ICES CIEM
International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

Model-informed classification of broadband acoustic backscatter from zooplankton in an *in situ* mesocosm

Muriel Dunn [1,2,*,†], Chelsey McGowan-Yallop [3,†], Geir Pedersen [4], Stig Falk-Petersen [5],
Malin Daase [6], Kim Last [3], Tom J. Langbehn [7], Sophie Fielding [8], Andrew S. Brierley [9],
Finlo Cottier [3,6], Sünnje L. Basedow [6], Lionel Camus [1], Maxime Geoffroy [2,6]

[1]Akvaplan-niva AS, Fram Centre, Postbox 6606, Stakkevollan, 9296 Tromsø, Norway
[2]Centre for Fisheries Ecosystems Research, Fisheries and Marine Institute of Memorial University of Newfoundland, St. John's A1C 5R3, Canada
[3]Scottish Association for Marine Science, Oban, Argyll PA37 1QA, United Kingdom
[4]Institute for Marine Research, 5005 Bergen, Norway
[5]Independent Scientist, 9012 Tromsø, Norway
[6]Department of Arctic and Marine Biology, UiT The Arctic University of Norway, 9036 Tromsø, Norway
[7]Department of Biological Sciences, University of Bergen, 5020 Bergen, Norway
[8]British Antarctic Survey, Natural Environment Research Council, High Cross, Cambridge CB30ET, United Kingdom
[9]Pelagic Ecology Research Group, School of Biology, Scottish Oceans Institute, Gatty Marine Laboratory, University of St Andrews, St Andrews KY16 8LB, United Kingdom
*Corresponding author. Tel: +47 46744522, E-mail: muriel.dunn@sintef.no
†These authors contributed equally: Muriel Dunn and Chelsey McGowan-Yallop.

Abstract

Classification of zooplankton to species with broadband echosounder data could increase the taxonomic resolution of acoustic surveys and reduce the dependence on net and trawl samples for 'ground truthing'. Supervised classification with broadband echosounder data is limited by the acquisition of validated data required to train machine learning algorithms ('classifiers'). We tested the hypothesis that acoustic scattering models could be used to train classifiers for remote classification of zooplankton. Three classifiers were trained with data from scattering models of four Arctic zooplankton groups (copepods, euphausiids, chaetognaths, and hydrozoans). We evaluated classifier predictions against observations of a mixed zooplankton community in a submerged purpose-built mesocosm (12 m³) in-sonified with broadband transmissions (185–255 kHz). The mesocosm was deployed from a wharf in Ny-Ålesund, Svalbard, during the Arctic polar night in January 2022. We detected 7722 tracked single targets, which were used to evaluate the classifier predictions of measured zooplankton targets. The classifiers could differentiate copepods from the other groups reasonably well, but they could not differentiate euphausiids, chaetognaths, and hydrozoans reliably due to the similarities in their modelled target spectra. We recommend that model-informed classification of zooplankton from broadband acoustic signals be used with caution until a better understanding of *in situ* target spectra variability is gained.

Keywords: machine learning; zooplankton; classification; broadband acoustics; cage experiment

Introduction

Acoustic target classification of zooplankton is needed to improve our understanding of variability in zooplankton spatio-temporal distribution and community composition. In the past decade, the commercial availability of broadband echosounders has made it possible to characterize the backscattering spectra of aquatic targets over a continuous frequency range (Bassett *et al.* 2018). Compared to conventional narrowband echosounder methods, the wider bandwidth of frequency-modulated (FM) echosounders offers the potential for improved classification of fish and zooplankton (Benoit-Bird and Waluk 2020). In addition, pulse-compression signal processing of broadband data improves the range resolution and the signal-to-noise ratio, enabling weak zooplankton targets to stand out above the stochastic background noise (Chu and Stanton 1998, Ehrenberg and Torkelson 2000). These improvements have made it possible to distinguish smaller and acoustically weaker individual targets, such as

mesozooplankton (0.2–20 mm), offering the potential for target classification using the target strength (TS [dB re 1 m²])—frequency response spectra [TS(f), hereafter 'target spectra'] as a predictive feature (Bandara *et al.* 2022).

Machine learning, a field of artificial intelligence (AI), is an increasingly popular tool for target classification in fisheries acoustics, reflecting a broader trend of AI applications in the marine sciences (Beyan and Browman 2020, Malde *et al.* 2020). Machine learning algorithms are objective, efficient, and can handle the large, complex datasets associated with broadband sampling (Malde *et al.* 2020). In short, supervised classification algorithms are trained to predict the class of new, unidentified samples with reference to scattering spectra from labelled training samples (i.e. samples for which the class is known) to optimize the classification function. In a fisheries acoustics context, the class is typically the species (or a broader functional group, e.g. based on gross anatomical properties) of the target or aggregation. The feature

variables used to predict the class of each target may include various acoustic features (e.g. backscattering strength and derived quantities), often in combination with geometric features (e.g. school length and height; Proud *et al.* 2020) or bathymetric features (e.g. distance from the seabed) (Korneliussen 2018). Machine learning algorithms improve the potential for real-time target classification and subsequent analysis (such as density estimates; Blackwell 2020) in fisheries acoustics, which is of interest given the increasing use of autonomous or remotely operated vehicles equipped with echosounders (e.g. Ludvigsen *et al.* 2018, De Robertis *et al.* 2019, Malde *et al.* 2020, Dunn *et al.* 2022). However, a major obstacle to applying supervised classification in broadband fisheries acoustics is the collection of labelled observations to train the algorithms (Handegard *et al.* 2021).

Labelled observations of target spectra have been obtained using various direct sampling or remote sensing methods, all of which have limitations. For example, directed trawl sampling of acoustic targets in areas with high densities of the species of interest has been used for jellyfish (Brierley *et al.* 2001), Antarctic krill (Hewitt and Demer 1996), and mesopelagic fish (Sobradillo *et al.* 2019), but this method is prone to sampling biases like net avoidance and acoustic shadowing of weaker targets (Peña 2018). Optical verification has been used to validate acoustic targets, for example, krill (Lawson *et al.* 2006) and salps (Wiebe *et al.* 2010), but has limited range resolution, especially for small targets (Trenkel *et al.* 2011) and is further limited by avoidance of the external light source (Geoffroy *et al.* 2021). Controlled tank experiments with zooplankton (e.g. McGehee *et al.* 1998, Pauly and Penrose 1998, Stanton *et al.* 1998) have typically relied on purpose-built or laboratory sonars (Conti *et al.* 2005, Amakasu and Furusawa 2006) because there are physical limits associated with (large and powerful) commercially available echosounders (i.e. beam angle and near-field range; Simmonds and MacLennan 2005). Controlled cage experiments have been used to measure the acoustic signal of large Antarctic krill (e.g. Foote *et al.* 1990), jellyfish (Monger *et al.* 1998), and fish (e.g. Legua and Lillo 2017, Gugele *et al.* 2021), but measurements of mesozooplankton remain challenging because detection of weak scatterers requires a cage designed to minimize noise and reverberation (Knutsen and Foote 1997).

Model-informed classification theoretically removes the need to collect measurements of known targets for use as labelled training data (e.g. Cotter *et al.* 2021). Validated scattering models (e.g. Korneliussen and Ona 2003, Peña 2018, Cotter *et al.* 2021) provide theoretical target spectra for each class (e.g. species) expected to be present in the acoustic data. Sound scattering models are considered validated when predictions of acoustic backscatter are comparable to benchmark models (Gastauer *et al.* 2019). Benchmark models are predictions of acoustic backscatter from exact or approximate analytical models and serve to find the limitations and validity domain of sound scattering models (Jech *et al.* 2015). These modelled spectra are then used as labelled training data for machine learning classification algorithms (hereafter, 'classifiers'). This approach has been used to classify scatterers into gross anatomical groups based on their acoustic properties for mesopelagic species (Cotter *et al.* 2021) and reef fish (Roa *et al.* 2022). However, to our knowledge, model-informed classification of target spectra has not yet been validated for any species.

This study aims to evaluate the validity and reliability of model-informed classification for the target spectra of zooplankton species with similar gross anatomical properties and size distributions. We applied model-informed classification to a mixed assemblage of Arctic mesozooplankton that was dominated by fluid-like species, i.e. animals with sound scattering properties similar to water (e.g. euphausiids, copepods, and salps) (Stanton and Chu 2000). The objectives were threefold: (i) to design an *in situ* mesocosm experiment to insonify zooplankton in a near-natural environment with minimal background noise and reverberation; (ii) to evaluate the performance of classifiers trained with scattering models for differentiating weakly backscattering mesozooplankton groups; and (iii) to validate the classifier predictions on a known community of zooplankton. We conclude by providing recommendations for model-informed classification of target spectra.

