

Article

Body Length, Tow Conditions, and Guide-Net Angle as Joint Predictive Contributors to Catch Loss in a Marine Mammal Bycatch Reduction Device: An Interpretable Machine Learning Analysis

Kyu-Suk Choi , Jung-Mo Jung and Hyun-Young Kim *

Fisheries Engineering Division, National Institute of Fisheries Science, Busan 46083, Republic of Korea; kysukchoi@korea.kr (K.-S.C.); jjm6538@korea.kr (J.-M.J.)

* Correspondence: sys9318@korea.kr

Abstract

Guide-net marine mammal bycatch reduction devices (BRDs) are intended to reduce protected species interactions in trawl fisheries, but their use may also cause catch loss of commercial fish and cephalopods. We restructured June and September 2024 sea trial records into a final six-taxon BRD/COD modeling dataset containing 5641 individual records from 24 unique haul clusters. Logistic GLM, random forest, and XGBoost models were evaluated using haul-grouped cross-validation, and random forest SHAP and permutation analyses were used as exploratory variable contribution diagnostics. Model discrimination was modest and split-sensitive, but Brier score and accuracy were interpreted cautiously because the catch loss prevalence was 0.163 and trivial baselines were strong. Body length and headline height were the leading non-species contributors, while nominal guide-net angle remained a relevant design variable but ranked below several biological, tow condition, and realized gear geometry variables. Additional haul-clustered GEE sensitivity analyses treated nominal guide-net angle as a two-level categorical factor ($30^\circ/45^\circ$). Across four GEE specifications, the 45° angle term was consistently positive, but confidence intervals were wide, p -values were not significant, and the full species-adjusted model showed suspected quasi-separation in the hairtail term. These results indicate that nominal angle contribution cannot be cleanly isolated from species composition and realized gear geometry in this field dataset. We therefore interpret guide-net angle as an exploratory predictive contributor rather than as a causal factor of catch loss.

Keywords: marine mammal bycatch; bycatch reduction device; guide-net BRD; midwater trawl; catch loss; body length; SHAP; interpretable machine learning



Academic Editor: Junbo Zhang

Received: 9 June 2026

Revised: 13 July 2026

Accepted: 14 July 2026

Published: 16 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Key Contribution: This study provides an individual-level predictive reanalysis of sea trial data from a guide-net marine mammal bycatch reduction device. The revised analyses distinguish individual records from haul-level clusters, compare model performance with trivial baselines, and use haul-clustered GEE sensitivity models to show that nominal guide-net angle is a relevant predictive contributor rather than a standalone causal factor.

1. Introduction

Marine mammal bycatch in fishing gear such as trawls, gillnets, and pots remains a persistent issue in fisheries resource management and ecosystem conservation [1]. Under

the import provisions of the Marine Mammal Protection Act (MMPA), the United States restricts imports of fish and fish products from foreign fisheries that do not meet standards for reducing marine mammal bycatch [2]. In response to the National Oceanic and Atmospheric Administration (NOAA) comparability finding process, Korea has also developed bycatch reduction devices (BRDs) for five fisheries: gillnet, pot, trawl, stow net, and set net fisheries [3]. For stow nets, Lee et al. [4] developed a BRD to reduce bycatch of the finless porpoise (*Neophocaena phocaenoides*) and analyzed species-specific catch loss according to exit direction. That study also emphasized the need to evaluate bycatch reduction and catch loss of commercial species together [4], indicating that balancing protected species exclusion with commercial catch retention is a central challenge in BRD design.

Midwater trawls are designed to operate in the water column above the seabed, but their large net openings and long towing durations make it difficult to completely avoid encounters with non-target organisms, including marine mammals. In Korea, a guide-net BRD was developed by adapting the principle of turtle excluder devices (TEDs) but using a netting panel instead of a rigid metal grid to address crew safety concerns [5]. Sea trials have subsequently quantified catch loss and mesh selectivity associated with two guide-net angles, 30° and 45° [6,7]. Because fishers are unlikely to voluntarily adopt BRDs if reductions in protected species interactions are accompanied by substantial losses of commercial catch, BRD evaluations should consider catch loss as well as bycatch reduction [8]. No marine mammals were captured during the sea trials reanalyzed in the present study; therefore, this study focuses on the effects of a guide-net BRD on catch loss of commercial fish and cephalopods.

Previous studies have analyzed the same sea trial datasets separately by cruise. Jung et al. [6] used the June 2024 dataset to evaluate catch loss and selectivity curves for common flying squid, sailfin sandfish, and pearlside, whereas Jung et al. [7] analyzed the September 2024 dataset to assess catch loss in additional species and the reduction of jellyfish bycatch. Both studies reported that guide-net angle significantly affected catch loss and that under the tested conditions the 30° configuration resulted in lower catch loss than the 45° configuration.

However, BRD performance evaluations centered primarily on guide-net angle have two limitations. First, species-specific logistic selectivity analyses are effective for quantifying length-dependent selectivity [9,10], but they are less suited to comparing the relative importance of multiple predictive contributors, such as body length, tow conditions, and realized gear geometry, within a single unified model. Whether an individual passes through the BRD pathway or reaches the codend can be viewed as the combined outcome of body length, towing speed, gear depth, net opening, and the realized gear configuration during towing, in addition to the nominal guide-net angle [11]. These variables may also act nonlinearly and interact with one another. Second, in field sea trials, species composition, season, and guide-net angle are difficult to allocate independently because of operational constraints. As a result, apparent angle effects may be confounded with differences in species composition and tow conditions.

Rather than repeating the cruise-specific, species-specific catch loss and selectivity analyses of Jung et al. [6,7], the present study restructures the same two sea trial datasets at the individual level and applies an interpretable machine learning framework to compare the relative predictive contributions of body length, tow conditions, realized gear geometry, species structure, and guide-net angle within a single integrated model.

To address these limitations, this study adopted an interpretable machine learning (IML) approach. Machine learning methods have increasingly been applied across fisheries research, from biological and ecological applications to fisheries monitoring and decision support contexts [12]. Tree-based ensemble models, such as random forest and XGBoost,

can capture nonlinear relationships and interactions among predictors without requiring them to be specified a priori [13,14]. SHAP (SHapley Additive exPlanations) and permutation importance further allow model predictions to be decomposed into contributions from individual predictors, enabling quantitative comparison of which factors contribute more strongly to catch loss [15–17]. Based on this framework, this study had two objectives. First, we evaluated the main predictors contributing to individual-level catch loss among guide-net angle, body length, tow conditions, and gear geometry variables. Second, considering the imbalanced distribution between species and guide-net angle conditions, we identified the species most suitable for within-species comparison of guide-net angle effects and assessed the extent to which such comparison was supported by the data.

2. Materials and Methods

2.1. Sea Trials and Data Structure

This study reanalyzed sea trial datasets previously reported by Jung et al. [6,7], which were collected in coastal waters of the East Sea during two periods: 10–21 June 2024 and 2–13 September 2024. A guide-net marine mammal BRD was installed in front of the codend of a midwater trawl, and the nominal guide-net angle set on deck was configured at two levels, 30° and 45°. In both cruises, the two angle conditions were allocated in approximately equal proportions; the 30° condition was tested first in June, whereas the 45° condition was tested first in September. Detailed haul-level allocation is provided in Supplementary Table S1. The June cruise comprised 17 hauls in total, including 1 initial trial haul that was not counted among the 16 hauls reported by Jung et al. [6]; this trial haul was retained in the present analysis because valid BRD/COD measurements were available. Of the 34 hauls listed in Supplementary Tables S1 and S3 were excluded from the haul-level operational summaries linked to the analytical catch records: J07 and J08, because the catch sheets contained only box/basket aggregate rows and yielded no directly measured individual records; and S17, an auxiliary camera-validation haul for which no catch-measurement sheet was available. This yielded the 31 valid hauls used for the haul-level operational summaries. No marine mammals were captured during either of the two cruises reanalyzed in this study.

This study used two primary data sources from the sea trials: haul logs and catch records. The haul logs contained haul-level variables, including date, time, coordinates, warp length, gear depth measured at the headrope, headline height, otter-board spread, wing-end spread, speed over ground (SOG), speed through water, and seawater temperature. Among these variables, five variables that had relatively low rates of missing or unmeasured values and were directly related to BRD operation were used as model predictors: gear depth, headline height, otter-board spread, SOG, and guide-net angle. Variables recorded at multiple time points within a haul were averaged over the towing period, from completion of net setting to immediately before hauling, and used as haul-level representative values. Coordinates recorded in degrees–minutes–seconds format were converted to decimal degrees. Coordinates from two hauls with minute or second values of 60 or greater were treated as input errors and set to missing.

