part of the inventory and specified that we use ReCiPe 2016 Hierarchist model.

- l. 363: what is defined as "LCA models"; is it a study? Is it a compared alternative? You need to clarify!

Response: We should have said life cycle inventory data. We also clarified this in the introduction. The sentence referred to now reads: "For each study, where available, we compiled inventory data on economic feed conversion ratios, general feed composition (soy, other crop, fishery, and livestock-derived ingredients), electricity use, inputs of diesel and other energy carriers, and data relating to stocking densities and annual yields (t m-2)." (Lines 386-489).

- Remark: ecoinvent 3.0 is quite old (from 2014/2015) and had some bugs, incl. water flow balance. Any reason why a more updated version was not used? (current version accessible is 3.7)

Response: This should be ecoinvent 3.6 and was just used for the energy sources. We corrected the version number in the methods and SI text.

- l. 419: how did you handle gaps, e.g. species or technologies not covered by emission factors --> hence no impacts? Also in studies reviewed: how to deal with studies only covering part of impact categories, e.g. not providing LCI for all 4 impacts? How did you handle those?

Response: Gaps for emission factors and impacts categories is not problematic for our approach since we use the underlying inventory data to estimate the parameters used within our models. The handling of missing parameter values is described within the model description itself. Since we took a Bayesian approach, instances with few data points for a parameter generally results in greater uncertainty in the final estimates. We clarified this workflow in the new methods flowchart figure (Fig S1).

- l. 422-423: what is the coverage of species within each taxa group? Again information of what is actually covered in terms of species, geography and technology, etc. is lacking, rendering any discussion of representativeness impossible to check.

Response: The species included in each taxa group along with the number of farms and number of studies within each are presented in table S1. The full breakdown of species, geography and production method is included in the data file which has now been released.

SI file:

- There is a general confusion of concepts between emission and impact in the study in both the manuscript and SI, yielding ambiguity when reading the paper. Table S1 suggests that some impact assessment is performed but the authors often refer to emissions and even units are sometimes incorrect (see Figure Sx, where the units for eutrophying impacts should be expressed in kg-Neq and kg-Peq). Scientific rigor on this aspect should be expected at least in a study addressing environmental performances.

Response: One challenge we face is that we recognize these terms have specific meaning within the LCA literature, but we aim to use them in a way that is understandable for a broader audience. We use "emissions" to refer to GHG, N and P emissions, representing three of the five stressors (with the other two representing resource use - land use and water use). We distinguish this use from impacts at the beginning of the "From stressors to ecosystem impacts" section: "Emission and resource use stressors are valuable for comparing environmental performance across foods but cannot fully capture final ecosystem and biodiversity consequences (i.e., impacts)." We revisited the main text and SI to ensure this use of the terms is consistent throughout.

We assume that by Table S1 you mean Figure S1 (now Fig S2). We did perform an impact analysis, defined in the Methods section through the model description. This figure defines the system boundaries for the analysis. This complements the new Fig S1.

We also ensured that "-eq" is included in the units in all SI figures.

- l.42-43 (and rest of paragraph): more details on system boundaries and scope of the assessment are required. This should describe and argue for what was included/excluded, data quality and representativeness for specific processes and implications. There is a brief outline of how the "simplified LCA" approach was conducted, with some references to selected papers and LCI databases, but there is no discussion about data representativeness (e.g. how are the data from Agri-footprint representative of all feed production processes in use today? In terms of time, technology and geographical coverage, cf. scope definition of LCA studies).

Response: We expanded on our goal and scope description, especially to describe and justify the attributional approach, allocation method and functional unit. We previously responded to points about data representativeness and the inclusion criteria but will reiterate a few key points here. The species and geographic representation are illustrated through improved figures (Fig S3 and S4) and the time period is all post 1995, with all but one study (on mussels) coming from after 2000. All studies represent

commercial scale production systems, as experimental and hypothetical systems were excluded. This is included in the methods description and illustrated in Fig S1.

Crop inputs selected from Agri-footprint were informed by the major crops identified as relevant to global aquaculture in Troell et al. 2014., and identifying key source countries based on rates of production and export by crop. As for the processes represented in the Agri-footprint data, we refer to the Agri-footprint documentation: "Agri-footprint contains attributional LCIs, so generally average mixes are considered that are representative for the specific crop, process, transport modality, product or location... The main baseline data source is the public domain (Scientific literature, FAOstat, Eurostat, etc.). Data from the public domain are assessed based on representativeness (time-related coverage, technical coverage and geographical coverage), completeness, consistency and reproducibility. When data from public or confidential research initiated by the industry and conducted by Blonk Consultants are more representative, complete or consistent, these data were used. Where possible, the data have been reviewed by industry experts.... "

- l. 51: please specify the correct version of ecoinvent 3. In the manuscript, it is 3.0 (which in my opinion is rather outdated now).

Response: The correct version is 3.6 and we corrected this throughout the manuscript.

- l.55-61: the use of allocation needs to be defended (relative to the use of system expansion, which according to ISO14044 precedes the use of allocation). An attributional modelling as applied in the analysis-type study by the authors may warrant including system expansion to account for secondary functions that may differ across the different species/taxa groups (e.g. different byproducts) and across different technologies. This may be considered a better reflection of actual market mechanisms. The authors discusses the choice of different allocation methods, which is fine, but before doing so, I encourage them to justify the choice of allocation first.

Response: We have expanded our justification for the allocation method in the SI:

> **"Co-product allocation is the division of impacts among products originating from the same process (e.g., livestock meat and by-products). Allocation is an artificial problem and can therefore only be solved in artificial ways14. One should, however, seek to maintain a consistent allocation strategy throughout the modeling and try to adhere to well**

justified allocation strategies14. In attributional LCAs, as in this case, the most common allocation strategies include: cut-off, where all impacts are attributed to the main product; allocation based upon mass; allocation based upon economic value; and allocation based on gross energy content.

Our global scope also complicates economic allocation due to regional price differences and price fluctuations over time. In terms of edible yields, cultural preferences strongly influence the market price. Fish heads or racks are, for example, more expensive than the fillets in some markets. We consequently choose to use mass allocation for our primary results, as it is simple to understand, can be consistently applied to all regions, and remains static over time. It is also the only allocation option possible for capture fisheries, as details on all types of fish landed were limited6.

To test the robustness of our conclusions, we also recalculated results using allocation based on monetary value (economic allocation) and gross energy content (Fig S11-S12). These results landed in slightly different absolute impacts, but their relative ranking remained largely consistent. Thus, we argue that in the relative realm of relative LCA results, changing allocation strategy does not undermine our study's conclusions."