Materials and methods

Study area and zooplankton collection

Zooplankton were collected in Kongsfjorden, Svalbard, from the R/V *Helmer Hanssen* using a Tucker trawl (1 m^2 opening and 1000 μm mesh size, 10 min at 3 m s^{-1}) on the night of 15 January 2022 (Fig. 1). Twelve Tucker trawl tows were taken at the depth of the strongest sound scattering layer (~150 m) as seen from the vessel's echosounder (Kongsberg Discovery AS, Horten, Norway; Simrad EK60, 18–38 kHz, 1.024 ms pulse duration, 2 Hz ping rate). Samples from all tows were combined and kept alive for up to 15 h in running seawater, and delivered unsorted to the wharf in Ny-Ålesund on 16 January. The zooplankton samples were stored overnight in three 100-L holding tanks with a low-pressure flow system of filtered ambient seawater (~2°C) at the Kings Bay Marine Laboratory. An additional Tucker trawl sample collected on 15 January was preserved in a 4% formaldehyde-in-seawater solution buffered with hexamine and stored for species shape analysis.

Mesocosm design and experiment

Acoustic data were collected on 17 January 2022 from a mesocosm deployed from a wharf in Ny-Ålesund (Fig. 1). The mesocosm, or AZKABAN (Arrested Zooplankton Kept Alive for Broadband Acoustics Net experiment), was formed by a cuboid of zooplankton net (3 m high, 2 m wide, and 2 m long) with a 500 μm mesh, enclosing a volume of 12 m^3 (Fig. 2a). The net was mounted on the top section of an 8 m high by 2 m wide and 2 m long aluminium frame oriented vertically (Fig. 2a). Ropes attached eyelets on the net to the frame at each corner and along the edges.

A 200 kHz nominal frequency transducer (ES200-7CDK-Split; Kongsberg Discovery AS) was mounted on a plate centred inside the mesocosm through a hole on the top panel of the net with the acoustic axis pointing directly down. A Wideband Autonomous Transceiver (WBAT; Kongsberg Discovery AS) was fastened to the frame to operate the transducer (Fig. 2). The AZKABAN frame was purpose-built by Havbruksstasjonen (Ringvassøya, Norway) and designed to contain the entire main lobe of a 7° opening beam angle transducer inside the net.

The AZKABAN mesocosm was deployed by crane and lowered into the fjord (Fig. 2b). Zippers on the top and bottom
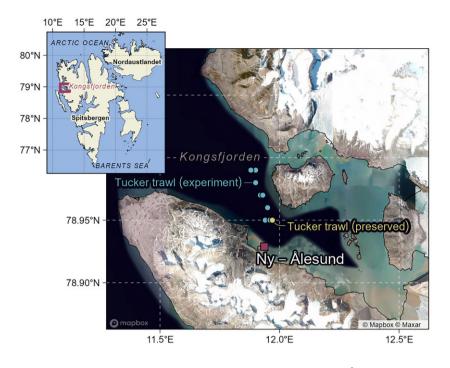
Figure 1: Study area in Kongsfjorden with locations of the mesocosm experiment from the wharf in Ny-Ålesund (red square) and Tucker trawl deployments for the experiment (blue circles with some overlap; $n = 12$). The yellow circle indicates the Tucker trawl deployment from which zooplankton was preserved for morphometric analyses (yellow circle; $n = 1$). The red box in the inset shows the location of the study area within the Svalbard archipelago.
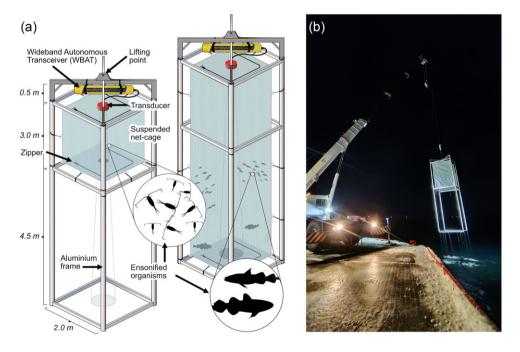

Figure 2: (a) Schematic of the AZKABAN mesocosm with the small zooplankton net (left) and large fish net (left). Only the configuration with a small net (left) was used for this study to limit the volume of insonified mesozooplankton. The acoustic transceiver (yellow cylinder) is attached to the frame and the transducer (orange cylinder). There is a hole at the top of the net for the transducer face to be unobstructed inside the net. (b) The AZKABAN mesocosm lifted with the crane at the end of the experiment.

panels of the net were used to add the alive and active zooplankton from the holding tanks into the submerged net. The frame was lowered such that the depth of the transducer face was ~0.5 m below the surface. The mesocosm was recovered after 3 h of data collection (Supplementary Fig. S1). The zooplankton were rinsed off the net and collected for species com-

position analysis. The species composition of the recovered mesocosm sample was analysed by identifying and counting 10% of the total sample for all species with >1000 individuals. All other species were counted for the entire sample.

The mesocosm experiment was conducted on an unsorted assemblage to maintain a high detection probability (i.e. with

Table 1: Scattering model parameters distributions for each zooplankton group.

Parameters	Copepods	Euphausiids	Chaetognaths	Hydrozoans
Modelled species	*Calanus glacialis*	*Thysanoessa inermis*	*Parasagitta elegans*	*Aglantha digitale*
Length (mm)	N (3.3, 0.7)[a]	L (2.4, 0.3)[d]	$\mathbb{\Gamma}$ (10.6, 0.6)[a]	L (2.4,0.4)[a]
Length-to-width ratio	N (5.3, 0.9)[a]	N (11.0, 2.0)[a]	N (26.0, 8.0)[a]	N (2.8,0.5)[a]
Density contrast (*g*)	N (0.997, 0.005)[b]	N (1.037, 0.005)[b]	N (1.03, 0.005)[e]	N (1.007, 0.005)[f]
Sound speed contrast (*h*)	N (1.027, 0.007)[b]	N (1.026, 0.005)[b]	N (1.03, 0.005)[e]	N (1.007, 0.005)[f]
Orientation (°)	N (90, 30)[c]	N (20, 20)[e]	N (0, 30)[e]	N (90, 30)[g]

[a]Measurements from the preserved sample with the distribution assessed as the best fit based on a 1:1 line between theoretical and empirical quantiles in Q-Q plots.
[b]Kögeler *et al.* (1987); February–March measurements.
[c]Blanluet *et al.* (2019)
[d]Measurements from a subsample of the mesocosm experimental sample. The distribution was assessed as the best fit based on a 1:1 line between theoretical and empirical quantiles in Q-Q plots.
[e]Lavery *et al.* (2007)
[f]Inferred from a comparison of measurements of hydrozoans from Monger *et al.* (1998) and Brierley *et al.* (2001, 2004) to model predictions.
[g]Monger *et al.* (1998) from swimming shape analysis.
The distributions are log-normal: L (meanlog and sdlog), normal: N (mean and SD), and gamma: $\mathbb{\Gamma}$ (shape and rate), where SD is the standard deviation.

large numbers of target animals in the enclosure). The sampling effort required to obtain sufficient animals for single-species experiments was deemed too great in time and hence expense. In addition, separating the live mesozooplankton from a mixed assemblage (as caught) into single species groups would have risked injuring or killing individuals. Using the unsorted mixed population meant that individual animals were handled minimally and that stress to them was minimized; this left it likely that natural swimming behaviour was preserved.

Acoustic data collection and calibration

During the AZKABAN experiment, the WBAT was programmed to transmit frequency-modulated pulses covering the entire available bandwidth from 185 to 255 kHz. The transmitted pulses had fast ramping, a pulse duration of 512 μs with 75 W transmit power, and a ping interval of 0.35 s. Simultaneous pinging of two split-beam transducers is not possible with a WBAT, so we had to restrict the bandwidth to that achievable by one transducer alone for the experiment. The simultaneous pinging of two or more transducers would improve the classification potential of broadband signals (Benoit-Bird and Waluk 2020). Of the available transducers with a 7° beam width (120, 200, and 333 kHz), the 200 kHz transducer was chosen to have the greatest signal-to-noise ratio of the targeted species (mesozooplankton) while achieving a small wavelength to detect smaller zooplankton (7 mm; Simmonds and MacLennan 2005). We used a short pulse length to resolve targets near the net boundary and reduce reverberation volume (Soule *et al.* 1997).

The acoustic system was calibrated on 19 January 2022 with two spheres made of tungsten carbide (WC) with 6% cobalt binder and diameters of 38.1 and 22 mm (Demer *et al.* 2015). Calibrations were processed with the EK80 software (version 21.15; Kongsberg Discovery AS). The calibration parameters were calculated for each sphere (Supplementary Fig. S2) and combined.

