ISC/21/SHARKWG-1/I-01

Cluster analysis used to re-examine fleet definitions of North Pacific fisheries with spatiotemporal consideration of blue shark size and sex data

Kinney MJ[1], Carvalho F[1], Kai M[2], Semba Y[2], Liu K-M[3], Tsai W-P[4], Leonardo C-G J[5], Horacio H-A[6], Daniel C-C L[7], and Teo SLH[8]

1. *NOAA Fisheries, Pacific Islands Fisheries Science Center, Honolulu, Hawaii, USA*
2. *Fisheries Resources Institute, Shizuoka, Japan*
3. *National Taiwan Ocean University, Keelung, Chinese Taipei*
4. *National Kaohsiung University of Science and Technology, Kaohsiung, Chinese Taipei*
5. *Centro Regional de Investigación Pesquera (CRIP), Ensenada, Instituto Nacional de Pesca, Mexico*
6. *Centro Regional de Investigación Acuícola y Pesquera de Ensenada, Baja California, Mexico*
7. *Universidad Autónoma de Sinaloa, Sinaloa, Mexico*
8. *NOAA Fisheries, Southwest Fisheries Science Center, San Diego, California, USA*

Email: michael.kinney@noaa.gov

PIFSC Working Paper WP-22-001.
Issued 3 May 2022.
https://doi.org/10.25923/zet2-sk13

Abstract

This study looked at re-examining the North Pacific fleets that have been used for previous assessments of blue shark by investigating the size and sex composition data from observer records, port and scientific samples in greater detail. Our goal is to provide information that can be used by the ISC shark working group to more appropriately define fleet structure for the assessment based on size and sexual composition of the catch. Ultimately, refining fleet structure within the model with greater consideration for the spatiotemporal characteristics of blue shark catch may help reduce model misspecification in future assessments. We analyzed nearly 600,000 individual records of blue shark size and sex information divided across 240 5 x 5° grid cells covering the North Pacific. A clustering approach was taken to discern areas with related size and sex compositions. Results suggested four distinct clusters, where Clusters 1 and 4 (made up primarily of smaller immature animals) predominate in the catch at higher latitudes (north of ~25°N), especially in the eastern and western edges of the North Pacific (waters nearer the coasts). While Cluster 2 (mature males and females) and Cluster 3 (mostly males, both mature and immature) predominate in a band from ~ 20°N to near the equator. During fall and winter (seasons 1 and 4) this band of mature animals expands north in central Pacific waters, loosely around Hawaii, as high up as ~40°N. We suggest that this work, along with several other studies carried out by various members of the ISC shark working group over the years, be used to better define the fleets used in future assessments of blue sharks in the North Pacific.

Introduction

In previous assessments of North Pacific stocks of both blue (*Prionace glauca*) and shortfin mako sharks (*Isurus oxyrinchus*), catch has been apportioned to nation-specific fleets, typically distinguished from one another by some combination of fishing gear, and their primary target species as shark (especially blue shark) is almost never the main target species.

While a fleet definition based on gear and target species is reasonable, we believe that it could be improved by examining the spatiotemporal characteristics of catch in greater detail. Building off the work of Sippel et al. (2015) with shortfin mako, Teo (2016) and Ochi et al. (2016) with albacore, and Kinney et al. (2018) with blue sharks in Hawaii, we attempted to refine the North Pacific longline fleet definitions used to assess blue sharks by examining spatiotemporal differences in size and sex using a clustering approach. A re-examination of the biological data from these fisheries could improve our understanding of the differences in spatiotemporal distributions of male and female, as well as adult and juvenile blue sharks. The goal is to use cluster results to define spatial areas for fleets rather than nation and target species, thus reducing model misspecification by allowing the working group to produce indices and size compositions that more appropriately consider the spatiotemporal characteristics of blue shark catch.

Data and Methods

Blue shark catch and comp data were sourced from 14 total fishing fleets operating in the North Pacific (Table 1). Japanese and Taiwanese longlines constitute the bulk of the data used in this study, as catch from the other fleets is small in comparison. Reliable size and sex data for blue sharks in the majority of these fisheries vary over time but generally improve as it approaches the present. Concerns over reliability as well as issues with spatial coverage when examining the entire North Pacific led us to focus our analysis on the most recent 10 years of data, from 2009 to 2018, a time period with nearly 600,000 individual blue shark measurements (Figure 1).

All size data used in this analysis were in precaudal length (PCL). Lengths recorded in either fork length or total length were converted PCL to be consistent with length observations in the 2017 blue shark assessment (ISC 2017). Length conversation equations were taken from Fujinami et al. (2016b) Table 3.

To evaluate the spatiotemporal differences in size and sex compositions of North Pacific blue sharks, we divided the North Pacific into two hundred and forty (240) 5 x 5° grid cells. All cells with <5 measured blue sharks were discarded from the analysis as leaving cells with such small amounts of data in the analysis can lead to issues where such grid cells are seen as comparatively pure nodes, resulting in inflated estimates of the number of cluster groups needed to properly describe the data. We used a clustering approach to discern areas with consistent size and sex compositions. To reduce the dimensionality of the problem and autocorrelation between length bins, the size composition data were aggregated to several maturity group compositions:

1) 4 groups, sex-specific mature, and immature individuals
2) 6 groups, sex-specific immature, sub-adult, and adult individuals
3) 2 groups, mature and immature individuals with sexes grouped
4) 3 groups, immature, sub-adult, and adult individuals with sexes grouped

