

Reviewer 1:

- 1) **Reviewer Comment (RC):** The paper needs an explanation of the megafauna of nodule locations. This is critical because each survey could return very different results due to the sparse nature of the animals. The individual surveys are located in the same area but are not necessarily overlapping so could be expected to show different results.

Author Response (AR): Please see section 1.1 “Polymetallic nodules and associated fauna”.

Modified Text (MT): No changes.

- 2) **RC:** I do not understand why there are conflicting statements about how well the observers did at recognizing categories of animals – there are conflicting statements about them doing well and poorly. The gold standard process, that is not described, seems to improve their performance.

AR: Please see lines 294 to 303 on how the gold standard annotations were created and section 3.2 on the annotator performance. While the annotators themselves showed poor observer agreement, grouping the annotations to gold standard annotations created robust data.

MT: No changes.

- 3) **RC:** The conclusions are that the method is highly dependent on equipment, methodology and observer, which may not be very helpful for the industry but indicates caution in adopting new monitoring practices.

AR: Correct. This outcome is expressed in the five guidelines mentioned in lines 403 – 409.

MT: No changes.

- 4) **RC:** The English of the manuscript is not always very good even though there are English speaking authors. I have picked out some instances but there are many more.

AR: Thank you for pointing this out. We have corrected your specific comments (see below) and have subjected the document to another proof-reading by the native speaking co-authors.

MT: Many, see detailed points below.

- 5) **RC:** Specific comments Line 1 It is not “again” as they have never been commercially viable.

- 6) **AR:** We modified the sentence slightly to present our intended message.

MT: With the mining of polymetallic nodules from the deep sea seafloor again ~~approaching~~ producing commercial interest ~~viability~~, ...

- 7) **RC:** Line 13 Mention minimum standards here e.g. When and if commercial exploitation of these marine resources commences, minimum standards should be set in the regulations on how camera-based monitoring surveys should be conducted, and with what equipment so as to ensure a high standard and unambiguous results.

AR: We would like to go further and recommend not just “minimum standards” but “robust and testable standards that foresee future technological advances”.

MT: ~~When and if commercial exploitation of these marine resources commences, to ensure best monitoring practice, unambiguous rules on how camera-based~~

~~monitoring surveys should be conducted, and with what equipment, must be put in place.~~ When and if commercial exploitation of these marine resources commences, robust and testable standards that foresee future technological advances should be set in the regulations on how camera-based monitoring surveys should be conducted, and with what equipment so as to ensure a high standard and unambiguous results.

- 8) **RC:** Line 19 The increasing interest has driven the technological development
AR: Correct, thank you. The manuscript was modified accordingly.
MT: This increasing interest, ~~simultaneously driving~~ ~~coupled with~~ the technological ~~improvement~~ development ...
- 9) **RC:** Line 20 Contracts are awarded by the ISA – not licences.
AR: Correct, thank you. The manuscript was modified accordingly.
MT: ... granting of exploration ~~contracts~~ ~~licenses~~ ...
- 10) **RC:** Lines 19-25 This sentence is too long – split after MIDAS
AR: Done.
MT: ... has resulted in several recent European research projects (...). ~~Those were being~~ funded to focus ...
- 11) **RC:** Line 27 It is density of nodules and their mineral grade. No contracts have been awarded in the Peru Basin.
AR: The manuscript does not state that contracts were awarded in the Peru basin. We modified the manuscript anyhow to clarify.
MT: ... both of which have been considered ~~likely initial~~ as regions for ~~early~~ commercial exploitation ~~at some point in history~~.
- 12) **RC:** Line 39 The fauna on the nodules was destroyed as well
AR: Correct, thank you. The manuscript was modified accordingly.
MT: As a result, fauna that lived attached to the nodules were removed ~~and thus destroyed as well~~.
- 13) **RC:** Line 39 The soft-bottom community . . . Start a new paragraph here and explain that the site has been monitored subsequently.
AR: Done.
MT: ~~Several monitoring cruises of the impacted areas commenced in the following years and decades.~~ The repopulation ...
- 14) **RC:** Line 43 The ploughing activities also . . .
AR: We added “also” to the sentence.
MT: The ploughing activities ~~also~~ created a sediment plume ...
- 15) **RC:** Line 44 Change not directly to indirectly
AR: Done.
MT: In these ~~not directly~~ ~~indirectly~~ impacted areas, ...
- 16) **RC:** Line 46 Change is still to date . . . to remains . . .
AR: Done.