The catch records contained individual-level measurements for each haul, including species name, catch location, body length, maximum girth, and body weight. Body length was defined using the length measurement appropriate for each taxon, including fork length, total length, anal length, mantle length, or carapace width. For the integrated analysis, body length was used as a common variable name for the taxon-appropriate length metric rather than as a uniform total length measurement across all taxa. Catch location was recorded as BRD, COD, or COV. Individuals collected at the BRD location were classified as catch loss, representing individuals guided out of the gear through the

BRD pathway, whereas individuals collected in the COD location were classified as codend retention. COV records represented covernet samples collected in some June hauls; these data were used for descriptive review but were excluded from the binary BRD/COD classification models.

2.2. Data Preprocessing and Variable Definition

The haul logs and catch records were integrated to construct two analytical datasets. The haul-level dataset was compiled from the haul logs and included cruise, date, guide-net angle, towing period averages of gear depth, headline height, otter-board spread, and SOG, as well as available temperature records. The individual-level dataset was created by combining individual measurements distributed across multiple sheets of the catch records into a single table. Sheets were linked to hauls using date information in the sheet headers and the haul sequence recorded in the haul logs. Rows that did not represent individual measurements, such as box-level aggregate rows after hauling or summary labels, were excluded.

Previous studies estimated catch numbers and catch loss using expanded individual counts based on box- or basket-level subsampling. In contrast, the machine learning analysis in the present study was restricted to directly measured individual-level records for which reliable BRD/COD labels could be assigned. Accordingly, aggregate estimated records that could not be linked reliably to individual catch location and body length information were excluded from the analysis. When the same species was recorded under different Korean names in the field records, these entries were consolidated under a standardized species name. The full harmonization list is provided in Supplementary Table S2.

For hairtail (*Trichiurus lepturus*), anal length (AL) was used as the body length variable to maintain consistency with the previous selectivity analysis [7]. When both AL and total length (TL) were recorded for the same individual, AL was used as the body length metric for analysis, whereas TL was retained only as an auxiliary total length record. Hairtail records for which the distinction between AL and TL was unclear were excluded from the final machine learning dataset. As a result, 92 hairtail individuals were retained for analysis, with AL used as the single body length metric.

After extraction and standardization, the integrated individual-level dataset contained 6372 individuals from 22 species. Among these, six taxa with sufficient samples in BRD and COD locations were selected for modeling. Although 31 valid hauls were available for broader haul-level operational summaries, the final six-taxon BRD/COD modeling dataset contained 5641 individual records from 24 unique haul clusters after taxon selection, exclusion of COV records, and filtering. Before outlier removal, the modeling subset contained 5674 individuals (Table 1).

Table 1. Data flow from raw data extraction to the final modeling dataset.

Stage	Number of Individuals
Integrated extraction	6372
Machine learning target and BRD/COD subset	5674
Removal of biologically implausible body length records	33
Final modeling dataset	5641

Note: Data flow was preserved as 6372 individuals extracted, 5674 machine learning target and BRD/COD subset, 33 biologically implausible body length records removed, and 5641 individuals in the final modeling dataset. The machine learning subset comprised binary BRD/COD records for six target taxa; COV, non-target, ambiguous, and biologically implausible records were excluded before modeling.

To remove body length records suspected to reflect input errors or unit conversion errors, we applied the intersection of two criteria. The statistical criterion was the species-specific interquartile range rule, defined as $Q1 - 1.5 \times IQR$ to $Q3 + 1.5 \times IQR$. The biological criterion was a plausible body length range defined from the observed species-specific length distributions and known biological size ranges: 1–10 cm for pearlside, 1–15 cm for firefly squid, 1–30 cm for common flying squid, 3–30 cm for sailfin sandfish, 5–40 cm AL for hairtail, and 5–30 cm for Pacific herring. Records were removed as outliers only when they violated both criteria. This approach was used to reduce the risk of excluding biologically normal individuals at the extremes of natural variation when applying the statistical criterion alone. In addition, one common flying squid individual smaller than 1 cm was manually removed because it was considered a likely unit entry error. In total, 33 individuals were removed: 28 pearlside, 4 firefly squid, and 1 common flying squid. Removed records were retained separately in an outlier log (Table 1).

The response variable was defined as a binary BRD/COD label: $Y = 1$ for BRD catch loss and $Y = 0$ for COD retained catch. Nominal guide-net angle was treated as a two-level categorical factor ($30^\circ / 45^\circ$), not as a continuous per-degree predictor. Realized gear geometry was represented by headline height and otter-board spread, while gear depth and speed over ground (SOG) represented tow condition variables. Species indicators were included only as taxon-level control variables in selected models and were not interpreted as modifiable BRD performance variables.

2.3. Model Development, Validation, and Interpretation

Three classification models were compared. Logistic GLM was used as an interpretable baseline model linked to conventional logistic selectivity analysis. Random forest was included as a tree-based ensemble model capable of capturing nonlinear relationships and interactions among predictors without prior specification [13], and XGBoost was included as a gradient-boosting-based comparator [14]. Although random forest did not achieve the highest predictive performance among the three models, it was used as a tree-based interpretation model because it can capture nonlinearity and predictor interactions while providing stable estimates of SHAP TreeExplainer values and permutation importance. Therefore, the random-forest-based variable importance results in this study were interpreted not as an explanation of the best-performing predictive model but as interpretable machine learning results used to explore the structure of predictor contributions in the integrated dataset.

The analysis workflow was implemented in Python 3.13 [18] using pandas 2.3.1 and NumPy 2.2.6 for data handling [19,20], statsmodels 0.14.6 for the GLM and GEE analyses [21], scikit-learn 1.7.1 for machine-learning model fitting and evaluation [22], XGBoost 3.2.0 for gradient boosting [14], and SHAP 0.48.0 for model interpretation [16]. SOG missingness was reported at two levels. At the haul summary level, valid SOG values were available for 27 of the 31 valid hauls. At the individual record level used for modeling, SOG was missing for 1402 of 5641 records (24.85%) because each haul-level SOG value was linked to multiple individual catch records. Median imputation using 3.2 kt was used to retain SOG as a tow condition predictor while avoiding further loss of individual BRD/COD records, and SOG-related effects were interpreted cautiously. Random forest and XGBoost followed the model-specific missing value handling described in the analysis scripts.

All models were evaluated using haul-grouped cross-validation to avoid leakage of haul-level tow conditions, following recommendations for structured cross-validation [23]. For grouped cross-validation and GEE clustering, haul_id was defined as a unique cruise–haul identifier rather than as the within-cruise haul sequence alone. Model performance

was evaluated primarily using AUC [24], while Brier score and accuracy were interpreted against trivial baselines [25–27] because catch loss prevalence was 0.163. The prevalence-based constant forecast had a Brier score of 0.1366, and the always-retained baseline had an accuracy of 0.8367. Repeated haul-grouped cross-validation was also used as a stability check for the random forest model. For reproducibility, the random forest used for cross-validated model performance comparison was fitted with 300 trees and unrestricted depth, the random forest interpretation model used for the SHAP and permutation-importance analyses used 200 trees with a maximum depth of 10, and the random forest sensitivity models used 300 trees with unrestricted depth.

SHAP values and permutation importance were calculated for the random forest interpretation model as exploratory variable contribution diagnostics [15–17]. These analyses were not interpreted as causal effects of individual predictors. Directional patterns in SHAP dependence plots were interpreted as model-specific conditional patterns and compared with marginal and GEE-adjusted directions in a separate consistency check.

To examine how species–angle imbalance affected interpretation, four haul-clustered GEE logistic specifications were fitted using haul_id as the clustering unit: Spec_full included species indicators and realized gear geometry; Spec_nospec excluded species indicators but retained realized gear geometry; Spec_nogeom retained species indicators but excluded headline height and otter-board spread; and Spec_nospec_nogeom excluded both species indicators and realized gear geometry. GEE models used a binomial family and logit link [28] with an independence working correlation and bias-reduced robust covariance [29,30]. An independence working correlation was used so that clustering was accommodated through bias-reduced robust (sandwich) covariance rather than through a modeled correlation structure. Because the final modeling dataset contained only 24 haul clusters, the resulting standard errors and confidence intervals were interpreted as approximate cluster-aware sensitivity estimates rather than definitive inference. Because the full species-adjusted GEE showed suspected quasi-separation in the hairtail term, the same full specification was also refitted after excluding hairtail records as a robustness check.

3. Results

3.1. Analytical Dataset and Species–Angle Imbalance

The final six-taxon BRD/COD modeling dataset consisted of 5641 individual records from 24 unique haul clusters (Table 2). This cluster count differs from the 31 valid hauls available for broader haul-level operational summaries because the modeling dataset was restricted to six target taxa, BRD/COD records, and filtered individual measurements. Species and nominal guide-net angle were strongly imbalanced, limiting causal interpretation of angle effects across the integrated dataset. The species-specific distributions of BRD catch loss and codend retention are shown in Figure 1.