Composition 1 was established using the sex-specific size at 50% maturity for blue sharks in the North Pacific (Fujinami et al. 2016b). Size at 50% maturity for males was 160.9 cm PCL, while female size at 50% maturity was 156.6 cm PCL. Composition 3 used the same sex-specific size at 50% maturity to separate sub-adults and adults and then established a second sex-specific point at which to distinguish immature individuals from sub-adults. This second separation point was based on the accepted growth model for blue sharks which indicates that both male and female blues grow rapidly, >25% of their total PCL for each of the first 2 years of life, up to 92.1 and 90.7 cm PCL for males and females respectively. Following this rapid growth period, average sizes between ages begins to decline (Fujinami et al. 2016a, Table 3). This second separation point is somewhat arbitrary; however, it allows the analysis to distinguish between rapidly growing young animals and larger, yet still immature sub-adults, groups, which based on visual inspection of spatial catch data, can sometimes appear to be distinct. Composition 3 was established by averaging the sex-specific size at 50% maturity for male and female blue sharks in the North Pacific, resulting in an average sex-unspecific size at maturity of 158.7 cm PCL. Composition 4 is the same as Composition 2 but with the sexes grouped together, length at 50% maturity was set at 158.7 cm while the split between immature and sub-adult was set at 91.4 cm, both the averages between males and females for each split.

Apart from the various maturity group compositions, clustering tests were conducted on different combinations of area, season, and year. As Teo (2016) found with albacore, when 5 x 5° areas were further disaggregated by years, the clustering pattern became more complex. In the case of blue sharks, this also resulted in many areas no longer meeting the criteria for minimum number of individuals to be analyzed, thus making results hard to interoperate with large blanks between clustered areas from year to year. Combining years into 1 super year was necessary. This allowed a clustering approach with various maturity group compositions to be tested seasonally across the entire North Pacific. With seasons here represented as season 1: Jan – Mar (winter); season 2: Apr – Jun (spring); season 3: Jul – Sep (summer); and season 4: Oct – Dec (fall).

Our analysis followed the same approach as Teo (2016) in his work on albacore, using the k-means clustering algorithm described in Hartigan and Wong (1979), but with k (number of clusters) expanded to range from 1 to 10, and 100 random sets of initial centers each.

Determining the appropriate k from a cluster analysis can be a convoluted and somewhat subjective process. Many studies have established the appropriate value of k by examining the resulting clusters from the random initial set that lead to the most abrupt change in the within cluster sum of squares. Basically, looking for a value of k after which the improvement in total within sum of squares moderates (a point of diminishing returns), this method is referred to as the elbow method. The qualitative nature of this approach has led many researchers to attempt to develop their own, hopefully more quantitative and less subjective, methods; however, one single test that can best establish the appropriate value for k in a given analysis remains elusive. Here we used a method developed by Charrad et al. (2014) where 30 different indices (drawn from the scientific literature) for determining the appropriate number of clusters in a data set are run as a meta-analysis, the number of clusters with the most indices in agreement is preferred (majority rule). Additionally, we also took a closer look at three of the more predominantly used indices found in the scientific literature, the elbow method described above, as well as the gap statistic method, and the silhouette method, both of which are described in detail in Charrad et al. (2014).

Finally, once the appropriate k was established using k-means and the indices described above, we used agglomerative hierarchical clustering with complete linkage and Euclidean distance to examine the clusters across each of our four maturity group compositions. The resulting clusters were examined by visually inspecting the resulting dendogram from the cluster analysis, mapping the areas in each cluster, and breaking down the size and sex composition of each resulting cluster. A pairs plot was also used to examine the differences between the resulting clusters (Teo 2016).

Results and Discussion

Results from across the multiple maturity group compositions tested indicated reasonably consistent spatial and temporal clustering patterns across the North Pacific, with more differentiated maturity group compositions (maturity group 2) leading to more complex clustering results, while less differentiated grouping led to more simplified patterns. As the broad patterns remained intact across the different tested compositions for simplicity and consistency with past studies, we will focus our discussion on the results from maturity group compositions 1, where sex and size data were aggregated into 4 groups, sex-specific mature and immature individuals.

The various indices used to establish the appropriate k for our analysis showed a strong preference for 4 total cluster groups, with majority-rule results indicating that 9 of the tested indices determine 4 clusters as the best number of clusters for this data set (the next closest result indicated agreement between only 3 indices, which advocated for either 3 or 6 total cluster groups) (Tables 2 and 3). All 3 of the predominant indices used in the literature (elbow, gap, and silhouette) converged on the same value for k, 4 cluster groups (Figure 2).

Results of the hierarchical clustering approach with the number of clusters set to four indicated that Clusters 1 and 4 were predominantly made up of immature animals, with Cluster 1 consisting chiefly of immature females, while Cluster 4 had more immature males (Figure 3). Measurements of PCL also indicated that Clusters 1 and 4 consisted of smaller blue sharks than the other cluster groups (Figure 4). Based on PCL, Clusters 2 and 3 consisted of comparatively larger sharks, with Cluster 3 having the largest average size across all cluster groups. In terms of sex, Cluster 3 consisted predominantly of mature males, while Cluster 2 had a relatively even split of mature males and females. When mapped some broad patterns were observed (Figure 5). Regardless of season, Clusters 1 and 4 (made up primarily of smaller immature animals) predominate in the catch at higher latitudes (north of ~25°N), especially in the eastern and western edges of the North Pacific (waters nearer the coasts). While Cluster 2 (mature males and females) and Cluster 3 (mostly males, both mature and immature) predominate in a band from ~ 20°N to near the equator. During fall and winter (seasons 1 and 4) this band of mature animals expands north in central Pacific waters, loosely around Hawaii, as high up as ~40°N (Figure 4), a pattern that may be related to pupping but more information on maturity stage would be needed to validate this. Patterns which are in general agreement with the findings of Sippel et al. (2014), who indicated that the smallest mean-sized females and males were found in the northwestern and northeastern regions of the Pacific, while mature males and females predominated in the more central waters of the North Pacific.