We provide economic and energy allocation estimates for aquaculture in the SI and include the scenario analysis based on economic allocation as well.

We also added paragraphs to the Goal and Scope section of the SI related to the attributional approach:

"The goal and scope of our study sets out to define the current environmental impact of different blue foods globally. Thus, our simplified modeling approach adopts an attributional LCA approach whereas consequential LCA modeling warrants an understanding of substituting food products that would need to be grossly simplified at the global scale and broad taxa groups within our study. Even at a regional level it is hard to predict which food commodities blue foods substitute, and preference for different blue foods is in rapid transition12. Market substitutions will be further influenced by realities such as: not all fish species can be farmed; wild and domesticated fish can have different fillet yields; consumer perceptions of wild and farmed fish differ, and, maybe most importantly, price remains the main determinant for food

procurement choices13. Subsequently, system substitution (system expansion), was not a viable option for our model."

- l. 77: please indicate ReCiPe 2016 (full name of the methodology)

Response: We have added the full name of the methodology here.

- Table S1: This table is not informative at all because it does not provide any explicit references for the data described, which is a major flaw (see above comment on lack of transparency). It does not specify how many distinct studies there are either (e.g. the three "1" for aquatic plants: are they the same study or three different studies?; this can be important in terms of methodological consistency across studies). Meta data on time, technology and geographic representativeness associated with each species/taxa group is also absent, albeit essential here.

Response: We disagree that this table is not useful but agree that it does not provide every aspect a reader may be interested in (given that it is a summary table). The complete study-level details are now provided in the data file and accompanying "SI_reviewed_studies" csv file, which includes full references for all studies reviewed and notes for any studies excluded from the analysis for not meeting the inclusion criteria.

Following up on previous comment: discussion should also be included on taxa groups associated with very few studies (in terms of uncertainties). For example, taxa group with very few studies of different nature (technology/species wise) could explain some of the high variance observed in l. 143-144 (e.g. miscellaneous marine and diadromous fish stressors)?

Response: There are two distinct points here that should be differentiated. Within a Bayesian modeling framework, groups with few data points will have greater uncertainty in the mean because there is less information to overcome the weakly informative priors. Put another way, there is a lot of uncertainty in the estimate of the mean because there is not much data informing what the true mean of the distribution is. This is meaningful because it provides insight into where we are least certain about what the true mean of the group is. We also want to reiterate that species which represent small production volumes are given lower weight (proportional to global production) in the group-level mean.

The above statistical issue is distinct from high variability within a group, which is what is depicted in the observation-level distributions by taxa (Fig S7), system type (Fig S8) and system intensity (Fig S9) as one can have a lot of confidence in the mean of a wide distribution. We changed "variance" to "variability" to help distinguish this in the text. The groups that are highly variable are likely to some extent a function of the number of species and production methods included across the observations in the group, as we now emphasize in the text: "While some variability within a taxa group is due to differences in on-farm practices, production technology is an important factor across stressors49 as variability in stressors for a given species reared in different farming systems can be considerable (e.g., 50)." (Lines 187-190).

- Fig. S2: it is difficult to discern what happens in the cluster at the bottom left corner of the graph. Additional information in the caption should help appraise the situation there (may not be relevant once data representativeness is appropriately addressed). In addition, are those data points only the LCA studies on full aquaculture systems, or do they include the feed data points added (which should be separated, in my opinion)? Or is it representing the 1590 farms mentioned in the manuscript? There is a lack of clarifications on what is represented on this graph.

Response: We revised this figure (now Fig S3) and the following figure. To improve the readability, we used a log scale on the y-axes and added labels. We also scaled the points by the number of studies from which the observations were drawn. The number of farms refers to the number of farms included in that data point. As defined in this paper, it is the number of farms recorded in each study, where we default to one if the number of farms is not reported, thereby representing a conservative estimate of the number of farms included in the study. We updated the figure caption to now read: "Number of farm-level observations from which life cycle inventory data was recorded for each taxa group compared to total global production (2014–2018). Points are scaled by the number of studies from which the observations are recorded." The caption for Fig S4 is similarly edited.

- l. 94: for transparency, I think the background data for all calculated stressor estimates should be disclosed.

Response: This is provided in the data file, which we have now made available on our GitHub repository so that it links directly with the analysis code.

- Fig. S4: I may have missed something, but is it not surprising not to have contributions to eutrophication impacts (negligible) from off-farm processes (especially in light of feed production that may consume fertilizers?)? Maybe worth checking, if not done already.

Response: We also found this a little surprising, but the estimates are in line with estimates from the farms in Henriksson et al. (2017), while acknowledging a strong influence by different agricultural nutrient run-off models. We also re-checked the N and P content data as well as the N and P emissions data for the crops from ecoinvent 3.6 and did not identify any errors in these data inputs.

References:

Henriksson, P.J., Dickson, M., Allah, A.N., Al-Kenawy, D. and Phillips, M., 2017. Benchmarking the environmental performance of best management practice and genetic improvements in Egyptian aquaculture using life cycle assessment. Aquaculture, 468, pp.53-59.

Figure 4 caption: have units on graphs not in captions. Also have the units right (Neq, Peq, etc.)

Response: We believe the reviewer is referring to Fig S4. We have moved the units onto the graph and ensured that all units are displayed correctly.

- Table S3: what is reported? Is it r-square here, or? Please refer to conventional ways of reporting statistical results in scientific publications.

Response: This is a correlation matrix, which is Pearson's r and is a standard statistical presentation for scientific publications. We expanded the table caption to read "Correlations among stressors. The correlation matrix among the stressors represents Pearson's correlation across all stressor observations.".

- Captions for Figure S14 and others: complete lack of explanations to understand what the figure shows. Even if some explanations are in the manuscript for some figures, each figure should be self-standing and the reader should not have to piece things together to understand how to read the figure.

Response: We have added the full caption description to Fig S14, which now matches that of Fig 4 in the main text (apart from the fact that this is run for the estimates based on economic allocation).

Recap of references cited:

Avadi A, et al. (2015) Comparative environmental performance of artisanal and commercial feed use in Peruvian freshwater aquaculture. Aquaculture 435, 52–66.

Bohnes FA, et al. (2019) Life cycle assessments of aquaculture systems: a critical review of reported findings with recommendations for policy and system development. Reviews in Aquaculture 11, 1061–1079.