Scattering models

The training dataset for the classification was created with scattering models for the most abundant taxonomic groups in the Tucker trawl samples (≥1000 individuals). The most abundant were calanoid copepods, euphausiids, chaetognaths, and pelagic hydrozoans. All these groups are considered fluid-like scatterers with sound speed contrast (*h*) and density con-

trast (*g*) of 1 ± 5% (Stanton and Chu 2000). Near-unity sound speed and density contrasts imply that the material properties of the scatterers are not significantly different from the surrounding medium (seawater). To model the scattering of the zooplankton groups, we chose the phase-compensated distorted wave Born approximation (PC-DWBA) model because the parameters of this model are flexible to geometry, material properties, and acoustic frequency ranges, which makes the model adequate for the broad range of fluid-like zooplankton groups in this study (Chu and Ye 1999, Gastauer *et al.* 2019). The DWBA has been extensively tested (Lavery *et al.* 2007), and the PC-DWBA model has been used to infer length or material properties for Antarctic krill, *Euphausia superba* (Amakasu *et al.* 2017), decapod shrimp, *Palaemonetes vulgaris* (Chu *et al.* 2000), and eggs of North Atlantic cod, *Gadus morhua* (Chu *et al.* 2003) by comparison of model outputs with measurements of known species (in controlled laboratory experiments or concurrent trawl sampling). We ran 1000 model simulations for each zooplankton group using the ZooScatR package (version 0.5, Gastauer *et al.* 2019) with R (version 4.1.2) with shape, size, and material properties parameters chosen from distributions selected on the basis of the mesocosm-experiment samples, the preserved sample, or literature (Table 1). The modelled spectra were calculated with a 0.5 kHz frequency resolution.

The preserved Tucker Trawl sample was diluted and subsampled on 22 June 2022 for imaging of copepods ($n = 70$), euphausiids ($n = 20$), chaetognaths ($n = 70$), and hydrozoans ($n = 70$). Images were taken with a Leica M205 C stereomicroscope fitted with a Leica MC170 HD camera, and shape analysis was performed with an image processing software, ImageJ (version 1.53, National Institutes of Health, USA). The shapes were processed with ZooScatR with R to calculate the length and length-to-width ratio. Large individuals (>16 mm) were measured with a ruler. For the euphausiids, the length distribution was calculated from a subsample of 77 individuals from the mesocosm experiment sample. The processed images were used to create a shape input for each zooplankton group and its scattering model (Supplementary Fig. S3).

Material properties of copepods vary geographically and seasonally, predominantly because of their lipid reserves required to sustain the winter season (Sakinan *et al.* 2019). We selected *g* and *h* from Kögeler *et al.* (1987) because of the availability of measurements from the winter season (February–March) and the proximity to the Arctic of their measurements

of *Calanus* spp., thereby Arctic copepods. For hydrozoans, literature values for density and sound speed contrast were limited; therefore, we inferred the values for *g* and *h* from a comparison of the measurements from Brierley *et al.* (2001, 2004) and Monger *et al.* (1998) to the model predictions, a method used by Lavery *et al.* (2007).

Acoustic data processing

All acoustic data were processed in Echoview 13.0 (Echoview Software Pty Ltd, Hobart, Tasmania). Data analysis was restricted to samples within the 1.0–2.25 m range to exclude the near-field region (0.5 m) (Simmonds and MacLennan 2005) and the echo from the bottom of the net (2.5 m) (Supplementary Fig. S1). The 'Single Target Detection—wideband' operator was applied to the pulse-compressed wideband data (Supplementary Table S1). The minimum value for the compensated TS threshold was set to the minimum allowable value, -120 dB re 1 m^2, to enable the detection of weak scatterers. The identified single targets were grouped into tracks using the 'Detect Fish Tracks' algorithm. We used conservative parameters to increase the likelihood of each track containing targets from only one individual (Supplementary Table S2). Tracks were visually assessed to remove outlier targets to further ensure that each track originated from only a single zooplankton target.

The target spectra of all single targets assigned to a track were exported from Echoview for analysis. All target spectra were calculated using a Fourier transform window size of 0.33 times the pulse length (0.25 m) with a 0.5 kHz resolution. The Fourier transform window size was selected as a compromise to maximize frequency resolution while minimizing the likelihood of incorporating backscattering from multiple targets (Benoit-Bird and Waluk 2020).

Noise level

The noise level inside the mesocosm affected the minimum backscatter detectable from organisms. In this case, noise level is considered all unwanted signals, including background noise and reverberation from the cage. The noise level within AZKABAN was calculated using a 1-min segment of data collected during a period of low single echo detections (11:25–11:26 UTC). First, single target detection was applied to the pulse-compressed TS with less stringent detection thresholds (Supplementary Table S2) to identify all possible targets. Second, targets were removed from the dataset using a mask. The target masks covered entire pings to avoid contamination by side lobes associated with pulse compression from targets. The remaining signal was designated as noise. Weak targets that were not identified by the single target detection algorithm were included in the noise level estimation (see example of unidentified targets in Supplementary Fig. S1). Finally, the noise level was calculated by exporting the median target strength frequency response profile for increments of 0.1 m depth bins.

Thereafter, when selecting single targets for the spectra analysis, targets were flagged (i.e. excluded from the analysis) if their target strength at nominal frequency (200 kHz) had a signal-to-noise ratio (SNR) of <10 dB (Simmonds and MacLennan 2005) when compared to the noise level at nominal frequency at the range of the target. We calculated the proportion of flagged targets below the SNR threshold relative to the total amount of targets. The full spectrum was not assessed because of the peaks and nulls in the target spectra.

Classifier training

Various algorithms have been used for acoustic target classification in previous fisheries acoustics studies, including k-Nearest Neighbours (Cotter *et al.* 2021), decision trees (Fernandes 2009, D'Elia *et al.* 2014), random forests (e.g. Proud *et al.* 2020, Gugele *et al.* 2021), gradient boosting (Escobar-Flores *et al.* 2019), support vector machines (Roberts *et al.* 2011, Roa *et al.* 2022), and neural networks (e.g. Simmonds *et al.* 1996, Cabreira *et al.* 2009, Brautaset *et al.* 2020). Here, three supervised classifiers that take different approaches to classification were compared (Table 2). The algorithm k-Nearest Neighbours (kNN; Goldberger *et al.* 2004) was chosen because it has been used for model-informed classification of broadband acoustic measurements previously (Cotter *et al.* 2021). LightGBM (Ke *et al.* 2017), an implementation of gradient boosting (Friedman 2001), was chosen because it was considered representative of decision tree-based ensemble methods with the potential for improved performance compared to random forest (Fernández-Delgado *et al.* 2014), which is widely used in fisheries acoustics (Fernandes 2009, Gugele *et al.* 2021). Finally, the support vector machine algorithm (SVM; Cortes and Vapnik 1995) was chosen because it is another widely used algorithm that, together with gradient boosting, has been identified as among the best-performing classification algorithms based on comparisons of performance on large data set collections (Fernández-Delgado *et al.* 2014).

Training on modelled target spectra

Training of each machine learning classifier was conducted in Python (version 3.9) using the Scikit-learn library (version 1.1.1, Pedregosa *et al.* 2011). An L^2-normalization was applied to each target spectra from individuals modelled with the PC-DWBA model simulations so that if the values were to be squared and summed, the sum would equal one (Komer *et al.* 2014). The target variable (i.e. the classification output) was the zooplankton group: copepod, euphausiid, chaetognath, or hydrozoan.

For each classifier described in Table 2, we optimized the hyperparameters (Supplementary Information Codes S1, S2, and S3) and estimated the classifiers' performance on a hold-out dataset through cross-validation (CV; Stone 1974). Nested CV (Wainer and Cawley 2021) was used to optimize the hyperparameters and evaluate the performance of the classifiers (Supplementary Fig. S4). Nested CV ensured that separate data were used to train, validate, and test the classifier and provided an estimate of the classifier's true error with minimal bias (Varma and Simon 2006). We compared the classifiers' success using mean class-weighted F1 score (Equation 1; Pedregosa *et al.* 2011) because that is appropriate for scenarios where both false positives and false negatives are equally undesirable.

The F1 score is a measure of overall accuracy calculated as the harmonic mean of precision and recall, defined as

$$F1_i = \frac{2 \times (precision_i \times recall_i)}{(precision_i + recall_i)}. \tag{1}$$

Table 2: Overview of the machine learning algorithms (i.e. classifiers) compared in this study.

Classifier	Description	Strengths	Limitations
k-Nearest Neighbours (kNN)	Predicts the class of new samples by taking a majority vote of the k training samples, which are closest in distance by some metric (e.g. Euclidean distance) (Fix and Hodges 1951).	Few hyperparameters therefore easy to implement. Interpretable; an 'explainable artificial intelligence' algorithm (Islam *et al.* 2021, preprint: not peer reviewed). Computationally inexpensive, facilitates repeat classification, and examining the effects of parameters.	Limited ability to deal with noise and outliers (Korneliussen *et al.* 2018). Limited ability to identify low abundance groups (Peña 2018). Vulnerable to overfitting.
LightGBM	Implementation of gradient boosting, a decision tree-based ensemble method similar to random forest (Breiman 2001, Friedman 2001). Gradient descent is used to minimize a loss function with the addition of each new tree to the ensemble. Thus, each new tree attempts to correctly classify samples that were previously misclassified (Hastie *et al.* 2009).	Suitable for large datasets (Ke *et al.* 2017) and robust to outliers (Hastie *et al.* 2009). Reduced risk of overfitting (Hastie *et al.* 2009).	Rarely used in fisheries acoustics. Many hyperparameters, optimization is computationally expensive.
Support Vector Machine (SVM)	Maps data to a higher-dimensional feature space in which classes are linearly separable; the optimal decision boundary (hyperplane) has the maximal distance between itself and the closest training data points (support vectors) of any class (Cortes and Vapnik 1995, Hastie *et al.* 2009).	Few hyperparameters therefore easy to implement. Results are consistent and reproducible between repeat implementations (Bennett and Campbell 2000).	Sensitive to outliers (Kanamori *et al.* 2017). Unsuitable for large datasets, as it is very computationally expensive (Cervantes *et al.* 2008).