Our cluster results agree with tagging and scientific survey data collected by the Southwest Fisheries Science Center, with juvenile blue sharks predominately captured inshore, with rarer appearances of adult and sub-adult animals (Runcie et al. 2016), which when tagged appeared to head east and south out of coastal waters (Madigan et al. 2021; unpublished SWFSC tagging data).

Using a different approach, our analysis shares many of the same conclusions presented in Sippel et al. (2014) and Carvalho and Sippel (2016). The resulting consensus from these three papers give justification for some level of fleet aggregations, especially in terms of the fleets operating in the northwestern areas of the North Pacific as was suggested by Carvalho and Sippel (2016); however, the feasibility of this will rely heavily on the filtering processes used when reporting catch from these fleets. Without a clear understanding of the raw data prior to filtering, it is difficult to suggest specific aggregations. Here we present analytical evidence, which could be used to justify decisions made at the raw data filtering level.

Fleet aggregation, if the working group decides to proceed in that manner, can help to create more appropriate fleet definitions that reduce model misspecification by allowing the working group to produce indices and size compositions for the assessment that more appropriately consider the spatiotemporal characteristics of blue shark catch. A final paper on this topic by Kanaiwa et al. (2019), which uses a clustering approach that also considers CPUE, should be considered by the working group along with this paper and the two works listed above to determine the appropriate level of fleet aggregation for future assessments of blue sharks.

PIFSC Working Paper WP-22-001.
Issued 3 May 2022.
https://doi.org/10.25923/zet2-sk13

References

Carvalho F, Sippel T. 2016. Direct Estimates of Gear Selectivity for the North Pacific Blue Shark Using Catch-at-Length Data: implications for stock assessment. ISC/16/SHARKWG-1/13

Charrad M, Ghazzali N, Boiteau V, Niknafs A. 2014. NbClust: an R package for determining the relevant number of clusters in a data set. J. Stat. Softw. 61, pp.1-36.

Fujinami Y, Semba Y, Ijima H, Tanaka S. 2016a. Age and growth estimation of the blue shark, *Prionace glauca*, in the western North Pacific Ocean. ISC/16/SHARKWG-1/02

Fujinami Y, Semba Y, Okamoto H, Ohshimo S, Tanaka S. 2016b. Reproductive biology of the blue shark, Prionace glauca, in the western North Pacific Ocean. ISC/16/SHARKWG-1/01

Hartigan JA, Wong MA. 1979. Algorithm AS 136: A k-means clustering algorithm. J. R. Stat. Soc. Ser. C. Appl. Stat. 28:100-108.

ISC. 2017. Stock assessment and future projections of blue shark in the North Pacific Ocean through 2015. Report of the Meeting of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean, Vancouver, Canada.

Kanaiwa M, Semba Y, Kai M, Leonardo C-G J, Horacio H-A, Daniel C-C L, Liu K-M, Tsai W-P, Kinney M, Teo SLH, and Carvalho F. 2019. Clustering analysis of size data for blue shark (*Prionace glauca*). ISC/19/SHARKWG-1/01

Kinney MJ, Carvalho F, Teo S. 2018. Preliminary results of redefining fleet definitions of Hawaiian longline fisheries with spatiotemporal consideration of blue shark size data. ISC/18/SHARKWG-3/02

Madigan DJ, Shipley ON, Carlisle AB, Dewar H, Snodgrass OE, Hussey NE. 2021. Isotopic tracers suggest limited trans-oceanic movements and regional residency in North Pacific blue sharks (Prionace glauca). Front. Mar. Sci. 8, p.489.

Ochi D, Ijima H, Kinoshita J, Kiyofuji H. 2016. New fisheries definition from Japanese longline North Pacific albacore size data. ISC/16/ALBWG-02/03, Nanaimo, BC, Canada.

Runcie R, Holts D, Wraith J, Xu Y, Ramon D, Rasmussen R, Kohin S. 2016. A fishery-independent survey of juvenile shortfin mako (*Isurus oxyrinchus*) and blue (*Prionace glauca*) sharks in the Southern California Bight, 1994–2013. Fish. Res. 183:233-243.

Sippel T, Ohshimo S, Kotaro Y, Kai M, Carvalho F, Liu K-M, Castillo-Geniz J, Kohin S. 2015. Spatial and temporal patterns of shortfin mako shark size and sex in the North Pacific Ocean. ISC-15 SharkWG, Shimizu, Japan.

Sippel T, Semba Y, Shiozaki K, Carvalho F, Castillo-Geniz JL, Tsai W, Liu KM, Kwon Y, Chen Y, Kohin S. 2014. Size and sex structure of blue sharks in the North Pacific Ocean. ISC/16/SHARKWG-1/14

Teo S. 2016. Spatiotemporal definitions of the US albacore longline fleets in the North Pacific for the 2017 assessment. ISC/16/ALBWG-02/08, Nanaimo, BC, Canada.

https://doi.org/10.25923/zet2-sk13

InPort Citations

USA-DGN—Southwest Fisheries Science Center, 2022: California Gillnet Fishery, https://www.fisheries.noaa.gov/inport/item/12935

USA-HI-LL-D, and USA-HI-LL-S—Pacific Islands Regional Office, 2022: Longline Observer Data System, https://www.fisheries.noaa.gov/inport/item/9027

USA-JuvySurvey—Southwest Fisheries Science Center, 2022: Pacific Highly Migratory Species Data, https://www.fisheries.noaa.gov/inport/item/1884