Bohnes FA & Laurent A (2019) LCA of aquaculture systems: methodological issues and potential improvements. The International Journal of Life Cycle Assessment 24, 324–337.

Henriksson PJG, et al. (2017) Indonesian aquaculture futures – evaluating environmental and socioeconomic potentials and limitations. Journal of Cleaner Production 162, 1482–1490.

Reverter M, et al. (2020). Aquaculture at the crossroads of global warming and antimicrobial resistance. Nature Communications 11, 1870.

Weidema BP & Stylianou KS (2020) Nutrition in the life cycle assessment of foods—function or impact? The International Journal of Life Cycle Assessment (2020) 25:1210–1216

Referee #5 (Remarks to the Author):

Fisheries and aquaculture are underrepresented in studies on the environmental impacts of foods. This manuscript uses secondary data at the production and input phases to calculate environmental impacts of live weight fisheries and aquaculture products (ie., farm gate /dock sales) using a mass allocation method.

The methods and approach are appropriate and the sample sizes are in line with what's available in the peer-reviewed literature. The main strengths of the paper are the use of life cycle inventory data instead of life cycle impact assessments, which gives this manuscript a distinct advantage over other similar papers. However, this may be lost on readers. It might help to emphasize this point as well as bring in the arguments made by co-authors in Henricksson et al. 2021. GFS https://doi.org/10.1016/j.gfs.2021.100508. The authors also take a global geographic scope including a large proportion of global production, but miss small scale fisheries and aquaculture which is unfortunate but I'm not sure if this is solvable due to data gaps.

Response: We agree that the novelty of our approach can be clarified. To do so, we added brief text to the introduction (as we still want to keep it oriented to a broad audience), which reads: "We identify a set of key life cycle inventory data (i.e., material and energy input and farm-level performance data) from published studies and datasets to which a harmonized methodology is applied." (Lines 40-42). More importantly, we expanded on this in the methods section and added a new methods flowchart (Fig S1) to clarify the overall process. We also added the Henriksson et al. (2021) reference to the introduction.

The point related to small-scale production is accurate. We went to some lengths to try to incorporate inland fisheries, which are predominantly small-scale, but were unable to identify sufficient data (see response to Referee #1 on inland fisheries). We added text to ensure that this gap with respect to inland and marine capture is more clearly communicated: "Our capture fishery data primarily represents commercial marine fisheries25. However, subsistence marine and inland catches often utilize non-motorized or no vessels which likely generate few emissions, but there is insufficient data on fuel use across the diversity of small-scale fishing methods to reliably estimate emissions." (Lines 93-96).

We do have representation of small-scale aquaculture production within our database, but it is difficult to evaluate the representativeness of this because not all studies report the farm size, there is not a standard definition of small-scale, and there is not global data on the size distribution of aquaculture farms

against which we can compare. Nevertheless, this is an important point and we have added it to the text on data representativeness in the methods: "Further, while we have both small- and large-scale producers represented in the data, we do not have complete data on the size of all farms or the size distribution of farms globally, so we cannot compare the representativeness with respect to producer size." (Line 544-547).

Fisheries and aquaculture are included together, however, the treatment of aquaculture in parts of the analysis and supplement is a bit richer. Is this because fisheries already has good synthesis w/ life cycle methods? Or that the bounadry conditions for aquaculture are broader, thus more room for analysis? or something else.

Response: As part of the Blue Food Assessment, this paper set out to cover the range of aquatic foods and to provide standardized estimates across capture and aquaculture production to ensure representation of aquatic foods in the broader food systems dialogue. However, it is true that more space is dedicated to aquaculture within this manuscript. This is for two primary reasons. First, we wanted to focus on the stressors that are most commonly reported for comparing terrestrial foods. Of these five (GHG, land, water, N and P), only GHGs are non-trivial within our system boundaries and stressor definitions (e.g., inland fisheries require environmental flows, which could be considered a passive use of water, but we only look at consumptive use here). This means that we discuss the results related to the non-GHG stressors for aquaculture, but not for capture. Second, the GHG results for capture production are based on the fuel use data reported in Parker et al. (2018). We reanalyzed this data so it would be directly comparable to the aquaculture estimates, but we wanted to avoid repeating findings previously reported in the literature.

Within this revision, we did add some additional text related to capture fisheries to improve the balance to some extent, but also ended up adding additional text and figures related to aquaculture in response to other reviewer comments, so the balance remains tilted toward aquaculture.

References:

Parker, R.W., Blanchard, J.L., Gardner, C., Green, B.S., Hartmann, K., Tyedmers, P.H. and Watson, R.A., 2018. Fuel use and greenhouse gas emissions of world fisheries. Nature Climate Change, 8(4), pp.333-337.

Specific comments

L9 "…but flatfish and crustaceans generating the highest." add greenhouse gas emissions to the end of the sentence?

Response: We added this to the sentence, which now reads: "Capture fisheries predominantly generate greenhouse gas emissions, with small pelagic fishes generating lower emissions than all fed aquaculture, but flatfish and crustaceans generating the highest greenhouse gas emissions." (Lines 7-9).

L77-9 "To compare with terrestrial foods, we estimated stressors for chicken…" I like that the authors selected chickens since they are the most similar in terms of environmental impacts to fish. As the authors note in the intro, granularity is important and something most LCAs miss with blue foods. In that spirit, it would be helpful to note here that the chicken data is from US/Europe and from intensive/vertically integrated farms, which gives readers key information with which to make the chicken/ blue food comparison.

Response: We agree and have updated the text accordingly to read: "To compare with terrestrial foods, we estimated stressors for industrial chicken produced in the US and Europe and find it falls in the middle of farmed blue foods, with similar stressors as tilapia (Fig 1, S14)." (Line 85-87).

L81-2 "Notably, the groups generating the lowest stressors (e.g., bivalves and small pelagic fishes) also provide the greatest nutritional quality across all forms of aquatic foods2,9." I'm not sure I agree. They are higher in omega-3 and vitamin B12, and zinc for bivalves but similar for most other micronutrients listed in Fig 1 of (Golden et al. in rev).

Response: Fig 1 of Golden et al. (in revision) is ordered from best (top) to worst (bottom) using weighted averages across all nutrients, with bivalves and small pelagics at the top. This supports the sentence about overall nutritional quality.

L125-8 I like the cross-cutting look at marine mammal risk x GHG for fishing gear types. It's important to show where synergies exist. Are decisions about gear ever made with fuel use in mind instead of what gear is most efficient for max sust. yield? I see this comes up again in L171-2 without refs and L215 supported with refs. In L171 I was left wondering how serious this idea was?