Table 2. Summary of target taxa in the final modeling dataset.

Species	Scientific Name	<i>n</i>	BRD	COD	Catch Loss Rate	30° <i>n</i>	45° <i>n</i>
Sailfin sandfish	<i>Arctoscopus japonicus</i>	3515	454	3061	0.129	2611	904
Pearlsides	<i>Maurolicus muelleri</i>	1056	239	817	0.226	1	1055
Common flying squid	<i>Todarodes pacificus</i>	477	107	370	0.224	155	322
Firefly squid	<i>Watasenia scintillans</i>	259	17	242	0.066	7	252
Pacific herring	<i>Clupea pallasii</i>	242	48	194	0.198	0	242
Hairtail	<i>Trichiurus lepturus</i>	92	56	36	0.609	90	2
Total	—	5641	921	4720	0.163	2864	2777

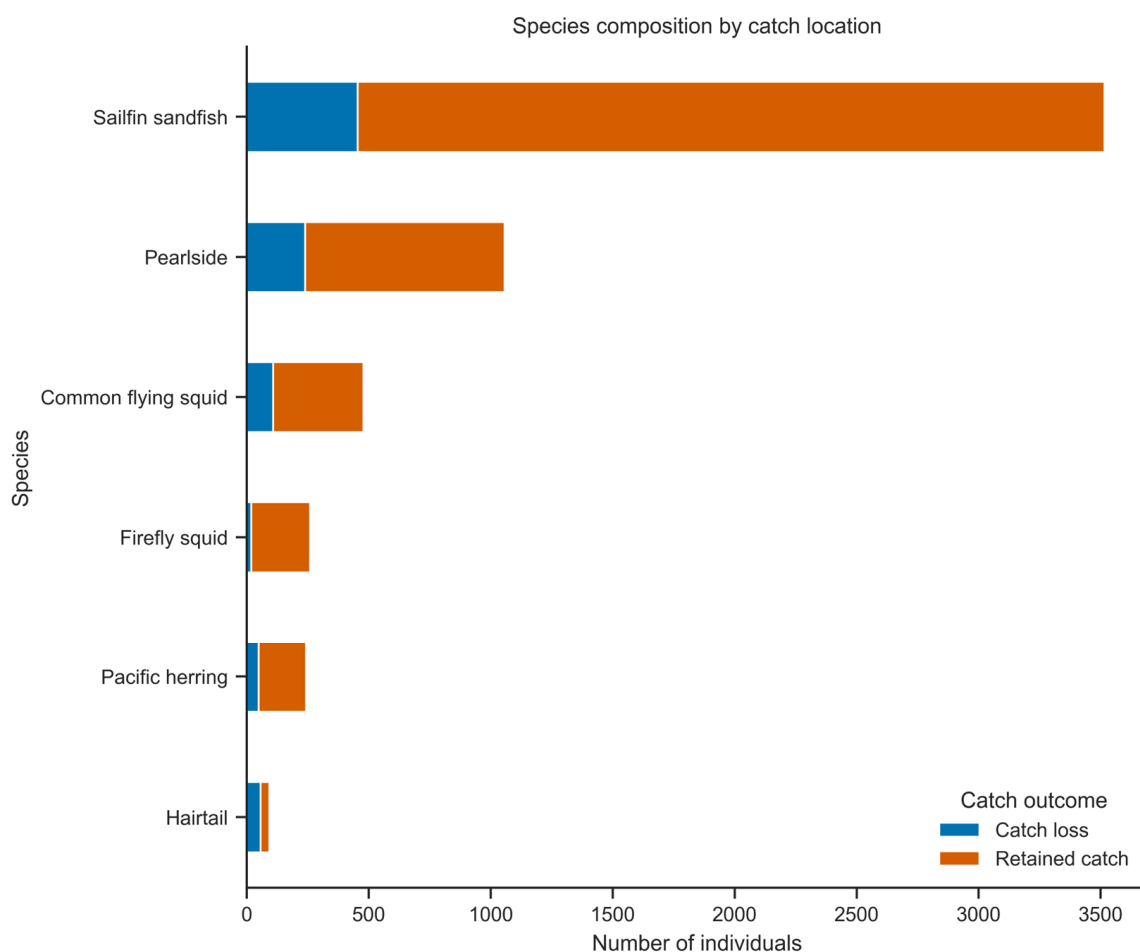


Figure 1. Species $\times$ catch location stacked bar plot showing the numbers of individuals classified as catch loss (BRD) and codend retention (COD) for each target taxon. COV (covernet) records were excluded from binary BRD/COD modeling and are therefore not shown.

Sample distribution across BRD guide-net angle conditions showed strong species-specific imbalance (Table 2). Sailfin sandfish was the only taxon with substantial sample sizes under both angle conditions (30° : $n = 2611$; 45° : $n = 904$), whereas the other five taxa were concentrated almost entirely in one angle condition. Approximately 98% of hairtail individuals occurred under the 30° condition, while all Pacific herring individuals, more than 99% of pearlsides individuals, and approximately 97% of firefly squid individuals occurred under the 45° condition. Common flying squid was also more frequent under the 45° condition, with approximately 67% of individuals observed at 45° . In particular, Pacific herring was observed only under the 45° condition, and only one pearlside individual was observed under the 30° condition, making it difficult to isolate guide-net angle effects for these taxa. This distribution indicates varying degrees of imbalance among species occurrence, cruise timing, fishing ground, and guide-net angle. As shown in Figure 2, several species $\times$ guide-net angle $\times$ catch outcome cells were sparse or empty, limiting data support for isolating within-species guide-net angle effects for most taxa. Therefore, except for sailfin sandfish, guide-net angle effects were not interpreted causally at the species level and were considered only in terms of predictive contribution.

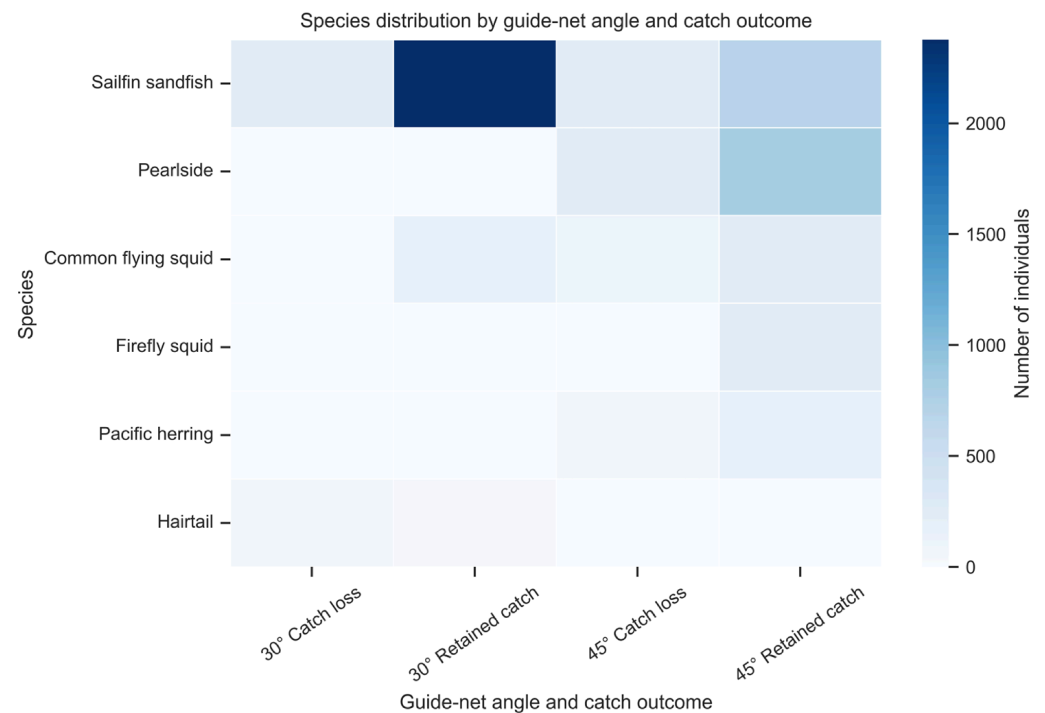


Figure 2. Heatmap showing the distribution of species $\times$ guide-net angle ($30^\circ/45^\circ$) $\times$ catch outcome (BRD/COD). Sparse or empty cells indicate limited data support for interpreting guide-net angle effects within the corresponding species.

This imbalance was also evident in species- and angle-specific catch loss rates (Figure 3).