USA_Trawl-survey—Southwest Fisheries Science Center, 2022: FRD Trawl Database, https://www.fisheries.noaa.gov/inport/item/20693

PIFSC Working Paper WP-22-001.
Issued 3 May 2022.
https://doi.org/10.25923/zet2-sk13

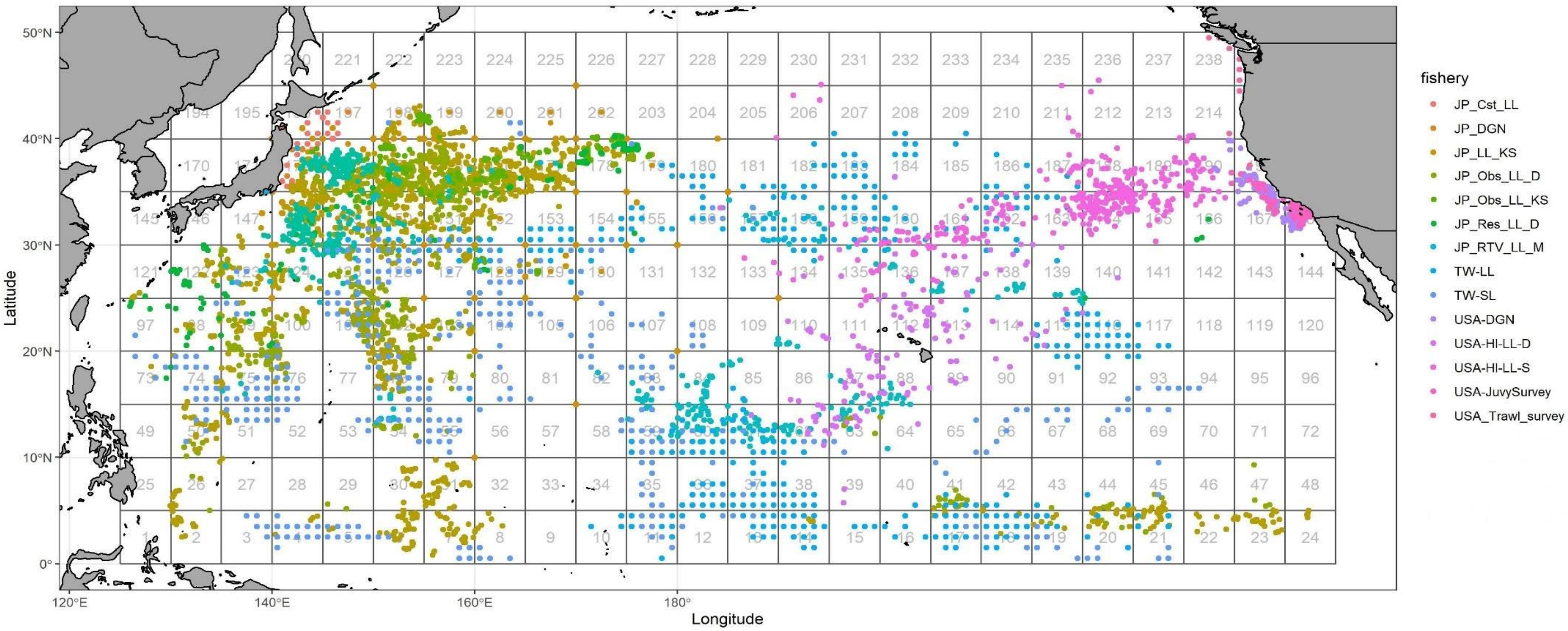

Figure 1. Map of the North Pacific with numbered 5 × 5° grid cells used in cluster analysis. Points indicate individual records of captured blue sharks with recorded size and sex information covering the 2009−2018 period. Colors indicate from which of the 14 different fishing fleets data the captured animals were sourced.