Response: We are not aware of the two dimensions influencing decision making together at the policy or industry level. Fuel use is often an important operating cost, so there are incentives to reduce fuel use, whereas gear choice depends on tradition, management and species. If it is a highly valuable species, the fuel cost of an energy-intensive gear may also not be as important today (such as bottom-trawling for crustaceans), especially if fossil fuels are subsidized. If fuel prices go up, fishing patterns often change (such as species targeted, fishing grounds closer to shore) – but only within the opportunities that lay within the management regulations, where change of gears is not always possible (fisher may be "locked in" a sector). Given the important context dependency of biodiversity impacts (which we now detail further), we don't provide this as a clear recommendation for a universal win-win, but rather use the Box to illustrate a potential opportunity. This box text and figure is what supports the previous line, while the references in the original Line 215 are in relation to management improvement opportunities and the fuel use improvements that can occur with improved management. We certainly agree that this cross-cutting approach is an important frontier for informing sustainable blue food production and hope that raising this issue as an example can seed further work in this space (which we also point to in the text).

L184-5 "...fed aquaculture still outperforms many other animal-source foods." I see that the point of this paper was not to compare aquaculture and animal sourced food, but it might help readers if you specify which ones, for example beef and pork, but not eggs and poultry? Eat Lancet with a Blue Lens has a nice comparison across species groups, but not with the granularity of this study.

Response: This work (and the work of the entire Blue Foods Assessment) builds off of the EAT Lancet report and the indicated scoping report, with a goal of adding much more resolution to aquatic foods in order to better represent the vast diversity of production systems. We agree that it would be great to include such information, but this would require a substantial expansion of the data collection that was not possible during the allotted revision time and we likely would not have sufficient space to fully discuss this broader range of comparisons. Given that we could not expand the scope of the analysis to cover the full range of terrestrial foods that are of interest, we decided to modify this sentence rather than speculate beyond the analysis conducted: "Blue foods already have great potential for reducing food system environmental stressors. Unfed aquaculture results in negligible values for most considered stressors, and many fed aquaculture groups outperform industrial chicken, the most efficient major terrestrial animal-source food." (Lines 230-233). This does imply that blue foods typically outperform beef and pork, but there is wide variability in these production systems (Poore and Nemecek 2018) so we are hesitant to generalize without the available space to provide such important context for the comparison.

References

Poore, J. and Nemecek, T., 2018. Reducing food's environmental impacts through producers and consumers. Science, 360(6392), pp.987-992.

L192 "Further, blue foods provide the highest nutrient richness across multiple micronutrients, vitamins, and fatty acids" as mentioned above, I think this overgeneralizes and oversimplifies the comparison. Should the authors specify what type of fatty acids (polyunsaturated)? If the comparison is between two groups (blue foods vs terrestrial foods) should it be "higher" instead of "highest"?

Response: We revised the text to be more specific. The sentence now reads "Further, blue foods provide the highest nutrient richness across multiple micronutrients (e.g., iron, zinc), vitamins (e.g., B12), and long-chain polyunsaturated fatty acids (e.g., EPA and DHA) relative to terrestrial animal-source foods10, which may provide greater incentive to shift demand since consumers generally prioritize seafood freshness, food safety, health, and taste over sustainability56." (Lines 241-245).

L144 "High variance in milkfish and miscellaneous marine and diadromous fish stressors points to large potential performance gains per unit…". Was sample size a cause of higher variance?

Response: There are two distinct points here that should be differentiated. Within a Bayesian modeling framework, groups with few data points will have greater uncertainty in the mean because there is less information to overcome the weakly informative priors. Put another way, there is a lot of uncertainty in the estimate of the mean because there is not much data informing what the mean of the distribution is. This is meaningful because it provides insight into where we really are least certain about what the true mean of the group is. We also want to reiterate that species which represent small production volumes are given lower weight (proportional to global production) in the group-level mean.

The above statistical issue is distinct from high variability within a group, which is what is depicted in the observation-level distributions by taxa (Fig S7), system type (Fig S8) and system intensity (Fig S9) as one can have a lot of confidence in the mean of a wide distribution. We changed "variance" to "variability" to help

distinguish this in the text. The groups that are highly variable are likely to some extent a function of the number of species and production methods included across the observations in the group, as we now emphasize in the text: "While some variability within a taxa group is due to differences in on-farm practices, production technology is an important factor across stressors49 as variability in stressors for a given species reared in different farming systems can be considerable (e.g., 50).)" (Lines 187-190).

L152 "…a 10% yield improvement reduces…" I thought you meant yield in processing and then reread the supplement the boundary conditions and scenarios. It might be good to clarify this point.

Response: This is a good point. We modified the text to now read: "Holding all else constant, a 10% production yield improvement (t ha-1) reduces land and water use for freshwater pond systems by 1–9% (Fig 4a)." (Lines 217-218).

L206 and L212 could these thoughts be combined? I agree that there is a need for more than just technology (L206) and was wondering when that would be mentioned.

Response: We were not entirely sure how the reviewer was hoping for these to be combined as they are currently part of the same paragraph, with the technology pieces followed by the policy changes. We did however expand the discussion of the policy changes and key actors within this paragraph.

L207 in addition to husbandry more generally, it would be useful to call out disease and mortality specifically as important factors.

Response: We added this so that the sentence now reads: "These include policy or technological interventions that improve husbandry measures (especially reducing disease and mortality) and lower FCRs" (Lines 263-265).

Figures.

Fig 1. It may be helpful to define what "credible intervals" are for readers.

Response: This is a good suggestion. We edited the Fig 1 caption to now read: "Dot indicates the median, colored regions show credible intervals (i.e., range of values that have a 95% (light), 80%, and 50% (dark) probability of containing the true parameter value)"

Fig 3. The authors switch from median and credibility intervals in Fig 1 to mean and 95% CI in Fig 3. Is this an oversight or is there a reason for switching?

Response: Thank you for noting this. We changed Figure 3 to represent credible intervals so that this is consistent across figures.

Fig 4. The x-axis is a little bunched. Could the author stretch the plot wider for ease of reading Fig4 A and to distinguish differences in the lollipop plot in Fig 4 B.

Response: We made a number of improvements to Figure 4 to improve its readability: 1) we changed the color from the yellow-blue color scheme to red-blue (still drawing upon the colors from the Blue Foods Assessment color palette though); 2) we shortened the labels in b and aligned the stressor categories in parts a and b; 3) we reduced the margin sizes and added arrows for any change greater than 50% to make the other lines more visible. We made the same adjustments to the analogous figures in the SI.