The strengths and limitations are detailed for use in fisheries acoustics.

Precision reports the relative success of the classifier, expressed as

$$precision_i = \frac{T_i}{(T_i + FP_i)}, \quad (2)$$

where T is the number of true positives and FP is the number of false positives for each class i (each zooplankton group). Whereas recall is a measure of the sensitivity from repeat detections, expressed as

$$recall_i = \frac{T_i}{(T_i + FN_i)}, \quad (3)$$

where FN is the number of false negatives for each class i (each zooplankton group). An F1 score of 1.0 would indicate that a classifier could correctly classify all samples. The F1 score was only calculated for the classifier training on the modelled target spectra because the label of the modelled target spectra was known. There was only one individual per target spectra; therefore, we did not need to account for mixtures.

Hyperparameter optimization was repeated on the entire modelled dataset (1000 target spectra for each of the 4 zooplankton groups) without subsampling to obtain the final trained classifiers.

Classifier sensitivity

To determine the optimal frequency bandwidth for model-informed classification of copepods, euphausiids, chaetognaths, and hydrozoans, kNN classifiers were trained and evaluated with modelled target spectra over the bandwidths commonly used in fisheries acoustics (Simmonds and MacLennan 2005). The selected bandwidths were the individual bandwidths from the 70, 120, 200, and 333 kHz transducers produced by Kongsberg Discovery AS (45–90, 90–170, 185–255, and 283–383 kHz) and their continuous bandwidth (45–383 kHz). Only kNN was used for this analysis as it is less computationally expensive than the other algorithms.

A kNN classifier was also trained using modelled cross-sectional backscattering strength—frequency spectra $\sigma_{bs}(f)$, the linear scale of TS(f), for the bandwidth of 185–255 kHz—to examine the effect of the logarithmic scale of the modelled target spectra on the classification performance.

Additionally, we evaluated the classifiers' sensitivities to the parameterization of material properties in the scattering models because this can strongly influence backscattering intensity (Chu and Ye 1999, Sakinan *et al.* 2019). A PC-DWBA model was parameterized using material property values for Antarctic copepods drawn from the literature (*Calanus* spp.) ($g = 0.995 \pm 0.001$ and $h = 0.959 \pm 0.010$; Chu and Wiebe 2005). These values are from spring (2 May 2002) but from similar water temperatures (-0.8–$0.4°$C) as those used for Arctic copepods in this study. All other model parameters for copepods and the other zooplankton groups remained the same.

Classification

The trained and optimized classifiers were used to classify the measured *in situ* target spectra from AZKABAN into zooplankton groups (copepods, euphausiids, chaetognaths, or hydrozoans). The classifier predictions were evaluated by comparing: (i) the predicted class distributions to the species composition of the zooplankton sample recovered from AZKABAN; (ii) the class predictions from each classifier (classifier agreement); and (iii) the class predictions for targets from the same track (within-track consistency).

Table 3: Taxonomic group, species, count, and proportion of the sample retrieved from the net after the experiment.

Taxonomic group	Species	Total individuals	Proportion of sample (%)	Median length (mm) (±SD)
Copepoda	*Calanus* spp.	13 380	50.61	3.3 (±0.7)
Copepoda	*Metridia* spp.	6310	23.87	
Copepoda	*Paraeuchaeta* spp.	710	2.69	
Copepoda	Other copepods	160	0.61	
Euphausiacea	*Thysanoessa inermis*	2485	9.40	11.0 (±4.0)
Chaetognatha	*Parasagitta elegans*	2220	8.40	17.0 (±5.0)
Hydrozoa	*Aglantha digitale*	1000	3.78	11.0 (±5.0)
Decapoda	Juvenile *Pandalus* spp.	76	0.29	
Decapoda	Benthic shrimp	2	0.01	
Pteropoda	*Clione limacina*	40	0.15	
Amphipoda	*Themisto* spp.	27	0.10	
Amphipoda	Undetermined	14	0.05	
Fish (larvae)	*Leptoclinus maculatus*	7	0.03	
Mysidacea	Undetermined	4	0.02	

Samples with <1000 individuals were counted for the entire recovered mesocosm sample. The groups in grey were modelled to create the labelled training dataset for the classification algorithms.

Results

Species composition

The zooplankton sample collected from AZKABAN after the experiment showed that copepods were numerically dominant. Over 20 000 copepods were in AZKABAN, mostly *Calanus* spp. (>13 000 individuals; Table 3). The second most abundant group was euphausiids, which were an order of magnitude less abundant in the samples than copepods. The most common euphausiid was *Thysanoessa inermis*, and the population consisted mainly of small juveniles (median length of 11 mm; Table 3). The sample contained almost as many chaetognaths as euphausiids, predominantly *Parasagitta elegans*. The fourth most abundant group in the sample were hydrozoans, predominantly *Aglantha digitale*. All other zooplankton and fish sampled had <100 individuals; therefore, we did not include these species in the classification analysis due to the low likelihood of repeated detections. During the experiment, the AZKABAN mesocosm had a total density of 2203 individual zooplankton per m^3.

Scattering models

Copepods were the smallest scatterers in this experiment with a median total length (± SD) of 3.3 ± 0.7 mm and an average modelled TS of −113 dB re 1 m^2 across the frequency spectrum. The amplitude of the modelled target spectra was typically lower for the copepods than the other three groups. Modelled Antarctic copepods had similar target spectra results but with a 5 dB mean increase across the spectra compared to the Arctic copepods, with an average TS of −107 dB re 1 m^2 (Fig. 3a; blue).

Euphausiids and hydrozoans had the same median total lengths of 11 mm (±4 mm for euphausiids and ±5 mm for hydrozoans). Despite their similar length distributions, euphausiids had a higher average TS (−89 dB re 1 m^2 for euphausiids and −94 dB re 1 m^2 for hydrozoans) due to differences in their material properties. However, both groups had relatively flat average spectra over the measured bandwidth (Fig. 3b and d). Lastly, chaetognaths had the longest median length (17 ± 5 mm) but had a relatively low median TS (−98 dB re 1 m^2). The target spectra of chaetognaths had a slight positive slope and a large dispersion of TS intensity (Fig. 3c and g).

Noise level

The noise level inside AZKABAN was low, being below −100 dB re 1 m^2 throughout the mesocosm (Fig. 4) and across the frequency bandwidth. There were peaks in the noise level profile at 1.1, 1.6, and 1.9 m range from the transducer (Fig. 4). The noise profile followed a similar magnitude and trend across the bandwidth, with ∼5 dB re 1 m^2 variability. We found that the signal-to-noise ratio at 200 kHz was <10 dB re 1 m^2 for 10.6% of the single targets used for classification, as shown by the overlaid detected target used for classification analysis in Fig. 4. This was deemed adequate, and all targets were retained for subsequent analyses.

Mesocosm target detections

A total of 7722 tracked single targets were detected during the 3-h AZKABAN mesocosm experiment. The mesocosm target detections were from a mixed zooplankton assemblage, and individual detections were from targets of unknown identity. There were 777 distinct tracks, with a mean of 10 single target detections per track. The minimum number of detections in a track was 4, and the maximum was 178.

Evaluation of classifier training

The optimized kNN classifier used the KDTree algorithm (Pedregosa *et al.* 2011) and Euclidean distance as the distance metric. For the kNN classifier, the optimized value for the number of training samples closest in distance to the query sample used for predictions, k, was 1. The optimized SVM classifier used a radial basis function kernel, and the optimized LightGBM comprised 3400 trees with a maximum tree depth of 7. Full details of the optimized classifiers are provided in Supplementary Information Codes S1, S2, and S3.

