ICES Journal of Marine Science, 2025, Vol. 82, Issue 6, fsaf094
https://doi.org/10.1093/icesjms/fsaf094
Received: 31 October 2024; **revised:** 14 May 2025; **accepted:** 29 May 2025
Original Article

Effects of fish behaviour on abundance and length frequency estimates from in-trawl stereo cameras

Taraneh Westergerling [1,2,*], Maria Tenningen [1], Katja Enberg [2], Shale Pettit Rosen [1]

[1]Institute of Marine Research, P.O. box 1870 Nordnes, 5817 Bergen, Norway
[2]Department of Biological Sciences, University of Bergen, 5006 Bergen, Norway

*Corresponding author. Department of Biological Sciences, University of Bergen, Thormøhlens gate 53B, 5006 Bergen, Norway. E-mail: taraneh.westergerling@uib.no

Abstract

In-trawl stereo cameras can provide fine-scale spatial and temporal information on species along the trawl path and record small-sized and fragile organisms typically absent from catches. Reliable estimates of abundance and length frequency from in-trawl cameras will improve ecosystem understanding and lessen the need for physical catches on scientific surveys. However, determining these estimates from camera footage is challenging since the same individual can appear in multiple frames and swim repeatedly in and out of the camera's field of view. The manual image analysis performed in this study provides important information on how the swimming behaviour of three abundant pelagic taxa in the Norwegian Sea, along with a camera's field of view and frame rate, affect the number of repeated appearances. Moreover, these manual annotations serve as a valuable dataset for validating automatic image analyses. Our results show that, depending on the taxa, swimming orientation, length, density of individuals, and distance to the camera affect the extent of time an individual is observed. If the repeated appearance of individuals is not accounted for, taxa or length classes with fewer appearances are under-represented in relative abundance and lead to skewed length frequency distributions. Compared to herring and blue whiting, a large fraction of mesopelagic fishes remains undetected during automatic analysis (RetinaNet). Assessing the factors driving repeated appearances improves our understanding of in-trawl camera data and highlights the importance of integrating tracking with automatic image analysis.

Keywords: Deep Vision; image analysis; machine learning; mesopelagic fish; pelagic fish; pelagic survey; RetinaNet; video-trawl survey

Introduction

Aside from being a fundamental tool in commercial fishing, trawls are a standard sampling gear on fisheries scientific surveys (Clark et al. 2016). In acoustic abundance estimation, trawl samples are used to obtain population parameters such as species, size, age, and sex composition of the target organisms recorded by the echosounder (Gunderson 1993). For swept-area surveys, trawl catches are used directly to monitor long-term fluctuations in abundance (Gunderson 1993, Fréon and Misund 1999). However, trawls do not provide fine-scale spatial information since individuals entering the net at different time points and locations along the trawl path are accumulated and mixed in the codend (Kracker 1999). Furthermore, trawl nets can introduce large selectivity and catchability biases depending on the mesh size, sampling volume, species, and size-dependent avoidance behaviour (Pope 1975, Gunderson 1993).

To address some of these limitations, advance scientific trawl sampling, and improve catch information in commercial fisheries, several research institutions and companies have developed in-trawl cameras (Williams et al. 2010, Rosen and Holst 2013, Stokesbury et al. 2017, Fernandes et al. 2021, Krag et al. 2023). These systems can gather quantitative data on species composition, body lengths, and densities. Applying in-trawl cameras to fisheries scientific surveys could facilitate innovative sampling strategies, such as open codend trawling, which allows for the release of fish at the depth of capture and increases survival rates (Rosen et al. 2013, DeCelles et al. 2017, Trenkel et al. 2019). Moreover, in-trawl cameras can provide catch data at a finer resolution than traditional trawls while recording additional information on small and fragile species (e.g. gelatinous organisms) usually lost through the codend meshes (Wileman 1996, Rosen and Holst 2013, Allken et al. 2021). Such sampling methods can provide valuable information for ecosystem-based management (Underwood et al. 2014).

Accurate estimates of abundance and length frequencies are essential for image data to replace measurements from physical samples in fisheries scientific surveys. However, obtaining these estimates from in-trawl cameras is more challenging than from the catch, as the same individual can be imaged multiple times (hereafter: recurrence count), resulting in repeated measurements. Among fish, endurance and maximum swimming speed differ between species and generally increase with body length and temperature (Wardle 1980, He 1993). During the capture process, fish encounter a range of stressors such as exhaustion, injury, crowding, hypoxia, and exposure to artificial light when using in-trawl cameras (Nguyen and Winger 2019, Breen et al. 2020). All these have the potential to induce changes in swimming behaviour (He 1993, Underwood et al. 2018), affecting the number of times an individual is imaged by an in-trawl camera.

Regardless of swimming capacity, recurrence count is also affected by the physical principles on which cameras operate. A camera's field of view (FOV) increases with distance. Therefore, an individual that is farther away will be imaged

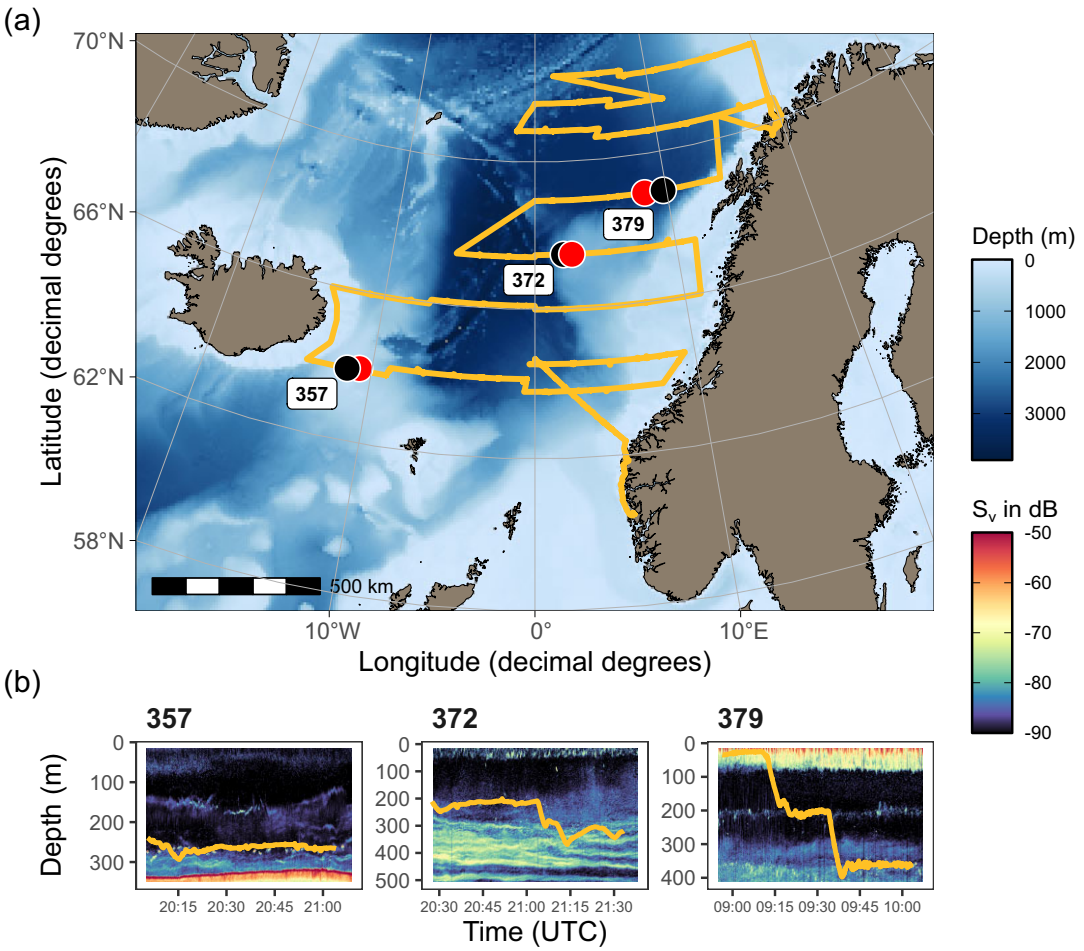

Figure 1. International Ecosystem Survey in the Nordic Seas (IESNS) in 2018: (a) map of the Norwegian Sea featuring acoustic transect lines (yellow), locations of pelagic hauls (black circles) used for behavioural observations inside the Deep Vision camera system, and positions of corresponding CTD casts (red circles). (b) 38 kHz echograms of the three pelagic trawl stations (357, 372, 379) used in this study with trawl path (yellow line) superimposed over acoustic backscatter. Colour scale refers to volume backscattering (S_v), with blue-black representing the weakest and red the strongest echoes.

more times than a closer individual passing at the same velocity. Moreover, high frame rates contribute to fewer individuals being missed and increase the likelihood of recording an image suitable for measurement or identification, but also lead to a higher recurrence count.

The volume of collected image data makes manual image analysis unfeasible (Rosen et al. 2013, Underwood et al. 2014). Previous studies on automating the extraction of counts have either applied linear regressions with the catch (Allken et al. 2021) or used tracking, which involves linking objects from consecutive frames to a single individual (Chuang et al. 2015, Avsar et al. 2023) to handle recurrence counts. Linear regression models rely on catch data, which is not available for species which are lost through codend meshes and may become inaccessible once open-codend trawling is implemented. Tracking can be challenging at low frame rates (Chuang et al. 2015) or in high-density scenarios where objects occlude each other (Blackman and Popoli, 1999).

The aim of this study was, therefore, to investigate the factors driving recurrence counts (environment, swimming behaviour, FOV, and frame rate) and to evaluate how this recurrence count impacts the abundance and length frequency estimates derived from in-trawl cameras for three taxa commonly captured in an ecosystem survey for pelagic fishes in the northeastern Atlantic Ocean. This investigation enhances the overall understanding of in-trawl camera data and may contribute to improving tracking methods.

Materials and methods

Acoustic-trawl survey

Acoustic, catch, image and environmental data were collected in the Norwegian Sea between the 5th and the 20th of May 2018 as part of the International Ecosystem Survey in the Nordic Seas (IESNS). Following a stratified transect design model (Jolly and Hampton 1990), the Norwegian research vessel "G.O. Sars" collected acoustic data along parallel, equal-distance transects (Fig. 1a). Pelagic trawls were deployed upon acoustic registration of fish schools and/or sound scattering layers (SSL: layers of fish and/or zooplankton as observed by the echosounder). At each trawling station, CTD (conductivity, temperature, depth meter) casts were taken to measure oceanographic conditions at every meter from the surface to 10 m above the bottom or until 1000 m depth.

