Reviewers' Comments to Author:  
Reviewer: 1  
  
Comments to the Author  
Overall Assessment:  
  
This study could be significantly strengthened through the following key improvements:  
  
1. Clearly define the study’s rationale and implications: The manuscript lacks a strong justification for its importance. The authors should clearly explain why this work is needed and what practical or scientific implications the results will have, especially in the context of fishery management on the U.S. West Coast.

The authors agree with this statement. The authors clearly define the rationale and implications of this study, as well as it’s place in the current landscape of management of the WCG fishery, at the end of the introduction. Additional context on current multispecies management in the fishery is added into the introduction as further context to the implications of this analysis to this specific system.  
  
2. Improve organization and readability: The text would benefit from a structural reorganization to enhance clarity and logical flow. Reducing redundancy and improving transitions between sections would aid reader comprehension.  
The authors improved topic sentences and ensured delineations between topics are clear; this is something noted by multiple reviewers. Additional section headings were also added to improve clarity.

3. Substantially revise and expand the methodology: The methods need a more rigorous and transparent presentation. Specific details should be provided on:  
   • Model inputs and outputs  
   • Parameters choices (e.g., sampling fraction, iterations for CLARA)  
   • How potential issues like subsampling bias in cluster analyses were addressed  
   • Consistency in reported filtering criteria (e.g., 99% of ex-vessel revenue [abstract] vs. 99% of total catch [methods])  
The model input is clearly defined in the methodology ax ex-vessel revenue, with an accompanying formula for the calculation. Units have been added for consistency. The specific outputs from the CLARA run. Parameter selection was also explicitly stated in the updated methodology.

The model was rerun with 100 randomly-generated seeds to detect potential differences in clustering, but no change was found across all random seeds. This has been added to the methodology, and the specific random seeds were added to the GitHub page associated with the manuscript.

Care was taken to ensure filtering criteria was correct across the manuscript and the specific error mentioned by the reviewer was fixed, and as well as a handful of others in the results section and methods.

Additionally, the authors further expanded on the methodology in the supplemental material  
4. Streamline the results section: The results should focus on core findings, supported by concise text, summary tables, and high-quality visuals. Avoid excessive narrative detail.

The authors will limit the discussion of the 17 clusters down to short 3-4 sentences per cluster. Specific details about each cluster will be moved to the supplemental material.  
  
5. Improve figure quality and relevance: Figures should be clearer, professionally formatted, and include descriptive legends. Limit the number to those that directly support key points.  
The authors will adjust the figures to increase legibility and clarity. Another reviewer suggested removing Figure 1, as Figure 2 shows the same data in a different way. The authors will move Figure 1 to the supplemental and have updated the figure numbers accordingly.

Section-by-Section Feedback:  
  
Introduction:  
   • Reads more like a literature summary than a narrative justifying the study

The authors added specific justifications of the study and its applicability to the WCG fishery to the end of the introduction, after the rationale of the study is given.  
   • The rationale for identifying métiers is underdeveloped. Simply stating that no métiers have been identified for this region is not sufficient.

See above; the rationale is included and supported by specific ways this métier analysis can advance (multispecies) management in the fishery  
   • Claims about the importance of characterizing fleets and species interactions require references or clearer reasoning.

Paragraphs were added to the end of the introduction, one on the importance of characterizing fleets and another on species interactions with support from the published literature.

   • The stated objective—identifying métiers—is not well justified in terms of added value over existing management approaches.

The authors will reinforce the limitations of the current single-species approach to management and how métiers can mitigate these weaknesses  
   • The authors mention applying these métiers to reference point calculations “in future work,” leaving a gap in understanding the value added by this current study.

The authors will include text highlighting the importance of this current work to the field. This analysis is the first chapter of a thesis and the reference point calculation will be a subsequent chapter and publication.   
   • Consider restructuring this section to:  
      o Justify the need for métier identification in this region  
      o Highlight what is novel or improved over prior work  
      o Clearly state the research questions or hypotheses

The first two bullet points have been addressed previously in the responses to Reviewer 1; the authors revised the original hypothesis to be clearer and more explicit.  
Methods:  
   • Background on the FMP and regional fishery context would be more appropriately placed in the introduction.

The authors have moved this section to the introduction meet the requests of the reviewer and have also expanded the section to provide a more comprehensive overview of the fishery.  
   • Define "U.S. West Coast" for clarity to a broader audience.

The geographic boundaries of this region were defined at the first mention of “US West Coast’ in the main text.  
   • There is confusion around the aggregation level (“shots” vs. coarser scale). This should be clarified.

The authors will adjust the use of ‘shots’ in the manuscript, as the difference between ‘shots’ used in other studies (i.e. Pascoe et al., 2022) and the aggregated shots used here are not clearly delineated at many points in the manuscript.  
   • Streamline and clarify data filtering steps. If the analysis focuses on 2011–2023, omit discussion of earlier data.