MT: ... megafaunal community composition in these areas ~~is still to date~~ remains significantly different ...

17) **RC:** Line 49 I suggest remove between regions sampled differently . . .

AR: The point of the different sampling is important here. We split and altered the sentence to accommodate the message.

MT: ... can greatly influence assessments. ~~This challenges the rendering~~ direct comparison ~~of between~~ regions sampled differently ~~problematic~~ ...

18) **RC:** Line 87 In this paragraph you explain the results of the previous disturbances but you are aiming the document at helping monitoring post mining. It would therefore be better to describe the post mining surface that is likely to be planed flat with track marks and then overlain by plume de-deposits – described in Weaver and Billett 2019 in: Deep-sea mining and Environment–issues, consequences and management Edited by Rahul Sharma. Springer International

AR: We have added this perspective and additionally cite “Sharma 2019 Environmental Issues of Deep-Sea Mining: Impacts, Consequences and Policy Perspectives”.

MT: Further, depending on the removal technique, the seafloor will likely be ~~highly~~ perturbed, with ~~a range of depressions, ploughs or~~ compaction tracks potentially formed, ~~and all overlaid by plume deposits~~.

19) **RC:** Line 98 As we approach This is an odd statement when you are discussing monitoring nodule mining that will impact 100’s of km² per year per mining activity – to be concerned about a few box cores or trawls, both of which scientists do all the time. I suggest removing these two sentences. The importance of the technique is that it is inexpensive and can cover much larger areas than seabed sampling. The paper will have more impact if it follows this line.

AR: We removed the first sentence but keep the second as it is a general statement on monitoring impacts that should influence future mining regulations.

MT: ~~As we approach the third decade of the 21st century, and experience a continuous increase in public interest in maintaining as near to intact as possible even remote and inaccessible ecosystems, non-invasive monitoring is required (-).~~

20) **RC:** Line 104 This is a better argument about data availability. If used commercially could you verify the image data is genuinely from a given location at a given time?

AR: Good point. We added a sentence on data protection.

MT: ~~To assure reliable monitoring, contractors need to publish data with location and timing metadata uncorrupted and the acquisition technology of that metadata needs to be fraud-proof (e.g. by incorporating navigation data right the imagery).~~

21) **RC:** Line 129 loss of information at the image extremities Presumably the best option is only to look at the central portion of the image, ignoring the periphery?

AR: Yes, this is an option – depending on the severity of the information loss. We added a sentence with that regard.

MT: ... lens correction. ~~Image analysis could therefore focus on central parts of the image and the boundary area of images be used to display e.g. navigation metadata.~~

Lenses of a more ...

22) **RC:** Lines 142-145 The English needs attention here

AR: We modified the sentence.

MT: ~~A vignetting of illumination in an image can be partially addressed prior to analysis by excluding the edges of collected images from analysis.~~ An illumination vignetting can be addressed partially prior to analysis by excluding the image edges from analysis.

23) **RC:** Line 150 The illumination system is though in direct relation with the target altitude of survey above the seabed. Not sure what this means?

AR: Thank you for pointing out this incomplete sentence.

MT: The illumination system ~~needs to be setup to accommodate the target altitude of the camera platform above the seafloor as well as the expected altitude variation is though in direct relation with the target altitude of survey above the seabed.~~

24) **RC:** Line 182 The suggestions should be targeted at the CCZ or Peru Basin which have unique and peculiar faunas – low numbers of individuals and very high species diversity with many unknown species. A brief review would be useful – there are a growing number of papers including some referenced here - Simon-Lledo, A et al 2019, Vanreusel et al, 2016 etc. This review would then point to the need for extensive baseline surveys including repeat surveys to capture the spatial and temporal variation. I'm not sure of the value for surveys during the mining activity when the area would be covered by plumes that would hide the seabed. Maybe these could be targeted to find the edge of the plume or to examine effects in the far field when the plume has settled.