Supplement

The fisheries component is not as well characterize in the supplement as the aquaculture data. For example, after reading Fig S1 I was wondering if there was similar table for fisheries?

Response: The system boundaries and methodological approach are much simpler for capture fisheries given that fuel use was the only contribution to the stressor calculations, along with the weightings for production and gear type. The calculations for capture fisheries are also largely a re-expression of results from Parker et al. (2018) so that the estimates would align with similar taxa groups as aquaculture. We wanted to include capture fisheries to provide a more complete picture of blue foods, all calculated in a standardized way, but we wanted to avoid duplicating any of the content from Parker et al. (2018). The result is that we included much less content on capture fisheries than aquaculture. However, to make this point clearer, we added text to the SI to emphasize this point.

Fig S4. Blue and green colors are difficult to tell apart

Response: We updated the colors in Figure S4 to improve the contrast between the two colors. We also changed the order of the bars to align with Figure 1 and removed the duplicated y-axis labels to utilize more of the space for the data. We similarly revised Fig S11 to align the colors for on- and off-farm sources of stressors across the two figures.

Reviewer Reports on the First Revision:

Referee #1 (Remarks to the Author):

The authors addressed my comments satisfactorily

Referee #2:

The authors have done a great job addressing the comments raised during the previous round of review.

The sole remaining comment I have is to mention the functional unit earlier in the main text. It is a bit hidden right now because it follows a long string of numbers and is in a parentheses, and I'm concerned that readers might not see it.

Adding something similar to "reporting outcomes per tonne of edible weight" to the end of the sentence starting "We leverage data from over 1,690 farms..." in the introduction is one idea. A different option is to add a short sentence to the first paragraph of the section titled "Blue food environmental stressors".

Referee #4 (Remarks to the Author):

Starting with a side remark, I still cannot understand why the full documentation, now accessible via the GitHub link, was not made available in the 1st round of review for the paper. As I wrote previously, this is a must for any sensible peer review process, and it would have saved quite some time and efforts – and unnecessary comments – in the previous round, where I tried to piece together what had been done in the study, which then definitely looked more like an advanced draft than a submitted paper.

That said, the study has now greatly improved and it has gained a lot in clarity and relevance. Several/most of my earlier comments have been addressed. I still have some left though, including a few new ones:

Major/Main comments:
* l.457-473: please explain how the model constructed has been validated. Have you performed some cross-validation of the model? and/or have you compared the results obtained from the simplified LCA with results from the original studies when this could be done (e.g. for climate

change impacts when the study had similar methodological choices)? This is an important point to document in both Methods and SI to justify the validity of the results you display.

* In relation to the scenarios performed (Figure 4b, Table S8), the authors should indicate in the manuscript (along with Figure 4b analysis), methods and SI the limitations of the explorative study on decreasing environmental impacts of blue foods. Doing such assessments, which are in fact consequential in nature (as clear when reading Table S8), with an attributional/allocation-based model has little meaning, in my opinion. As a bare minimum, the authors should be open and discuss/comment in the text about the inconsistency of this approach and the implied uncertainties/limitations.

* l. 49- 103: For a given taxa group, the variability in Figure 1 reflects the variability of input data (if I understood correctly). Therefore, it would be nice for the reader to add 1-2 sentences, e.g. in l.92-101, commenting on how representative of the different locations and technologies for each species/taxa the data are, so the reader has an impression of how comprehensive/reliable each "box plot" in the different Fig. 1.x is. I think these are relevant information because most of the text before l.92 in this part is focused on comparing across species groups for each impact category.

* l. 219: I recommend nuancing the statement, as it could be misread/misinterpreted by the readers. It may create co-benefits but may also create new impacts as well, and only a full coverage of environmental problems can reveal such trade-offs. That is a general point – and a very important one – to make in this paragraph that any decision to change current production modes should be accompanied by a holistic environmental sustainability and risk assessment to avoid burden-shifting (across impacts/risks). This message is only touched upon in Box 2, although with another purpose/context. It should be reinforced in the text, especially since the title of the paper reads "environmental performance" (which signals a broad scoping).

* Data representativeness: the added text in the Methods responds to my earlier comment, and it is now clear what the authors did, and I understand the data limitation preventing to quantify the actual representativeness. That said, I recommend 2 additional adjustments to be open about this: (i) In l. 43, I would still suggest to alleviate any ambiguity as to what the 70% coverage relate to: I suggest to split the sentence to make clear that you refer to the taxa coverage, and not the actual representativeness of the production processes (which are the drivers of environmental impacts). (ii) in Figure S3 and S4, I recommend complementing the captions to indicate what is meant by "representativeness" in those graphs. As I mentioned previously, the study may cover 99 farms that capture 1% of the actual production volume of a given species in the country while the 100th farm producing 99% of the species in that country – and which would not be covered – may drive all environmental impacts of the production of that species (with a different intensity than the mean across the 99 farms). Although the study is certainly the most comprehensive to date and no data exist to quantify the actual representativeness, I think it is important that the authors are open about this potential bias.

* Units for most Fig. Sx (e.g. S6, S7, etc.) and figs in the manuscript, incl. Fig. 1, etc. (+ Table 1): unlike what has been replied to my earlier comment, units are still wrong on several/most graphs, incl. missing "eq" for climate change (should be "kg-CO2eq"), incomplete units for land use (should be "m2a"), etc. The unit for land use is also wrong as it should be "m2a-crop-eq", which is the actual unit if you applied ReCiPe2016 for land use impact assessment (right now the "crop-eq" is systematically missing). I therefore do repeat my comment from the previous round of reviews: please do a *thorough* check of all units for all figures and tables; this was obviously not done last time.