The authors disagree with this assertion by the reviewer. The authors believe that it is important to highlight the original dataset (e.g. total number of species and gears) to understand the scale as to which the dataset was refined. The sentence (original submission lines 186-188, current submission lines XX-XX) will be kept.  
   • Address the inconsistency between filtering based on ex-vessel revenue (abstract) and total catch (methods).

The authors note this inconsistency and have taken care to adjust the   
   • The authors rely heavily on Briton (2019) for defining methodological decisions for their cluster analysis. However, this reference is not readily available and its contents cannot be verified. Either provide key information or use verifiable alternatives.

The authors agree that Briton (2019) is not easily accessible online and is not publicly accessible. Burch et al., (2021) summarizes the approach of Briton (2019) (Briton is a coauthor of Burch et al., 2021). A note about Briton (2019)’s lack of availability and a reference to Burch et al., (2021) was made in the main text.

Burch P, Sutton, C, Cannard, T, Briton, F and Sporcic, M (2021). An investigation of the bycatch of rebuilding species and other selected species in the Southern and Eastern Scalefish and Shark Fishery. December 2021, CSIRO, Australia.  
   • Figure A.1 content differs between the manuscript and the supplementary materials.

The authors note the error and ensured the figures are identical for the resubmission.

   • More detail is needed on:  
      o CLARA parameters (e.g., sampling fraction, number of iterations)

Specific parameters are now included in the methods of the text, including sampling fraction, number of iterations, and the number of samples taken per run.  
      o How silhouette scores were computed and used to select k. Were independent CLARA runs from k=2 to k=n used? Was this done multiple times to generate a distribution for each k or just once?

The authors note that the discussion on the calculation to compute k is given in lines 253-261 in the original text and XX-XX in the current manuscript. The authors added text that the silhouette analysis was iterated upon using 100 random seeds and did not see any change across seeds.  
      o Additional methods for assessing clustering robustness beyond silhouette scores

Reviewer 3 requested many additional sensitivities to clustering robustness that have been conducted and were included in both the response letter and the supplemental material.  
   • Selection of k based only on statistical indices can be ambiguous; Fig A.1 suggests a smaller k (e.g., 4) may be justified, aligning with the final métier count.

The authors believe that a higher k allows for the selection of final métiers to be done beyond statistical indices. Given the high dispersion between clusters at high k (see the dendrogram; Figure 5), halting the analysis at k = 4 would increase the reliance on statistical indices when converting to métiers and would limit the inclusion of fishery-specific inclusions (e.g., widow rockfish and Pacific whiting are midwater species often caught together in midwater trawls and this fishery, split between clusters, are likely to be grouped into the same final métier).

   • Consider referencing Gao et al. (2023) for a broader framework on cluster validation and the limitations of relying on a single metric.  
The authors have reviewed Gao et al., (2023) and agree that it provides a helpful framework for validation. A handful of the R packages (but none of the specific functions) cited in the paper were used in sensitivity runs, and those packages have been cited appropriately.  
Results:  
   • Cluster descriptions are overly detailed; focus on key trends and insights.

This has been addressed above; the authors cut down on the   
   • Use well-organized summary tables and visualizations instead of lengthy narrative.  
Mentioned above; the authors removed additional detail and ensured tables/visualizations can support the claims asserted.

Figures:  
   • Many figures are missing legends and are of poor quality.

The authors will add a legend to Figure A.2 and rearrange plots within figures to improve readability  
   • Eliminate redundant or low-value figures.

The authors will move Figure 1 to the supplemental materials and include new figures as suggested by Reviewer 2  
   • Improve formatting for readability and clarity.

The authors will adjust the figures to improve clarity and legibility, specifically with font sizes and graph layout  
  
References:  
   • Some links are broken, and citation formatting is inconsistent (e.g., "PMFC" vs. "PFMC"). Ensure that all in-text citations correspond correctly to full references.

The authors ensured all links work and formatting is consistent across citations. The sole mention of PMFC was fixed to PFMC.  
  
The results and discussion sections were not fully evaluated, as substantial revisions to the introduction and methods are necessary before the findings can be appropriately interpreted or assessed.  
  
Gao et al. (2023). An overview of clustering methods with guidelines for application in mental health research. Psychiatry Research. 327: 115265  
  
  
Reviewer: 2  
  
Comments to the Author  
Overview:  
Ex-vessel prices are used to look for patterns in species and gear types in order to define métiers for the U.S. West Coast groundfish fishery. A clustering, then dendrogram approach is used to identify 4 main métiers in this fishery.  
For the most part, the methods were clear, though I had many questions on how best to carry out the analysis. I did wonder if ex-vessel if the best metric to cluster by. I also had some suggestions on how best to present and ultimately identify the clusters.  
I also think that there are incomplete presentations on the idea of how the métiers can be used to define reference points. This seems to be a hanging idea that never is developed, so it needs more consideration.  
I hope these comments are constructive and useful in revising the manuscript.  
  