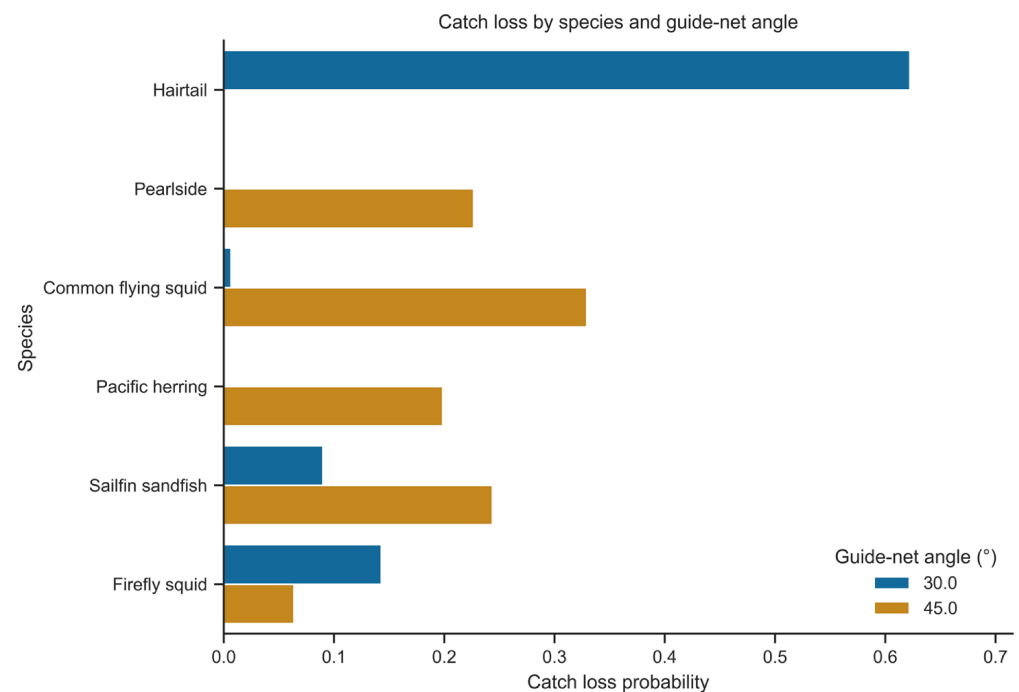


Figure 3. Catch loss rates by species and guide-net angle. Catch-loss rates in cells with small sample sizes should be interpreted cautiously because of high sampling variability.

Tow and gear geometry variables also varied substantially across the valid hauls (Table 3). Gear depth ranged from 14.53 to 146.75 m, headline height from 8.00 to 38.36 m, otter-board spread from 27.90 to 65.28 m, and SOG from 3.10 to 5.20 kt. Temperature variables were excluded from the main models because of high rates of missing or unmeasured values. Comparison of predictor distributions by catch outcome indicated that BRD catch loss individuals tended to be associated with relatively shallow gear depth and higher towing speed, whereas COD-retained individuals tended to be associated with relatively deeper gear depth (Supplementary Figure S1).

Table 3. Summary of haul-level operational and environmental variables.

Variable	<i>n</i> Haul	Mean	SD	Min	Max
Gear depth (m)	31	67.43	41.97	14.53	146.75
Headline height (m)	31	23.89	9.68	8.00	38.36
Otter-board spread (m)	31	46.89	12.07	27.90	65.28
Speed over ground (kt)	27	4.01	0.55	3.10	5.20
Surface temperature (°C)	22	19.92	9.30	2.20	29.00
Bottom temperature (°C)	12	4.93	4.02	1.40	15.80

Note: Table 3 summarizes the 31 valid hauls used for operational summaries; the final six-taxon BRD/COD modeling dataset contained 24 unique haul clusters after taxon selection, exclusion of COV records, and filtering.

3.2. Model Performance and Variable Importance

The primary haul-grouped stratified five-fold cross-validation was used for cross-model comparison (Table 4; Figure 4). Logistic GLM achieved the highest pooled OOF AUC among the three models (0.717), random forest had a pooled OOF AUC of 0.672, and XGBoost had a pooled OOF AUC of 0.703 and the lowest raw Brier score (0.162). As a stability check rather than a cross-model comparison metric, repeated haul-grouped cross-validation was additionally conducted for the random forest model and yielded a pooled AUC mean of 0.649 (SD = 0.060; range = 0.537–0.724), confirming that discrimination was sensitive to the limited number and composition of haul clusters. Because catch loss prevalence was 0.163, the prevalence baseline Brier score was 0.1366 and the always-retained baseline accuracy was 0.8367. Therefore, Brier score and accuracy were treated as diagnostic metrics rather than evidence of model superiority over trivial baselines. Post hoc isotonic calibration reduced apparent Brier scores (GLM = 0.123, random forest = 0.127, XGBoost = 0.126), but this was treated as an optimistic diagnostic based on pooled OOF predictions and not as a primary performance estimate.

Table 4. Model performance in the primary haul-grouped stratified 5-fold cross-validation, with repeated haul-grouped cross-validation reported as a random forest stability check.

Model	Pooled OOF AUC	Brier Score	Accuracy	Mean fold AUC (Primary Split)	Fold AUC SD (Primary Split)
Logistic GLM	0.717	0.170	0.769	0.598	0.232
Random forest	0.672	0.176	0.756	0.607	0.187
XGBoost	0.703	0.162	0.753	0.635	0.205

Note: Logistic GLM: highest pooled OOF AUC; quasi-separation in some species indicators; random forest: tree-based interpretation model used for SHAP and permutation importance; XGBoost: lowest raw Brier score among fitted models; Brier score and accuracy interpreted diagnostically because trivial baselines were stronger. The primary 5-fold split was used for cross-model comparison. Repeated haul-grouped cross-validation was additionally conducted for the random forest model as a stability check and yielded a pooled AUC mean of 0.649 (SD = 0.060; range = 0.537–0.724), indicating split-sensitive discrimination.

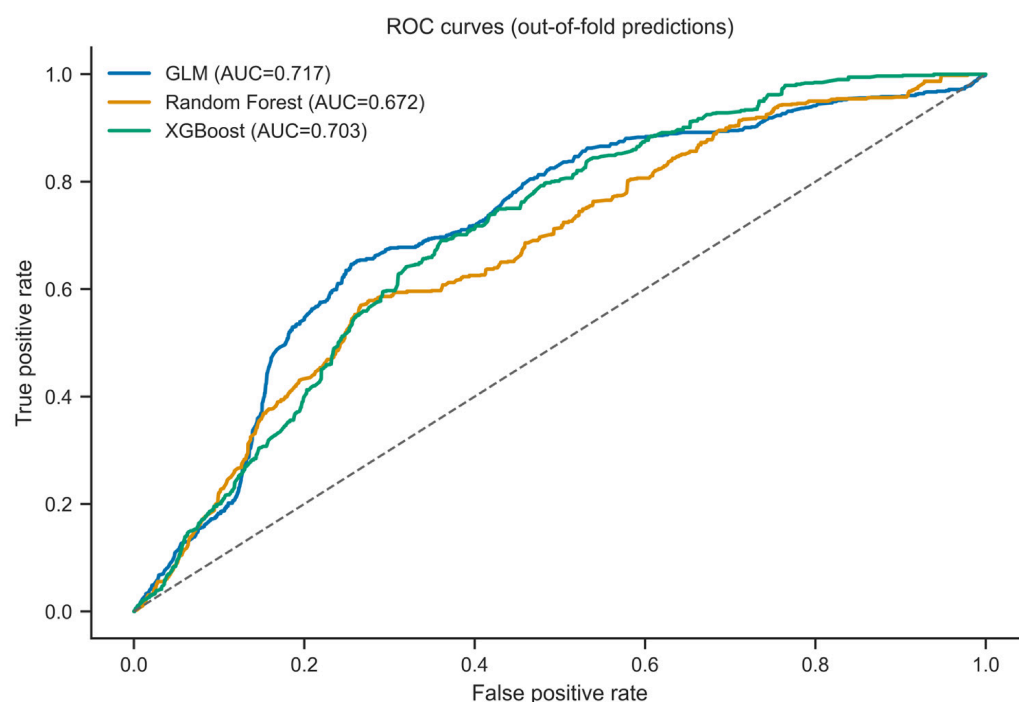


Figure 4. Pooled out-of-fold ROC curves for the GLM, random forest, and XGBoost models based on haul-grouped stratified 5-fold cross-validation. The dashed line indicates the random classification reference line (AUC = 0.5).

The logistic GLM and species-adjusted GEE analyses indicated instability in species terms because several taxa were concentrated within specific nominal guide-net angle conditions. In the four-specification GEE sensitivity analysis, the full species-adjusted model showed suspected quasi-separation in the hairtail term. Consequently, species-adjusted angle estimates were interpreted cautiously, and no single GEE specification was treated as a definitive primary inference model.

In the global SHAP analysis of the random forest model, body length was identified as the predictor with the largest exploratory contribution (Figure 5; Table 5). The highest mean absolute SHAP value was observed for body length (0.090), followed by headline height (0.075), SOG (0.046), otter-board spread (0.045), gear depth (0.044), and nominal guide-net angle (0.035). These values should be interpreted as model-based variable contribution diagnostics, not as causal effect sizes.