A pelagic trawl (Multpelt 832) was towed for ~1.5 h at a constant speed over ground of 1.6 to 1.8 m s^{-1}, sampling one or two depth layers (hereafter: trawl layers,

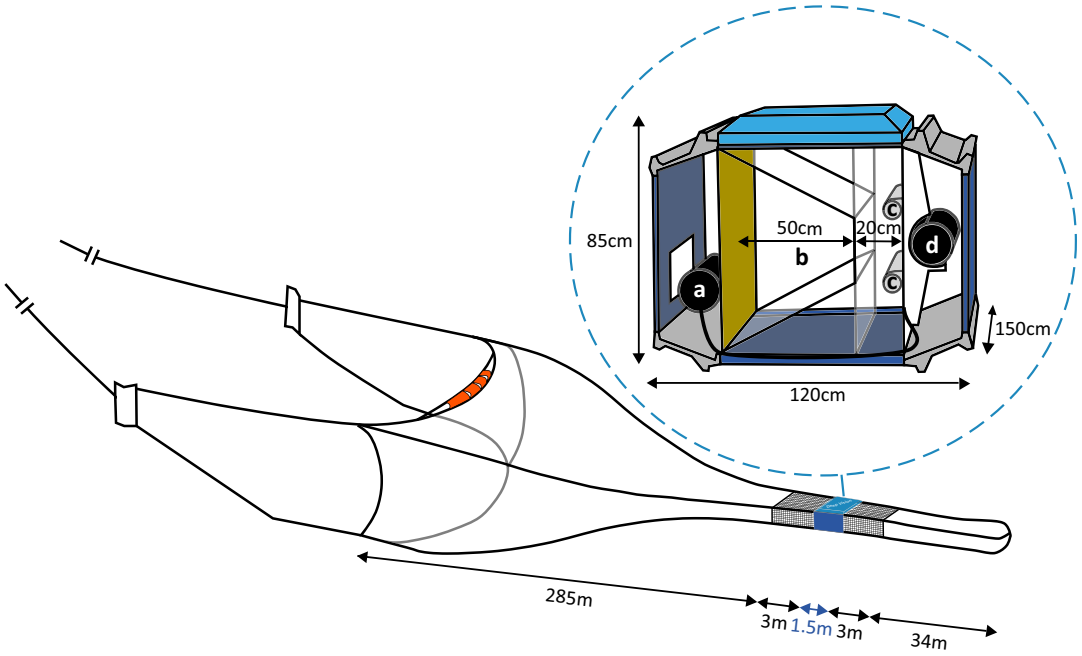

Figure 2. Position of the Deep Vision in-trawl camera system inside the Multpelt 832 pelagic trawl. Inset: detail of the Deep Vision model deployed during the 2018 IESNS shown from the side that faces the codend: (a) battery, (b) imaging chamber, (c) two stroboscopic lights, and (d) stereo camera including a depth sensor. All catch is guided in and out of the imaging chamber by leading panels 3 m in front and behind the Deep Vision system (shaded grey in overview drawing).

Fig. 1b). The Multpelt 832 has a vertical opening of 30–35 m, with mesh size reducing from 16 m inside the trawl opening to 40 mm in the codend (ICES 2018). For catches considered to be large on this survey (>150 kg), total catch weight was estimated from the volume of catch in the codend before subsampling. For medium-sized catches (100–150 kg), the total weight was measured accurately before subsampling, while for small catches (<100 kg), the entire catch was weighed and processed. Since the IESNS is an acoustic-trawl survey, the main information from the catch is relative species abundance and length frequency, not total catch size.

Image collection

In-trawl images were collected using the Deep Vision camera system (Scantrol Deep Vision AS, Bergen, Norway) mounted between the extension and codend of the trawl (Fig. 2). The box-shaped frame is 85 cm high × 120 cm wide × 150 cm long and holds digital colour cameras in a stereo arrangement, synchronised LED strobes for illumination, and a depth sensor (Fig. 2). The FOV of each camera has a horizontal angle of ca. 94.4° underwater. The two LED strobes together have a flux of 30 000 lm, with peak output at 445 and 555 nm wavelengths. Images were collected at a rate of 5 frames per second and recorded to internal memory with depth and time stamp. All fish are guided through a trapezoidal imaging chamber within the vertical FOV of the cameras. The system is described in greater detail in Allken et al. (2021).

Three of 35 trawl stations were selected for further image analysis. These stations represented a range of latitudes and bathymetric structures (Iceland–Faroe Ridge, Vøring Plateau, and Lofoten Basin) (Fig. 1a) and sampled several distinct acoustic features (fish schools and/or two or more sound scattering layers) (Fig. 1b). The limited number of stations resulted from the lengthy processing time needed for manual image analysis.

Manual image analysis

Manual image analysis (hereafter: manual IA) was performed with the Deep Vision Analysis Software (DVAS) version 3.3RC9 (Scantrol Deep Vision AS, Bergen, Norway). The DVAS is a centreline annotation tool, primarily developed to identify and length-measure objects using both images from the left and right cameras. The software reads a calibration file and rectifies each stereo image pair following the procedures described by Hartley and Zisserman (2003) and Garcia et al. (2020). Since most individuals were recorded over several consecutive images, the image most suitable for length measuring was chosen for manual annotation (fish fully within the image, presented side-on to the camera) (Fig. 3).

Four types of information were recorded at each annotation: (i) identification to the lowest possible taxonomic unit (Atlantic herring (*Clupea harengus*, hereafter: herring), blue whiting (*Micromesistius poutassou*), or mesopelagic fishes as a group containing *Maurolicus muelleri*, *Benthosema glaciale*, and other species of the Myctophid family present in the Norwegian Sea), (ii) body length (fork length, FL, for herring and blue whiting; standard length, SL, for the mesopelagic fishes), (iii) swimming orientation when first entering the imaging chamber (towards the codend or facing the opening of the trawl), and (iv) recurrence count from right camera image (Fig. 3).

In fishes with a straight body posture, length measurement used two points: one at the snout and one on the tail fork, or the caudal peduncle for mesopelagic fishes. For curved body postures, additional measuring points along the fish's centre-

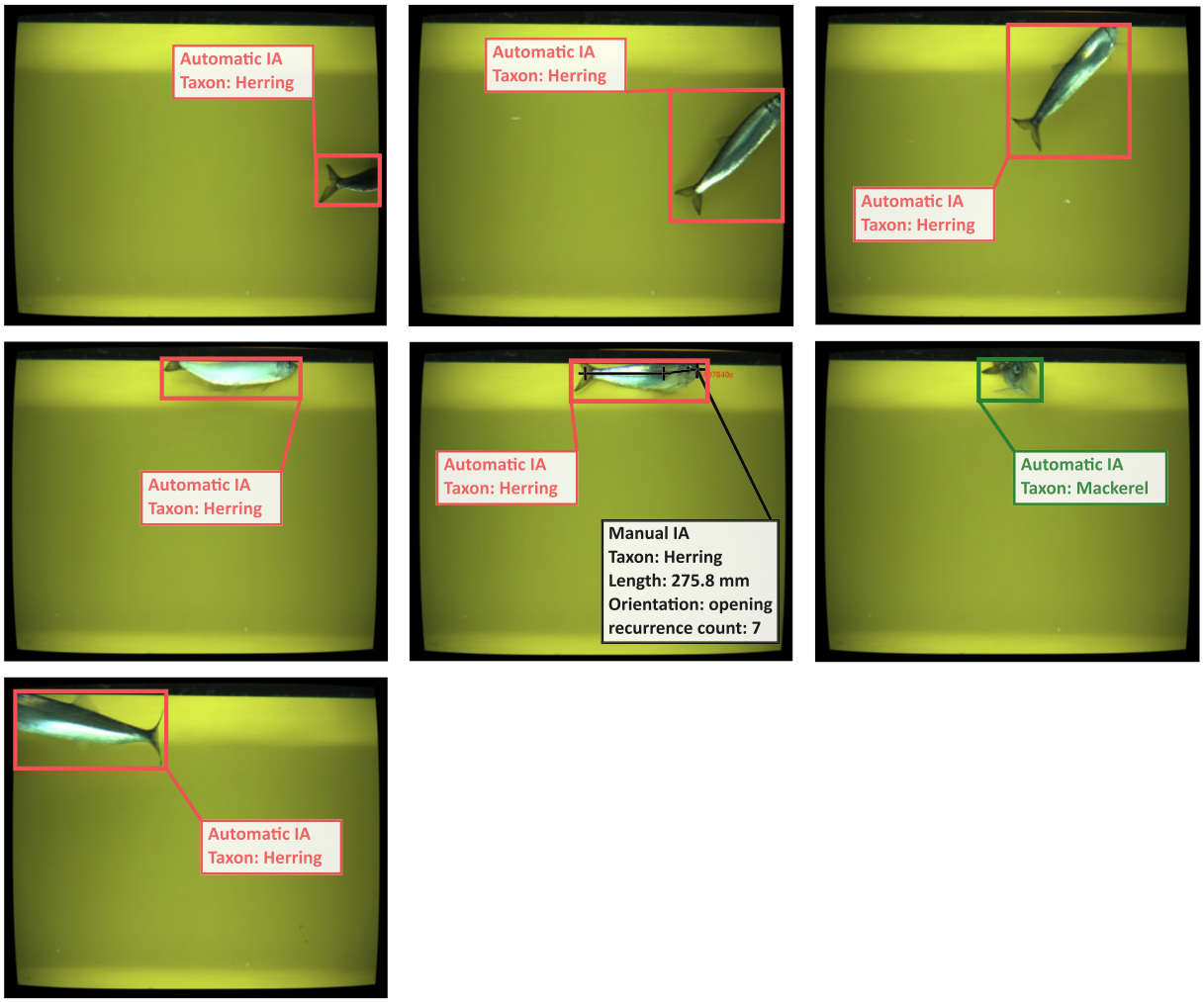

Figure 3. Example sequence of consecutive rectified images of a single herring. Centreline annotation and information in black are from manual image analysis (manual IA): taxon; fork length; swimming orientation when the fish first entered the camera system (trawl opening/codend); number of times a fish was photographed. Boxes from the automatic image analysis (automatic IA) of detected objects are labelled and coloured according to taxon assigned (red: herring, green: mackerel). In the sixth image of the sequence, when the herring was photographed head-on, the automatic IA mistakenly classified it as mackerel.

line were necessary. Sometimes an individual entered but did not fully pass through the imaging chamber before swimming out of the FOV, or an individual which had previously passed fully through the imaging chamber swam into the FOV again. In both cases, this individual was assumed to be the previous fish of that taxon for the purpose of calculating recurrence counts.

Automatic image analysis

Automatic image analysis (hereafter: automatic IA) was conducted with a deep learning algorithm developed by Allken et al. (2021) based on the RetinaNet object detection network. The trained algorithm works on an image-by-image basis, generating detections in the form of bounding boxes around every object in an image and classifying them as either blue whiting, herring, Atlantic mackerel (*Scomber scombrus*, hereafter: mackerel), or mesopelagic fishes (Fig. 3). Since the automatic IA does not integrate tracking and is not trained to differentiate between individuals, repeated detections of the same individual in consecutive images result in inflated counts (Allken et al. 2021). The automatic IA was performed on both the left and the right camera. However, only the output from the left cameras was used in this study.

Datasets

Individuals were identified to species level, and length was measured from the trawl catch using standard sampling protocols for the survey. Images were reviewed by a trained operator, and all individuals were identified to the lowest possible taxonomic group, length measured, and annotated with swimming orientation and recurrence count (Manual IA). The same images were also analysed using the RetinaNet deep learning algorithm developed by Allken et al. (2021) to detect and identify each object (Automatic IA).