Other comments:
- I am now very glad to see a full transparency on the study, incl. background data and codes. It is a non-negligible added value to the work, in my opinion. I would suggest however to combine the

information of CSV file “LCI_compiled_for_SI” (from Github repository) with the information from the file “384293_2_extended_data_3667030_qtgbcb” (provided as SI) and to have this into a single Excel file as SI to the paper. I think it is an important file to accompany the paper (instead of going through the Github repository to find this) and the information in both are very relevant (raw LCI data in the former and full referencing with justifications for inclusion/exclusion in the latter).
- To document your modelling approach, you may consider including some of the information of the workflow figure from Github in the SI as well (simplifying the figure to drop the orange boxes that are not necessary for understanding). I just found the blue boxes quite useful to understand whats was done, i.e. more explicit/clearer than the methods flow chart in Fig. S1. The workflow also seems to cover the entire study scope, while Fig. S1 is limited to aquaculture (although not explicitly written in the caption).
- l.25: please add “for example” before referring to the specific 3 goals you list. Others could also be listed with direct link to aquaculture/seafood (nr. 3, 6, 13, 15, etc.).
- l.93-96: how important are the non-commercial fisheries (subsistence) relative to commercial fisheries? I have not checked and I do not know if it is possible, but my initial thought would be that they are negligible. Therefore, if so, the data gaps would not be a major concern here.
- l. 97: grammar to revisit
- l.99: “snapshot of the knowledge of current production” (or alike) instead of “snapshot of current production”, which, as correctly stated in the previous sentences, does contain gaps (this still relates to my previous comment on data representativeness).
- Fig. 1e: remove “farmed” (or add that information to all graphs systematically)
- l.115: which processes (give 1-2 examples very briefly)
- l.119: “from” instead of “form”
- l.122: please correct to “tropospheric” and not “stratospheric” for ozone. These are two very different types of impacts.
- Figure 2: I like the figure, which reads better now with the text changes. However, it mixes different concepts in it and I recommend to “clean” it to only retain actual environmental stressors as understood in the study. For example, energy consumption is not a stressor, but emissions and resource use deriving from it are (and those are already included). Another example may be fish-derived resources: not clear to me how this is an environmental stressor. Other symbols are also redundant, e.g. by-catch (twice represented).
- l. 165: “sustainability messaging” into “sustainability communication” or alike (?), if that is what you meant. Please clarify the wording.
- l.169: Please add “As an example, integrating…”. You only give an example here.
- The difference between the 2 headings “Reducing blue food environmental impacts” (l.182) and “moving toward lower-impact blue foods” (l.229) is too subtle in my opinion (the different focus of each underlying part is clear though). I recommend you better distinguish the 2 headings (maybe reformulate the latter).
- l.262: suggest another word to “charismatic” (mainly used for persons).
- l. 506: as already asked previously, please indicate the version of ReCiPe2016 used (“Hierarchist” set of characterization factors is relevant to mention, but it says nothing of the versioning).
SI file:
- Indicate which ecoinvent db you used (there are three).
- l. 88-89: I would like that you explicitly refer to ISO14044:2006 in your argumentation about systematically using allocation in your model (which is a last resort in ISO hierarchy). This is important, e.g. for the by-product handling when converting from live weight to edible weight.
- Fig. S4: legend is missing some items (small dots representing 1 study, I guess?).
- Fig S5: should there be some numerical values on the axes? Or else indicate which species is taken as a baseline (does not seem any is though..).
- Table S4: please add “Pearson correlation coefficient” (coefficient missing)

Referee #5 (Remarks to the Author):

The manuscript is an important contribution to the literature and a definitive study of the environmental performance of aquaculture. The authors have addressed all of my comments. I have no further comments.

Author Rebuttals to First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

The authors addressed my comments satisfactorily

Response: Thank you for your review and we are glad we have addressed all of your comments.

Referee #2:

The authors have done a great job addressing the comments raised during the previous round of review.

The sole remaining comment I have is to mention the functional unit earlier in the main text. It is a bit hidden right now because it follows a long string of numbers and is in a parentheses, and I'm concerned that readers might not see it.

Adding something similar to "reporting outcomes per tonne of edible weight" to the end of the sentence starting "We leverage data from over 1,690 farms..." in the introduction is one idea. A different option is to add a short sentence to the first paragraph of the section titled "Blue food environmental stressors".

Response: We added this to the last paragraph of the introduction, with the sentence now reading "Here, we provide standardized estimates of greenhouse gas (GHG) emissions, consumptive freshwater use (water use), terrestrial land occupation (land use), and nitrogen (N) and phosphorus (P) emissions for blue foods, reported per tonne of edible weight."

Referee #4 (Remarks to the Author):

Starting with a side remark, I still cannot understand why the full documentation, now accessible via the GitHub link, was not made available in the 1st round of review for the paper. As I wrote previously, this is a must for any sensible peer review process, and it would have saved quite some time and efforts – and unnecessary comments – in the previous round, where I tried to piece together what had been done in the study, which then definitely looked more like an advanced draft than a submitted paper.

That said, the study has now greatly improved and it has gained a lot in clarity and relevance. Several/most of my earlier comments have been addressed. I still have some left though, including a few new ones:

Response: We are glad that our revision has clarified the value of the study and we respond to the remaining comments below.

Major/Main comments:

* l.457-473: please explain how the model constructed has been validated. Have you performed some cross-validation of the model? and/or have you compared the results obtained from the simplified LCA with results from the original studies when this could be done (e.g. for climate change impacts when the study had similar methodological choices)? This is an important point to document in both Methods and SI to justify the validity of the results you display.

Response: We developed the simplified models around the inclusion of processes that have been identified as most important in other life cycle assessments and reviews of life cycle assessments. Because our estimates use published inventory data to estimate the parameters, differences between our estimates and published values are due to differences in the models themselves. Consequently, we feel such comparisons would be inappropriate in the text as they would require extensive explanation of the differences in the models themselves. Nevertheless, we can provide some ballpark comparisons. Our median estimate of the mean for GHG (live weight equivalent) ranges from about 0.6 tonnes CO2-eq for bivalves to 7 tonnes CO2-eq for miscellaneous diadromous fishes and the GHG emissions from Bohnes et al. (2019) presents a median around 2.7 tonnes CO2-eq (live weight equivalent) across all studies. The other metrics in Bohnes et al. (2019) are not in units that can be directly compared to our units. Poore and Nemecek (2018) reports values for farmed crustaceans and farmed fish, but they are presented in terms of 100 g protein and apply economic allocation. That paper uses 230 g protein per kg of fish (edible portion) and 150 g protein per kg of head free shrimp (for crustaceans) and it appears that their estimates are then roughly 23000 kg CO2 per tonne edible for crustaceans and 7500 kg CO2 per tonne edible for fish. Our shrimp values for economic allocation and edible weight are a little less than 10000 kg

CO2, with the difference likely being related to pond emissions and possibly including things other than shrimp in crustaceans, as our average value for fish is very close to 7500 kg CO2, depending on how the different fish taxa are weighted (most medians are between 5000 and 1000 kg CO2). Our GHG values also compare reasonably well to MacLeod et al. (2020), which presents the results as kg CO2-eq/kg carcass weight (note that there are also some differences in system boundary and other methodological choices). They estimate bivalve emissions of 3-4 compared to our 1.4 [0-4 95% credible interval], catfish as 3-6 compared to our 7.2 [6.3-8.2 95% credible interval], cyprinids as 2-5 compared to our 3.6-7 [2-5.5 for silver/bighead and 5.9-8.2 for misc carps 95% credible intervals], misc marine fish as 6-18 compared to our 11.5 [9.2-13.9 95% credible interval], salmonids as 3-6 compared to our ~4 [4.2-6.3 for salmon and 4.1-6.7 for trout 95% credible intervals], 6-14 for shrimps and prawns compared to our 9.1 [8.1-10.0 95% credible interval], and 3-6 for tilapia compared to our 10.7 [9.7-11.8 95% credible interval].