Major considerations:  
\*How different are the results if you did you use landed weight instead of ex-vessel price? Ex-vessel price just seems more unstable and/or subject to market change (sablefish has seen this over the last 5 years). I would suggest an analysis using landed weight so this can be addressed. I am not sure why targeting choice would not also be seen in the landed weights (more caught as they are targeted more).

The authors reran the clustering analysis using landed weight instead of ex-vessel revenue. Without a set k, the clustering resulted in 14 clusters with a lower silhouette score than the ex-vessel revenue run with 17 clusters. Similarly, with k set to 17, the clusters had a lower silhouette score, and 92.5% of the aggregated shots remained in the same clusters. Most shots with high sablefish landings were moved and became less aggregated/spread over more clusters, as sablefish have the highest price in the fishery. Note that much of the literature supports ex-vessel revenue as a metric (Briton (2019); Burch et al., (2021); Pascoe et al., (2022).

Briton, F., 2019. Defining metiers and fleets for the SESSF using multivariate statistical methods. Unpublished document presented to the December 2019 SERAG meeting.

Burch P, Sutton, C, Cannard, T, Briton, F and Sporcic, M (2021). An investigation of the bycatch of rebuilding species and other selected species in the Southern and Eastern Scalefish and Shark Fishery. December 2021, CSIRO, Australia.

Pascoe, S., Punt, A. E., Hutton, T., Burch, P., Bessell-Browne, P., & Little, L. R. (2022). Estimating economic-based target reference points for key species in multi-species multi-métier fisheries. Canadian Journal of Fisheries and Aquatic Sciences, 80(4), 732-746.  
  
\*It seems Figure 1 and 2 are showing similar things, just differently. I find Figure 2 much easier to interpret, so would suggest deleting Figure 1 (unless it has something unique to it).

The authors agree that Figures 1 and 2 show the same data in different ways and will move Figure 1 to the supplemental per the request of both Reviewers 1 and 2 to reduce redundant figures.   
\*The percentage of total landing plots are the same for each cluster whether you are showing species, gears or state. I would only show it once and would sort the clusters by % of total data (highest to lowest). It also makes me think whether it is even worth showing some of the clusters that have <5% (or 2.5% or whatever is a useful cutoff) of the data. I know you have already used a filter, but it seems like a secondary filter to really highlight the most prominent clusters is more useful than showing many other small clusters that 1) make it harder to track the results and 2) maybe be less robust because they are based on much less data. This would also simplify the text, as there is a lot that is dedicated to walking through 17 clusters. Looking at silhouette plot (Figure A.1) looks like not much is gained after 5-7 clusters, which also goes with the simplification suggestion above.  
\*Later you do simplify down to 4 métiers. It would seem the simplification process should happen first, then a description of the final métiers could happen rather than explaining on 17 clusters in detail.

The authors will create visualizations to reflect the final 4 métiers selected and move cluster-specific visualizations to the supplemental. Given the number of clusters, the authors believe it is important to include the percentage plots above each visualization to improve comprehension.  
\*How different are the clusters you get from the simplification process versus if you just used the 5 recognized by the silhouette approach? It seems like a lot of additional work when the cluster selection process using silhouettes gets you to a similar place?

The differences between the clusters at k = 4 to k= 7 is different compared to the refined clusters when k = 17. For instance, the midwater trawl gear is split between two clusters starting at k = 5. Distinctions between bottom trawl gears and catch composition also become more explicit at greater k, allowing for revision to match species ecology and fishery targeting information.

\*One obvious question is how do these métiers persist through time. Heery and Cope (2014; [https://urldefense.com/v3/\_\_http://fishbull.noaa.gov/1121/heery.pdf\_\_;!!K-Hz7m0Vt54!jDRM-rIoR4BVzUhG8Rabx7AAn2aT-V42\_b4wTN8Wz3AOuQstN4Ng5yuSsovYPVs5O1HPJ394qbuIahS\_6CgKIKuoUAkkWz0$](https://urldefense.com/v3/__http:/fishbull.noaa.gov/1121/heery.pdf__;!!K-Hz7m0Vt54!jDRM-rIoR4BVzUhG8Rabx7AAn2aT-V42_b4wTN8Wz3AOuQstN4Ng5yuSsovYPVs5O1HPJ394qbuIahS_6CgKIKuoUAkkWz0$) ) used a cluster analysis to analysis pacific coast groundfish bottom trawl clusters over time. Would such a consideration be useful in this analysis?

The authors suggest the need for revisions to the métiers across time but do not provide a framework or recommendation for the frequency. Barring unforeseen updates to the stock (see Kinneen et al., 2025 for the drastic change in ACL estimates for the widow rockfish fishery), these métiers should be robust for significant periods of time, but should be constantly reevaluated to ensure best available science is used.