Effect of swimming behaviour on observation period

Response variable

To investigate how fish behaviour affects the repeated appearance of an individual, recurrence count was standardised with respect to FOV and the frame rate. The trapezoid cross-section of the Deep Vision imaging chamber already accounts for the vertical increase in the camera's sampling area with distance (Fig. 2). Therefore, only the horizontal FOV needed to be considered. First, the FOV (mm) of the right camera was determined for each individual based on the 94.4° angle (θ) and the individual's distance to the camera (Z) (Equation 1). The

Table 1. Description of parameters used in the analysis of the observation period of herring, blue whiting, and mesopelagic fishes inside the Deep Vision in-trawl camera

Parameter	Categorical/continuous	Unit	Description
Response variable			
Observation period	Continuous	s	Time a fish is observed, standardized for the camera's FOV
Predictors			
Station	Categorical		Sampling station (357, 372, 379)
Elapsed time	Continuous	s	Elapsed trawling time for each sampled layer where depth and vessel speed over ground were constant.
Distance	Continuous	mm	Distance of a fish to the camera's lens
Orientation	Categorical		Direction in which the fish faced when first imaged (towards the codend or facing the opening of the trawl).
Length	Continuous	mm	Total length (TL) for herring and blue whiting, standard length (SL) for mesopelagic fishes.
Density	Continuous		Average number of automatic detections across the frames in which an individual was observed.

method for calculating Z is detailed in the next section. Then, a correction factor was determined by calculating the ratio of FOV to the total width of the imaging chamber (150 cm) (Equation 2). Finally, a standardised observation period (hereafter: observation period) was calculated for every individual by dividing the recurrence count by the correction factor and the frame rate (5 images s^{-1}) (Equation 3).

$$FOV_i(mm) = 2 \times Z_i \times \tan\left(\frac{\theta}{2}\right), \quad (1)$$

$$correction\ factor_i = \frac{FOV_i}{width\ of\ the\ imaging\ chamber(mm)} \quad (2)$$

$$observation\ period_i(s) = \frac{recurrence\ count_i}{correction\ factor_i} \times frame\ rate, \quad (3)$$

where i is the target individual. This correction ensures that any remaining differences in observation period can be attributed to environmental factors or fish behaviour.

Predictor variables

Station, elapsed trawling time, distance to the camera, swimming orientation, body length, and fish density served as predictor variables for observation period (Table 1).

Using rectified images, the distance of an object from the camera lens (Z) was calculated as

$$Z = \frac{f \times B}{x_l - x_r}, \quad (4)$$

where f is the fixed focal length (set to 906 pixels), B (baseline, 60 mm) is the distance between the centres of the left and right camera lenses, and $x_l - x_r$ is the disparity (d) in the x coordinates of the first length measuring point (typically the snout of the fish) between the left and right image (Fig. 4).

Body length in herring and blue whiting (species with forked tails) was measured as FL because the DVAS relies on matching the same landmark on both the left and right images. Before the analysis, FL was converted to total length (TL) to conform to catch measurements. The conversion factor for herring was retrieved from "FishBase" (www.fishbase.org) ($TL = FL/0.917$) (Ojaveer et al. 2003). For blue whiting, FL and TL were measured from 100 individuals ranging 20.1–32.0 cm (TL) collected during a different survey in the Norwegian Sea in 2021, and a regression was calculated from those ($TL = 0.0233 + FL \times 1.03$) (Fig. S2). A comparison of length frequency distributions estimated from the catch and images showed that measurements using more than two centreline points resulted in an overestimation in length (Fig. S5). Therefore, this analysis focused solely on annotations with two points, where the length frequency showed no significant difference from the catch.

Fish density was estimated from the average number of automatic detections per frame within the time interval the target individual spent in the system since the number of fish per frame was not recorded during the manual IA (Fig. 3). The time interval was calculated based on the recurrence count of an individual, assuming that the manual annotation occurred midway as the fish passed through the imaging chamber.

Statistical analysis

Using R version 4.3.1 (R Core Team 2020), separate statistical models were developed to identify the factors affecting the observation period of each taxon. Data exploration followed the protocol described by Zuur et al. (2010). The analyses focused on the data collected inside each trawling layer (Fig. 1b), where constant speed could be assured (minimal spooling in or out of the trawl warps). A generalised linear model (GLM) following a gamma (link = "log") distribution was chosen:

$$observation\ perod_i \sim Gamma(\mu_i, \nu), \quad (5)$$

$$E(observation\ period_i) = \mu_i, \quad (6)$$

$$var(observation\ period_i) = \mu_i^2/\nu, \quad (7)$$

where i represents the target individual, μ the mean observation period, ν the shape parameter of the distribution, and E the expected observation period (Zuur et al. 2009).

Station, water temperature, trawling depth, vessel speed over ground, elapsed time, distance to the camera, swimming orientation, body length, and density were initially considered as predictors of the observation period (Table 1). Both water temperature, trawling depth, and vessel speed over ground were excluded due to limited contrast in those variables. Nonetheless, the predictor variable station still captured any differences in temperature and vessel speed between the stations. No mixed effects were included because of the low number of stations ($n = 3$).

Data exploration revealed a strong correlation between station and density and between station and elapsed time for all taxa. For herring, there was also a correlation between sta-

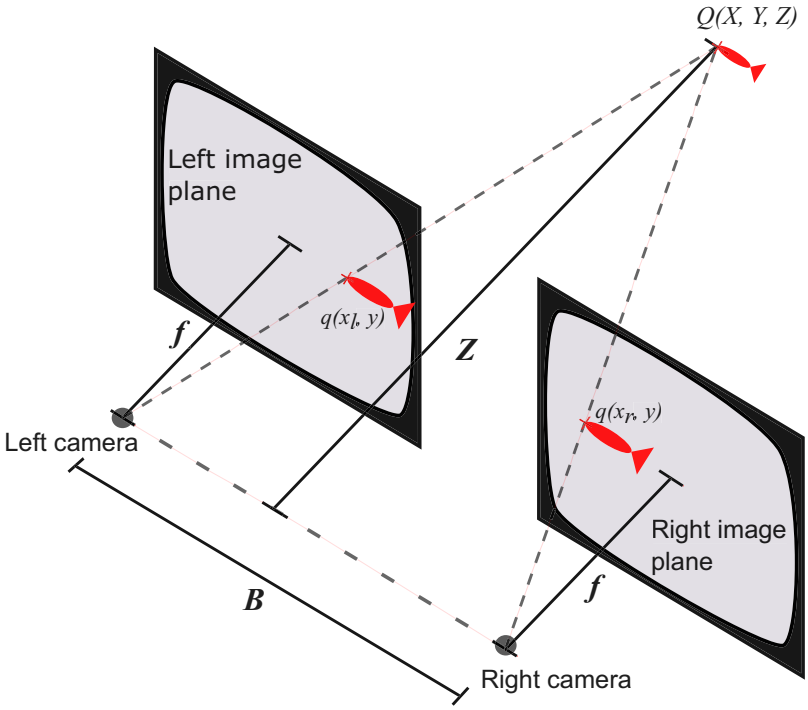

Figure 4. Stereo triangulation scheme on rectified images: (*B*) baseline distance between the centre of the left and right camera lenses; (*f*) fixed focal length; (*Z*) perpendicular distance of an object to the baseline; (*Q*) Cartesian coordinates of the fish in the imaging chamber; (*q*) first length measuring point on the snout of the fish here used as the location of the fish in right and left image planes [modified from Adil et al. (2022)].

tion and length, and between with elapsed time and density. Blue whiting exhibited an additional correlation between station and length. To avoid multicollinearity in the full model, two or three separate models were formulated for each taxon (Table 3, M1a, M1b, M1c, M2a, M2b, M3a, M3b). For all full models that included an interaction between orientation and length, its significance was tested using the "drop1()" function from the package "stats" (R Core Team 2020). The full model with the lowest Akaike information criterion (AIC) was then chosen for model selection (Table 3, M1a, M2a, M3b). Backward stepwise selection was performed using the "step()" function from the package "stats" (R Core Team 2020). Based on this method, the term that led to the largest reduction in AIC was dropped at each step until no further reduction was achieved.

Model assumptions were verified by plotting residuals against fitted values and against each covariate in the model by means of functions in the package "car" v. 3.1–2 (Fox and Weisberg 2019) (Fig. S3). Simultaneously, a lack-of-fit test for each continuous covariate was computed, where $P < .05$ indicated nonlinear patterns in the Pearson residuals. Additional diagnostics plots were produced using the "check_model()" function from the package "performance" v. 0.10.8 (Lüdecke et al. 2021). This step of the model validation process involved a posterior predictive check, testing for homogeneity of variance, normality of residuals and the detection of influential observations via Cook's distance (Fig. S4).

Effect of field of view and frame rate on recurrence count

Underwater, the horizontal FOV of the Deep Vision was estimated to increase by 2.16 cm for every centimetre of increased distance from the camera (Equation 1). Therefore, at any given velocity and frame rate, individuals passing further away are imaged more times. To test whether this change in sampling area affected all three taxa equally, we assessed the empirical cumulative distribution functions (ECDFs) of individual counts along distance.

Frame rate and recurrence counts are directly proportional until a recurrence count of one is reached (individual imaged just once). Beyond this, a further decrease in frame rate either results in an individual being missed or being imaged once. Each individual was manually annotated in only one of the frames it appeared in (Fig. 3). Therefore, to assess the impact of frame rate, we created an "artificially populated" dataset from the original one provided by the manual IA. This involved populating all the frames based on the recurrence count, assuming the manual annotation of each individual occurred midway through its passage through the FOV. A simulated reduction in frame rate from 5 images per second to 3 images per minute was then achieved by increasing the time interval in 0.2-s intervals between images selected. To assess whether the decrease in frame rate affected the detection probability of all three taxa equally, we calculated the proportion of individuals recorded at each frame rate relative to the number of individuals counted at 5 frames per second.

Effect of recurrence count on abundance

To compare count estimates from the manual IA with the catch and automatic IA, the total number of individuals and the total number of objects (sum of all recurrence counts) were extracted from the manual IA. Throughout the study, the output from the manual IA was assumed to reflect the true number of individuals and objects.

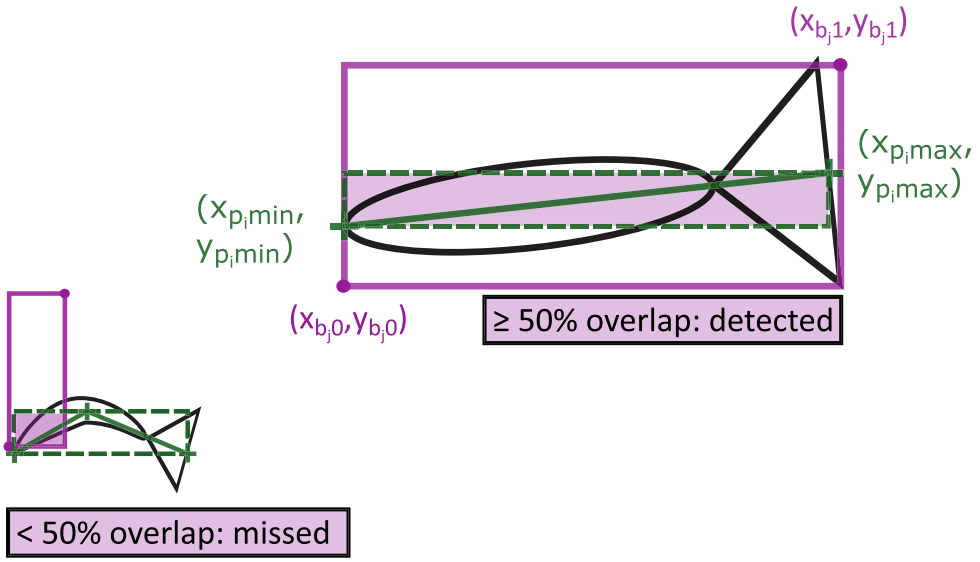

Figure 5. Example of bounding boxes from manual annotations (dashed green, where p_i represents the manual measuring points ($i = 0, \ldots, m$) of the target individual) and automatic detections (solid purple, where b_j represents the automatic detection box j). Solid green lines are centrelines of fish determined during manual labelling and define the vertices of the manual annotation bounding boxes. The shaded area represents the overlap between the two boxes (light purple).