Classifier performance

The F1 scores reflect the classifiers' performance at classifying the modelled target spectra. The highest class-weighted F1 score was achieved using LightGBM (0.71 ± 0.02), followed by kNN (0.70 ± 0.03) and SVM (0.59 ± 0.03) for the 185–255 kHz bandwidth. Per-class F1 scores showed consistently highest scores for copepods (0.71–0.87). The lower per-class F1 scores for euphausiids (0.64–0.72), hydrozoans (0.58–0.67), and chaetognaths (0.44–0.58) indicated that the
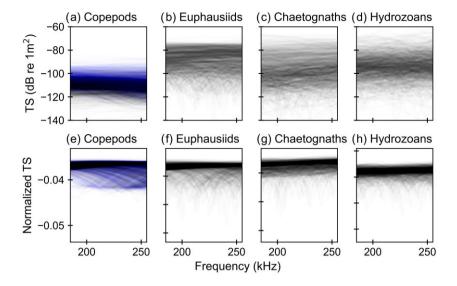
Figure 3: (a–d) All PC-DWBA model simulation results for each dominant zooplankton group. For copepods (a), the model results are shown for Arctic species (black; Kögeler *et al.* 1987) and Antarctic species (blue; Chu and Wiebe 2005). (e–h) L^2-normalized PC-DWBA model simulation results for each dominant zooplankton group.
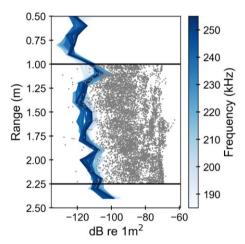

Figure 4: Background noise profile inside AZKABAN across the available bandwidth (185–255 kHz; blue lines). The grey dots indicate the TS of each detected tracked target detection at 200 kHz. The detection zone is delimited by the horizontal black lines at 1–2.25 m. The transducer face and top of the net are at 0 m range and the bottom of the net is at 3 m range.

classifiers had limited precision and/or recall in classifying these groups. The limited precision and recall of the classifiers were reflected in the confusion matrices for each classifier (i.e. the high numbers of misclassifications; Supplementary Tables S3, S4, and S5).

Classifier sensitivity

The nested CV procedure was conducted for modelled target spectra across five different frequency bandwidths (45–90, 90–170, 185–255, 283–383, and 45–383 kHz) to test the effect of bandwidth selection on classifier performance. The comparisons were only run with kNN because it was the least computationally expensive algorithm of those used in this study and, based on the results in Table 4, provided similar performance to LightGBM. The mean class-weighted F1 score for kNN with the full bandwidth (TS$_{45–383 kHz}$) was 0.92 (±

0.02) (Supplementary Table S6). The best score for a single 'transducer' was 0.86 (± 0.01), using modelled spectra at the centre bandwidth of the 120 kHz transducer (TS$_{70–190 kHz}$).

The cross-sectional backscatter spectra ($\sigma_{bs\,185–255\,kHz}$) (i.e. the linear domain representation of the target spectra) were also used to train a kNN classifier. Using the linear scale of the target spectra brought a slight improvement to classifier performance (mean class-weighted F1 score: 0.73 ± 0.03 in the linear domain compared to 0.70 ± 0.02 in the logarithmic domain).

The performance of the kNN classifier trained with modelled target spectra of Antarctic copepods (Supplementary Code S5) (mean class-weighted F1 score: 0.69 ± 0.03; Supplementary Table S7) was not significantly different from the classifier trained with modelled target spectra of Arctic copepods (mean class-weighted F1 score: 0.70 ± 0.02).

Classification of *in situ* measurements

All classifiers predicted a different class distribution to the species composition of the zooplankton sample recovered from AZKABAN (Fig. 5). For kNN, hydrozoans were predicted to be the most abundant class, followed by chaetognaths, euphausiids, and copepods, which was the inverse of the recovered sample (Fig. 5). For LightGBM, chaetognaths were predicted as the most abundant class with no copepod detections. The SVM predictions implied a majority of hydrozoans, followed by euphausiids, chaetognaths, and copepods.

The measured *in situ* target spectra for each class, as classified by kNN and LightGBM, were generally consistent with each other and the modelled spectra (Fig. 6). However, the measured *in situ* target spectra classified as copepods by kNN had a higher target strength than the copepods' modelled target spectra (Fig. 6). Of the mesocosm targets, those with high intensity and flat target spectra were labelled as copepods by the SVM classifier. However, the target spectra for euphausiids, chaetognaths, and hydrozoan predictions from SVM were in general agreement with the modelled results.

Only 18.13% of the measured target spectra (1400 samples) were classified as the same zooplankton group by all

Table 4: Classifier F1 scores estimated through nested cross-validation (mean ± SD) for the 185–255 kHz bandwidth.

Classifier	kNN	LightGBM	SVM
Mean class-weighted F1 score	0.70 ± 0.03	0.71 ± 0.02	0.59 ± 0.03
Mean F1 score for copepods	0.87 ± 0.02	0.87 ± 0.02	0.71 ± 0.03
Mean F1 score for euphausiids	0.70 ± 0.03	0.72 ± 0.03	0.64 ± 0.03
Mean F1 score for chaetognaths	0.58 ± 0.04	0.58 ± 0.05	0.44 ± 0.03
Mean F1 score for hydrozoans	0.66 ± 0.04	0.67 ± 0.03	0.58 ± 0.04

A score of 1.0 indicates that a classifier could correctly classify each sample (100% classification success).

Figure 5: (a) Composition of the zooplankton sample used in the mesocosm experiment as a proportion of the total sample for the four most abundant groups ($n = 26\ 435$). (b–d) the proportion of predicted targets of the total detections for tracked single targets ($n = 7722$) assigned to each group by k-nearest neighbours (kNN), LightGBM, and support vector machine (SVM) classifiers.

three classifiers: 10.09% were consistently classified as hydrozoans, 5.93% as chaetognaths, 1.29% as euphausiids, and 0% for copepods because no target spectra were labelled as copepods by LightGBM. Pairwise comparisons of classifiers show that 50.62% of tracked single target spectra (3909 samples) were classified as the same zooplankton group by kNN and LightGBM, compared to 42.55% (3286 samples) by kNN and LightGBM and 29.31% (17 103 samples) by LightGBM and SVM.

SVM had the highest within-track prediction consistency: on average, 75% of targets within a track were assigned the same class label. However, 70% of tracks included at least two different classes. For LightGBM, 67% of detections within a track were assigned to the same class, and 100% of tracks included at least two classes, compared to 62 and 93%, respectively, for kNN.

Discussion

AZKABAN: a mesocosm for *in situ* broadband acoustic backscatter measurements

AZKABAN was designed to facilitate *in situ* broadband acoustic backscatter measurements of caged fish and zooplankton. The estimated noise level of AZKABAN was sufficiently low to enable the detection of mesozooplankton. Noise and reverberation from mesocosm walls have been a major challenges in past experiments with weak scatterers (Knutsen and Foote 1997). The successful detection of weak targets in the AZKABAN mesocosm was partly due to the improvements in signal-to-noise ratio and range resolution associated with pulse compression of the broadband received signal.

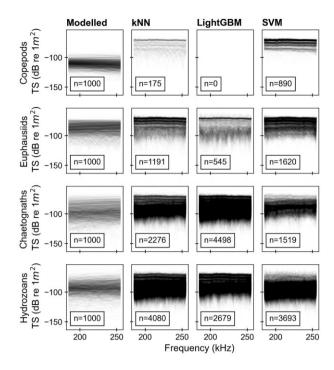The purpose-built mesocosm offered a practical platform for broadband measurements of mesozooplankton. The

Figure 6: Modelled: PC-DWBA model simulations (theoretical) target spectra for each zooplankton group. kNN, LightGBM, and SVM: measured target spectra of tracked single targets from the mesocosm experiment as classified by k-Nearest Neighbours (kNN), LightGBM, and support vector machine (SVM). All panels include the number (*n*) of target spectra in each panel.

design enabled the zooplankton sample to be added to the submerged net from a small boat, minimizing stress on the animals. It was also possible to recover the samples after the experiment for enumeration and morphometric analysis. Therefore, this mesocosm could be an effective experimental setup for controlled behavioural experiments, such as reactions to different sources and intensities of light and sound.

Performance of classifiers trained by modelled target spectra

Of the three conceptually different classifiers trained on modelled target spectra, the best-performing classifier was Light-GBM, with a mean class-weighted success rate of 0.71. Copepods consistently had the highest mean F1 score (0.71–0.87), indicating that copepods' modelled target spectra could be discriminated from the other modelled target spectra. The sensitivity analysis with the copepods parametrized with Arctic or Antarctic material properties demonstrated that changes in *g* and *h* have little effect on the normalized target spectra (Fig. 3a and e) or classification success (Table 4, Supplementary Table S7). All the classifiers were limited in their ability to discriminate between euphausiids, chaetognaths, and hydrozoans. Despite parametrizing the scattering models with representative parameters and shapes of the different zooplankton group organisms, these groups had overlapping modelled target spectra. Presumably, the overlap in the modelled target spectra of euphausiids, chaetognaths, and hydrozoans is due to the close similarity of the parameter distributions. The model's inability to resolve the target spectra of different fluid-like zooplankton directly introduces consequences for target detection and classification. This suggests that thresholds should be estab-

lished to determine possible taxonomic resolution for classification of species with overlapping model parameter distributions. Ross *et al.* (2013) report a similar effect with juvenile euphausiids and pteropods and conclude that the similarity in frequency responses of these groups may render them indistinguishable.