* In relation to the scenarios performed (Figure 4b, Table S8), the authors should indicate in the manuscript (along with Figure 4b analysis), methods and SI the limitations of the explorative study on decreasing environmental impacts of blue foods. Doing such assessments, which are in fact consequential in nature (as clear when reading Table S8), with an attributional/allocation-based model has little meaning, in my opinion. As a bare minimum, the authors should be open and discuss/comment in the text about the inconsistency of this approach and the implied uncertainties/limitations.

Response: We disagree that the scenario analysis is inconsistent with our attributional approach. Sensitivity analyses and scenario analyses aimed at identifying meaningful opportunities to target changes for improvement are common in LCAs and are in fact recommended components of the interpretation stage of LCA according to ISO 14040/44 regardless of attributional/consequential approach. The scenarios we outline are largely an extension of sensitivity analysis by considering what realistic changes to systems may be; to extend these scenarios into a consequential approach would require an economic analysis that is not an objective of this paper.

* l. 49- 103: For a given taxa group, the variability in Figure 1 reflects the variability of input data (if I understood correctly). Therefore, it would be nice for the reader to add 1-2 sentences, e.g. in l.92-101, commenting on how representative of the different locations and technologies for each species/taxa the data are, so the reader has an impression of how comprehensive/reliable each “box plot” in the different Fig. 1.x is. I think these are relevant information because most of the text before l.92 in this part is focused on comparing across species groups for each impact category.

Response: We modified this text to read “Our results represent the most comprehensive and standardized blue food stressor estimates to date. Overall, data availability is correlated with global aquaculture production across these taxa groups, but there are still notable taxonomic and geographic gaps (Fig S3, S4).” We then point to these gaps later in the paragraph.

* l. 219: I recommend nuancing the statement, as it could be misread/misinterpreted by the readers. It may create co-benefits but may also create new impacts as well, and only a full coverage of environmental problems can reveal such trade-offs. That is a general point – and a very important one – to make in this paragraph that any decision to change current production modes should be accompanied by a holistic environmental sustainability and risk assessment to avoid burden-shifting (across impacts/risks). This message is only touched upon in Box 2, although with another purpose/context. It should be reinforced in the text, especially since the title of the paper reads "environmental performance" (which signals a broad scoping).

Response: This sentence is already doubly caveated, reading "In some cases, this could create co-benefits for biodiversity impacts (Box 2)" with the Box pointing to both a co-benefit and a trade-off. We only had space to further detail this point in the box text, where it was intended to make this point (which is why we reference it in this sentence).

* Data representativeness: the added text in the Methods responds to my earlier comment, and it is now clear what the authors did, and I understand the data limitation preventing to quantify the actual representativeness. That said, I recommend 2 additional adjustments to be open about this: (i) In l. 43, I would still suggest to alleviate any ambiguity as to what the 70% coverage relate to: I suggest to split the sentence to make clear that you refer to the taxa coverage, and not the actual representativeness of the production processes (which are the drivers of environmental impacts). (ii) in Figure S3 and S4, I recommend complementing the captions to indicate what is meant by "representativeness" in those graphs. As I mentioned previously, the study may cover 99 farms that capture 1% of the actual production volume of a given species in the country while the 100th farm producing 99% of the species in that country – and which would not be covered – may drive all environmental impacts of the production of that species (with a different intensity than the mean across the 99 farms). Although the study is certainly the most comprehensive to date and no data exist to quantify the actual representativeness, I think it is important that the authors are open about this potential bias.

Response: We split the sentence into two sentences, which now read: "We draw on studies that collectively report data from over 1,690 farms and 1,000 unique fishery records around the world. The 23 species groups represented in our results cover over 70% of global blue food production."

We also added the following sentence to the Figure caption of Fig S3 and S4: "Farm-level production volumes varied substantially but were not consistently available across source studies."

We also note that an important consideration for data inclusion was that the studies were broadly commercially representative (excluding all hypothetical and experimental systems).

* Units for most Fig. Sx (e.g. S6, S7, etc.) and figs in the manuscript, incl. Fig. 1, etc. (+ Table 1): unlike what has been replied to my earlier comment, units are still wrong on several/most graphs, incl. missing "eq" for climate change (should be "kg-CO2eq"), incomplete units for land use (should be "m2a"), etc. The unit for land use is also wrong as it should be "m2a-crop-eq", which is the actual unit if you applied ReCiPe2016 for land use impact assessment (right now the "crop-eq" is systematically missing). I therefore do repeat my comment from the previous round of reviews: please do a *thorough* check of all units for all figures and tables; this was obviously not done last time.

Response: We added these units to all main text and SI figures.

Other comments:

- I am now very glad to see a full transparency on the study, incl. background data and codes. It is a non-negligible added value to the work, in my opinion. I would suggest however to combine the information of CSV file "LCI_compiled_for_SI" (from Github repository) with the information from the file "384293_2_extended_data_3667030_qtgbcb" (provided as SI) and to have this into a single Excel file as SI to the paper. I think it is an important file to accompany the paper (instead of going through the Github repository to find this) and the information in both are very relevant (raw LCI data in the former and full referencing with justifications for inclusion/exclusion in the latter).

Response: We now include the file "LCI_compiled_for_SI.csv" as part of the SI.

- To document your modelling approach, you may consider including some of the information of the workflow figure from Github in the SI as well (simplifying the figure to drop the orange boxes that are not necessary for understanding). I just found the blue boxes quite useful to understand whats was done, i.e. more explicit/clearer than the methods flow chart in Fig. S1. The workflow also seems to cover the entire study scope, while Fig. S1 is limited to aquaculture (although not explicitly written in the caption).