Effect of recurrence count on length frequency

In fisheries scientific surveys, a sample size of 100 individuals per species is commonly considered sufficient for estimating the length frequency distributions of the catch (Gerritsen and McGrath 2007). Therefore, we investigated whether reducing the frame rate could prevent multiple measurements of the same individual and yield accurate estimates of the length frequency. This analysis was conducted on the artificially populated dataset previously created to test frame rate, but was limited to frame rates and stations containing at least 100 length measurements per taxa to guarantee sufficient sample size. Consequently, the data was restricted to herring from station 357 with frame rates of ≥0.11 images per second, blue whiting from station 379 with rates of ≥0.3 images per second, and mesopelagic fishes from stations 372 and 379, with frame rates of ≥0.24 and ≥0.39 images per second, respectively. Measurements were confined to intervals when trawling at a constant depth and speed. The object-based length frequency distributions were visually compared to the individual-based length frequencies obtained from the manual IA.

Recall of automatic image analysis

Standard evaluation metrics for object detection models require manual annotation of each object in an image to extract true positives (TP), false positives (FP), and false negatives (FN), which are then used for calculating precision and recall (Allken et al. 2021). During manual IA, individuals were annotated only in one of the multiple frames they appeared in (Fig. 3). Consequently, images could contain objects that were not annotated, limiting the analysis to TP, FN, and recall values. Moreover, the centreline annotations used in this analysis were initially produced as length measurements during the manual IA and required the conversion to bounding boxes (defined as $x_{p_imin}, y_{p_imin}, x_{p_imax}, y_{p_imin}$) (Fig. 5). Since the generated manual bounding boxes were significantly smaller than the automatic detections (defined as $x_{b_j0}, y_{b_j0}, x_{b_j1}, y_{b_j1}$), an alternative method to the standard Intersection over Union (IoU) was applied (Allken et al. 2021).

A total of 3674 manual annotations were compared with the automatic detections. First, the automatic detection with the shortest distance to the manual bounding box was matched to the manual annotation. Then we calculated the overlap between the two boxes. Only automatic detections overlapping at least 50% with the manual bounding box were considered to have successfully "detected" an object (Fig. 5). Manual annotations with a lesser degree of overlap were considered "missed." Detected objects where the manual and automatic IA assigned the same taxa were deemed "correct", and ones where the classification differed were considered "wrong."

Recall measures the model's effectiveness at finding positive instances and is calculated as the ratio between TPs ("correct") and the sum of TPs and FNs ("missed" plus "wrong"). We calculated one recall value for each taxon (n):

$$Recall_n = \frac{TP_n}{TP_n + FN_n}. \quad (8)$$

Results

Environmental parameters and catch composition of the three pelagic trawl stations are presented in Table 2.

Effect of Swimming behaviour on observation period

For all three taxa, the histogram of the observation period resembled that of a gamma distribution, with most individuals spending only a short amount of time in front of the camera (Fig. 6). The parameters of the distribution, however, differed between the three taxa. Herring and blue whiting displayed higher median observation periods (1.9 s, 1.6 s) than mesopelagic fishes (0.7 s) (Fig. 6). Furthermore, herring exhibited a larger spread of values (max: 35 s) than blue whiting (max: 14.1 s) or mesopelagic fishes (max: 2.8 s, Fig. 6).

Table 2. Overview of the three pelagic trawl stations included in this study

Station	Day–month (2018)	Start time	End time	Speed (m s^{-1})	Depth (m)	Temp (°C)	Number of images	Taxa	Number of ind.	Length range (cm)
357	09–05	19:39	21:54	1.6	263	5.5	39 802	Herring	818[a]	30.5–39.0
372	16–05	20:03	22:14	1.6, 1.7	222, 338	5.3, 4.4	38 748	Herring	52	31.0–38.5
								Blue whiting	62	24.5–31.0
								Mesopelagic fishes	421	4.0–11.5
379	20–05	08:34	11:00	1.8, 1.8	206, 363	5.7, 4.5	41 518	Blue whiting	271	20.5–30.5

Speed: average vessel speed over ground at each depth layer; Depth: average depth of each sampled trawl layer (one layer in station 357, two in stations 372 and 379); Temperature: average temperature at each trawl layer; Number of images: number of images recorded by the Deep Vision in-trawl camera system at each station; Taxa: taxa retained in the codend of the pelagic trawl; Number of ind.: number of individuals caught in the codend; Length range: minimum and maximum standard length of mesopelagic fishes and total length of herring and blue whiting in the catch.
[a]Volumetric estimate.

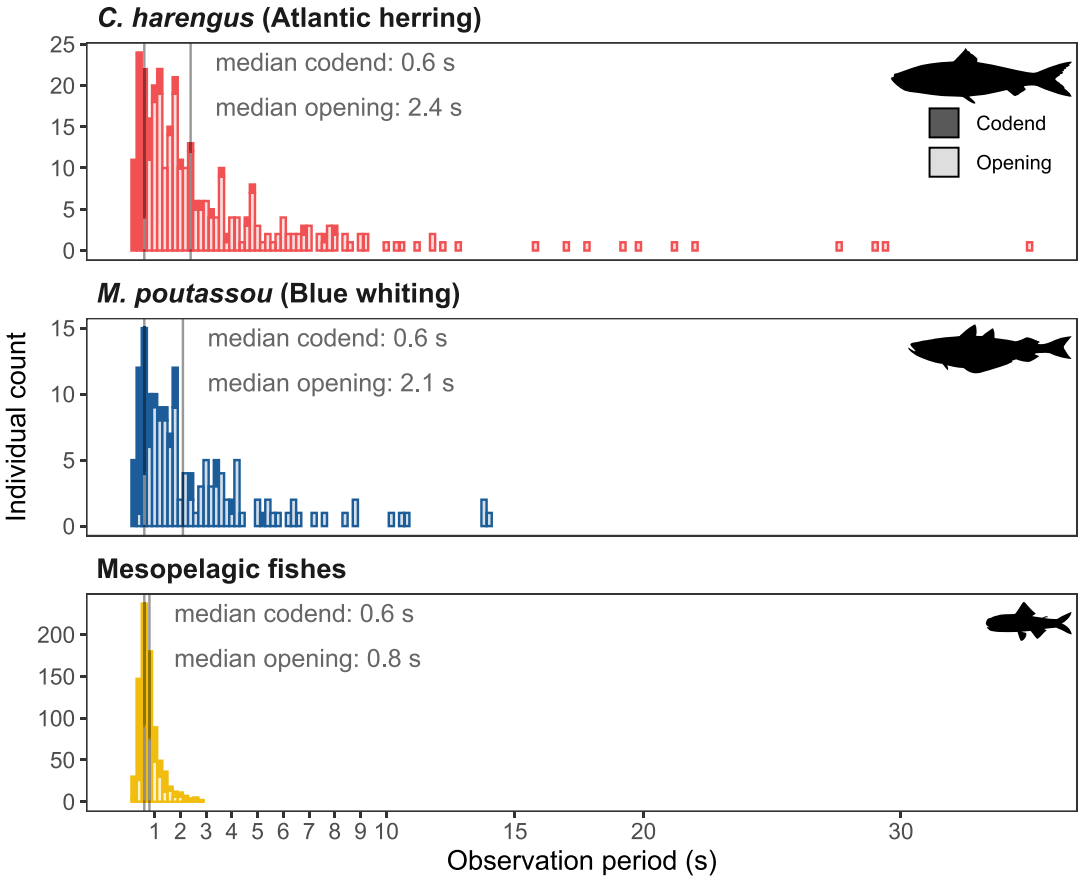

Figure 6. Stacked histogram of observation period for each examined taxon (herring, blue whiting, mesopelagic fishes). Filled bars represent individuals facing the trawl opening and empty bars represent those facing the codend. Data from all three stations were pooled together. The solid lines indicate the median observation periods (s) when facing the codend or the opening of the trawl.

Herring

The observation period of herring within the in-trawl camera system was influenced by the interaction between swimming orientation and body length, as well as density (Table 3, M1d). Herring were oriented predominantly towards the trawl opening (75.3%) and displayed a higher median observation period when facing this direction (2.4 s) than when facing the codend (0.6 s) (Fig. 6). The effect of body length on observation period depended on their orientation ($P = .021$) (Fig. 7a): Fish swimming towards the trawl opening showed a 1% increase in observation period for every additional cm in body length, whereas individuals facing the codend showed a 14% decrease per cm. Density ranged between 0 and 4.33 (median 1.2). With increasing densities, the observation period decreased significantly ($P < .001$) at a rate of 45% with each additional individual (Table 4, Fig. 7d). Station, elapsed trawling time, and distance to the camera were dropped during model selection.

Blue whiting

The most parsimonious model predicting the observation period of blue whiting included swimming orientation and body length (Table 3, M2c). Like herring, the majority of individuals were oriented towards the trawl opening (71.1%) and

Table 3. Description of full and final models predicting observation period inside the Deep Vision in-trawl camera system for each taxon separately

	Model	Covariates	AIC
Clupea harengus **(Atlantic herring)**			
Full model	M1a	distance + length*orientation + density	1299.8
Full model	M1b	distance + orientation + station	1325.6
Full model	M1c	distance + length*orientation + elapsed time	1328.2
Final model	M1d	length*orientation + density	1298.4
Micromesistius poutassou **(blue whiting)**			
Full model	M2a	distance + length*orientation + elapsed time + density	554.8
Full model	M2b	distance + orientation + station	554.5
Final model	M2c	length + orientation	549.5
Mesopelagic fishes			
Full model	M3a	distance + length*orientation + elapsed time + density	575.0
Full model	M3b	distance + length*orientation + station	567.3
Final model	M3c	distance + length + orientation + station	565.6

displayed significantly ($P < .001$) longer median observation periods than when facing the codend (2.1 and 0.6 s, respectively) (Table 4, Fig. 6). Body length had no significant influence on the residence time of blue whiting ($P > .05$) (Table 4). Station, elapsed trawling time, and density were dropped during model selection.

Mesopelagic fishes

For mesopelagic fishes, the final model included swimming orientation, body length, distance to the camera, and station (Table 3, M3c). Contrary to herring and blue whiting, most mesopelagic fishes were oriented towards the codend (61.2%). Mesopelagic fishes also showed significant differences ($P < .001$) in observation period depending on their orientation, with a median observation period of 0.8 s when facing the trawl opening and 0.6 s when facing the codend. (Table 4, Fig. 6). Observation period decreased significantly ($P < .001$) by 11% with every 10 cm of distance to the camera (Fig. 7e). Neither body length nor sampling station had a significant effect on the observation period of mesopelagic fishes (Table 4). Elapsed trawling time and density were dropped during model selection.