Previous studies on supervised classification of target spectra have used coarse taxonomic resolution to manually label measured target spectra to create a training set based on model-informed classes. Cotter *et al.* (2021) achieved a class-weighted F1 score of 0.90 for the classification of manually labelled fluid-like and gas-bearing targets detected with a broadband echosounder (25–40 kHz) using k-Nearest Neighbours. Roa *et al.* (2022) classified six reef fish using scattering models with a wide bandwidth (30–200 kHz) and found high classification accuracy (F1 score > 80%). We also found a wide bandwidth (45–383 kHz) resulted in high classification performance (class-weighted mean F1 score of 0.92 ± 0.02). However, the wide bandwidth (45–383 kHz) results were not possible to validate with the mesocosm experimental setup. Furthermore, we achieved higher accuracy with a lower bandwidth (90–170 kHz; class-weighted mean F1 score of 0.86 ± 0.01) than the one used in the mesocosm (185–255 kHz). Despite the higher F1 scores at a lower bandwidth (90–170 kHz), we used 185–255 kHz for its smaller wavelength for a better target size resolution. For the classification of *in situ* measurements, physical and practical limitations of target size and echosounder properties (beamwidth, wavelength, and transmit power) must be considered in addition to the F1 score of the classifier. While previous model-informed classification studies (Cotter *et al.* 2021, Roa *et al.* 2022) may have achieved better classification performances because the classes they used had distinct acoustic properties, in contrast to our study, their model-informed classifiers were not validated with *in situ* measurements.

Discrepancies between classifiers predictions and *in situ* measurements

We used the AZKABAN mesocosm experiment to validate the performance of three model-trained classifiers using measurements of a mesozooplankton community sample for which the species composition was known. Overall, the zooplankton community composition determined by the classifiers differed from the actual composition in the mesocosm. Copepods were overwhelmingly the most abundant group in the mesocosm (Fig. 5a) but were consistently the least abundant class in the classifier predictions (Fig. 5b–d). Hydrozoans were the least abundant group in the mesocosm but the most abundant predicted class for kNN and SVM. Whereas for LightGBM, chaetognaths were the most abundant class. These major discrepancies show that model-informed classification was not successful on *in situ* target detections.

Small copepods in the mesocosm (3 mm in length) were probably not detected, given the spatial resolution of the wavelength (7 mm at 200 kHz; Simmonds and MacLennan 2005). The classifier predictions reflected this, with few copepod predictions and the target strength mismatch between the modelled and predicted results (Fig. 6), despite the relatively high F1 scores for copepods. Therefore, while the larger copepods were detected acoustically, the majority of copepods were likely not identified by the echosounder. A higher frequency range (283–383 kHz) with a shorter wavelength

would theoretically resolve the issue, but previous noise level tests (not presented here) showed that the increased noise level at higher frequencies would not allow the detection of individual weak scatterers (<-110 dB re 1 m^2). Future studies testing model-informed classification of zooplankton with the 200 kHz transducer should only select the zooplankton fraction >7 mm. In that case, if the model-informed classification is successful, the proportion of each zooplankton group calculated from *in situ* target detections may match the real proportion in the zooplankton sample.

The *in situ* target detections were also used to assess the within-track consistency of predictions between classifiers. The target tracking algorithm associates many single target detections to an individual organism as it travels across the acoustic beam. There was a high variability of zooplankton groups assigned to each track, highlighting the large variability in target spectra from an individual organism. The inhomogeneity of predictions per track and poor agreement between classifiers provide compelling evidence that model-informed classification of fluid-like mesozooplankton is unreliable.

Recommendations to improve model-informed classification of zooplankton

Our results on classifier training with modelled target spectra suggest that the classification performance is highly dependent on the choice of algorithm when the groups cannot be reliably differentiated. Good practice for machine-learning-based science typically requires that a classifier's performance is evaluated on a test set 'drawn from the distribution of scientific interest' (Kapoor and Narayanan 2022, preprint: not peer reviewed). Model-informed acoustic target classification is appealing because it avoids the practical challenges and cost of obtaining labelled measurements of known species empirically (by sampling in the field or tank). However, using model-informed classification inevitably means that the samples used to train, validate, and test a classifier are not drawn from the distribution of scientific interest.

This study used a scattering model flexible to geometry, material properties, and acoustic frequency changes to generate training data for supervised machine learning classifiers. For future studies, we suggest that model-informed classification could be useful in assessing the theoretical classification potential of different bandwidths. However, classifier performance must be considered in the context of factors such as the target strength of the species of interest at a given frequency and the frequency's range resolution for the classification of *in situ* measurements. We conclude that a better understanding of the variability in the acoustic measurements from individuals is required before model-informed classification of target spectra can be implemented reliably. Features in broadband spectra, such as the locations of nulls and peaks, can provide insight into morphological characteristics of individuals (Reeder *et al.* 2004, Kubilius *et al.* 2020). A better understanding of these features could increase classification potential and the information we can extract from target spectra.

Summary and conclusions

This study evaluated a model-informed classification of zooplankton from broadband echosounder data using *in situ* measurements (185–255 kHz) of a mixed Arctic mesozooplankton assemblage in a purpose-built mesocosm. Acoustic scattering models generated modelled target spectra for the four most abundant zooplankton groups in the mesocosm: copepods, euphausiids, chaetognaths, and hydrozoans. Three different supervised machine learning algorithms were trained using modelled target spectra, and then compared in terms of their ability to classify the *in situ* measured target spectra obtained from the mesocosm experiment. Investigations of the classifier training using modelled target spectra showed that kNN and LightGBM classifiers could not differentiate euphausiids, chaetognaths, and hydrozoans reliably. The classifier training results were confirmed by their inconsistent predictions of euphausiids, chaetognaths, and hydrozoans within-track and between classifiers for the *in situ* mesocosm measurements. The lack of consistent predictions within a track suggests that the variability in target spectra per class is greater than in the target spectra between the different zooplankton groups from the sound scattering models. The outstanding challenge remaining is to understand the ping-to-ping variability in the spectra of individual scatterers (Martin *et al.* 1996, Dunning *et al.* 2023).

Another remaining challenge for this method is the requirement to model the dominant taxa that are expected to be found in the study area. An Arctic fjord was selected as the study location in part because of the low species diversity. For regions with higher diversity, similar taxa could be grouped based on their material properties and shape to expand model-informed classification. In addition, *in situ* imaging could complement the acoustic measurements to increase the taxonomic resolution of model-informed classification (Ohman *et al.* 2019).

The mesocosm design used in this study was an effective platform for measurements of fluid-like scatterers, which could be used to develop a better understanding of measured variability in target spectra. For example, mesocosm experiments with fewer individuals or a series of single-species experiments could improve model validation for broadband echosounder measurements of freely swimming individuals. However, a semi-permanent installation for longer experiment periods, visual validation through video or imaging for swimming behaviour information, and repeat experiments would be required to complete such comparative studies.

Supplementary data

Supplementary data is available at *ICES Journal of Marine Science* online.

Conflict of interest statement: The authors declare that they have no known comflict of financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

Data availability

The data underlying this article are available in Zenodo at https://doi.org/10.5281/zenodo.7501067. Additionally, the implementation of the sound scattering model is available at https://github.com/mbdunn/phd/tree/master/model-informed_classification.

Author contributions

M.D.: Conceptualization, Investigation, Methodology, Software, Formal analysis, Visualisation, Funding Acquisition, Writing—original draft

C.M.Y.: Conceptualization, Investigation, Methodology, Software, Formal analysis, Visualisation, Funding Acquisition, Writing—original draft

G.P.: Conceptualization, Methodology, Supervision, Writing—Review & editing

S.F.P.: Investigation, Writing—Review & editing

M.D.: Investigation, Funding Acquisition, Writing—Review & editing

K.L.: Conceptualization, Supervision, Writing—Review & editing

T.L.: Investigation, Visualization, Writing—Review & editing

S.F.: Methodology, Resources, Supervision, Writing—Review & editing

A.B.: Supervision, Writing—Review & editing

F.C.: Supervision, Writing—Review & editing

S.B.: Resources, Writing—Review & editing

L.C.: Funding Acquisition, Resources

M.G.: Conceptualization, Methodology, Resources, Supervision, Writing—Review & editing

References

Amakasu K, Furusawa M. The target strength of Antarctic krill (*Euphausia superba*) measured by the split-beam method in a small tank at 70 kHz. *ICES J Mar Sci* 2006;**63**:36–45. https://doi.org/10.1016/j.icesjms.2005.07.012

Amakasu K, Mukai T, Moteki M. Measurement of the volume-backscattering spectrum from an aggregation of Antarctic krill and inference of their length-frequency distribution. *Polar Sci* 2017;**12**:79–87. https://doi.org/10.1016/j.polar.2017.02.007

Bandara K, Basedow SL, Pedersen G *et al*. Mid-summer vertical behavior of a high-latitude oceanic zooplankton community. *J Mar Syst* 2022;**230**:103733. https://doi.org/10.1016/j.jmarsys.2022.103733

Bassett C, De Robertis A, Wilson CD. Broadband echosounder measurements of the frequency response of fishes and euphausiids in the Gulf of Alaska. *ICES J Mar Sci* 2018;**75**:1131–42. https://doi.org/10.1093/icesjms/fsx204

Bennett KP, Campbell C. Support vector machines: hype or hallelujah? *ACM SIGKDD Explor Newslett* 2000;**2**:1–13. https://dl.acm.org/doi/abs/10.1145/380995.380999

Benoit-Bird KJ, Waluk CM. Exploring the promise of broadband fisheries echosounders for species discrimination with quantitative assessment of data processing effects. *J Acoust Soc Am* 2020;**147**:411–27. https://doi.org/10.1121/10.0000594

Beyan C, Browman HI. Setting the stage for the machine intelligence era in marine science. *ICES J Mar Sci* 2020;**77**:1267–73. https://doi.org/10.1093/icesjms/fsaa084

Blackwell R *Real-time Reporting of Marine Ecosystem Metrics from Active Acoustic Sensors (Doctoral Dissertation)*. Norwich: University of East Anglia, 2020.