Table 5. Global SHAP importance of non-species predictors from the random forest interpretation model.

Rank	Variable	Mean Absolute SHAP
1	Body length (cm)	0.090
2	Headline height (m)	0.075
3	Speed over ground (SOG, kt)	0.046
4	Otter-board spread (m)	0.045
5	Gear depth (m)	0.044
6	Nominal guide-net angle (30°/45°)	0.035

Species indicators were included in the fitted model as taxon-level control variables but are not shown in this main text table because they are not environmental, operational, or gear design variables. Their importance values are provided in Supplementary Table S4.

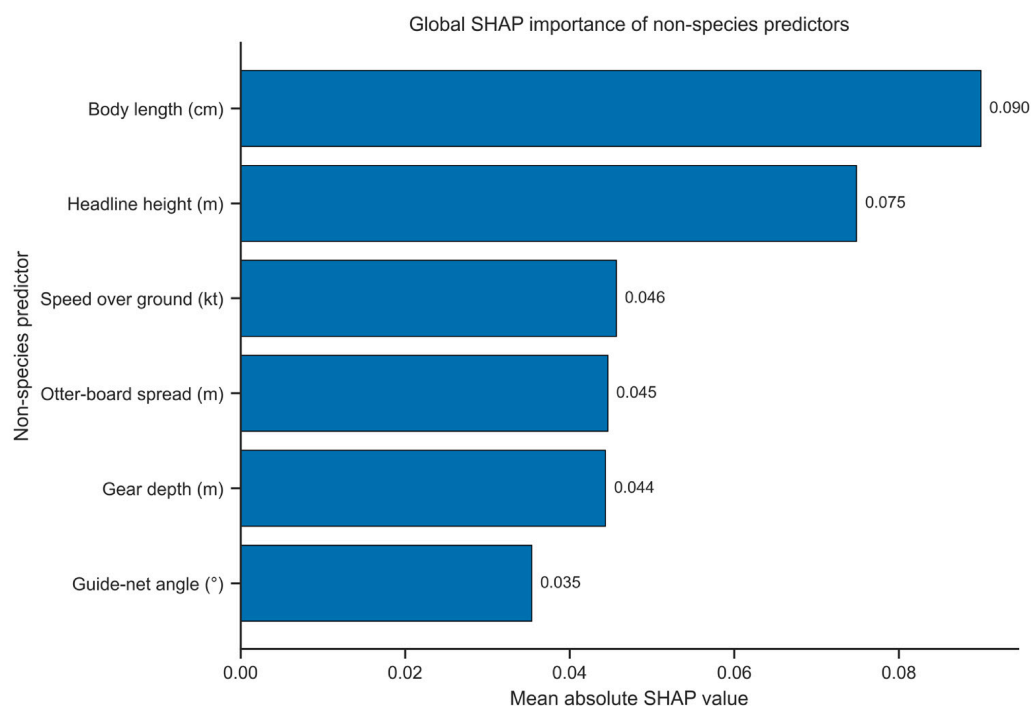


Figure 5. Global SHAP importance of non-species predictors from the random forest interpretation model, based on mean absolute SHAP values across all individuals in the final modeling dataset ($n = 5641$). Species indicators were included in the fitted model as taxon-level control variables but are reported separately in Supplementary Table S4.

Permutation importance also identified body length as the most important variable and headline height as the second most important variable (Table 6; Figure 6). The AUC decrease for body length was 0.134, approximately 2.4 times larger than that for headline height (0.055). Among non-species predictors, the remaining variables were ranked as follows: SOG (0.020), gear depth (0.016), otter-board spread (0.011), and guide-net angle (0.009). In the full fitted model, some species indicators also contributed to prediction, reflecting taxon-level baseline structure; their importance values are provided in Supplementary Table S4 and were not interpreted as modifiable BRD performance variables. The two importance methods were consistent in identifying body length and headline height as the top two predictors, and guide-net angle ranked below body length, headline height, and several towing or gear geometry variables in both analyses.

Table 6. Permutation importance of non-species predictors based on AUC decrease.

Rank	Variable	Mean AUC Decrease
1	Body length (cm)	0.134
2	Headline height (m)	0.055
3	Speed over ground (SOG, kt)	0.020
4	Gear depth (m)	0.016
5	Otter-board spread (m)	0.011
6	Nominal guide-net angle (30°/45°)	0.009

Species indicators were included in the fitted model as taxon-level control variables but are not shown in this main text table because they are not environmental, operational, or gear design variables. Their importance values are provided in Supplementary Table S4.

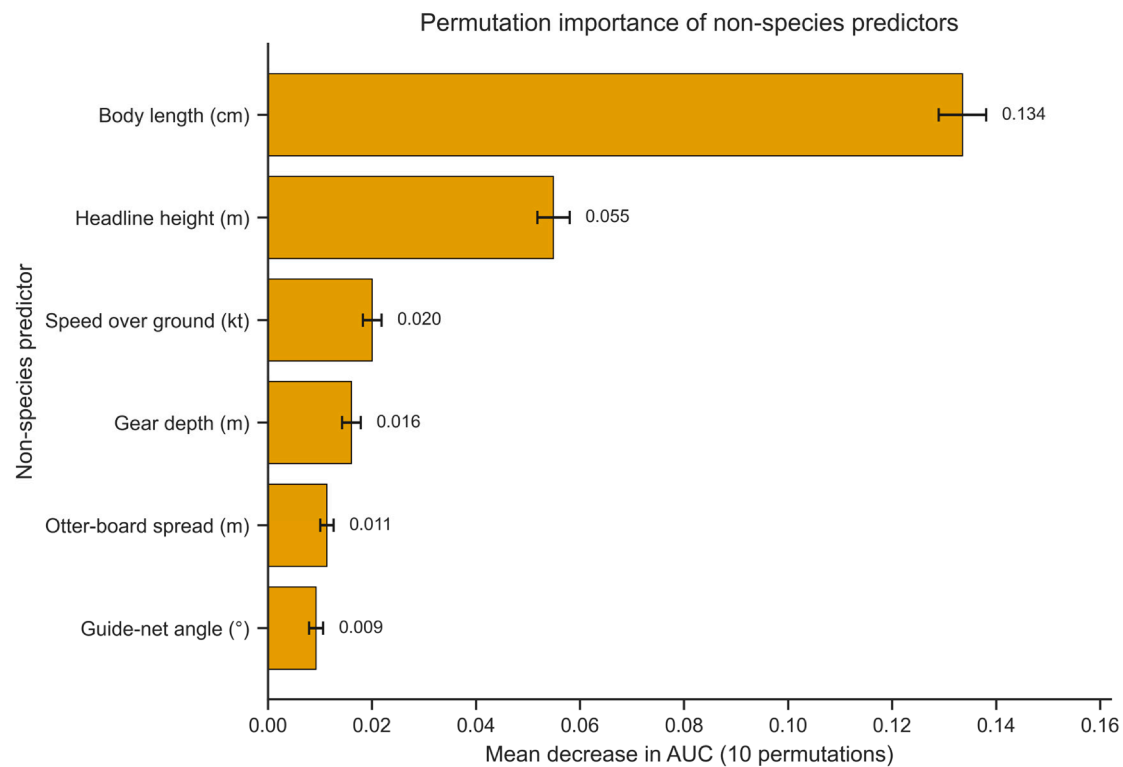


Figure 6. Permutation importance of non-species predictors from the random forest interpretation model based on the decrease in AUC after variable permutation. Species indicators were included in the fitted model as taxon-level control variables but are reported separately in Supplementary Table S4.

3.3. SHAP-Based Effect Interpretation and Species-Specific Patterns

The nonlinear effects of the main predictors were examined using SHAP dependence plots (Figure 7). A directional consistency check showed marginal-to-conditional reversals for body length and gear depth and model-specific directional differences for otter-board spread and SOG. These differences were interpreted as consequences of species composition, haul-level covariation, variable correlation, and nonlinear model structure rather than contradictions among analyses. Accordingly, the SHAP dependence plots were treated as exploratory model-specific diagnostics rather than estimates of isolated directional effects.

Species-specific mean absolute SHAP values showed that body length and headline height were among the top-ranked variables for most taxa, although the highest-ranked variable differed among taxa (Figure 8). Body length showed the largest contribution in sailfin sandfish, Pacific herring, and firefly squid; for example, the mean absolute SHAP value for body length was 0.173 in firefly squid and 0.079 in both sailfin sandfish and Pacific herring. In pearlside and common flying squid, headline height contributed similarly to or more strongly than body length: 0.120 vs. 0.110 in pearlside and 0.101 vs. 0.099 in common flying squid. In hairtail, SOG had the largest contribution (0.140 vs. 0.048 for body length). However, because the hairtail subset was small ($n = 92$) and strongly concentrated in the 30° condition, this result was interpreted as a descriptive outcome reflecting the distributional structure of the hairtail data. Thus, although body length was the largest contributor at the integrated dataset level, it was not the top-ranked variable in every taxon; depending on the species, headline height or SOG contributed similarly to or more strongly than body length.