During the validation processes, the final models for all three taxa indicated homoscedasticity, a lack of influential observations, and no collinearity. Posterior predictive checks pointed out discrepancies between real and simulated observation periods for herring and mesopelagic fishes but fit well for blue whiting (Fig. S4).

Effect of field of view and frame rate on recurrence count

All three taxa tended to occupy the wider section of the imaging chamber, with 50% of herring, blue whiting, and mesopelagic fishes passing at a distance greater than 62, 60, and 55 cm, respectively (Fig. 8a). Data exploration revealed no correlation between distance and length, nor between distance and density.

A simulated reduction in the frame rate from 5 to 1 image per second led to a loss of 12%, 17%, and 39% of the individuals for herring, blue whiting, and mesopelagic fishes, respectively (Fig. 8b). At a frame rate of 3 images per minute, only 9.5%, 9.3%, and 2% of individuals from the same three taxa were retained.

Effect of recurrence count on abundance

The total number of individuals reported in the catch was 818, 535, and 271 in station 357, 372, and 379, respectively. For stations 372 and 379, the manual IA recorded approximately the same number of herring and blue whiting as in the catches (Table 5). In station 357, where, due to a large catch (>150 kg), weight was estimated from the volume of the codend, herring counts from the manual IA exceeded the codend catch by 49.5% (Table 5). This discrepancy reflects the survey's focus on relative species abundance and length frequency rather than total catch size. Furthermore, all stations showed a greater presence of mesopelagic fishes in the images than what was retained in the catch (Table 5, Fig. 9a). The discrepancy between detections from the automatic IA and individuals from the manual IA was highest for herring (8–13 times higher), followed by blue whiting (6–7 times higher) and mesopelagic fishes (2–9 times higher) (Table 5). The number of objects detected by the automatic IA was slightly lower than the total objects from the manual IA (Fig. 9a).

The catch, manual IA, and automatic IA also differed in their relative species compositions (Fig. 9b), with a higher proportion of herring and blue whiting in the catch, manual object counts, and automatic IA detections compared to the manual individual counts.

No mackerel were registered in the catch or the manual IA, however, the automatic IA misidentified a number of objects as mackerel (Table 5, Fig. 9).

Effect of recurrence count on length frequency

Across the three stations, the lengths of herring, blue whiting, and mesopelagic fishes estimated from the manual IA ranged between 27–42 cm, 19–31 cm, and 1.8–8.2 cm, respectively. This range remained well-represented in object-based length frequency distributions, as long as at least 100 measurements were taken (greyscale curves in Fig. 10). However, depending on the taxon, the frequency with which each length was represented could vary. The greatest difference between individual-based and object-based length frequency distributions was observed for herring (Fig. 10a), where repeated measurements of the same individual led to an overestimation of the mid-range sizes and an underestimation of the extremes (≤33 cm or ≥37 cm). This error persisted regardless of the frame rate. For blue whiting and mesopelagic fishes, repeated measurements did not greatly alter the length frequency distribution (Fig. 10b–c). The individual-based distribution fell

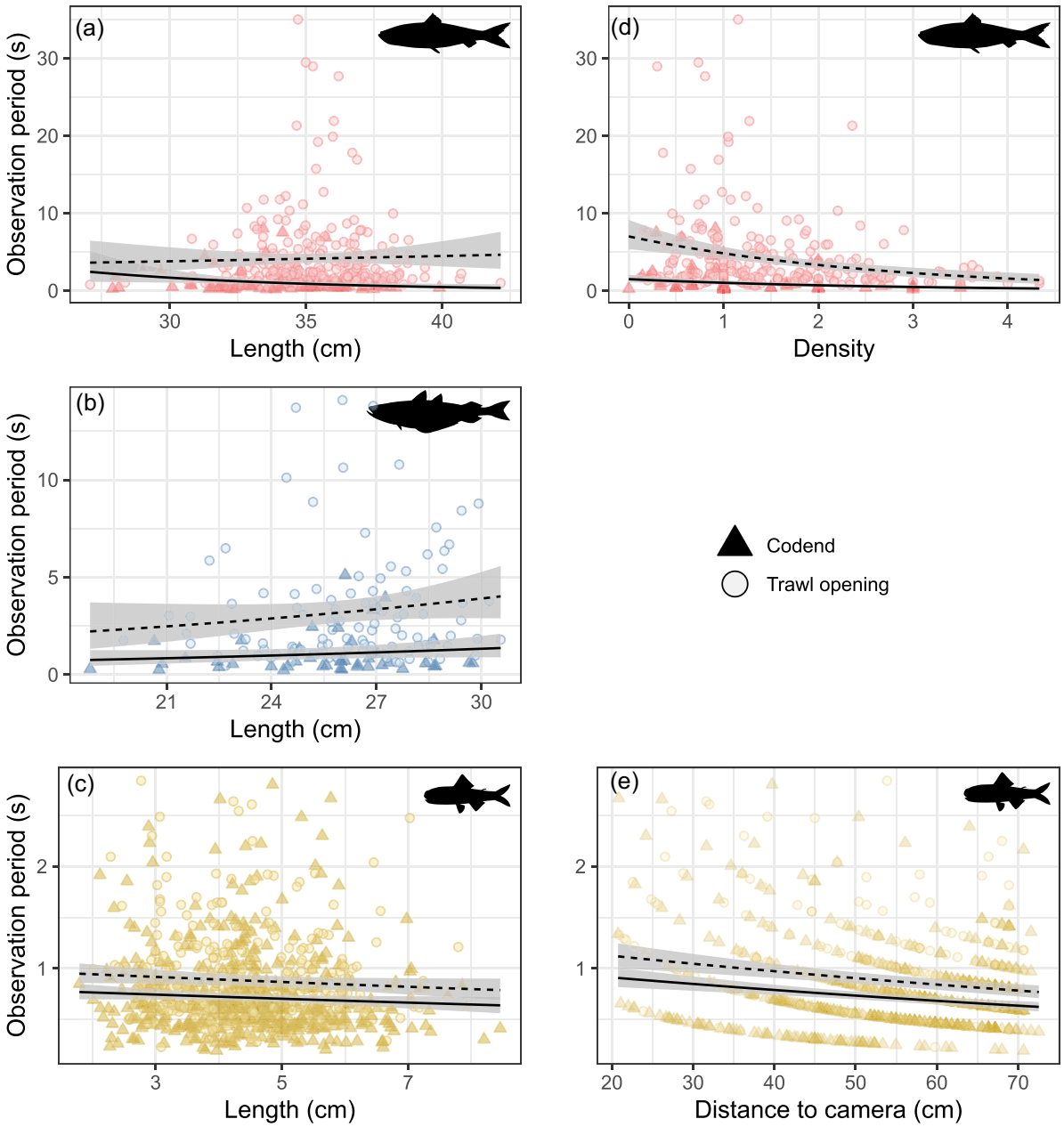

Figure 7. Fit of the Gamma GLMs for fish oriented towards the opening of the trawl (dashed black lines) and facing the codend (solid black lines): (a) effect of body length on the residence time of herring (M1d, density of 1.44) (b) effect of body length on the residence time of blue whiting (M2c), (c) effect of body length on the residence time of mesopelagic fishes (M3c, distance of 53.8 cm, station 372), (d) effect of density on the residence time of herring (M1d, length of 35.0 cm), (e) effect of distance to the camera on the residence time of mesopelagic fishes (M3c, length of 4.5 cm, station 372). Real data projected in the background (filled triangles: facing codend, empty circles: facing trawl opening). NB: the "line pattern" in panel (e) is due to the resolution of the *y* variable (observation period, 0.2 s increments); this is only apparent when the observation period is plotted against distance, as the latter was used to calculate the correction factor (Equations 1–3).

within the object-based distributions irrespective of the frame rate.

Recall of automatic image analysis

Recall values were higher for herring (88%) and blue whiting (87%) than for mesopelagic fishes (46%, Table 6). For all taxa, the greatest source of error was caused by fish being missed by the automatic IA, not by misidentification. The lowest detection rate was observed for mesopelagic fishes, which were missed 54% of the time.

Discussion

In-trawl cameras combined with automatic IA allow for new and improved ways to collect data from marine ecosystems. However, estimating abundance and length frequency is challenging since the same individuals can appear in multiple frames and swim repeatedly in and out of the camera's FOV (Williams et al. 2016, Allken et al. 2021, Avsar et al. 2023). As such, it is important to understand the factors driving recurrence count. The manual IA undertaken in this study provides valuable information on the behaviour of fish near in-trawl

Table 4. Summary of the three most parsimonious GLMs predicting observation period inside the Deep Vision in-trawl camera system for each taxon separately (herring, blue whiting, and mesopelagic fishes): sample size (*n*), estimated regression parameters (beta), standard errors, *z*-values, and *P*-values

Model terms	Beta	Std. error	*z*-value	*P*-value
***Clupea harengus* (Atlantic herring) $n = 316$**				
(Intercept)	4.877	1.761	2.769	.006
Orientation-trawl opening	−3.503	2.148	−1.631	.104
Length	−0.013	0.005	−2.492	.013
Density	−0.372	0.075	−4.945	<.001
Orientation-trawl opening:length	0.014	0.006	2.319	.021
***Micromesistius poutassou* (blue whiting) $n = 152$**				
(Intercept)	−1.237	0.854	−1.449	.149
Orientation-trawl opening	1.079	0.169	6.389	<.001
Length	0.005	0.003	1.533	.127
Mesopelagic fishes $n = 827$				
(Intercept)	0.133	0.130	1.024	.306
Distance	−0.001	0.000	−5.140	<.001
Orientation-trawl opening	0.209	0.036	5.881	<.001
Length	−0.003	0.002	−1.823	.069
Station372	0.043	0.089	0.486	.627
Station379	0.147	0.092	1.607	.109

cameras and a useful dataset for validating automatic image analyses. The results show that taxa behave differently in the trawl and that mesopelagic fishes, because of their lower recurrence count and small size, are under-represented in relative abundance estimates and often remain undetected by the automatic IA.

Effect of swimming behaviour on observation period

Our observation that most herring and blue whiting were oriented towards the trawl opening (against the water flow) is consistent with previous findings based on videos from the aft section of a pelagic trawl (Suuronen and Millar 1992, Suuronen et al. 1997, Skúvadal et al. 2011). Positive rheotaxis, or the reaction of fish to face an oncoming current, is mediated either directly by the flow of water over the body or indirectly through visual, tactile and inertial stimuli (Arnold 1974). Swimming towards the opening of a midwater trawl is also facilitated by a reduction in current velocities. Kroeger (1984) and Skúvadal et al. (2011) measured water velocities of up to two-thirds lower than the trawling speed inside the aft section of pelagic trawls. Mesopelagic fishes, in contrast, tended to face the codend when passing by the in-trawl camera. Due to their limited swimming abilities inside the gear, mesopelagic fishes are not able to maintain position and are therefore displaced opposite to the towing direction (Grimaldo et al. 2022). This is also reflected by their median observation period, which was ca. 60% shorter than that of herring and blue whiting. All three taxa exhibited significantly longer observation periods when facing the opening of the trawl, as swimming in the direction of trawling slowed their passage through the camera system (Fig. 6).