To see cluster stability across time, a split assignment test was conducted with n-1 from 2020-2022. Given that many species (e.g., widow rockfish) did not have their fishery open for the entirety of the dataset (2011-2023), care was taken to ensure all species had landings when doing the split-test assignment. This was also suggested by Reviewer 3.

Clustering 2011-2022 and assigning 2023 (with k set at 17) produced similar results, with 96.3% of all shots being assigned to the same cluster as the original dataset. The majority of the 3.7% of shots that were not assigned to the same cluster came from ports with the lowest catch or were dominated by the gears and species with the lowest 15% of ex-vessel revenue kept in the study. When clusters were assigned to the 2023 data, there was more variance in assignments, with 89.0% of clusters being assigned to the same cluster as the original analysis. There was no clear delineation across species or gear for the shots that were assigned differently, but the shots with the lowest catch across all species were (71% of all misassigned shots came from ports with the lowest 25% of ex-vessel revenue in 2023). Given that this is just one year of data removed (so the assignment data is relatively small) and the other stability metrics, the authors are not concerned with over 10% of the assignment data being assigned to a different cluster.

Clustering 2011-2021 and assigning 2022-2023 (with k set at 17), the proportion of shots remaining in the same cluster was lower, with 93.6% of shots remaining in the same cluster as the original analysis. A similar pattern held, with low-value species, ports, and gears as the primary culprits to shifted assignments. Interestingly, the other major shift were shots dominated by widow rockfish and/or or yellowtail rockfish shifting from cluster 7 to 17 (which were combined into Métier A in the analysis), making cluster 7 a significantly smaller cluster, but with the same characteristics as the original analysis output (shelf rockfish in midwater trawl gear). Assignment of future years was improved, with 92.7% of future year shots being assigned to the correct cluster. Shots with widow and/or yellowtail returned to cluster 7 if landings of other midwater trawl rockfish (e.g., redstripe rockfish) were high enough, but there were similarly no other patterns of misassigned future year clusters.

Clustering 2011-2020 and assigning 2021-2023 (with k set at 17) had slightly better results compared to assigning 2022-2023, as this run resulted in 94.7% of all shots being assigned to the same cluster. Similarly, small-catch ports, species, and gears were often misassigned; unlike previous iterations, another pattern emerged. Fish pot (which primarily catch sablefish, a high-value species) shots were also misassigned and condensed into a two clusters instead of the four clusters from the original analysis with landings from the gear. Given that more than 99.9% of fish pot landings were combined into Métier D in the original analysis, and the landings did not move to a cluster that was not characterized by fish pot landings, the authors did not see this as a cause for concern. Assignment to future years stayed relatively constant, with 94.9% of clusters assigned to the same cluster. It was clear that the clusters with a low proportion of ex-vessel revenue were most likely to have future years misassigned (83.1% from clusters with >1% of the total ex-vessel revenue) and most of the remaining misassignment came from fish pot shots being assigned to the other cluster characterized by fish pots. Kinneen, M., Goodman, M. C., Sulc, A., Balstad, L., Diaz, R., Randrup, K., Patrone, W., Spencer, L., Morell, A., Rovellini, A., Dedrick, A., Grunloh, N., Bargas, M., Hopkins, S., Gersteva, V., Oken, K., Taylor, I., Haltuch, M., & Hamel, O. (2025) Status of widow rockfish stock off the U.S. West Coast in 2025. Pacific Fishery Management Council, Portland, Oregon.

\*lines 556-558: I am not sure anyone is arguing that single-species setting of catch limits is a reflection of the fisheries operating in a single-species manner. Instead, it is to indicated limits of catch per species. How these are prosecuted in what are inherently multi-species fisheries is another thing. I would look out for wording that suggests the multispecies nature of the fisheries are not strongly considered in the management decisions of the fisheries. It is also notable that métiers are not set in stone and can change due to a variety of reasons, and thus not stable entities.

The authors agree that the wording of lines 556-558 can be improved, specifically the ‘managed independently’ statement (lines 556-557). This has been amended  
\*The matter of reference points is brought up, but not really delved into in the paper. I am not sure what is this analysis brings to that discussion. It may be worth cutting that issue out of the paper. If not, much more detail and discussion is needed.  
The authors note that multiple reviewers found this extraneous and have revised the manuscript to focus on the implications of the current work rather than projections to future work,   
Minor considerations:  
\*line 89: The transition from identifying metiers to EBFM/multispecies considerations could be better. I suggest considering a topic sentence that transitions from talking about identifying metiers to how that information could play into EBFM, instead of switching abruptly to the topic.  
The authors agree that the transition is very abrupt. We have added subheadings to the introduction, as Reviewer 1 suggested moving the information regarding the fishery and FMP to the introduction and the authors support this structure for reducing the abrupt transition and create clear delineation between topics.  
  