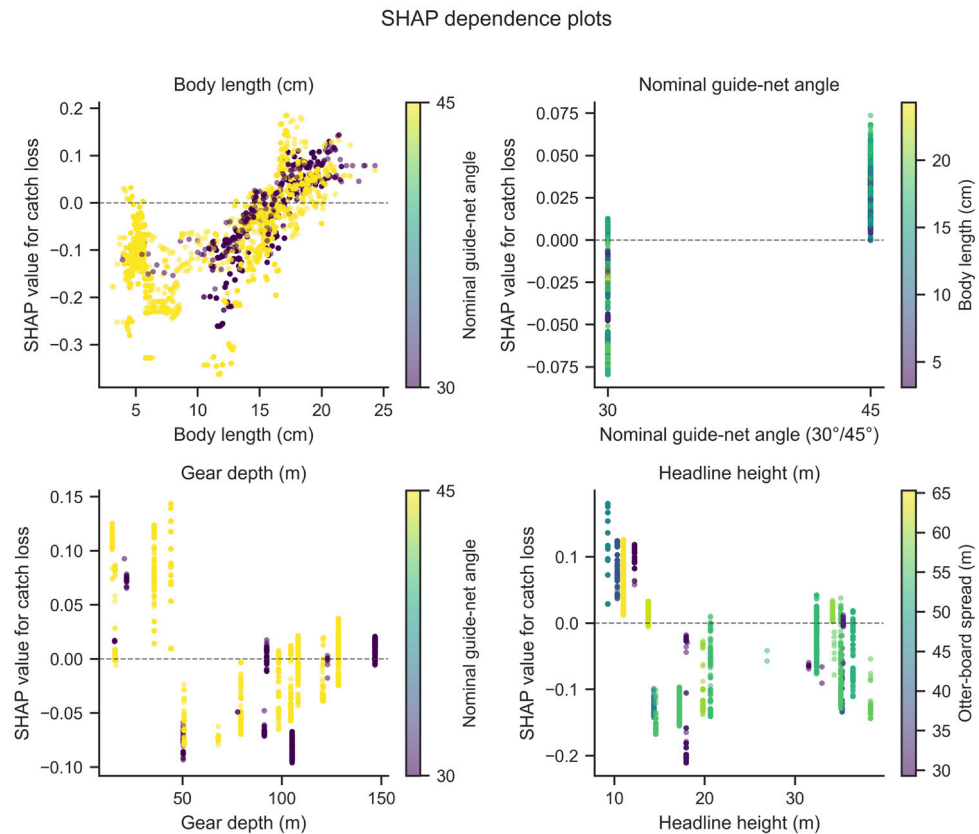


Figure 7. SHAP dependence plots for major predictors, including body length, nominal guide-net angle, gear depth, and headline height. For hairtail, body length was based on anal length (AL).

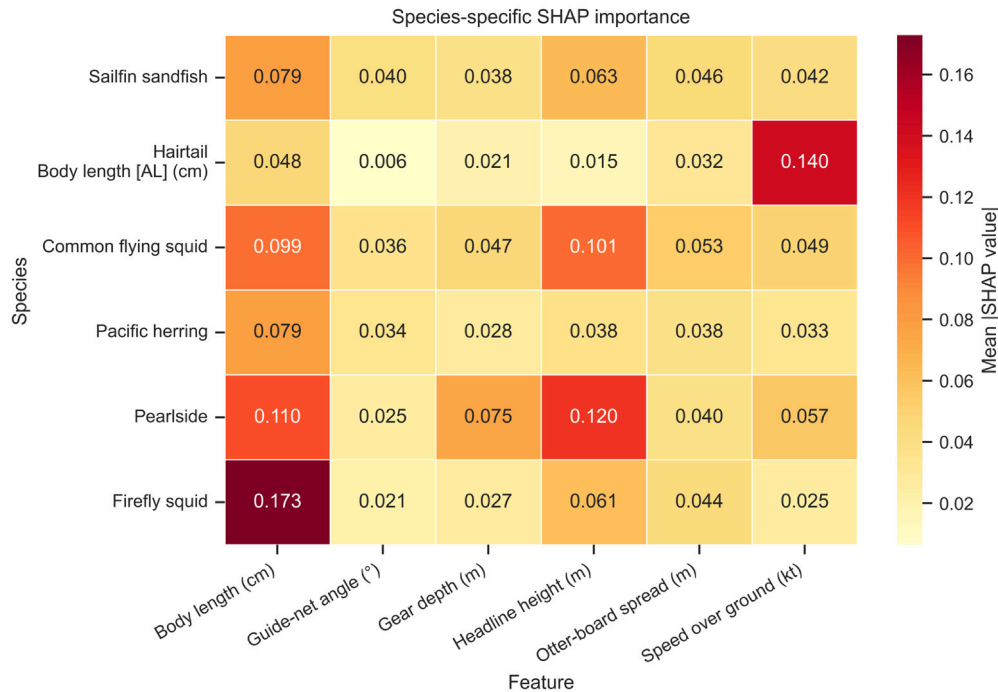


Figure 8. Species-specific heatmap summarizing mean absolute SHAP values for the six target taxa. For hairtail, anal length (AL) was used as the body length variable; therefore, the corresponding label in the figure is shown as body length [AL] (cm).

The additional GEE sensitivity analyses showed that the nominal 45° guide-net angle term was positive in all four specifications but highly uncertain. The 45° odds ratios were 15.835 in Spec_full (95% CI = 0.150–1672.429, $p = 0.2453$), 3.330 in Spec_nospec (95% CI = 0.070–158.022, $p = 0.5413$), 2.826 in Spec_nogeom (95% CI = 0.100–79.747, $p = 0.5421$), and 2.366 in Spec_nospec_nogeom (95% CI = 0.189–29.595, $p = 0.5039$). Spec_full showed suspected quasi-separation in the hairtail species term, and all confidence intervals were wide. Removing realized gear geometry reduced the angle OR from 3.330 to 2.366. Thus, this sensitivity analysis did not support the interpretation that realized gear geometry had masked a stronger nominal angle association. As an additional robustness check for the suspected quasi-separation in the hairtail term, the full species-adjusted GEE specification was refitted after excluding hairtail records. The 45° term remained positive but imprecise (OR = 21.750, 95% CI = 0.208–2270.963, $p = 0.1941$), with no separation warning, and the interpretation that nominal guide-net angle could not be isolated as a causal effect was unchanged.

In the sailfin sandfish subset, which had the largest within-species sample sizes under both angle conditions ($n = 3515$; 30°: $n = 2611$; 45°: $n = 904$), the raw catch loss rate was higher under the 45° condition than under the 30° condition: 24.34% vs. 8.96%. These within-species raw catch loss rates differ slightly from those reported by Jung et al. [6] because the present analysis was restricted to directly measured individual-level BRD/COD records, whereas the previous cruise-specific analysis used expanded catch estimates based on box- or basket-level subsampling. The sailfin sandfish-only random forest model had a pooled OOF AUC of 0.559 and a Brier score of 0.168. For guide-net angle, the mean absolute SHAP value was 0.036, and the AUC decrease based on permutation importance was 0.015. Within the model, guide-net angle ranked fifth by SHAP and fourth by permutation importance. Although the sailfin sandfish subset provided the most direct within-species comparison of the two guide-net angles in the present dataset, the pooled OOF AUC of 0.559 indicated limited predictive separation, and some validation folds may have had small sample sizes. Therefore, the sailfin sandfish subset results were interpreted as a supplementary check of whether the raw difference in catch loss rates was also reflected in model-based variable importance rather than as strong causal evidence for a standalone angle effect.

The random forest with species indicators reported in Table 7 was refitted within the sensitivity analysis pipeline using the same predictor structure as the main interpretation model but with the main random forest settings ($n_{\text{estimators}} = 300$, unrestricted tree depth) rather than the constrained settings used for SHAP and permutation importance in Tables 5 and 6 ($n_{\text{estimators}} = 200$, $\text{max_depth} = 10$). The SHAP and permutation importance values therefore differ slightly from the corresponding values in Tables 5 and 6.

Table 7. Sensitivity analysis results for species–angle imbalance.

Model	<i>n</i>	Pooled OOF AUC	Brier Score	Guide-Net Angle Mean Absolute SHAP	Guide-Net Angle AUC Decrease
RF full with species	5641	0.672	0.176	0.029	0.007
RF without species	5641	0.609	0.215	0.024	0.007
RF sandfish only	3515	0.559	0.168	0.036	0.015

Note: RF full with species: main model including species indicators as control variables; RF without species: species indicators removed; RF sandfish only: single-species subset with the largest within-species samples under both angle conditions.