AR: This manuscript does not aim to provide information on how much area to survey depending on the equipment at hand and the expected fauna composition. Also, we do not have the data to do so and we feel it would have fallen out of the scope of this manuscript. We modified the section to stress that a particular area has to be monitored several times in a consistent manner.

MT: To determine the level of impact an event has had on a **specific** region of seafloor, repeat visits to ~~a~~ **this** locale are required. **It is important to conduct baseline and impact monitoring surveys in a region-specific manner to accommodate differences in faunal composition. Baseline information acquired in one nodule area (e.g. the CCZ) cannot directly be transferred to another (e.g. the Peru basin).**

25) **RC:** Line 196 (ca. 600 x 150 m²) Is this per surveyor total?

AR: This is the area of the bounding box around the individual deployments. We modified the document to make this more clear.

MT: ... (see table x) **were** collected. **All data sets were acquired in ~~from~~ a discrete area of seafloor ~~were compared of~~ (ca. 600 x 150 m²) size.**

26) **RC:** Line 198 SO106 and SO242/1 Is this a cruise? If so what ship and when?

AR: Yes, we added this information to the manuscript.

MT: One dataset (DS_c) was acquired during **RV Sonne cruise SO106 in 1996**, the other seven during **RV Sonne cruises SO242/1 ...in 2015.**

27) **RC:** Line200 Describe the faunal catalogue – what is a category (family level?). Why 20? Do these cover all the megafaunal groups in the Peru Basin. Are some categories more common than others? Does each cover many species and are these similar to each other? How difficult is it to recognise each category? Is there a size cut-off? Did you test to get the best categories? Etc

AR: We added the following sentences to incorporate the missing information.

MT: The term category refers to an arbitrary object type, not necessarily related to taxonomic levels and also includes the category “litter”. The group of annotators selected the 20 categories by including fauna that is frequent enough for statistical interpretation. The 20 categories neither cover all object visible in the images nor represent all the fauna known to occur in the area. Categories mostly represent morphotypes and could thus potentially include different species of indiscernible visual appearance. Numbers of annotations per category and per data set vary. No size cut-off was defined for annotation, rather the image resolution determines which size of objects are still discernible.

28) **RC:** Line 212 coupled with their moderate personnel requirements . . . You have forgotten the ship crew! The difference in cost will be small.

AR: The ship’s crew is always needed, independent of the deployed gear. For an OFOS one person is needed to operate the winch (could be a crew member), everything else is done by the data interpreters. For a deep sea AUV or ROV operations, 4+ technicians are needed, next to the data interpreters. While AUVs can save ship time during deployments, hardware costs are at least one order of magnitude higher than for OFOS.

MT: None.

29) **RC:** Line 276 Figure 4(d) is extremely blurred. I assume you have selected one of the best images – so is this process pushing technology too far?

AR: This blurring is an artefact of the mosaicking procedure.

MT: None

30) **RC:** Line 281 fauna individual . . . Animal

AR: Sentence modified as below.

MT: ... annotated **fauna objects** by placing a circle around ~~each fauna individual instance~~ using the BIIGLE interface ...

31) **RC:** Line 287 gold standard annotations . . . What is this?

AR: A gold standard is used when no ground truth exists. It is the best-possible truth that can be achieved. It is frequently used in image interpretation.

MT: None.

32) **RC:** Line 288 Grouping was conducted by fusing overlapping annotations of similar size to one grouped annotation. The location and size of this grouped annotation was computed as the average of the annotation position and radius of the single expert annotations. I do not understand these two sentences

AR: We streamlined these two sentences.

MT: Grouping of **annotations** was conducted by fusing **overlapping** annotations **which overlap within one image and are** of similar size to one grouped annotation.

The ~~position location~~ and ~~radius size~~ of a ~~this~~ grouped annotation ~~was represent~~ ~~computed as~~ the ~~mean average~~ of the ~~annotation~~ positions and radii~~us~~ of the single, ~~overlapping expert~~ annotations.

33) **RC:** Line 293 Cohen's kappa was computed . . . What is this and what was the result of the computation?