Edits/suggestions  
\*line 128: “…relative to a proxy for the biomass…”  
\*lines 190-192: Specify in these catches that these are over the full time series (2011-2023), and not annual.  
The authors have made these edits

Reviewer: 3  
  
Comments to the Author  
  
Overall assessment.  
The manuscript presents a timely and policy-relevant clustering of groundfish landings into operational metiers using CLARA, followed by consolidation to a smaller management set. The approach is promising; however, several methodological choices are insufficiently justified, and key validation steps are absent. As written, the credibility of the inferred structure and the transparency of the consolidation are not yet at the level expected for publication in a fisheries analytics venue.  
  
Major methodological concerns and required clarifications  
  
1. Reliance on silhouette alone.  
The study appears to use silhouette widths both to select k and to assert structure. Silhouette is an internal index that favors roughly spherical clusters, is sensitive to scale, and does not assess stability. The manuscript should explain why stability diagnostics (e.g., bootstrap Jaccard per cluster) and between-run agreement (e.g., Adjusted Rand Index across seeds) were not undertaken. Without these, the risk of CLARA sampling/initialization artifacts remains, and the reported partition cannot be deemed robust.

The authors note that Reviewer 1 and Reviewer 3 both asked for cluster validation diagnostics and recognized the importance of including these diagnostics.

The CLARA algorithm was run across 100 random seeds and there was no change across seeds, meaning there was no need to compute ARI across seeds. This was not true across all clarifications run (e.g., HAC) but none of these supplemental runs produced more favorable outputs from clustering, so ARI was not run for these diagnostics as well.

The bootstrapped Jaccard was conducted using the clusterboot function in the *fpc* package. Per Hennig (2024), the recommended 100 bootstraps were run and recover (the Jaccard coefficient required to keep a cluster) was set at 0.75, and the threshold to dissolve a cluster was set at 0.5.

No cluster was dissolved more than half of the iterations, and only three were dissolved in a third or more runs. Five clusters were recovered at least 90% of the time with 10/17 clusters recovered at least 75% of the time. Smaller clusters had the largest variation in Jaccard similarities *J*, which makes sense given the higher variability with fewer data points, and were often the ones dissolved.

No clear stability thresholds are given in the literature, but Henning (2024) sets interpretation guidelines for *J*, which have been provided in Table 1. The number of clusters from the clusterboot run in each category is also reported in Table 1.

|  |  |  |
| --- | --- | --- |
| Threshold | Jaccard similarity *J* | Number of clusters |
| Dissolved cluster | *J* < 0.5 | 0 |
| Unstable cluster | 0.5 < *J* < 0.6 | 2 |
| Neither stable of unstable cluster | 0.6 < *J* < 0.75 | 4 |
| Valid and stable cluster | 0.75 < *J* < 0.85 | 3 |
| Highly stable cluster | 0.85 < *J* | 8 |

Table : Range of Jaccard similarity indexes for each stability threshold reported by Hennig (2024) and the number of clusters from our analysis fitting into each bin

Hennig C (2024). fpc: Flexible Procedures for Clustering. R package version 2.2-13,

<https://CRAN.R-project.org/package=fpc>.  
  
2. Method dependence and cross-algorithm agreement.  
The study adopts CLARA with Euclidean distance and no PCA, following prior evidence that Euclidean better recovers single-species fisheries and that PCA can obscure targeting patterns. Within this dataset, no cross-method baseline (e.g., HAC/Ward) or in-situ metric/scale sensitivity (Manhattan vs Euclidean; standardized vs raw revenue) is presented. Adding these concise checks would reduce the risk that results reflect algorithm–metric idiosyncrasies rather than underlying structure.

Manhattan distance was run without a set k, resulting in 11 clusters and a lower silhouette score. When k was fixed at 17, 93.8% of shots remained in the same cluster, with most of the movement coming from ports and gears with lower levels of ex-vessel revenue. The silhouette score was also lower than the original CLARA run with Euclidean distance.

A HAC was run across the 100 random seeds and resulted in a plethora of different clustering results, showing decreased stability of the method compared to the CLARA clustering analysis. Across the 100 random seeds, the number of clusters ranged from 9-15, each with a lower silhouette score than the CLARA runs. Reviewing the three clustering outputs with the highest silhouette scores for each k, many of the high-value and flatfish species were split across more clusters compared to the CLARA run, despite the lower k. Clusters dominated by bottom trawls did not differentiate well between the different bottom trawl gears, suggesting HAC is less suitable for the analysis. When k was set to 17, a similar trend was observed.

Briton, F., 2019. Defining metiers and fleets for the SESSF using multivariate statistical methods. Unpublished document presented to the December 2019 SERAG meeting.

Burch P, Sutton, C, Cannard, T, Briton, F and Sporcic, M (2021). An investigation of the bycatch of rebuilding species and other selected species in the Southern and Eastern Scalefish and Shark Fishery. December 2021, CSIRO, Australia.