Response: We believe the workflow described in the GitHub vs the SI are appropriately separated. As a repository for programming code, the workflow in GitHub is meant to describe how individual chunks of R scripts are executed in order to reproduce our analyses and figures. The workflow as described in the Methods overview will be more useful to the average reader interested in the

relationship between data, model, and variable outputs. This is a matter of opinion and different readers will require different levels of detail, but we believe we have correctly separated these details into data science and computer programming (i.e., GitHub) vs data sources and methodology (i.e., the manuscript SI).

- l.25: please add "for example" before referring to the specific 3 goals you list. Others could also be listed with direct link to aquaculture/seafood (nr. 3, 6, 13, 15, etc.).

Response: While true, the sentence is correct as it is currently written, and we are trying to avoid additional words where possible.

- l.93-96: how important are the non-commercial fisheries (subsistence) relative to commercial fisheries? I have not checked and I do not know if it is possible, but my initial thought would be that they are negligible. Therefore, if so, the data gaps would not be a major concern here.

Response: Since there is limited data available, this cannot be known. Nevertheless, it is relevant if it is a low greenhouse gas food production system that is undervalued as such. This point was also raised by other reviewers, so we will retain the text as is.

- l. 97: grammar to revisit

Response: We did not find a grammatical error here.

- l.99: "snapshot of the knowledge of current production" (or alike) instead of "snapshot of current production", which, as correctly stated in the previous sentences, does contain gaps (this still relates to my previous comment on data representativeness).

Response: We added this suggestion to the indicated sentence.

- Fig. 1e: remove "farmed" (or add that information to all graphs systematically)

Response: We disagree with this aesthetic suggestion, as best practices for data visualization suggest removing repeated text/information wherever possible.

- l.115: which processes (give 1-2 examples very briefly)

Response: The next sentence provides an example of this already.

- l.119: “from” instead of “form”

Response: Thank you, we corrected this.

- l.122: please correct to “tropospheric” and not “stratospheric” for ozone. These are two very different types of impacts.

Response: This was a suggestion from a reviewer and based on papers suggested by that reviewer. However, we have removed stratospheric from the sentence.

- Figure 2: I like the figure, which reads better now with the text changes. However, it mixes different concepts in it and I recommend to “clean” it to only retain actual environmental stressors as understood in the study. For example, energy consumption is not a stressor, but emissions and resource use deriving from it are (and those are already included). Another example may be fish-derived resources: not clear to me how this is an environmental stressor. Other symbols are also redundant, e.g. by-catch (twice represented).

Response: The two by-catch icons are included since they represent different types of by-catch and the energy consumption is a resource input that can be separate from emissions, as a system could consume energy from solar or wind, where the emissions would not be the primary concern (but the energy as a resource could still be limited and could come with other types of environmental impacts). More generally, this figure is conceptual to illustrate the broader range of considerations to highlight what is quantified within our study and what falls outside the bounds of our study.

- l. 165: “sustainability messaging” into “sustainability communication” or alike (?), if that is what you meant. Please clarify the wording.

Response: While the suggestion is synonymous, we believe the current phrasing is clearer.

- l.169: Please add “As an example, integrating…”. You only give an example here.

Response: The sentence is correct as currently written.

- The difference between the 2 headings “Reducing blue food environmental impacts” (l.182) and “moving toward lower-impact blue foods” (l.229) is too subtle in my opinion (the different focus of each underlying part is clear though). I recommend you better distinguish the 2 headings (maybe reformulate the latter).

Response: We revised the subsections to be:

Production levers for reducing blue food environmental impacts

And

Actions for improving the environmental performance of blue foods

- l.262: suggest another word to “charismatic” (mainly used for persons).

Response: Charismatic is very commonly used for non-human subjects (e.g., charismatic megafauna). We have therefore retained this phrasing as it is more compact than other alternatives we considered.

- l. 506: as already asked previously, please indicate the version of ReCiPe2016 used (“Hierarchist” set of characterization factors is relevant to mention, but it says nothing of the versioning).

Response: The sentence now reads: “Life cycle emission factors and resource use rates were calculated for all included feed inputs using characterization factors from ReCiPe 2016 Hierarchist method as implemented in the OpenLCA LCIA methods v2.0.5 by GreenDelta."

SI file:

- Indicate which ecoinvent db you used (there are three).

Response: The sentence now reads: "Proxies for all emissions related to electricity generation were derived from the ecoinvent 3.6 APOS database."

- l. 88-89: I would like that you explicitly refer to ISO14044:2006 in your argumentation about systematically using allocation in your model (which is a last resort in ISO hierarchy). This is important, e.g. for the by-product handling when converting from live weight to edible weight.

Response: ISO 14044 recommends avoiding allocation if possible, by either subdividing the system or expanding it to include the function of all outputs. Production of fish biomass cannot be subdivided by tissue and an expansion to include the many and varied functions of fish co-products would be impractical at such a global scale. ISO 14044 subsequently recommends allocation based on biophysical relationships between inputs and outputs, here determined to be the output of fish biomass from the input of appropriated productivity via ecosystems or feeds and reflected in the form of dry matter mass. This allocation method has been implemented in numerous seafood LCAs, defended and/or critiqued in multiple reviews, and is the recommended method of allocation in PAS 2050-2, ISO 22948, and NS 9418.

- Fig. S4: legend is missing some items (small dots representing 1 study, I guess?).

Response: The reviewer's suggestion would require a legend that depicts "1, 2, 4, 6, 8", or "1, 3, 5, 7, 8" or "1, 2, 3, 4, 5, 6, 7, 8". As the reviewer was able to correctly identify the meaning of the small dots given the provided legend, we don't think it is necessary to revise the legend.

- Fig S5: should there be some numerical values on the axes? Or else indicate which species is taken as a baseline (does not seem any is though..).

Response: The spider diagrams represent normalized stressors from 0 to 1 calculated as stressor / max(stressor) or in the case of N and P which are negative, we do a min/max normalization - i.e., (nitrogen - min(nitrogen)) / (max(nitrogen) - min(nitrogen))). We added this information to the figure caption.

- Table S4: please add “Pearson correlation coefficient” (coefficient missing)

Response: We added “coefficient” to the table caption.

Referee #5 (Remarks to the Author):

The manuscript is an important contribution to the literature and a definitive study of the environmental performance of aquaculture. The authors have addressed all of my comments. I have no further comments.

Response: We thank the reviewer for their time reviewing our manuscript and for their support in its publication.