4. Discussion

4.1. Catch Loss Drivers and Interpretation of Guide-Net Angle Effects

In this study, nominal guide-net angle contributed to catch loss prediction to some extent, but it was not a dominant variable and should not be interpreted as a causal effect. Both SHAP and permutation importance ranked body length first and headline height second, while SOG, gear depth, and otter-board spread showed contributions similar to or greater than that of nominal guide-net angle. The nominal angle contribution cannot be cleanly isolated from species composition and realized gear geometry in this field dataset.

Body length was identified as the highest-ranked variable by both SHAP and permutation importance. This result can be interpreted as reflecting the influence of fish and cephalopod body size on multiple processes leading to catch loss, including the probability of contact with the guide net, the likelihood of passing through mesh openings, and swimming ability during towing [9,10,31]. The monotonic pattern observed in the SHAP dependence analysis, in which smaller individuals tended to reduce the predicted probability of catch loss and larger individuals tended to increase it, is consistent with this interpretation. However, in the present dataset, larger individuals were strongly associated with hairtail-dominated hauls under the 30° condition; the body length contribution should therefore be interpreted as partially confounded with species composition and operational context, rather than as a pure size selectivity effect. Nevertheless, the consistent ranking of body length as the most important predictor agrees with the principle of selectivity theory that the physical relationship between effective gear opening and body size is a major factor influencing catch loss [9,11].

Headline height ranked second in both SHAP and permutation importance analyses, suggesting that realized gear geometry contributed meaningfully to catch loss prediction. However, this should not be read as evidence that realized gear geometry masked or absorbed a simple nominal angle effect. Even when headline height and otter-board spread were excluded from species-excluded GEE sensitivity models, the nominal 45° angle estimate remained positive but uncertain and decreased in magnitude.

Gear depth and SOG also contributed to prediction, but their directions differed across marginal, GEE-adjusted, and SHAP-based summaries. These differences likely reflect species composition, haul-level covariation, variable correlation, and nonlinear model structure. Therefore, gear depth, SOG, and otter-board spread should be described as predictive contributors under the observed multivariate data structure rather than as variables with single isolated directional effects.

Because of the imbalance between species and nominal angle distributions, it is difficult to interpret guide-net angle effects causally for all taxa in the present dataset. Under this structure, the nominal guide-net angle term cannot be fully separated from cruise timing, fishing ground conditions, species occurrence patterns, and realized gear geometry.

Sailfin sandfish was the only taxon with sufficient samples under both angle conditions and was therefore considered the most suitable species for within-species comparison of guide-net angle effects in the present dataset. In the raw data, the catch loss rate for sailfin sandfish was 8.96% under the 30° condition and 24.34% under the 45° condition, showing a difference between the two angle conditions in the same direction as reported in previous studies. These within-species raw catch loss rates differ slightly from those reported by Jung et al. [6] because the present analysis was restricted to directly measured individual-level BRD/COD records, whereas the previous cruise-specific analysis used expanded catch estimates based on box- or basket-level subsampling. However, the sailfin-sandfish-only random forest model had a pooled OOF AUC of 0.559, only slightly above random classification, and guide-net angle ranked fifth by SHAP and fourth by permutation importance within the model. Therefore, although the sailfin sandfish subset

provides the most direct data structure for examining guide-net angle effects in the present dataset, it should not be interpreted as strong causal evidence for a standalone angle effect. The observed angle-specific difference in catch loss rates for sailfin sandfish is more appropriately interpreted as a multivariate outcome under actual sea-trial conditions rather than an isolated angle effect under fully controlled experimental allocation.

The sensitivity analyses indicate that nominal guide-net angle should not be regarded as negligible; rather, it should be interpreted as a relevant design variable whose predictive contribution is entangled with species–angle structure and realized gear geometry. The GEE specifications did not provide statistically stable evidence for a causal nominal angle effect, and the full species-adjusted model was affected by suspected quasi-separation.

4.2. Implications for BRD Design and Operation, Limitations, and Future Research

Taken together, these results indicate that catch loss in BRD design and field application cannot be adequately evaluated using guide-net angle alone. The species-specific SHAP analysis showed that a single design criterion is unlikely to apply equally to all commercial species. In firefly squid, body length contributed more strongly than the other variables, suggesting that catch loss may depend strongly on length-dependent selectivity and the effective opening conditions of the gear. In pearlside and common flying squid, headline height contributed similarly to or more strongly than body length, indicating that changes in vertical opening may substantially affect the likelihood of individuals being guided into the BRD pathway. In sailfin sandfish and Pacific herring, body length was the top-ranked variable, but gear geometry variables such as headline height, gear depth, and otter-board spread also contributed to prediction. In hairtail, SOG showed the largest contribution in the model; however, given the small sample size and strong angle imbalance, the species-specific result alone is insufficient to draw a general conclusion about the factors associated with hairtail catch loss. Therefore, BRD evaluation and operational guidelines should move beyond single-angle recommendations. The target species context should be considered as a background condition, whereas body length, towing speed, gear depth, headline height, and otter-board spread represent the main biological, operational, and gear geometry predictors relevant to BRD performance.

The operational and gear geometry variables identified as important in this study can be obtained directly from vessel navigation records or gear monitoring sensors. These variables could therefore serve as input information for onboard decision support systems that estimate the likelihood of catch loss during towing and assist in adjusting tow conditions. In the longer term, interpretable machine learning results such as those presented here could be linked to research on adaptive gear operation and onboard decision support. However, extending this approach to practical gear operation policy will require external validation using independent datasets, acquisition of real-time sensor data, and additional experimental designs that systematically vary tow conditions.

The methodological contribution of this study lies in presenting an analytical procedure that combines individual-level sea trial data with interpretable machine learning, trivial baseline comparisons, repeated haul-grouped validation, and haul-clustered GEE sensitivity analyses. However, the empirical findings are based on a final modeling dataset comprising 5641 records from only 24 unique haul clusters, even though 31 valid hauls were available for broader operational summaries.

This study has several limitations. The original sea trials were designed to evaluate BRD guide-net angle and mesh selectivity, not to develop machine learning models or estimate causal effects. Most tow condition and gear geometry predictors were measured at the haul level; therefore, the effective independent information for these variables was closer to the number of haul clusters than to the number of individual records. The final modeling

dataset contained only 24 haul clusters, so GEE robust standard errors and confidence intervals should be interpreted as approximate cluster-aware sensitivity estimates.

Although haul-grouped cross-validation and repeated grouped cross-validation were applied, model performance should be interpreted as out-of-fold predictive performance within the final six-taxon BRD/COD modeling dataset, which contained 24 unique haul clusters. The broader 31 valid hauls were used for haul-level operational summaries, not as the final modeling cluster count. The generalizability of the results to other fishing grounds, seasons, vessels, or gear configurations should be evaluated through external validation using independent datasets. Finally, because the response variable was defined as a binary outcome based on collection location in the BRD or COD, this study could not distinguish individuals that actually encountered and escaped through the BRD pathway from those whose collection location reflected behavioral differences before entering the codend.

Future sea trials should allocate nominal guide-net angle conditions as evenly as possible within fishing grounds, seasons, and target species compositions. Designs that allow paired or systematically alternating comparisons of the 30° and 45° conditions under similar tow conditions would help reduce species–angle imbalance. Increasing the number of hauls, collecting direct underwater gear geometry measurements, and recording tension or video-based fish behavior would also help separate nominal angle effects from realized gear geometry and operational context.

5. Conclusions

This study conducted an interpretable machine learning analysis to predict individual-level catch loss in a guide-net marine mammal bycatch reduction device (BRD) attached to a Korean midwater trawl. From an integrated extraction dataset of 6372 individuals across 22 species, binary BRD/COD records for six target taxa were selected, resulting in a final modeling dataset of 5641 individual records from 24 unique haul clusters. The pooled out-of-fold AUCs of the GLM, random forest, and XGBoost models were 0.717, 0.672, and 0.703, respectively, but Brier score and accuracy were interpreted cautiously because trivial baselines were strong under the imbalanced outcome.

SHAP and permutation importance analyses identified body length as the largest exploratory contributor to catch loss prediction, with headline height ranked second by both methods. Other important variables included SOG, otter-board spread, gear depth, and nominal guide-net angle. These interpretation tools were used as exploratory variable contribution diagnostics rather than causal effect estimators.

Because of species–angle imbalance, sailfin sandfish was the most suitable taxon for within-species comparison of guide-net angle effects, but its subset model showed limited predictive separation. The four GEE sensitivity specifications showed consistently positive but highly uncertain 45° angle estimates, with wide confidence intervals and no statistically significant angle term. Therefore, the nominal angle contribution should be interpreted as exploratory and predictive rather than causal.