AR: Cohen's kappa is a measure to assess observer agreement. Results are given in table 2. We added a reference to the table in section 3.2 (see next point).

MT: None

34) **RC:** Line 314 (see also 2). What does this refer to?

AR: The word "Table" is missing here. Thanks for pointing this out.

MT: ... (see also [Table 2](#)).

35) **RC:** Line 314 It further highlights that the observer agreement drops with increasing image resolution reflecting the results in (c). This is a curious fact – the observers were worse at categorising better images! This seems to undermine the whole process. The rest of the paragraph seems to say the opposite. I am confused, and clearly more explanation is needed here.

AR: In fact, the group of observers was worse at categorizing better images. In essence, their disagreement grew with rising resolution. Hence the comment towards the necessity of grouping annotations in the next sentence and also in the discussion (section 4.3).

MT: None.

36) **RC:** Line 315 (c). Does this mean Figure 7c?

AR: Yes.

MT: ... reflecting the results in [Figure 7](#) (c).

37) **RC:** Line 335 the same area of seafloor According to Figure 1 this was not exactly the same area of seafloor, but areas of seafloor near to each other. Given the huge variability of faunal distributions in the Peru Basin would you expect the results to be similar? I am not sure I agree with your conclusion in line 344-345.

AR: The area of image acquisition is almost exactly the same for all data sets. 150x600 meters is less than 1 percent of the Discol Experimental Area for which previous studies have reported monitoring results as a whole. Looking at the photomosaic of the area (background in Figure 1) shows no obvious heterogeneity of the habitat despite the plough marks that occur throughout the area.

MT: ... imaging [almost exactly](#) the same area of seafloor ...

38) **RC:** Line 381 annotators with little or no experience can identify fauna within an image set with a degree of confidence This is in contradiction to lines 313-315. I am confused – there seems to be a step related to gold standard, that is not explained, but takes poor observer agreement to good observer agreement.

AR: We added the word "ample" and another sentence referring to the grouping of annotations to a gold standard (as described in section 2.3).

MT: ... can identify fauna within an image set with an [ample](#) degree of confidence. For complex studies of detailed community change trained scientific personnel would be required [that can create more trustworthy annotations. In either case, manual](#)

annotations need to be quality controlled, e.g. by creating a gold standard, to produce more reliable data. Employing several experts for the annotation task, which again would add a considerable financial cost to the monitoring program.

39) **RC:** Line 387 more traditional monitoring approaches – such as? Do you mean human recognition of photos/videos or box coring and other sampling?

AR: Both.

MT: ... approaches viability, more traditional monitoring approaches like manual image annotation or physical sampling are the only ones currently available for integration into regulatory frameworks and work plans. Anyhow, expected technological advances should be incorporated into the regulations.

40) **RC:** Figure 3 Better caption needed. Add indications of scale and other parameters.

AR: We added more information in the subcaptions.

MT: See figure.

41) **RC:** Table3 what do the column heads H, N, df and p mean. Improve the caption

AR: The column heads mean: H=Shannon-wiener diversity, N=total number of specimens, df=degrees of freedom, p= probability. This information was added to the table caption.

MT: See table caption.

42) **RC:** Figure 5 Better caption needed. What do multiple circles mean – multiple organisms or multiple operators?

AR: Multiple annotations. We added that information.

MT: Circular fauna identifications made by several operators using the BIIGLE software application. Each circle corresponds to one annotation by one annotator. Colors of circles correspond to categories.

43) **RC:** Figure 6 Caption needs more explanation What do the colour circles mean? List what each figure represents. Does this tie to table 3? L and m look to be very poor images – explain

AR: We extended the caption and also modified table 3.