Pascoe, S., Punt, A. E., Hutton, T., Burch, P., Bessell-Browne, P., & Little, L. R. (2022). Estimating economic-based target reference points for key species in multi-species multi-métier fisheries. Canadian Journal of Fisheries and Aquatic Sciences, 80(4), 732-746.

3. Treatment of data as amounts vs mixes.  
Because the scientific objective is metier identification (i.e., targeting preferences), the row vectors naturally read as species compositions. The manuscript should justify analyzing raw revenue vectors (which confound behavior with total scale and price variation) rather than compositions, or else explain why a compositional pipeline (zero handling + clr/Aitchison distance) was not considered. As currently framed, Euclidean distances on unstandardized revenue implicitly weight high-value species and market effects, potentially biasing cluster geometry.

Reviewer 2 also asked about using landed weight instead of ex-vessel revenue. Without a set k, the clustering resulted in 14 clusters with a lower silhouette score than the ex-vessel revenue run with 17 clusters. Similarly, with k set to 17, the clusters had a lower silhouette score, and 92.5% of the aggregated shots remained in the same clusters. Most shots with high sablefish landings were moved and became less aggregated/spread over more clusters, as sablefish have the highest price in the fishery. Note that much of the literature supports ex-vessel revenue as a metric (Briton (2019); Burch et al., (2021); Pascoe et al., (2022).

CLR/Aitchison distance was computed using the robCompositions package in R (Templ et al., 2011). All rows and columns have at least one zero, making this run impossible given the log ratio transformed data, making zeros an issue for computing. Lubbe et al., (2021) outlines zero replacement strategies for compositional data and suggests using BDLs (below detection limit) algorithms to replace zeros in a dataset. First, zero values were computed using the imputeBLDs function using the recommended presets in Lubbe et al., (2021) to improve convergence (maxit = 50, eps = 0.1), and increasing the number of bootstrapped samples to determine pls components to 50. The detection limit was set to the minimum non-zero value in the dataset (7.64 \* 10-6). Although imputeBDLs() takes hours to run, it is considered more robust than faster methods of zero-replacement (Lubbe et al., 2021).

The clustCoDa() function in the robCompositions package, which uses Aitchinson distance to compute clusters, was then used to calculate clusters with the CLARA algorithm and k = 17. The silhouette score was significantly lower than the original analysis (by over a full order of magnitude), and species dispersion across clusters was higher. This is likely to do with changing values away from zero when there was no catch on that species in the aggregated shot. Cluster size was more uniform, with no very small clusters and no large clusters (Cluster size ranged from 2.48%-14.7% of the total dataset. 10 of the 17 clusters in the original analysis are smaller than the smallest cluster from the clustCoDa() runs). The effects of location (which were already small) were less pronounced, as no cluster was isolated by state (while 4/17 clusters in the original analysis were California-only clusters). This is also likely caused by the new non-zero values, as species with geographic restrictions along the coast now had some landings at ports where it was extremely unlikely that the species in question would land.

Manhattan distance reduces weight to any dimension with large differences, but we want métiers to reflect differences in targeted species, which is reflected by the dimensions with higher values and therefore is less suited to our objectives than Euclidean distance. This is supported by Briton (2019) and Burch et al., (2021).

Briton, F., 2019. Defining metiers and fleets for the SESSF using multivariate statistical methods. Unpublished document presented to the December 2019 SERAG meeting.

Burch P, Sutton, C, Cannard, T, Briton, F and Sporcic, M (2021). An investigation of the bycatch of rebuilding species and other selected species in the Southern and Eastern Scalefish and Shark Fishery. December 2021, CSIRO, Australia.

Lubbe, S., Filzmoser, P., & Templ, M. (2021). Comparison of zero replacement strategies for compositional data with large numbers of zeros. *Chemometrics and Intelligent Laboratory Systems*, *210*, 104248.

Templ M, Hron K, Filzmoser P (2011). *robCompositions: an R-package for robust statistical analysis of compositional data*. John Wiley and Sons. ISBN 978-0-470-71135-4, [doi:10.1002/9781119976462.ch25](https://doi.org/10.1002/9781119976462.ch25).

4. External validation is qualitative only.  
Face validity (e.g., alignment with known gear practices) is described narratively, but no quantitative external validation is shown. The manuscript should state why simple measures such as gear-purity (or mutual information between cluster and gear) and a time-split assignment test (fit earlier years, assign later years by nearest medoid) were not reported. In their absence, it is difficult to judge whether clusters generalize beyond the fitting sample or merely echo gear labels.

The authors were unsure exactly what the reviewer meant by a gear-purity metric and hope that Reviewer 3 is satisfied with the other diagnostics run that showcase the robustness of the clustering.

A split assignment test was conducted with n-1 from 2020-2022. Given that many species (e.g., widow rockfish) did not have their fishery open for the entirety of the dataset (2011-2023), care was taken to ensure all species had landings when doing the split-test assignment. Reviewer 1 also suggested that cluster stability over time be assessed, which was examined with the clustering of n-1 years removed.