Blanluet A, Doray M, Berger L *et al*. Characterisation of sound scattering layers in the Bay of Biscay using broadband acoustics, nets and video. *PLoS One* 2019;**14**:e0223618. https://doi.org/10.1371/journal.pone.0223618

Brautaset O, Waldeland AU, Johnsen E *et al*. Acoustic classification in multifrequency echosounder data using deep convolutional neural networks. *ICES J Mar Sci* 2020;**77**:1391–400. https://doi.org/10.1093/icesjms/fsz235

Breiman L. Random forests. *Machine Learn* 2001;**45**:5–32. https://doi.org/10.1023/a:1010933404324

Brierley AS, Axelsen BE, Boyer DC *et al*. Single-target echo detections of jellyfish. *ICES J Mar Sci* 2004;**61**:383–93. https://doi.org/10.1016/j.icesjms.2003.12.008

Brierley AS, Axelsen BE, Buecher E *et al*. Acoustic observations of jellyfish in the Namibian Benguela. *Mar Ecol Prog Ser* 2001;**210**:55–66. https://doi.org/10.3354/meps210055

Cabreira AG, Tripode M, Madirolas A. Artificial neural networks for fish-species identification. *ICES J Mar Sci* 2009;**66**:1119–29. https://doi.org/10.1093/icesjms/fsp009

Cervantes J, Li X, Yu W *et al*. Support vector machine classification for large data sets via minimum enclosing ball clustering. *Neurocomputing* 2008;**71**:611–9. https://doi.org/10.1016/j.neucom.2007.07.028

Chu D, Stanton TK. Application of pulse compression techniques to broadband acoustic scattering by live individual zooplankton. *J Acoust Soc Am* 1998;**104**:39–55. https://doi.org/10.1121/1.424056

Chu D, Wiebe P, Copley N. Inference of material properties of zooplankton from acoustic and resistivity measurements. *ICES J Mar Sci* 2000;**57**:1128–42. https://doi.org/10.1006/jmsc.2000.0800

Chu D, Wiebe PH. Measurements of sound-speed and density contrasts of zooplankton in Antarctic waters. *ICES J Mar Sci* 2005;**62**:818–31. https://doi.org/10.1016/j.icesjms.2004.12.020

Chu D, Wiebe PH, Copley NJ *et al*. Material properties of North Atlantic cod eggs and early-stage larvae and their influence on acoustic scattering. *ICES J Mar Sci* 2003;**60**:508–15. https://doi.org/10.1016/S1054-3139(03)00047-X

Chu D, Ye Z. A phase-compensated distorted wave born approximation representation of the bistatic scattering by weakly scattering objects: application to zooplankton. *J Acoust Soc Am* 1999;**106**:1732–43. https://doi.org/10.1121/1.428036

Conti G, Demer DA, Brierley AS. Broad-bandwidth, sound scattering, and absorption from krill (*Meganyctiphanes norvegica*), mysids (*Praunus flexuosus* and *Neomysis integer*), and shrimp (*Crangon crangon*). *ICES J Mar Sci* 2005;**62**:956–65. https://doi.org/10.1016/j.icesjms.2005.01.024

Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;**20**:273–97. https://doi.org/10.1007/BF00994018

Cotter E, Bassett C, Lavery A. Classification of broadband target spectra in the mesopelagic using physics-informed machine learning. *J Acoust Soc Am* 2021;**149**:3889–901. https://doi.org/10.1121/10.0005114

D'Elia M, Patti B, Bonanno A *et al.* Analysis of backscatter properties and application of classification procedures for the identification of small pelagic fish species in the Central Mediterranean. *Fish Res* 2014;**149**:33–42. https://doi.org/10.1016/j.fishres.2013.08.006

Demer DA, Berger L, Bernasconi M *et al.* Calibration of acoustic instruments. *ICES Coop Res Rep* 2015;**326**:133. https://doi.org/10.25607/OBP-185

De Robertis A, Lawrence-Slavas N, Jenkins R *et al.* Long-term measurements of fish backscatter from Saildrone unmanned surface vehicles and comparison with observations from a noise-reduced research vessel. *ICES J Mar Sci* 2019;**76**:2459–70. https://doi.org/10.1093/icesjms/fsz124

Dunn M, Pedersen G, Basedow SL *et al.* Inverse method applied to autonomous broadband hydroacoustic survey detects higher densities of zooplankton in near-surface aggregations than vessel-based net survey. *Can J Fish Aquat Sci* 2023;**80**:451–67, https://doi.org/10.1139/cjfas-2022-0105

Dunning J, Jansen T, Fenwick AJ *et al.* A new in-situ method to estimate fish target strength reveals high variability in broadband measurements. *Fish Res* 2023;**261**:106611. https://doi.org/10.1016/j.fishres.2023.106611

Ehrenberg JE, Torkelson TC. FM slide (chirp) signals: a technique for significantly improving the signal-to-noise performance in hydroacoustic assessment systems. *Fish Res* 2000;**47**:193–9. https://doi.org/10.1016/S0165-7836(00)00169-7

Escobar-Flores PC, Ladroit Y, O'Driscoll RL. Acoustic assessment of the micronekton community on the Chatham Rise, New Zealand, using a semi-automated approach. *Fronti Mar Sci* 2019;**6**:1–22. https://doi.org/10.3389/fmars.2019.00507

Fernandes PG. Classification trees for species identification of fish-school echotraces. *ICES J Mar Sci* 2009;**66**:1073–80. https://doi.org/10.1093/icesjms/fsp060

Fernández-Delgado M, Cernadas E, Barro S *et al.* Do we need hundreds of classifiers to solve real world classification problems? *J Mach Learn Res* 2014;**15**:3133–81.

Fix E, Hodges Jr JL. Discriminatory analysis, nonparametric discrimination: consistency properties. *Int Statis Rev Rep* 1951;**4**:261–79.

Foote KG, Everson I, Watkins JL *et al.* Target strengths of Antarctic krill (*E uphausia superba*) at 38 and 120 kHz. *J Acoust Soc Am* 1990;**87**:16–24. https://doi.org/10.1121/1.399282

Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stats* 2001;**29**:1189–232. https://doi.org/10.1214/aos/1013203451

Gastauer S, Chu D, Cox MJ. ZooScatR—an R package for modelling the scattering properties of weak scattering targets using the distorted wave Born approximation. *J Acoust Soc Am* 2019;**145**:EL102–8. https://doi.org/10.1121/1.5085655

Geoffroy M, Langbehn T, Priou P *et al.* Pelagic organisms avoid white, blue, and red artificial light from scientific instruments. *Sci Rep* 2021;**11**:14941. https://doi.org/10.1038/s41598-021-94355-6

Goldberger J, Hinton GE, Roweis S *et al.* Neighbourhood components analysis. *Adv Neu Inform Process Sys* 2004;**17**:513–20.

Gugele SM, Widmer M, Baer J *et al.* Differentiation of two swim bladdered fish species using next generation wideband hydroacoustics. *Sci Rep* 2021;**11**:10520. https://doi.org/10.1038/s41598-021-89941-7

Handegard NO, Eikvil L, Jenssen R *et al.* Machine learning + marine science: critical role of partnerships in Norway. *J Ocean Technol* 2021;**16**:1–9.

Hastie T, Tibshirani R, Friedman JH *et al. The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edn. New York: Springer, 2009, 745. https://doi.org/10.1007/978-0-387-84858-7

Hewitt RP, Demer DA. Lateral target strength of Antarctic krill. *ICES J Mar Sci* 1996;**53**:297–302. https://doi.org/10.1006/jmsc.1996.0038

Islam SR, Eberle W, Ghafoor SK *et al.* Explainable artificial intelligence approaches: a survey. *arXiv* 2021. https://doi.org/10.48550/arXiv.2101.09429. (04 December 2023, date last accessed), preprint: not peer reviewed.