Maximum swimming speed and endurance in fish are generally considered a function of body length (He 1993). The presence of body length in the most parsimonious models of all three taxa highlights its importance in explaining the observation period of fishes. For herring, the effect of body length on observation period was significant ($P < .05$) and depended on orientation. Larger herring facing the codend moved faster with the current and tended to spend less time in the system than smaller individuals. Conversely, larger herring facing the trawl opening demonstrated greater endurance when swimming against the water flow, leading to extended observation periods. For blue whiting, the data showed no significant influence of body length nor an interaction between length and orientation. Skúvadal et al. (2011) observed that blue whiting displayed less stamina than herring and often passively drifted towards the codend upon reaching the lower aft section of a pelagic trawl. The nonsignificant effect of body length on the observation period of mesopelagic fishes supports the hypothesis that, by the time these smaller-sized organisms (1.8–8.2 cm) reached the aft section of the trawl, they were mostly passively transported and not actively swimming.

Herring were observed for significantly shorter periods as densities increased. Underwood et al. (2018), employing a camera with steady white light in the trawl section 5.2 m ahead of the Deep Vision, did not observe any alterations in the passage rates of herring regardless of density. However, studies have shown that strobe light can elicit strong avoidance behaviour and cause visual pigment bleaching in fish (Amaral et al. 2001, Patrick and Poulton et al. 2001, Novales Flamarique et al. 2006). The light emitted by the system could therefore disrupt swimming behaviour within the imaging chamber, while increased densities may amplify the disorientation caused by impaired vision. Furthermore, the confined volume of the imaging chamber might result in density levels higher than those observed ahead of the camera. Density was collinear with station and increased with elapsed trawling time. Therefore, environmental effects such as ambient light, temperature, and current velocities, as well as temporal effects, cannot be excluded from interpreting this result. A larger set of stations and a more targeted sampling design would be needed to quantify these factors.

Distance to the camera (corrected for FOV) was solely retained in the final model for mesopelagic fishes, which exhibited significantly shorter observation periods farther from the camera. Assuming these individuals were passively drifting, this could indicate uneven water flow within the imaging chamber due to boundary layer effects. Consequently, flow velocities inside the imaging chamber may be reduced nearest the camera where the three sides of the trapezoid converge (Fig. 2). Mesopelagic fishes were also more evenly distributed within the imaging chamber, while herring and blue whiting tended to occupy the wider section. This might be further evidence of herring and blue whiting avoiding the light source and mesopelagic fishes being passively transported.

During model validation, both herring and mesopelagic fishes showed some discrepancies between the observed and modelled observation periods (Fig. S4a). The modelled observation period appeared to be less peaked than the real data, indicating that the model was not able to account for the individuals with extreme observation periods (Fig. 6). Fish that left and re-entered the FOV may also have resulted in incorrect observation periods, although there seems to be no reason why this should be biased towards either long or short observation periods.

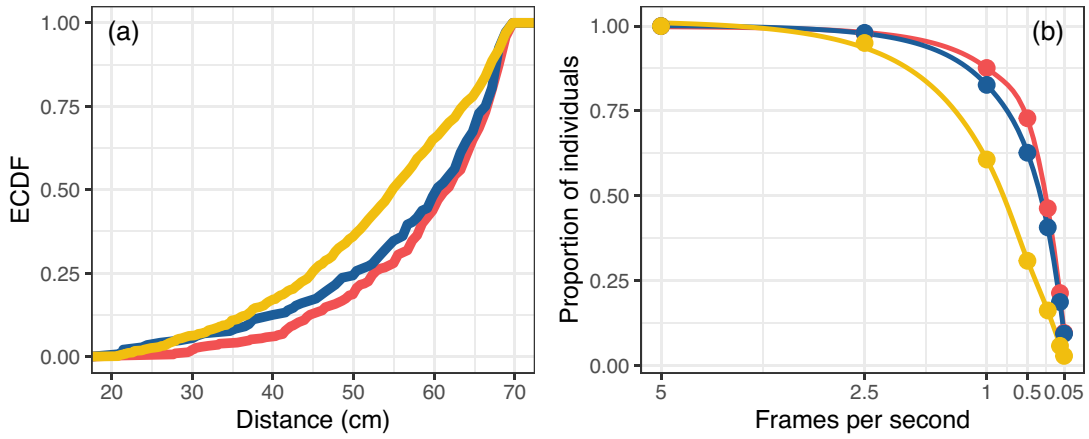

Figure 8. (a) Empirical Cumulative Distribution Functions (ECDF) of individual counts along distance from the camera. (b) The proportion of individuals recorded at each frame rate relative to the number of individuals counted at 5 frames per second. Colours refer to the three taxa herring (red), blue whiting (blue), and mesopelagic fishes (yellow). For clarity, only a selected number of frame rates (images s^{-1}) are exemplified here (5, 2.5, 1, 0.5, 0.05).

Table 5. Overview of the absolute number of individuals counted in the manual IA and the trawl catch, as well as total object counts from manual IA (each individual counted as many times as it was imaged) and number of detections from the automatic IA

Station	Individuals manual IA	Individuals catch	Objects manual IA	Detections automatic IA
Clupea harengus (Atlantic herring)				
357	1223	818[a]	12 591	9624
372	67	52	1122	866
379	2	0	4	209
Micromesistius poutassou (blue whiting)				
357	0	0[a]	0	432
372	74	62	526	500
379	297	271	2422	1706
Mesopelagic fishes				
357	81	0[a]	264	760
372	1187	421	3503	2785
379	722	0	2802	1520
Scomber scombrus (Atlantic mackerel)				
357	0	0[a]	0	495
372	0	0	0	142
379	0	0	0	103

Atlantic mackerel (*Scomber scombrus*) is included since it was mistakenly detected by the automatic IA.
[a]Volumetric estimate.

Effect of field of view and frame rate on recurrence count

FOV increases with the distance from the camera. The impact of this change in sampling area is partially reduced by the design of the Deep Vision imaging chamber, which prevents fish from occupying vertical spaces outside the camera's FOV. However, the horizontal effect persists, causing individuals that are further away to be imaged more often than closer individuals passing at the same velocity. Consequently, mesopelagic fishes, which tended to pass closer to the camera than herring and blue whiting, had lower recurrence counts (Fig. 8a).

A high frame rate increases the recurrence count but prevents the loss of rapidly passing individuals. Frame rate and recurrence counts are directly proportional until an individual is imaged just once. Beyond this, further decreasing the frame rate can only result in an individual being imaged once or missed completely. The estimated numbers of herring and blue whiting individuals from the in-trawl camera and the catch showed only minor discrepancies (except for herring at station 357 due to volumetric estimation). This indicates that these taxa tend to remain in the camera's FOV for >0.2 s, and a frame rate of 5 images per second is adequate to capture most individuals. Nonetheless, 4%, 5%, and 11% of herring, blue whiting, and mesopelagic fishes, respectively, had a recurrence count of only one. This would suggest that at a frame rate of 5 images per second, the likelihood of missing individuals persists and is greatest for mesopelagic fishes. Thus, it is advisable to increase the frame rate until all the individuals are imaged at least twice.

FOV and frame rate can also cause variations in recurrence counts between length classes. For example, if length and distance to the camera are positively correlated, the increase in FOV with distance would cause larger fish to have a higher recurrence count than smaller ones. The statistical data exploration conducted in this study, however, revealed no correlation between length and distance for any of the three taxa. Herring of ≤33 cm or ≥37 cm length showed overall lower recurrence counts, and would therefore be increasingly missed if the frame rate is reduced.

Effects of recurrence count on abundance and length frequency

The continuous footage from in-trawl cameras provides detailed records of the time and depth at which organisms were captured. However, the added information comes at the cost of the same individual being imaged multiple times. If not properly addressed, taxa with lower recurrence counts (e.g. mesopelagic fishes) become underestimated in relative abundance estimates (Fig. 9). For taxa where the observation period significantly varies with length (e.g. herring, Table 4), certain sizes become misrepresented in the length frequency distributions (Fig. 10a).

Lowering the frame rate to avoid the repeated appearance of individuals did not affect all taxa and length classes equally. Mesopelagic fishes were increasingly under-represented due to their lower recurrence count relative to the other species (Fig. 8b). For herring, the over-representation of certain length

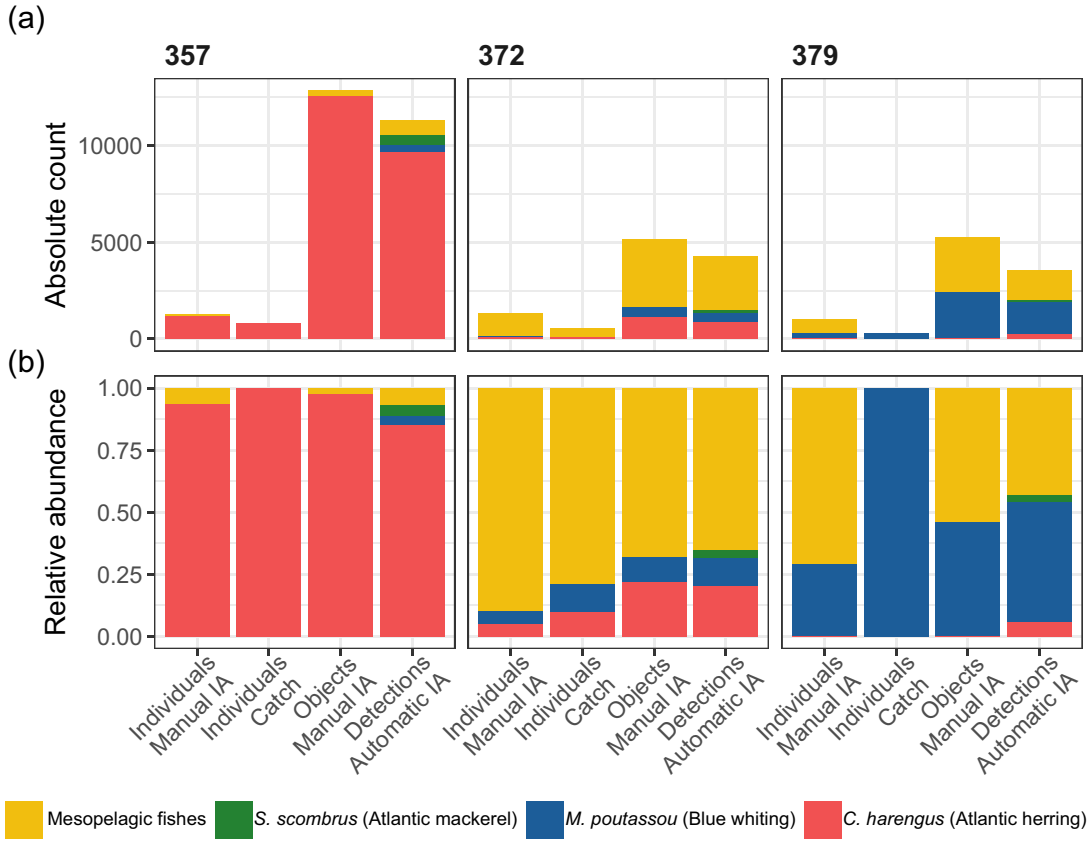

Figure 9. Absolute (a) and relative (b) composition of taxa in trawl stations 357, 372, and 379 calculated from three different methods: manual IA, trawl catches, and detections generated by RetinaNet during the automatic IA. For the manual IA, both the number of individuals and objects (each individual counted as many times as it was imaged) are provided for the comparison.

classes in object-based length frequency distributions persisted irrespective of the frame rate (Fig. 10a). Conversely, the object-based length frequencies of blue whiting and mesopelagic fishes exhibited minimal sensitivity to lowering the frame rate, likely because their observation period did not significantly vary with body length. This suggests that for these taxa, but not for herring, length frequencies can be manually extracted from a subset of images or automatically extracted from all the imaged objects.