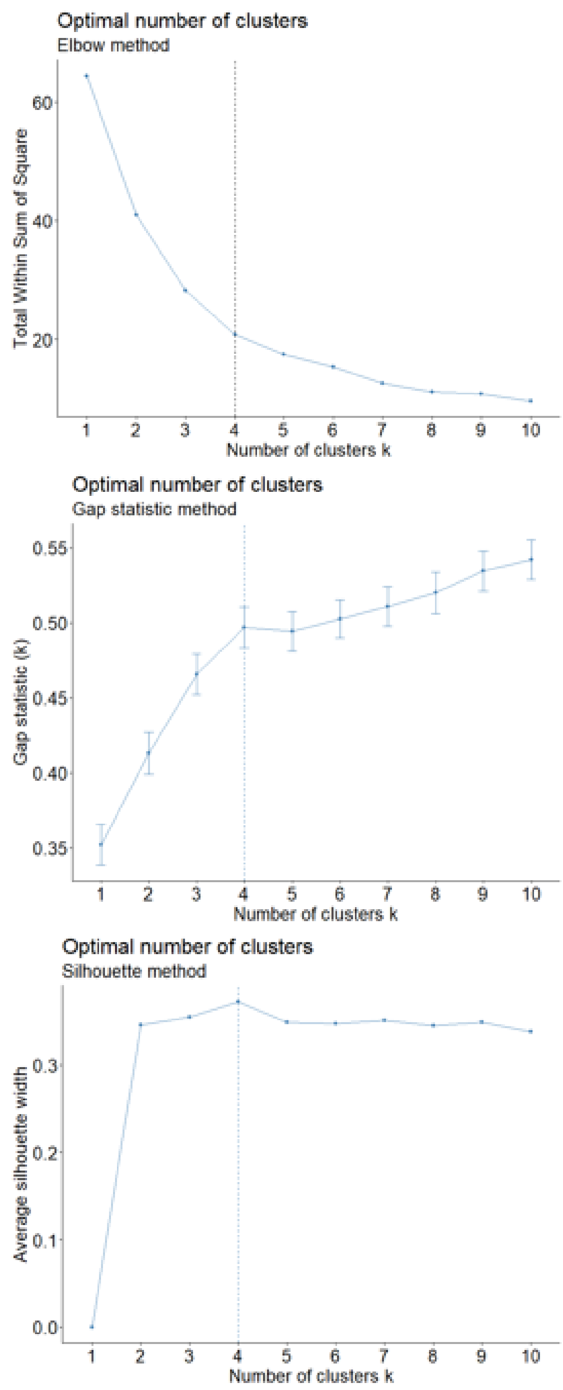

Figure 2. Upper plot, elbow method: change in the total within cluster sum of squares with increasing number of clusters. Middle plot, gap statistic method: change in within-cluster dispersion with that expected under an appropriate reference null distribution. Lower plot, silhouette method: a measure of how much a point is similar to its own cluster compared to other clusters.

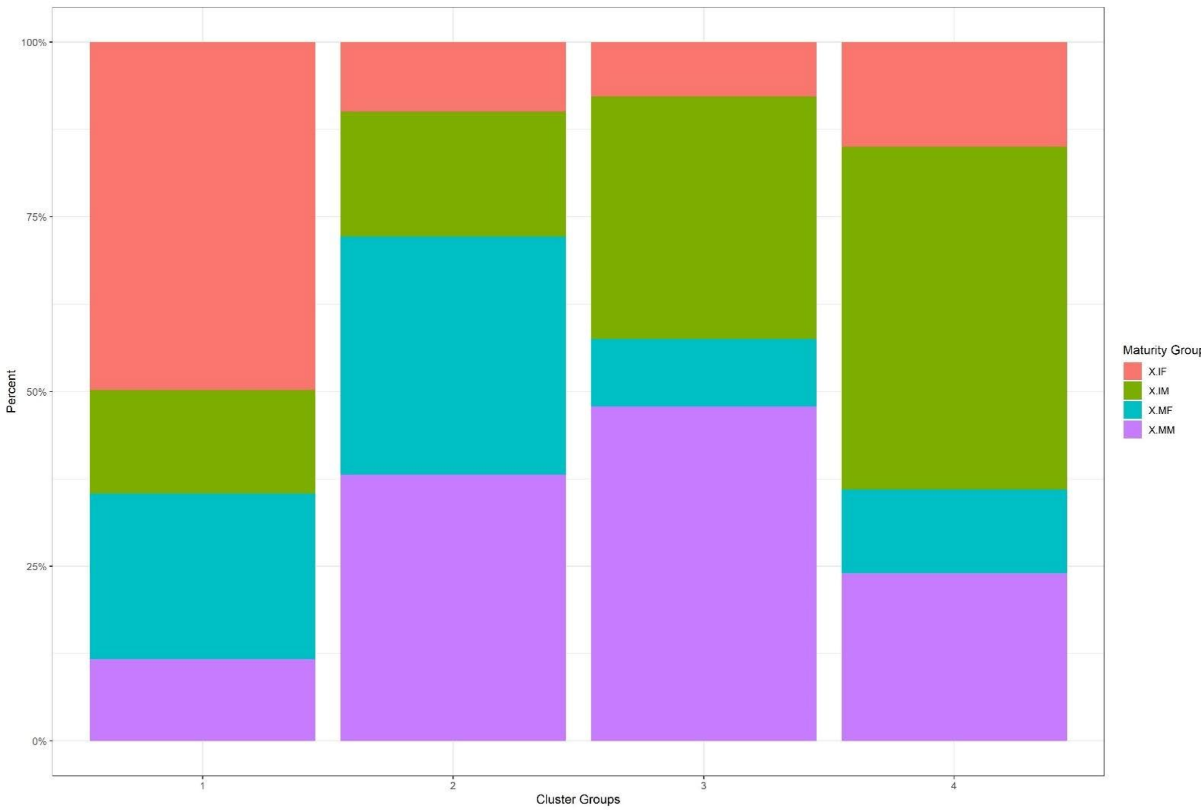

Figure 3. Stacked bar plot indicating the percent makeup of each of the four analyzed clusters by maturity group (X.IF—immature female, X.IM—immature male, X.MF—mature female, and X.MM—mature male).