MT: Fig 6 to fill in Table 3: a-b= Anemones (Actiniaria). C1-c2= corals, C3 looks like a bryozoan to me, d=Crustacea/Decapoda/Parapaguridae, e=epifauna, f=fish/Ipnops, g=Jellyfish, h=litter, i=Ophiuroidea or brittle stars, j=Cladorhizidae Porifera, l=Enteropneusta (different morphospecies than p!), k=fish, l=Polychaeta worms, m=tubeworms Polychaeta, n=Holothuroidea – Sea cucumbers, o=small encrusting, p=Enteropneusta, q= Porifera, r= stalked crinoid, s=stalked Porifera, t= Asteroidea/sea star

Based on this and to avoid confusion with biologists I'd leave J=Enteropneusta out of it as well as J=Cladorhizidae because they do fit in other categories (and are quite easily recognisable)

44)

Reviewer 2:

- 1) **RC:** Language is at times murky, specific suggestions included in comments (other examples not included in referee comments exist).
AR: We have had the document cross-checked again by our native-speaking colleagues.
MT: Several changes throughout the manuscript. Specific changes as proposed by the reviewer are outlined below.
- 2) **RC:** Sustainable and environmentally thoughtful method. Field methods clearly described. Statistical analysis description could be expanded, albeit references are included. **Gold standard process could be explained more obviously.**
AR: See Review 1, Comment 32.
- 3) **RC:** At times the writing becomes a bit too qualitative when describing the agreement of the annotators. The figures in Figure 6 support the authors but also include outliers that probably drive the regression, making the link to the conclusions weak.
AR: For future improvements of this first gear comparison study, perfect underwater navigation, repeated annotation category workshops and intra-observer agreement assessment should be conducted. In that case, some of the result figures might show less spread. Unfortunately, this goes beyond this study and what is currently available in terms of deep sea data.
MT: None.
- 4) **RC:** Second 1.4- to 1.4.1 Transition: This is a little Murky. Consider removing the bit about the limitations of box cores and just focusing on the considerations for photo data sets. Authors could move the limitations of box cores/trawls to Section 1.3 (Methodologies for fauna abundance assessment) if they wanted to keep it in the text.
AR: Not sure, to which line you refer. In case of section 1.4 before 1.4.1 we stress here, that other sampling gear has less factors influencing the acquisition of robust data compared to images.
MT: None.
- 5) **RC:** Section 1.4.5 is very Method Like. Consider removing first sentence to Section 2 at the end of second sentence (These 8 datasets were collected by three difference towed camera platforms and an AUV during three research cruises and were compared based on their resolution . . .). Start section 1.4.5 with 'Image resolution is controlled by the camera optics and the deployment altitude and allows to data sets to be compared numerically. The camera . . .'
AR: We modified section 1.4.5 to state general aspects rather than addressing this study.
MT: ~~Within this study, image datasets are compared with regard to their resolution. This factor~~ Image resolution is a combination ...
- 6) **RC:** Section 3.1: In Figure 7b, d, h, l, and l, there is a single point that appears to drive the regression. If this point is removed, particularly in figure 7d, it may not be a strong an argument that resolution and acquisition efficiency of an inverse relationship (Line308).
AR: Yes, anyhow, none of the data sets can be said to be an outlier in all of the plots,

hence we chose not to exclude any of them separately for the different correlation plots. It appears, that being an outlier in one data dimension is more often associated with data sets of low image resolution. In the case of (d), the correlation would still be negative when removing data point G. This is reasonable given the general tradeoff in surveying between resolution and coverage.

MT: In (d) it can be seen ~~is obvious~~ that the increase in resolution comes with a decrease in acquisition efficiency in terms of the area per hour (m^2 / h) that can be imaged. This negative correlation exists also when removing data point G.

- 7) **RC:** Consider including a greater discussion into how the inclusion and removal of outliers in figure 6 influence the regressions.

AR: We added a sentence to the results section.

MT: Removing single points that appear as outliers in the different data dimensions (Figure 7 (a) - (l)) does not change the general trends of the correlation lines.

- 8) **RC:** Section 4.1: Considering supporting this assumption by including information or references to previous studies on the spatial and temporal variability in the Peru Basin.

AR: We added references to manuscripts describing the spatial and temporal variability.

MT: (Bluhm 2001, Borowski 2001)

- 9) **RC:** throughout listing of citations is variable, sometimes in alphabetical order, sometimes in order of publication year.

AR: Thanks for pointing this out. Has been fixed.

MT: Several.