Given that lowering the frame rate introduces new biases and manual analysis of entire trawl stations is not feasible, automatic IA becomes essential for extracting data from in-trawl cameras. This study emphasises that for accurate abundance and length frequency estimates from images, automatic analysis must implement additional measures for recurrence counts.

Previous studies on automating counts from in-trawl cameras have employed two different methods: Allken et al. (2021) addressed recurrence count by comparing the number of automatic detections with the catch and fitting different linear regression models for three pelagic fish species: Atlantic herring, blue whiting and Atlantic mackerel. Alternatively, tracking, which involves linking objects from consecutive frames by minimising positional offset, has shown promising results for counting crustaceans (Avsar et al. 2023) and fish (Chuang et al. 2015) in trawls.

Tracking has the advantage of being independent of the catch, which is unavailable for small and fragile organisms or in open-codend trawling procedures. Moreover, unlike the linear regressions used by Alken et al. (2021) and in this study, tracking does not depend on fixed model coefficients and is likely to deal better with the high variability we detected in observation periods.

Tracking opens the possibility of obtaining real-time estimates of abundance and length, given that the required computational power can be met (Avsar et al. 2023, 2024). In fisheries scientific surveys, acoustic scrutinisation is ongoing during the cruise and does not require real-time data. However, real-time information is crucial in decision-making processes when considering open-closed codend sampling, where the goal may be to capture specific species or sizes for measurement of age, sex, and DNA.

To implement tracking for highly mobile species, automatic IA needs to account for individuals swimming in and out of the FOV multiple times (Chuang et al. 2015). For fast-swimming species and smaller fish, like herring and mesopelagic fishes, a high frame rate is necessary to minimise positional offset between consecutive frames. However, raising the frame rate does not address the issue of occlusion in high-density scenarios.

Objects occluding each other in a 2D image may be separated by distance from the camera (Z). Thus, integrating 3D

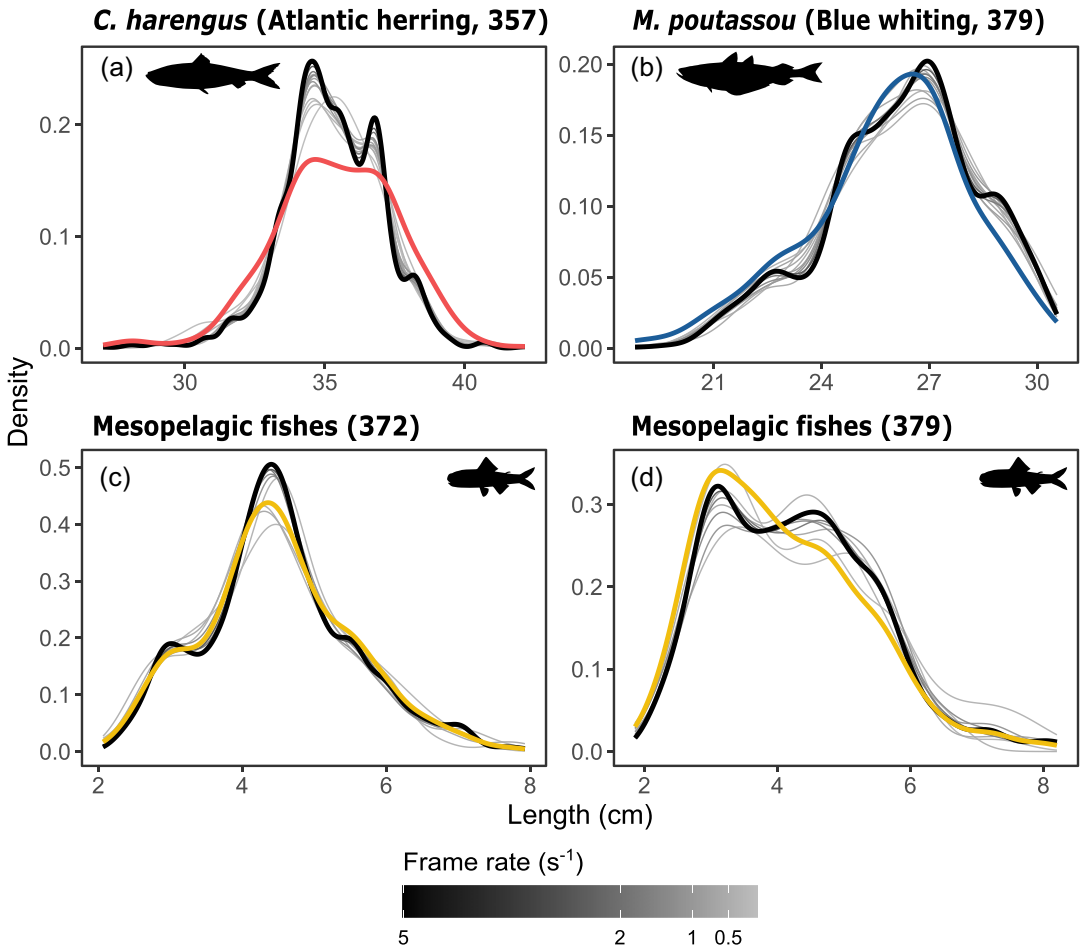

Figure 10. Length frequency distributions for each taxon and sampling station with at least 100 length measurements. Within each panel, the thick coloured curve represents the distribution when each individual is measured only once (red: Atlantic herring, blue: blue whiting, yellow: mesopelagic fishes). Greyscale curves illustrate the distributions of all imaged objects. The greyscale corresponds to the frame rate, with the thick black curve representing the highest frame rate (5 images s^{-1}). For clarity, only a subset of frame rates is exemplified here (herring: ≥0.5, 0.33, 0.25, 0.17, 0.11; blue whiting: ≥0.5, 0.41; mesopelagic fishes: ≥1, 0.56, 0.38, 0.26, 0.13).

Table 6. Confusion matrix of manual annotations (rows) and automatic detections (columns) with a score threshold of 0.47

Taxa	Herring	Blue whiting	Mesopelagic fishes	Mackerel	Missed	Total (manual)
Herring	1134 (**0.88**)	33 (0.03)	0 (0)	22 (0.02)	101 (0.08)	1290
Blue whiting	17 (0.05)	323 (**0.87**)	5 (0.01)	1 (<0.01)	25 (0.07)	371
Mesopelagic fishes	4 (<0.01)	3 (<0.01)	917 (**0.46**)	0 (0)	1089 (0.54)	2013
Total (automatic)	1158	358	1096	22	1040	3674

The matrix only includes true positives ("correct") and false negatives ("wrong" or "missed"). The automatic IA mistakenly classified some of the herring and blue whiting as mackerel, even though no mackerel was present in the images. The values between parentheses represent the proportion of automatic detections with respect to the row total. The recall value of each taxon is shown in bold.

data from a stereo camera could improve tracking in high-density scenarios. Chuang et al. (2015), however, found no performance gains from including distance, as their algorithm allowed for overlap in multiple frames. Another option would be to reduce packing density by moving the camera to a trawl section that has a greater diameter.

Moving the camera has other implications. First, as the camera is positioned further ahead in the trawl, the data collected from the images diverges more from what is collected in the codend, due to greater opportunities for organisms to escape between the camera and the codend. Second, the camera's FOV no longer covers the entire volume of the trawl section. As a result, the estimates derived from these images would assume that fish swimming outside the camera's FOV exhibit the same swimming behaviour and species, and size distribution as those within it. Our findings show that certain taxa are positioned nearer to the camera than others, suggesting this assumption may not hold true. Therefore, more research into the distribution and swimming behaviour is essential. Finally, the increase in volume allows objects to pass at a greater distance from the camera, which can complicate identification.

Sampling mesopelagic fishes with in-trawl cameras

The comparison of catch and manual IA showed that while herring and blue whiting were similarly represented in both methods, the smaller-sized mesopelagic fishes were often only present in the images (Fig. 9). Trawls targeting small mesopelagic organisms integrate small-mesh liners inside the trawl body and the codend to limit mesh selectivity (Grimaldo et al. 2022). The pelagic trawl (Multpelt 832) deployed in this study was not modified to retain small-sized fishes. Given that the mesh size of the codend was 40 mm and the length of mesopelagic fishes ranged between 1.8 and 8.2 cm, most of the smaller individuals were likely not retained. The patchy distribution of micronekton (including mesopelagic fishes) makes depth-specific records from in-trawl cameras highly valuable, as they can potentially expand our knowledge about their vertical distribution (Pearcy 1983) without increasing handling time on existing surveys. Understanding the catchability of mesopelagic species ahead of the in-trawl cameras would facilitate a better interpretation of the image-based data.

Only 46% of the manually annotated mesopelagic fishes were detected and correctly classified (Table 6). This starkly contrasts with herring and blue whiting, which showed recall values of 88% and 87%, respectively. Allken et al. (2021) reported similar percentages for herring (72%) and blue whiting (82%) but a much higher value for mesopelagic fishes (89%). As a mixed-species category, mesopelagic fishes are likely to differ in relative species composition between this study and the one conducted by Alken et al. (2021). Mesopelagic fishes are generally smaller than herring and blue whiting and can therefore be more challenging for object detection algorithms. Despite RetinaNet's implementation of multi-resolution classification (Lin et al., 2017), Zhou et al. (2021) showed that RetinaNet does not perform as effectively on small objects. According to the results of this study, the current object detection algorithm, as described in Allken et al. (2021), would require additional fine-tuning to improve the recall of mesopelagic fishes.

Conclusion

Our analyses showed that depending on the taxa of interest, swimming orientation, distance to the camera, body length, and density inside the imaging chamber can influence observation periods. As a result, recurrence counts varied within and between taxa and across length classes. This variation means that recurrence counts cannot be addressed by simply reducing the frame rate. Tracking, therefore, seems necessary for providing reliable abundance and length frequency estimates from in-trawl images. If accurate estimates are provided, in-trawl cameras can lessen the catch and handling time on fisheries scientific surveys. Refining the automatic detection to sample small and fragile organisms would allow for a wider range of species to be continuously monitored, ultimately enhancing our understanding of the ecosystem.

Acknowledgements

The authors thank Sindre Vatnehol (IMR) and Vaneeda Allken (IMR) for operating the in-trawl camera and providing the machine learning output, as well as the crew of the RV "G.O. Sars" for their assistance and hospitality at sea. We thank Richard James Telford (UiB) and Neil Anders (IMR) for their advice on statistical analyses. We also thank Ricard Prados Gutiérrez (Girona Vision Research) and Eirik Osborg (Scantrol Deep Vision AS) for their insights on the camera system and its calibration. Lastly, we are grateful to Terje Jørgensen (IMR), Anne Gro Vea Salvanes (UiB), and the three anonymous reviewers for providing valuable feedback to the manuscript.