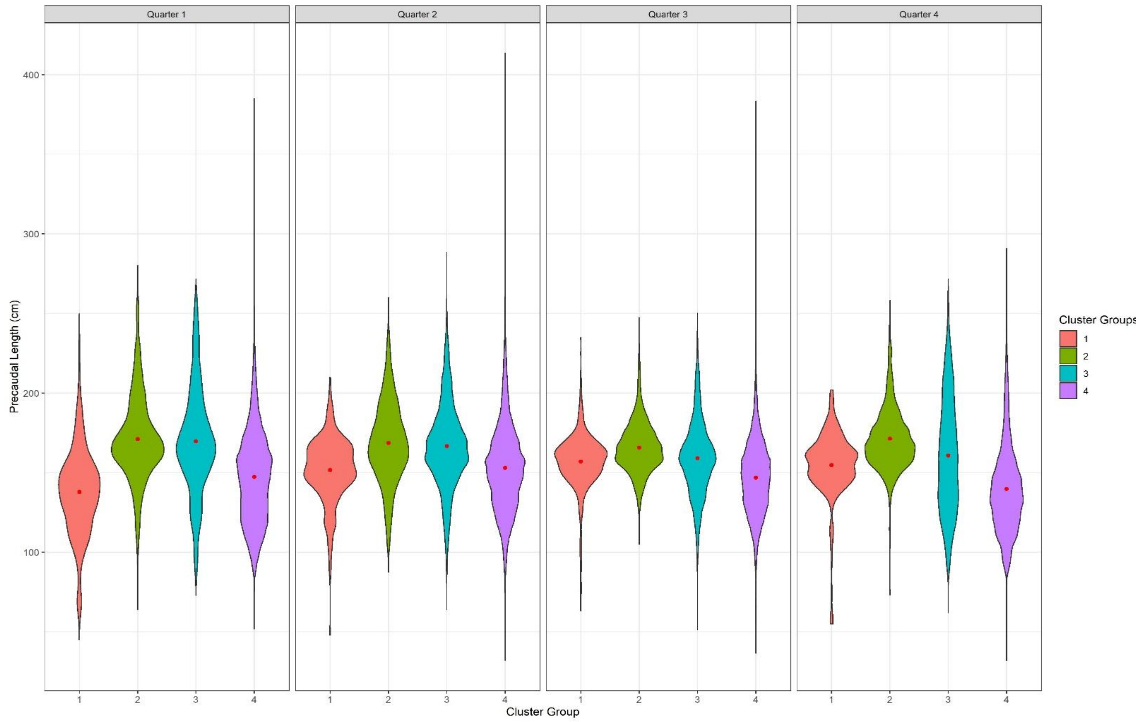

Figure 4. Size composition of each identified cluster (1−4) across season (season is represented by plot faceting, right to left, seasons 1−4).

PIFSC Working Paper WP-22-001.
Issued 3 May 2022.
https://doi.org/10.25923/zet2-sk13

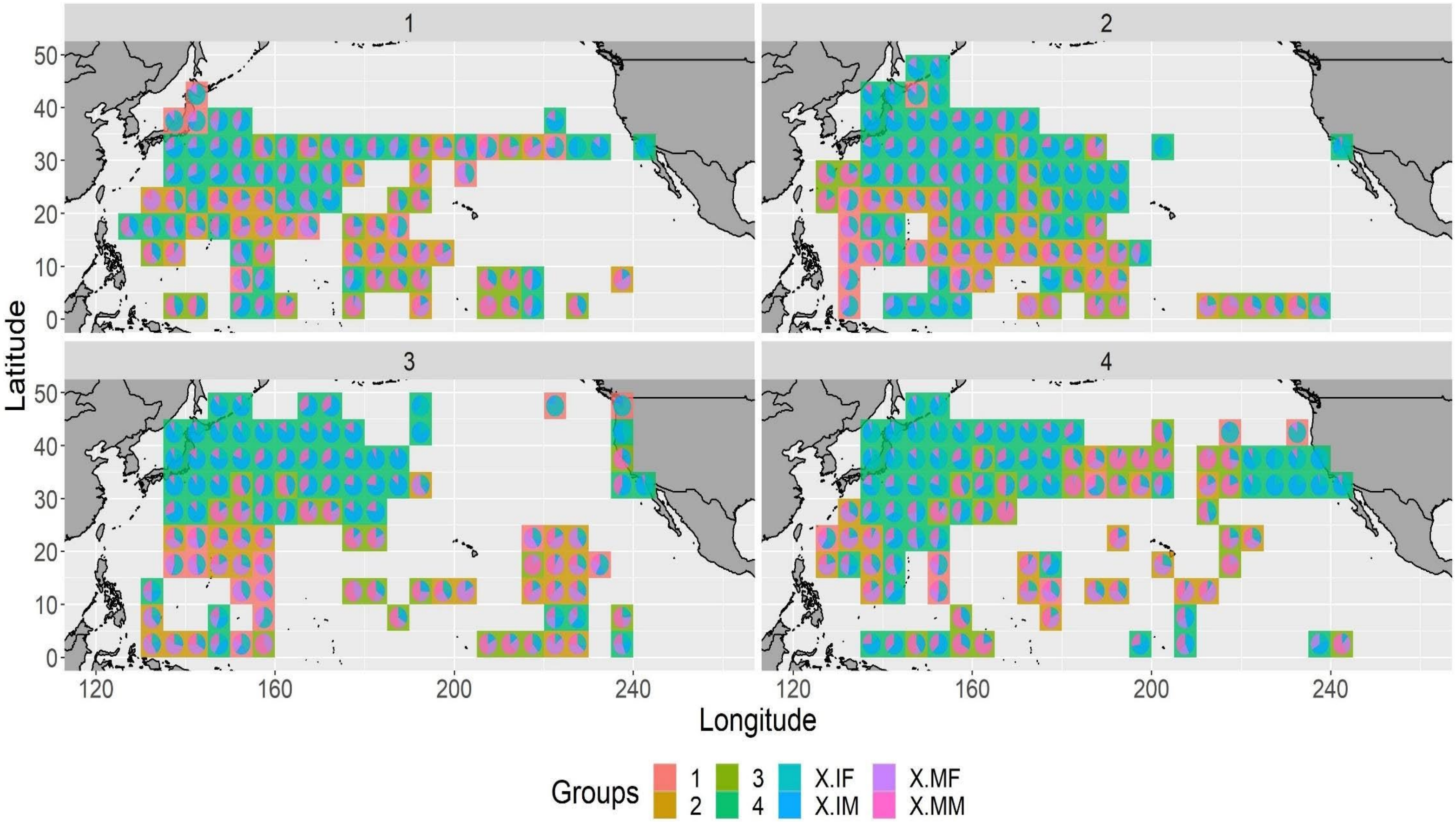

Figure 5. Map of the North Pacific with 5 × 5° grids. Groups 1−4 identify the cluster group to which each 5 × 5° box belongs, boxes without color (grey) did not meet the minimum criteria to be included in the cluster analysis (e.g. number of trips and number of individuals). Groups X.IF (immature females), X.IM (immature males), and X.MF (mature females), X.MM (mature males) identify the pie chart colors. The pie charts indicate the percent of each maturity group within each box. Plots are faceted by season (1−4).