Clustering 2011-2022 and assigning 2023 (with k set at 17) produced similar results, with 96.3% of all shots being assigned to the same cluster as the original dataset. The majority of the 3.7% of shots that were not assigned to the same cluster came from ports with the lowest catch or were dominated by the gears and species with the lowest 15% of ex-vessel revenue kept in the study. When clusters were assigned to the 2023 data, there was more variance in assignments, with 89.0% of clusters being assigned to the same cluster as the original analysis. There was no clear delineation across species or gear for the shots that were assigned differently, but the shots with the lowest catch across all species were (71% of all misassigned shots came from ports with the lowest 25% of ex-vessel revenue in 2023). Given that this is just one year of data removed (so the assignment data is relatively small) and the other stability metrics, the authors are not concerned with over 10% of the assignment data being assigned to a different cluster.

Clustering 2011-2021 and assigning 2022-2023 (with k set at 17), the proportion of shots remaining in the same cluster was lower, with 93.6% of shots remaining in the same cluster as the original analysis. A similar pattern held, with low-value species, ports, and gears as the primary culprits to shifted assignments. Interestingly, the other major shift were shots dominated by widow rockfish and/or or yellowtail rockfish shifting from cluster 7 to 17 (which were combined into Métier A in the analysis), making cluster 7 a significantly smaller cluster, but with the same characteristics as the original analysis output (shelf rockfish in midwater trawl gear). Assignment of future years was improved, with 92.7% of future year shots being assigned to the correct cluster. Shots with widow and/or yellowtail returned to cluster 7 if landings of other midwater trawl rockfish (e.g., redstripe rockfish) were high enough, but there were similarly no other patterns of misassigned future year clusters.

Clustering 2011-2020 and assigning 2021-2023 (with k set at 17) had slightly better results compared to assigning 2022-2023, as this run resulted in 94.7% of all shots being assigned to the same cluster. Similarly, small-catch ports, species, and gears were often misassigned; unlike previous iterations, another pattern emerged. Fish pot (which primarily catch sablefish, a high-value species) shots were also misassigned and condensed into two clusters instead of the four clusters from the original analysis with landings from the gear. Given that more than 99.9% of fish pot landings were combined into Métier D in the original analysis, and the landings did not move to a cluster that was not characterized by fish pot landings, the authors did not see this as a cause for concern. Assignment for future years stayed relatively constant, with 94.9% of clusters assigned to the same cluster. It was clear that the clusters with a low proportion of ex-vessel revenue were most likely to have future years misassigned (83.1% from clusters with >1% of the total ex-vessel revenue) and most of the remaining misassignment came from fish pot shots being assigned to the other cluster characterized by fish pots.   
  
5. Consolidation to metiers lacks auditable rules.  
The reduction from many clusters to a few metiers is central to the management utility of the work, yet the decision logic appears partly qualitative. The manuscript should present a transparent mapping (Cluster --> Metier) with explicit criteria (dominant species thresholds, gear compatibility, revenue share cutoffs) and clarify the role of dendrogram lineage (guidance vs determinism). Without a reproducible rule set, readers cannot evaluate potential subjectivity or replicate the consolidation in other datasets.

The authors elaborated on the process on the refinement of clusters to métiers, which is now included in the supplemental material.  
  
6. Price specification and scaling are unspecified.  
The construction of revenue vectors hinges on the treatment of species prices Ps. It is unclear whether Ps is constant or varies by time/port, and whether columns were scaled before distance computation. Both choices materially affect distances (and thus clusters). The manuscript should clarify the price model and any standardization steps, and discuss implications for sensitivity to price volatility.

The authors defined price in the manuscript at the first mention of the variable. The price of species *s* is the average price of all landings of that species in the specific port using a specific gear type in the associated month of the aggregated shot, and is reported directly in the PacFIN database. Prices were not scaled before computing, but the PacFIN database does have an API-adjusted price for inflation. This sensitivity was run prior to submission and found that 98.9% of the aggregated shots were grouped the same way without inflation-adjusted price, with mostly low-value species and gears being moved across the smallest clusters.  
  
7. Aggregation scale and potential masking.  
Using month×port×gear "aggregated shots" is practical, but it introduces the possibility of mixing heterogeneous trips within cells. The manuscript should explain why this level was preferred and provide a brief diagnostic (e.g., dispersion of species mix within cells or across months) to reassure readers that aggregation is not erasing finer-scale tactics that would alter the inferred metiers.

We added a within-cluster and between-cluster variance calculation into the CLARA algorithm. This was calculated as the percentage of the total variation contributed by both within and between cluster classifications. The within-cluster variance was 18.5% of total variance and between-cluster variance was 81.5% of total variance, which is sufficiently low for the purposes of this work. The choice to choose this level of scale has been added to the supplemental materials, along with the results of this sensitivity.