Author contributions

T.W. conceived and designed the study in discussion with S.P.R., M.T., and K.E.. T.W. conducted the manual image analysis and generated the figures. T.W. conducted the statistical analysis and interpretation with additional input from S.P.R. and M.T.. T.W. led the writing of the manuscript. All authors edited and provided critical contributions to drafts and gave final approval for publication.

Supplementary data

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

Conflict of interest: The authors declare they have no conflict of interest.

Funding

This work was financed by the Research Council of Norway project "CRIMAC," grant number 309512, and the REDUS project with funding from the Norwegian Ministry of Trade, Industry and Fisheries.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

References

Adil E, Mikou M, Mouhsen A. A novel algorithm for distance measurement using stereo camera. *CAAI Trans Intell Technol* 2022;**7**:177–86. https://doi.org/10.1049/CIT2.12098

Allken V, Rosen S, Handegard NO *et al.* A deep learning-based method to identify and count pelagic and mesopelagic fishes from trawl camera images. *ICES J Mar Sci* 2021;**78**:3780–92. https://doi.org/10.1093/icesjms/fsab227

Amaral SV, Winchell FC, Pearsons TN. Reaction of Chinook salmon, Northern pikeminnow, and smallmouth bass to behavioral guidance stimuli. *Behavioral Technologies for Fish Guidance: American Fisheries Society Symposium* 2001;125. https://hdl.handle.net/20.500.14394/26026

Arnold GP. Rheotropism in fishes. *Biol Rev* 1974;**49**:515–76. https://doi.org/10.1111/j.1469-185X.1974.tb01173.x

Avsar E, Feekings JP, Krag LA. Edge computing based real-time Nephrops (*Nephrops norvegicus*) catch estimation in demersal trawls using object detection models. *Sci Rep* 2024;**14**:9481. https://doi.org/10.1038/s41598-024-60255-8

Avsar E, Feekings JP, Krag LA. Estimating catch rates in real time: development of a deep learning based Nephrops (Nephrops norvegicus) counter for demersal trawl fisheries. *Frontiers in Marine Science* 2023;**10**:1129852. https://doi.org/10.3389/FMARS.2023.1129852/BIBTEX

Blackman S. S., Popoli R.. Design and Analysis of Modern Tracking Systems. *Radar Library* 1999.

Breen M, Anders N, Humborstad O-B *et al. Catch Welfare in Commercial Fisheries*. 2020;401–37. https://doi.org/10.1007/978-3-030-41675-1_17

Chuang M-C, Hwang J-N, Williams K *et al.* Tracking live fish from low-contrast and low-frame-rate stereo videos. *IEEE Trans Circuits Syst Video Technol* 2015;**25**:167–79. https://doi.org/10.1109/TCSVT.2014.2357093

Clark MR, Bagley NW, Harley B. Trawls. In: MR Clark, M Consalvey, Ashley A. Rowden (eds.), *Biological Sampling in the Deep Sea*. Hoboken, NJ: John Wiley & Sons, 2016, 126–58. https://doi.org/10.1002/9781118332535.CH7

DeCelles GR, Keiley EF, Lowery TM *et al.* Development of a video trawl survey system for new england groundfish. *Trans Am Fish Soc* 2017;**146**:462–77. https://doi.org/10.1080/00028487.2017.1282888

Fernandes PG, Chacko V, Polanski J *et al.* SMARTRAWL 2.5 Final Report. University of Aberdeen, 2021, 33. https://fisorg.uk/wp-content/uploads/2021/03/Smartrawl-Phase-2.5-Final-Report.pdf

Fox J, Weisberg S. *An R Companion to Applied Regression*, Thousand Oaks, CA: 3rd edn. Sage, 2019. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Fréon P, Misund OA. *Dynamics of Pelagic Fish Distribution and Behaviour: Effects on Fisheries and Stock Assessment*, Vol. **348**. Oxford: Fishing News Books, 1999.

Garcia R, Prados R, Quintana J *et al.* Automatic segmentation of fish using deep learning with application to fish size measurement. *ICES J Mar Sci* 2020;**77**:1354–66. https://doi.org/10.1093/icesjms/fsz186

Gerritsen HD, McGrath D. Precision estimates and suggested sample sizes for length-frequency data. *Fish Bull* 2007;**105**:116–20.

Grimaldo E, Herrmann B, Brčić J *et al.* Prediction of potential net panel selectivity in mesopelagic trawls. *Ocean Eng* 2022;**260**:111964. https://doi.org/10.1016/j.oceaneng.2022.111964

Gunderson DR. *Surveys of Fisheries Resources*. Hoboken, NJ: Wiley. 1993, 13–5.

Hartley R, Zisserman A. *Multiple View Geometry in Computer Vision*. Cambridge: Cambridge University Press, 2003.

He P. Swimming speeds of marine fish in relation to fishing gears. *ICES Mar Sci Symp* 1993;**196**:183–9.

ICES. International ecosystem survey in the Nordic Sea (IESNS) in May to June 2018. *Working Document No. 01 to Working Group on International Pelagic Surveys (WGIPS 2019) and Working Group on Widely Distributed Stocks (WGWIDE), Tórshavn, Faroe Islands, 28 August–3 September 2018*, 2018, 51. https://www.hav.fo/PDF/Ritgerdir/2018/IESNS_survey_2018.pdf

Jolly GM, Hampton I. A stratified random transect design for acoustic surveys of fish stocks. *Can J Fish AquatSci* 1990;**47**:1282–91. https://doi.org/10.1139/f90-147

Kracker LM. The geography of fish: the use of remote sensing and spatial analysis tools in fisheries research. *Prof Geographer* 1999;**51**:440–50. https://doi.org/10.1111/0033-0124.00178

Krag LA, Savina E, Lyngby MS *et al.* Real-time camera observation in the trawl fisheries (Technofish): final report. DTU Aqua-Rapport, 421-2023, 2023.

Kroeger M. *Some Results of Flow Measurements on a Full Scale Pelagic Trawl*. Copenhagen: ICES, 1984.

Lin T-Y, Dollar P, Girshick R *et al.* Feature pyramid networks for object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition* 2017;2117–25.

Lüdecke D, Ben-Shachar MS, Patil I *et al.* Performance: an R package for assessment, comparison and testing of statistical models. *J Open Source Softw* 2021;**6**:3139. https://doi.org/10.21105/joss.03139

Nguyen KQ, Winger PD. Artificial light in commercial industrialized fishing applications: a review. *Rev Fish Sci Aquac* 2019;**27**:106–26. https://doi.org/10.1080/23308249.2018.1496065

Novales Flamarique I, Hiebert S, Sechrist J. Visual performance and ocular system structure of kokanee and sockeye salmon following strobe light exposure. *N Am J Fish Manag* 2006;**26**:453–9. https://doi.org/10.1577/M04-215.1

Ojaveer E, Pihu E, Saat T. *Fishes of Estonia*. Estonian Academy Publishers, 2003.

Patrick PH, Poulton JS, Brown R. Responses of American Eels to strobe light and sound (preliminary data) and introduction to sound conditioning as a potential fish passage technology. 2001. https://hdl.handle.net/20.500.14394/25608

Pearcy WG. Quantitative assessment of the vertical distributions of micronektonic fishes with opening/closing midwater trawls. *Biol Oceanogr* 1983;**2**:289–310. https://doi.org/10.1080/01965581.1983.10749463

Pope JA. Manual of methods for fish stock assessment. Part III. Selectivity of fishing gear. *FAO Fish Tech Pap* 1975;**41**:9–14.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 2020. https://www.r-project.org/

Rosen S, Holst JC. DeepVision in-trawl imaging: sampling the water column in four dimensions. *Fish Res* 2013;**148**:64–73. https://doi.org/10.1016/j.fishres.2013.08.002

Rosen S, Jorgensen T, Hammersland-White D *et al.* DeepVision: a stereo camera system provides highly accurate counts and lengths of fish passing inside a trawl. *Can J Fish AquatSci* 2013;**70**:1456–67. https://doi.org/10.1139/cjfas-2013-0124

Skúvadal FB, Thomen B, Jacobsen JA. Escape of blue whiting (*Micromesistius poutassou*) and herring (*Clupea harengus*) from a pelagic survey trawl. *Fish Res* 2011;**111**:65–73. https://doi.org/10.1016/J.FISHRES.2011.06.012

Stokesbury KDE, Cadrin SX, Calabrese N *et al.* Towards an improved system for sampling new england groundfish using video technology. *Fisheries* 2017;**42**:432–9. https://doi.org/10.1080/03632415.2017.1342630

Suuronen P, Lehtonen E, Wallace J. Avoidance and escape behaviour by herring encountering midwater trawls. *Fish Res* 1997;**29**:13–24. https://doi.org/10.1016/S0165-7836(96)00523-1

Suuronen P, Millar RB. Size selectivity of diamond and square mesh codends in pelagic herring trawls: only small herring will notice the difference. *Can J Fish AquatSci* 1992;**49**:2104–17. https://doi.org/10.1139/F92-234

Trenkel V, Vaz S, Albouy C *et al.* We can reduce the impact of scientific trawling on marine ecosystems. *Mar Ecol Prog Ser* 2019;**609**:277–82. https://doi.org/10.3354/meps12834

Underwood MJ, Rosen S, Engås A *et al.* Deep vision: an in-trawl stereo camera makes a step forward in monitoring the pelagic community. *PLoS One* 2014;**9**:e112304. https://doi.org/10.1371/journal.pone.0112304

Underwood MJ, Rosen S, Engås A *et al.* Species-specific residence times in the aft part of a pelagic survey trawl: implications for inference of pre-capture spatial distribution using the deep vision system. *ICES J Mar Sci* 2018;**75**:1393–404. https://doi.org/10.1093/ICESJMS/FSX233

Wardle CS. Effects of Temperature on the Maximum Swimming Speed of Fishes. *Environmental Physiology of Fishes* Boston, MA: Springer, 1980;**35**:519–31. https://doi.org/10.1007/978-1-4899-3659-2_20

Wileman DA., Ferro RST, Fonteyne R, Millar RB. Manual of methods of measuring the selectivity of towed fishing gears. *ICES Cooperative Research Reports (CRR)* 1996. https://doi.org/10.17895/ices.pub.4628

Williams K, Lauffenburger N, Chuang M-C *et al.* Automated measurements of fish within a trawl using stereo images from a Camera-Trawl device (CamTrawl). *Methods in Oceanography* 2016;**17**:138–52. https://doi.org/10.1016/j.mio.2016.09.008

Williams K, Towler R, Wilson C. Cam-trawl: a combination trawl and stereo-camera system. *Sea Technol* 2010;**51**:45–50.

Zhou Z, Li Y, Peng C *et al.* Image processing: facilitating retinanet for detecting small objects. *J Phys Conf Ser* 2021;**1815**:012016. https://doi.org/10.1088/1742-6596/1815/1/012016

Zuur AF, Ieno EN, Elphick CS. A protocol for data exploration to avoid common statistical problems. *Methods Ecol Evol* 2010;**1**:3–14. https://doi.org/10.1111/J.2041-210X.2009.00001.X

Zuur AF, Ieno EN, Walker N *et al. Mixed Effects Models and Extensions in Ecology with R*, Vol. **574**. Berlin: Springer, 2009. https://doi.org/10.1007/978-0-387-87458-6

Handling editor: Michael Pol