Table 1. Description of fishing fleets which provided blue shark data to this study.

Fishery	Full Name	Data Source	Description
USA-DGN	United States Drift Gill-net Fishery	Onboard observers (coverage varies from 4 to 23%)	Drift gill-net fishery operating mostly near shore off the coast of the western United States targeting swordfish and thresher sharks.
USA-HI-LL-D	United States Hawaii Longline Deep-set Fishery	Onboard observers (20% coverage since 2001, less than 5% prior to that)	Deep-set longline gear operating in the waters around Hawaii using ≥15 hooks per float (Walsh et al. 2009) targeting tunas.
USA-HI-LL-S	United States Hawaii Longline Shallow-set Fishery	Onboard observers (100% coverage since 2004, less than 5% prior to that)	Shallow-set longline gear operating in the waters north of Hawaii using <15 hooks per float (Walsh et al. 2009) targeting swordfish.
USA-JuvySurvey	United States SWFSC Juvenile Shark Survey	Scientific samples	The NOAA Southwest Fisheries Science Center conducted an annual longline research cruise for juvenile blue and mako sharks off the Southern California Bight from 1993 to 2015.
USA_Trawl-survey	United States Trawl Survey	Scientific samples	The NOAA Southwest Fisheries Science Center conducts an annual trawl survey for coastal pelagic species (sardine, mackerel, etc.) and gets some incidental shark catch.
TW-LL	Taiwan Large-scale Tuna Longline Fishery	Onboard observers since 2004	The large-scale tuna longline fishery operates in two areas: north of 25°N and south of 25°N, with catching mainly albacore tuna, *Thunnus alalunga*, in more temperate waters, while targeting bigeye tuna, *Thunnus obesus*, in equatorial waters.
TW-SL	Taiwan Small-scale Tuna Longline Fishery	Onboard observers since 2012	The small-scale tuna longline fishery operates mainly in coastal and off-shore waters but some are in high seas.
JP_DGN	Japanese Large-mesh Drift Gill-net Fishery	Port sampling data available since 2011	Large-mesh drift gill-net fishery operating within Japan's economic exclusive zone (EEZ) targeting billfish and swordfish.

Fishery	Full Name	Data Source	Description
JP_LL_KS	Japanese Offshore Longline Shallow-set Fishery	Port sampling data available since 2008	Shallow-set longline fishery with number of hooks between floats (HPB) smaller than 5 or 6 operating in the Japanese off-shore water with smaller vessels (20–120 mt).
JP_Cst_LL	Japanese Coastal Longline Fishery	Port sampling data available since 2012	Longline fishery operating in the coastal waters near Japan with the vessels smaller than 20 mt.
JP_RTV_LL_M	Japanese Research and Training Vessel (Middle-set)	Measurement on board between 2016 and 2018	Same above. Note that data is from operation with gear depth between shallow and deep set (operation in coastal area).
JP_Res_LL_D	Japanese Research Offshore Longline Deep-set Fishery	Measurement on board available since 1993	Collected by NRIFSF-related research cruise with deep-set longline.
JP_Obs_LL_D	Japanese Deep-set Longline Observer Data	Onboard observers (coverage, less than 9% for distant water fishery and less than 5% for offshore water fishery) available since 2011	Deep-set longline fishery with number of hooks between floats (HPB) larger than 4 or 5 operating in the Japanese distant water with larger vessels (>120 mt).
JP_Obs_LL_KS	Japanese Offshore Shallow-set Longline Observer Data	Onboard observers (coverage, less than 9% for distant water fishery and less than 5% for offshore water fishery) available since 2011	Shallow-set longline fishery with number of hooks between floats (HPB) smaller than 5 or 6 operating in the Japanese off-shore water with smaller vessels (20–120 mt).

PIFSC Working Paper WP-22-001.
Issued 3 May 2022.
https://doi.org/10.25923/zet2-sk13

Table 2. Results from the NbClust R package described in Charrad et al. (2014). Table 3 lists descriptions of each of these indices.

Cluster #	KL	CH	Hartigan	CCC	Scott	Marriot	TrCovW	TraceW	Friedman	Rubin	Cindex	DB	Silhouette	Duda	Pseudot2
2	0.5059	97.4578	226.1254	18.8175	14376.84	0	384.0278	52.014	1.47E+15	3.2153	0.3653	1.0305	0.2771	0.6172	211.5025
3	1.6909	189.7724	153.3622	28.1708	14300.18	0	137.3704	32.8308	4.59E+14	5.0941	0.2873	1.0914	0.3185	0.5651	153.1362
4	2.7166	227.3111	14.5421	36.9993	14567.41	0	68.1988	23.4957	3.65E+14	7.118	0.332	1.0079	0.3485	0.8164	15.2896
5	0.6824	180.0893	86.1079	30.8643	14621.1	0	65.4057	22.6405	3.46E+14	7.3869	0.3294	0.9816	0.3276	0.5617	109.2242
6	1.6347	193.0954	2.9302	32.8301	15181.61	0	44.8855	18.4936	8.42E+14	9.0433	0.3326	0.9862	0.3208	1.0221	-2.7893
7	1.0729	162.2076	53.6185	28.9032	NaN	0	44.4738	18.3531	-5.80E+14	9.1125	0.2837	0.935	0.2806	0.3968	95.7613
8	1.2952	165.7751	3.317	29.9775	NaN	0	32.1796	16.0941	-5.30E+14	10.3915	0.2753	0.8427	0.3005	1.0063	-0.6569
9	2.5578	146.3456	13.0284	27.4553	NaN	0	31.875	15.9552	-5.12E+14	10.482	0.2745	0.8035	0.2826	0.6315	19.2599
10	0.1247	135.6347	89.0148	30.2029	NaN	0	30.3915	15.4263	-4.35E+14	10.8414	0.2733	0.8275	0.2769	0.5998	85.3927