Jech JM, Horne JK, Chu D *et al.* Comparisons among ten models of acoustic backscattering used in aquatic ecosystem research. *J Acoust Soc Am* 2015;**138**:3742–64. https://doi.org/10.1121/1.4937607

Kanamori T, Fujiwara S, Takeda A. Breakdown point of robust support vector machines. *Entropy* 2017;**19**:83. https://doi.org/10.3390/e19020083

Kapoor S, Narayanan A. Leakage and the reproducibility crisis in ML-based science. *arXiv* 2022. https://doi.org/10.48550/arXiv.2207.07048. (04 December 2023, date last accessed), preprint: not peer reviewed.

Ke G, Meng Q, Finley T *et al.* Lightgbm: a highly efficient gradient boosting decision tree. *Adv Neu Inform Process Sys* 2017;**30**:3149–3157.

Knutsen T, Foote KG. Experiences in making acoustic measurements in a mesocosm with *Calanus finmarchius*. *ICES CM* 1997;**11**:1–19.

Kögeler JW, Falk-Petersen S, Kristensen Å *et al.* Density-and sound speed contrasts in sub-Arctic zooplankton. *Polar Biol* 1987;**7**:231–5. https://doi.org/10.1007/bf00287419

Komer B, Bergstra J, Eliasmith C. Hyperopt-sklearn: automatic hyperparameter configuration for scikit-learn. *ICML Workshop AutoML* 2014;**9**:50.

R. J. Korneliussen (ed). Acoustic target classification. *ICES Coop Res Rep* 2018;**344**:104.

Korneliussen RJ, Ona E. Synthetic echograms generated from the relative frequency response. *ICES J Mar Sci* 2003;**60**:636–40. https://doi.org/10.1016/S1054-3139(03)00035-3

Kubilius R, Macaulay GJ, Ona E. Remote sizing of fish-like targets using broadband acoustics. *Fish Res* 2020;**228**:105568. https://doi.org/10.1016/j.fishres.2020.105568

Lavery AC, Wiebe PH, Stanton TK *et al.* Determining dominant scatterers of sound in mixed zooplankton populations. *J Acoust Soc Am* 2007;**122**:3304–26. https://doi.org/10.1121/1.2793613

Lawson GL, Wiebe PH, Ashjian CJ *et al.* Improved parametrisation of Antarctic krill target strength models. *J Acoust Soc Am* 2006;**119**:232–42. https://doi.org/10.1121/1.2141229

Legua JA, Lillo SD Target strength *ex situ* of *Sprattus fuegensis* in Chilean fjords. *IEEE/OES Acoustics in Underwater Geosciences Symposium (RIO Acoustics)* 2017;1–5. https://doi.org/10.1109/RIOAcoustics.2017.8349735

Ludvigsen M, Berge J, Geoffroy M *et al.* Use of an autonomous surface vehicle reveals small-scale diel vertical migrations of zooplankton and susceptibility to light pollution under low solar irradiance. *Sci Adv* 2018;**4**:1–9. https://doi.org/10.1126/sciadv.aap9887

McGehee DE, O'Driscoll RL, Traykovski LM Effects of orientation on acoustic scattering from Antarctic krill at 120 kHz. *Deep Sea Res Part II* 1998;**45**:1273–94. https://doi.org/10.1016/S0967-0645(98)00036-8

Malde K, Handegard NO, Eikvil L *et al.* Machine intelligence and the data-driven future of marine science. *ICES J Mar Sci* 2020;**77**:1274–85. https://doi.org/10.1093/icesjms/fsz057

Martin LV, Stanton TK, Wiebe PH *et al.* Acoustic classification of zooplankton. *ICES J Mar Sci* 1996;**53**:217–24. https://doi.org/10.1006/jmsc.1996.0025

Monger BC, Meir E, Billings S *et al.* Sound scattering by the gelatinous zooplankters. *Deep Sea Res Part II* 1998;**45**:1255–71. https://doi.org/10.1016/S0967-0645(98)00029-0

Ohman MD, Davis RE, Sherman JT *et al.* Zooglider: an autonomous vehicle for optical and acoustic sensing of zooplankton. *Limnol Oceanograph Methods* 2019;**17**:69–86. https://doi.org/10.1002/lom3.10301

Pauly T, Penrose JD. Laboratory target strength measurements of free-swimming Antarctic krill (*Euphausia superba*). *J Acoust Soc Am* 1998;**103**:3268–80. https://doi.org/10.1121/1.423077

Pedregosa F, Varoquaux G, Gramfort A *et al.* Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;**12**:2825–30. https://jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

Peña M. Robust clustering methodology for multifrequency acoustic data: a review of standardisation, initialisation and cluster geometry. *Fish Res* 2018;**200**:49–60. https://doi.org/10.1016/j.fishres.2017.12.013

Proud R, Mangeni-Sande R, Kayanda RJ *et al.* Automated classification of schools of the silver cyprinid *Rastrineobola argentea* in Lake Victoria acoustic survey data using random forests. *ICES J Mar Sci* 2020;**77**:1379–90. https://doi.org/10.1093/icesjms/fsaa052

Reeder DB, Jech JM, Stanton TK. Broadband acoustic backscatter and high-resolution morphology of fish: measurement and modeling. *J Acoust Soc Am* 2004;**116**:747–61. https://doi.org/10.1121/1.1648318

Roa C, Pedersen G, Bollinger M *et al.* Taxonomical classification of reef fish with broadband backscattering models and machine learning approaches. *J Acoust Soc Am* 2022;**152**:1020–34. https://doi.org/10.1121/10.0012192

Roberts PL, Jaffe JS, Trivedi MM. Multiview, broadband acoustic classification of marine fish: a machine learning framework and comparative analysis. *IEEE J Oceanic Eng* 2011;**36**:90. https://doi.org/10.1109/JOE.2010.2101235

Ross T, Keister JE, Lara-Lopez A. On the use of high-frequency broadband sonar to classify biological scattering layers from a cabled observatory in Saanich Inlet, British Columbia. *Method Oceanograph* 2013;**5**:19–38. https://doi.org/10.1016/j.mio.2013.05.001

Sakinan S, Lawson GL, Wiebe PH *et al.* Accounting for seasonal and composition-related variability in acoustic material properties in estimating copepod and krill target strength. *Limnol Oceanograph Method* 2019;**17**:607–25. https://doi.org/10.1002/lom3.10336

Simmonds EJ, Armstrong F, Copland PJ. Species identification using wideband backscatter with neural network and discriminant analysis. *ICES J Mar Sci* 1996;**53**:189–95. https://doi.org/10.1006/jmsc.1996.0021

Simmonds EJ, MacLennan DN. *Fisheries Acoustics*, 2nd edn. Oxford: Blackwell Science LTD, 2005, 437.

Sobradillo B, Boyra G, Martinez U *et al.* Target strength and swimbladder morphology of Mueller's pearlside (*Maurolicus muelleri*). *Sci Rep* 2019;**9**:1–14. https://doi.org/10.1038/s41598-019-53819-6

Soule M, Barange M, Solli H *et al.* Performance of a new phase algorithm for discriminating between single and overlapping echoes in a split-beam echosounder. *ICES J Mar Sci* 1997;**54**:934–8. https://doi.org/10.1006/jmsc.1997.0270

Stanton TK, Chu D. Review and recommendations for the modelling of acoustic scattering by fluid-like elongated zooplankton: euphausiids and copepods. *ICES J Mar Sci*, 2000;**57**:793–807. https://doi.org/10.1006/jmsc.1999.0517

Stanton TK, Chu D, Wiebe PH *et al.* Sound scattering by several zooplankton groups. I. Experimental determination of dominant scattering mechanisms. *J Acoust Soc Am* 1998;**103**:225–35. https://doi.org/10.1121/1.421469

Stone M. Cross-validatory choice and assessment of statistical predictions. *J R Stat Soc Series B Stat Methodol* 1974;**36**:111–33. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x

Trenkel VM, Ressler PH, Jech M *et al.* Underwater acoustics for ecosystem-based management: state of the science and proposals for ecosystem indicators. *Mar Ecol Progress Ser* 2011;**442**:285–301. https://doi.org/10.3354/meps09425

Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinf* 2006;**7**:1–8. https://doi.org/10.1186/1471-2105-7-91

Wainer J, Cawley G. Nested cross-validation when selecting classifiers is overzealous for most practical applications. *Expert Syst Appl* 2021;**182**:115222. https://doi.org/10.1016/j.eswa.2021.115222

Wiebe PH, Chu D, Kaartvedt S *et al.* The acoustic properties of *Salpa thompsoni*. *ICES J Mar Sci* 2010;**67**:583–93. https://doi.org/10.1093/icesjms/fsp263

Handling Editor: Alina Wieczorek