- 10) **RC:** Throughout the paper authors use numerical symbols i.e. Line 196: (8 as oppose to eight), except in this line the spell out three (instead of use 3). Suggestion to be consistent. Also occurs in lines 296/297 with five and 8.

AR: Thanks for pointing this out. Has been fixed.

MT: Several.

- 11) **RC:** Line 5 and Line 135: Consider using Autonomous in place of Automated, and in Line135 just use AUV since you've already defined the acronym

AR: Thanks for pointing this out. Has been fixed.

MT: Several

- 12) **RC:** Line 76: The etymological plural of octopus is octopuses or octopodes because the root is Greek. Note, the word octopi does not make the sentence any less clear.

AR: Has been changed to octopodes.

- 13) **RC:** Line 149: consider word order change to: (Finally, the colour spectrum of light also needs to be considered because the returned yellow . . .)

AR: Has been modified.

- 14) **RC:** Line 150: Confusing, I think the authors are saying that the set up of the lighting system is based on the altitude of the vehicle? Consider rewording for clarity.

AR: Yes, this was fixed.

MT: Please see reviewer 1 comment 23.

15) **RC:** Line190: Insert word 'they' after shelf, (Norwegian continental shelf, they all used . . .) OR remove 'these' from line 189: 'Despite 5 survey cruises . . .)

AR: We added "they".

16) **RC:** Line 232: Consider changing to: The 1989 data set was digitized by the JPIO . . .

AR: Done.

MT: ~~The photographs taken back then were recently digitised by the~~ The 1989 data set was recently digitized by the JPIO Mining Impact project and made available for this study.

17) **RC:** Line 316: Comma missing after any how

AR: Fixed.

18) **RC:** Line 336 and Line338: Two sentences in a row start with 'Given the', considering re-wording.

AR: Fixed.

MT: ~~Given~~ But due to the reasonably homogeneous ...

19) **RC:** Table 1: For Date Column, consider adding (dd/mm/yyyy) but not necessary as all days at greater than 12.

AR: Done.

MT: See table.

20) **RC:** Figure 6: In figure caption, note the reason for the circles on each sub-image

AR: Done.

MT: Circles correspond to annotations in BIIGLE. Colours of annotations visualise the category type.

21) **RC:** Figures 8 and 9: in X-axis labeling consider adding to the method description the dataset name as authors refer to them in the text (DSA etc.)

AR: We modified the figures accordingly.

Short comment 1:

- 1) **Short Comment:** Almost all of the fauna being observed in the images are either undescribed or lack local taxonomic records linked to vouchered specimens. This means it is not possible to identify the animals observed to species level. For this reason, image survey can-not replace actual sampling until such time that we have a reasonably good checklist of known species (verified from samples) from the region where the image survey is taking place. Once that is achieved, image survey will be a valuable tool for conducting biodiversity assessment. At the moment, the use is limited to the abundance of higher taxa. This is not the species-level assessment that the regulator, stakeholders and society should demand. The authors refer to this in this sentence only: "To label fauna to species level from imagery requires a certain amount of skill, and an awareness of fauna likely to occur in a particular survey region." I believe they could easily add to this section the obvious need to undertake

taxonomic descriptions to enable identification, and also make the caveat that species-level identification is not possible until this has been done.

AR: Thanks for the comment. You are absolutely right. Taxonomy and knowledge about what might be visible in the images is crucial for assessing megafauna communities. Our manuscript focusses on the technical aspects and comparison of gear which is actually rather independent of what we are looking at - in terms of the gear comparison only! Of course, when we consider the biological interpretation of what those gears see, then the taxonomy is important. We have nonchalantly left this out of the manuscript for now but will add your concerns to the next updated version.

MT: To label fauna to species level from imagery requires a certain amount of skill, and an awareness of fauna likely to occur in a particular survey region. In many cases, only self-chosen annotation categories will be annotated that can correspond to morphotypes. This is due to the fact, that most fauna in the areas are still unknown or impossible to discern from images alone. To properly assess fauna occurring in a habitat – especially when assessing human impacts – requires not only ecological expertise but also support from taxonomists. Even with knowledge on the fauna to expect in an image data set, such knowledge, inter-observer differences ...