Cluster #	Beale	Ratkowsky	Ball	Ptbiserial	Frey	McClain	Dunn	Hubert	SDindex	Dindex	SDbw	CritValue_Duda	CritValue_PseudoT2	Fvalue_Beale
2	1.493	0.2697	26.007	0.3159	0.4576	0.1909	0.0451	0.0271	6.8635	0.3354	1.099	0.7317	125.0077	0.2019
3	1.8485	0.3872	10.9436	0.4976	0.2208	0.8699	0.0423	0.0272	6.3513	0.2639	1.0906	0.6983	85.9659	0.1176
4	0.535	0.3994	5.8739	0.5563	0.3063	1.2605	0.0565	0.0313	6.1801	0.2263	0.8796	0.5993	45.4567	0.7102
5	1.8701	0.3617	4.5281	0.5571	0.3333	1.2786	0.0565	0.0328	6.6574	0.2222	0.5434	0.6713	68.5545	0.1142
6	-0.0518	0.3455	3.0823	0.5592	0.3382	1.5209	0.0631	0.0364	6.5956	0.2027	0.4274	0.6643	65.1858	1
7	3.6123	0.3203	2.6219	0.5591	0.395	1.5304	0.0631	0.0365	6.373	0.2017	0.358	0.5902	43.7378	0.007
8	-0.015	0.3069	2.0118	0.556	0.3993	1.6165	0.0631	0.0382	6.3174	0.1858	0.2597	0.6455	57.6609	1
9	1.3676	0.2898	1.7728	0.5556	0.5509	1.6257	0.0631	0.0382	6.1674	0.1848	0.2223	0.4993	33.0904	0.2486
10	1.5981	0.2767	1.5426	0.5544	0.4276	1.6426	0.0631	0.039	6.637	0.1821	0.179	0.6636	64.8773	0.1735

Table 3. Overview of indices implemented in the NbClust package to test for appropriate cluster number.[1,2]

Name of the index in NbClust		Optimal number of clusters
1	"ch" (Calinski and Harabasz 1974)	Maximum value of the index
2	"duda" (Duda and Hart 1973)	Smallest number of clusters such that index > criticalValue
3	"pseudot2" (Duda and Hart 1973)	Smallest number of clusters such that index < criticalValue
4	"cindex" (Hubert and Levin 1976)	Minimum value of the index
5	"gamma" (Baker and Hubert 1975)	Maximum value of the index
6	"beale" (Beale 1969)	Number of clusters such that critical value ≥ alpha
7	"ccc" (Sarle 1983)	Maximum value of the index
8	"ptbiserial" (Milligan 1980, 1981)	Maximum value of the index
9	"gplus" (Rohlf 1974; Milligan 1981)	Minimum value of the index
10	"db" (Davies and Bouldin 1979)	Minimum value of the index
11	"frey" (Frey and Van Groenewoud 1972)	Cluster level before index value < 1.00
12	"hartigan" (Hartigan 1975)	Maximum difference between hierarchy levels of the index
13	"tau" (Rohlf 1974; Milligan 1981)	Maximum value of the index
14	"ratkowsky (Ratkowsky and Lance 1978)	Maximum value of the index
15	"scott" (Scott and Symons 1971)	Maximum difference between hierarchy levels of the index
16	"marriot" (Marriot 1971)	Max. value of second differences between levels of the index
17	"ball" (Ball and Hall 1965)	Maximum difference between hierarchy levels of the index
18	"trcovw" (Milligan and Cooper 1985)	Maximum difference between hierarchy levels of the index
19	"tracew" (Milligan and Cooper 1985)	Max. value of second differences between levels
20	"friedman" (Friedman and Rubin 1967)	Maximum difference between hierarchy levels of the index
21	"mcclain (McClain and Rao 1975)	Minimum value of the index
22	"rubin" (Friedman and Rubin 1967)	Minimum value of second differences between levels
23	"kl" (Krzanowski and Lai 1988)	Maximum value of the index
24	"silhouette" (Rousseeuw 1987)	Maximum value of the index
25	"gap" (Tibshirani et al. 2001)	Smallest number of clusters such that criticalValue ≥ 0
26	"dindex" (Lebart et al. 2000)	Graphical method
27	"dunn" (Dunn 1974)	Maximum value of the index
28	"hubert" (Hubert and Arabie 1985)	Graphical method

[1] Source: Charrad M, Ghazzali N, Boiteau V, Niknafs A. 2014. NbClust: an R package for determining the relevant number of clusters in a data set. Journal of statistical software. 61:1-36.

[2] Gamma is not a retuned result in the current output of NbClust and so is not found in Table 2.

Name of the index in NbClust		Optimal number of clusters
29	“sdindex” (Halkidi et al. 2000)	Minimum value of the index
30	“sdbw” (Halkidi and Vazirgiannis 2001)	Minimum value of the index

PIFSC Working Paper WP-22-001.
Issued 3 May 2022.
https://doi.org/10.25923/zet2-sk13