Taken together, catch loss in the guide-net BRD was associated with body length, tow conditions, realized gear geometry, nominal guide-net angle, and species–angle structure. However, because the final modeling dataset contained only 24 haul clusters and showed strong species–angle imbalance, the nominal angle contribution cannot be cleanly isolated from species composition and realized gear geometry. Future BRD evaluations should combine balanced experimental allocation, more haul clusters, direct underwater gear geometry measurements, and external validation to support stronger causal interpretation.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/fishes11070416/s1>, Table S1: Haul-level allocation of guide-net angle conditions; Table S2: Species name harmonization list; Table S3: Logistic GLM coefficients; Table S4: Permutation importance values for species indicators included as control variables; Figure S1: Distribution of major predictors by catch outcome; Figure S2: Global SHAP beeswarm plot for non-species predictors; Figure S3: SHAP interaction plots for body length $\times$ guide-net angle and gear depth $\times$ guide-net angle.

Author Contributions: Conceptualization, K.-S.C., J.-M.J. and H.-Y.K.; methodology, K.-S.C.; software, K.-S.C.; validation, K.-S.C. and H.-Y.K.; formal analysis, K.-S.C. and H.-Y.K.; investigation, K.-S.C. and J.-M.J.; resources, K.-S.C., J.-M.J., and H.-Y.K.; data curation, K.-S.C.; writing—original draft preparation, K.-S.C.; writing—review and editing, K.-S.C., J.-M.J. and H.-Y.K.; visualization, K.-S.C.; supervision, H.-Y.K.; project administration, K.-S.C.; funding acquisition, K.-S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Institute of Fisheries Science (NIFS) of Korea, grant number R2026008.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available from the corresponding author upon reasonable request. The data are not publicly available because they include operational records from sea trials. The analysis code supporting the reported models, GEE sensitivity analyses, baseline comparisons, and SHAP/permutation-importance diagnostics is available from the corresponding author upon reasonable request.

Acknowledgments: The authors thank the captain and crew of the R/V Tamgu-20 and the staff who assisted with sea trials and catch measurements.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AL	Anal length
AUC	Area under the receiver operating characteristic curve
BRD	Bycatch reduction device
COD	Codend
COV	Covernet
GLM	Generalized linear model
IML	Interpretable machine learning
OOF	Out-of-fold
RF	Random forest
SHAP	SHapley Additive exPlanations
SOG	Speed over ground
TED	Turtle excluder device
TL	Total length
MMPA	Marine Mammal Protection Act
NOAA	National Oceanic and Atmospheric Administration
NIFS	National Institute of Fisheries Science
IQR	Interquartile range
FL	Fork length
CW	Carapace width
XGBoost	Extreme Gradient Boosting

References

- Read, A.J.; Drinker, P.; Northridge, S. Bycatch of Marine Mammals in U.S. and Global Fisheries. *Conserv. Biol.* **2006**, *20*, 163–169. [[CrossRef](#)] [[PubMed](#)]
- National Oceanic and Atmospheric Administration (NOAA). Implementation of Fish and Fish Product Import Provisions of the Marine Mammal Protection Act. *Fed. Regist.* **2016**, *81*, 54389–54413.
- National Oceanic and Atmospheric Administration (NOAA). Modification of Deadlines Under the Fish and Fish Product Import Provisions of the Marine Mammal Protection Act. *Fed. Regist.* **2022**, *87*, 63955–63957.
- Lee, G.-H.; Song, D.-H.; Kim, H.-Y. Characteristics of Catch Losses in Stow Nets with Finless Porpoise Excluder Devices and a Strategy to Reduce Them. *Reg. Stud. Mar. Sci.* **2022**, *50*, 102147. [[CrossRef](#)]
- Mitchell, J.F.; Watson, J.W.; Foster, D.G.; Caylor, R.E. *The Turtle Excluder Device (TED): A Guide to Better Performance*; NOAA Technical Memorandum NMFS-SEFSC-366; National Marine Fisheries Service: Pascagoula, MS, USA, 1995.
- Jung, J.-M.; Park, M.-S.; Choi, K.-S. A Study on the Catch Losses and Mesh Selectivity Related to the Attachment of Marine Mammal Bycatch Reduction Devices on Midwater Trawl Gear. *Fishes* **2024**, *9*, 391. [[CrossRef](#)]
- Jung, J.-M.; Kim, H.-Y.; Cha, B.-J.; Kim, S.-J.; Kim, T.-S.; Hyun, G.-C.; Choi, K.-S. Catch Losses and Reduction of Bycatch for Jellyfish Using Marine Mammal Bycatch Reduction Devices in Midwater Trawl Gear. *Fishes* **2025**, *10*, 276. [[CrossRef](#)]
- Broadhurst, M.K. Modifications to Reduce Bycatch in Prawn Trawls: A Review and Framework for Development. *Rev. Fish Biol. Fish.* **2000**, *10*, 27–60. [[CrossRef](#)]
- Wileman, D.A.; Ferro, R.S.T.; Fonteyne, R.; Millar, R.B. *Manual of Methods of Measuring the Selectivity of Towed Fishing Gears*; ICES Cooperative Research Report 1996; International Council for the Exploration of the Sea: Copenhagen, Denmark, 1996; No. 215.
- Herrmann, B.; Krag, L.A.; Krafft, B.A. Size Selection of Antarctic Krill (*Euphausia superba*) in a Commercial Codend and Trawl Body. *Fish. Res.* **2018**, *207*, 49–54. [[CrossRef](#)]
- Brčić, J.; Herrmann, B.; Sala, A. Predictive Models for the Size Selectivity of Bottom Trawls Using Gear and Species-Related Parameters. *ICES J. Mar. Sci.* **2018**, *75*, 1909–1919. [[CrossRef](#)]
- Kühn, B.; Cayetano, A.; Fincham, J.I.; Moustahfid, H.; Sokolova, M.; Trifonova, N.; Watson, J.T.; Fernandes-Salvador, J.A.; Uusitalo, L. Machine Learning Applications for Fisheries-At Scales from Genomics to Ecosystems. *Rev. Fish. Sci. Aquac.* **2025**, *33*, 334–357. [[CrossRef](#)]
- Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
- Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
- Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
- Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [[CrossRef](#)] [[PubMed](#)]
- Fisher, A.; Rudin, C.; Dominici, F. All Models Are Wrong, but Many Are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *J. Mach. Learn. Res.* **2019**, *20*, 1–81.
- Van Rossum, G.; Drake, F.L. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, USA, 2009.
- McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference 2010, Austin, TX, USA, 28 June–3 July 2010; pp. 56–61. [[CrossRef](#)]
- Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array Programming with NumPy. *Nature* **2020**, *585*, 357–362. [[CrossRef](#)] [[PubMed](#)]
- Seabold, S.; Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. In Proceedings of the 9th Python in Science Conference 2010, Austin, TX, USA, 28 June–3 July 2010; pp. 92–96.
- Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillerá-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schroder, B.; Thuiller, W.; et al. Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography* **2017**, *40*, 913–929. [[CrossRef](#)]
- Hanley, J.A.; McNeil, B.J. The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology* **1982**, *143*, 29–36. [[CrossRef](#)] [[PubMed](#)]
- Brier, G.W. Verification of Forecasts Expressed in Terms of Probability. *Mon. Weather Rev.* **1950**, *78*, 1–3. [[CrossRef](#)]
- Niculescu-Mizil, A.; Caruana, R. Predicting Good Probabilities with Supervised Learning. In Proceedings of the 22nd International Conference on Machine Learning (ICML 2005), Bonn, Germany, 7–11 August 2005; pp. 625–632. [[CrossRef](#)]
- Van Calster, B.; McLernon, D.J.; van Smeden, M.; Wynants, L.; Steyerberg, E.W. Topic Group Calibration and Performance Measures of the STRATOS Initiative. Calibration: The Achilles Heel of Predictive Analytics. *BMC Med.* **2019**, *17*, 230. [[CrossRef](#)] [[PubMed](#)]

28. Liang, K.-Y.; Zeger, S.L. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika* **1986**, *73*, 13–22. [[CrossRef](#)]
29. Mancl, L.A.; DeRouen, T.A. A Covariance Estimator for GEE with Improved Small-Sample Properties. *Biometrics* **2001**, *57*, 126–134. [[CrossRef](#)] [[PubMed](#)]
30. Kauermann, G.; Carroll, R.J. A Note on the Efficiency of Sandwich Covariance Matrix Estimation. *J. Am. Stat. Assoc.* **2001**, *96*, 1387–1396. [[CrossRef](#)]
31. Millar, R.B.; Fryer, R.J. Estimating the Size-Selection Curves of Towed Gears, Traps, Nets and Hooks. *Rev. Fish Biol. Fish.* **1999**, *9*, 89–116. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.