8. Absence of null-model benchmarking.  
No check is shown against chance structure. The manuscript should explain why a simple null (e.g., row-wise permutation of species columns or price randomization) was not used to demonstrate that internal indices materially exceed what arises under broken co-structure.

The authors used price randomization to provide a null-model benchmark. Prices were randomized across species (e.g., the price pulled for sablefish in aggregated shot A was selected from a list of sablefish prices). This model was run across the same 100 random seeds from the original analysis, and variations did exist across the model runs. The majority (61/100) resulted in 17 clusters with similar shot allocations to clusters (the average assignment matching across these runs was 87.4%), and most of the results without 17 clusters had shifts in high-value species where price is more variable across time (e.g., sablefish) that either broke them into more clusters or combined them together. This also broke any location-specific assignments, with only 23/100 runs having a cluster specific to one state (in the original clustering, 4/17 were California-only clusters).

When prices were not randomized across species (e.g., price for sablefish in aggregated shot A can be drawn from all available prices), the modeling clustering result does not produce any meaningful results. The number of clusters ranged from 3-8 across the 100 seeds, and no clear pattern emerged in species or gear composition in clusters. High value species with lower catch (e.g., sablefish) were especially impacted, as the higher weighting from being more expensive was removed and their impact on clustering was diminished. The highest silhouette score across the 100 runs was more than a full order of magnitude lower than the original run. Given these results and the fact that the CLARA clustering did not change across 100 random seeds in the original run, we are confident that chance structure is not occurring.  
  
 Minor but important items (methods reporting)  
  
 Algorithm description. Standardize terminology: CLARA is a PAM/medoids procedure; calling it a k-means variant is inaccurate and may confuse readers.

The authors note that CLARA is a k-medoids procedure, and this was adjusted in the manuscript accordingly.

 Filtering thresholds. Provide justification for species/gear inclusion cutoffs and a short sensitivity (e.g., whether borderline species materially change cluster membership).

The clustering analysis was rerun removing the lowest-EVR gear and two lowest-EVR species (flatfish trawl, and silvergray rockfish and starry flounder were the species removed), as well as adding the gear and two species on the border of the gear and species threshold (other hook and line gear, and both vermillion and sharpchin rockfish).

Border, non-set k: The clustering resulted in 19 clusters with primarily the same catch composition across clusters, with stronger differentiation between other bottom trawl gears and their landings within clusters. The silhouette score for this run was slightly higher (0.46 vs 0.45) than the original analysis. The two new clusters were primarily derived from clusters 2, 6 and 16 and clusters 4, 9, and 14 from the original analysis. However, these clusters were combined into métier B and C, respectively, and using this run would not alter the final métier designations.

Border species removed, k fixed at 17: There are slight changes in clustering with the removal of these borderline species and gear. The primary change is the movement of much of the bottom trawl gear to clusters 6 (19.1% of moved shots) and 9 (80.9%). Clusters 4 and 9 are both in the same métier, but cluster 6 is not. Given that most of the movement of these records are into the same métier (and the small proportion of moved records, 0.94% of the total fishery), the final métier clustering would not significantly change.

Border added, non-set k: Adding borderline species and gear reduced the number of clusters to 15, with clusters 2, 11, 12, 14 and 16 shuffled into three clusters. The five original clusters were generally similar, with cluster 14 grouped into a separate métier due to differing catch composition and more bottom trawl gear. These clusters are also less than 1% of the fishery’s ex-vessel revenue. The silhouette score is slightly lower as well. Much of the other line and pole additions are split between the revised five clusters; the same could be said for both rockfish species. The silhouette score for this run is slightly lower than the original analysis.

Border added, k fixed at 17: The clustering is almost identical to the original analysis, with nearly identical silhouette scores (down to the third decimal) and 98.8% of all shots remaining in the same cluster. The shots that moved generally moved between clusters that ended in the same métier and were primarily commercial pole gear. The new gear shots were split between the small-value pole and line clusters, with a small (7.5%) proportion of the new shots ending up in cluster 10, which is in a separate métier than the rest of the pole and line fishery.

Given these sensitivities, we are confident that the clustering is robust to borderline species and the thresholds are sufficient to accurately capture the dynamics of the fishery.

Reproducibility. Specify random seeds, software versions, and code commit/hash so that results are exactly reproducible.

The authors note that the code is available on the GitHub repository, which is linked at the bottom of the manuscript. Software versions will be added and seeds will be available in the repository as well. Will- still need to organize with new code from reviewers  
  
Editorial recommendation  
  
Accept with major revision. The contribution is compelling, but the manuscript currently lacks (1) stability evidence, (2) cross-method confirmation, (3) minimal quantitative external validation, and (4) transparent, auditable consolidation rules. The authors should either substantiate the present choices with the above clarifications or provide concise robustness evidence demonstrating that the principal conclusions do not hinge on a single algorithm–distance specification or on market-driven scaling.