Chapter 3

Image-Based Machine Learning Reveals Structural ‘Fingerprints’ of Pelagic Seascapes in Shipboard Echograms

This chapter was originally published in ICES Journal of Marine Science, Vol. 82, Issue 12, Article fsaf221, 2025, doi:10.1093/icesjms/fsaf221.

Abstract

Echograms generated using active shipboard acoustics data provide visually rich insights into the structure and behaviour of interior pelagic communities, yet current numerical approaches underuse this complexity. Here, we develop a novel approach for analysing diel echogram imagery using convolutional autoencoders (CAEs). We apply this approach to 1,596 daily echograms collected using uncalibrated Acoustic Doppler Current Profilers (ADCPs) from 306 cruises in the Pacific Ocean between 2001 and 2018 over nearly 500,000 km. By compressing each echogram into a low-dimensional feature space, we identify recurring structural patterns—such as diel vertical migrations and scattering layers—and use these features to cluster the echograms into six coherent regional modes across the basin. These modes, or acoustic "fingerprints", provide a new, interpretable framework for classifying pelagic ecosystems based on structure rather than backscattering intensity. When mapped and predicted to the entire Pacific basin using environmental covariates, our classification aligns closely with previous biogeographies derived from calibrated scientific echosounder data, demonstrating the potential of unsupervised deep learning to extend ecosystem monitoring using widely available ADCP archives. The Pacific ADCP archive used in this study is nearly 40% larger than any of the 38 kHz archives used in previous global studies: by enabling analysis of uncalibrated acoustic data, this approach unlocks a vast global archive for ecological classification, expanding opportunities for ecosystem monitoring, biogeography, and change detection.

3.1 Introduction

The ocean interior is the largest habitable space on the planet. It is invisible to satellites and too dark and vast to be efficiently observed with optical methods (St. John et al. 2016). Active acoustic devices, like shipboard echosounders, are the primary tool for large-scale surveillance of this region (Benoit-Bird and Lawson 2016). These devices emit pulses of sound and record backscatter from organisms in the water-column, generating vertical profiles of volume backscattering strength (S_v, mean backscattering intensity per m) that reflect the vertical arrangement of pelagic organisms (Benoit-Bird et al. 2009b; Cotter et al. 2021a). Consecutive profiles are typically composed into 2-dimensional echograms, which form image-like representations of biological structure when observed along the time dimension (e.g. across a transect). Over the past decade, availability of acoustic data has grown considerably due to widespread deployment across diverse platforms, offering unprecedented opportunity for large-scale analysis using 'big-data' approaches. However, the sheer volume and complexity of these data have outpaced traditional echosounder analysis pipelines, which depend heavily on manual processing, hindering timely advances in marine ecological research (Lee and Staneva 2020).

A class of big-data workflows has emerged in bioacoustics to leverage repeatable feature extraction methods on large, multi-cruise archives to group active acoustics observations in space and time (Ariza et al. 2022; Proud et al. 2017, 2018; Receveur et al. 2020c). Existing approaches are mainly distinguished by their feature extraction methods, which fall into two main categories: feature-specific (e.g., targeting peaks in profiles representing scattering layers or migrations; (Proud et al. 2017)) and unsupervised (e.g., dimensionality reduction of the full vertical profile; (Ariza et al. 2022)). Targeted, feature-specific methods offer clear biological interpretability but depend on prior assumptions regarding what features are important, while unsupervised approaches trade that interpretability for a more complete, data-driven representation of signal variability. The latter excels in capturing broad, emergent patterns crucial for large-scale analyses like biogeography or phenology, which is exemplified by the recent shift in large-scale bioacoustics workflows from targeted approaches to unsupervised machine learning (Ariza et al. 2023, 2022; Izard et al. 2024, 2025b; Lee and Staneva 2020). However, the efficacy of recent unsupervised approaches is inhibited by the use of isolated vertical profiles or short temporal segments rather than full diel echograms (Lee and Staneva 2020). Profile-based methods may obscure or eliminate biologically important features related to diel migrations from low-dimensional feature sets, and consequently from resulting biogeographic solutions. Existing approaches are further limited by their dependence on high-quality, single-frequency data—primarily 38 kHz echosounders—that preserve numerical scattering intensity for biomass estimation (Izard et al. 2025b) but precludes use of any other active acoustics data sets, such as from uncalibrated echosounders or Acoustic Doppler Current Profilers

(ADCPs). Thus, given the limited ability of profile-based approaches to incorporate migration dynamics into feature sets and models, and the fact that existing methods have not been able to leverage the growing archive of acoustics observations available from other acoustic instruments, new approaches are needed to expand our ability to observe and monitor the ocean interior.

Convolutional autoencoders (CAEs) are a form of unsupervised convolutional neural networks (CNNs) for nonlinear dimensionality reduction that are often used in image analysis. CAEs compress inputs into a lower-dimensional latent space and reconstruct them, preserving spatial structure through convolutional layers (Mao et al. 2016). This technology has demonstrated exceptional performance in other disciplines—including analogous sonic data from ultrasounds in medical imaging (Daoud et al. 2019) and spectrograms in radio astronomy (Gheller and Vazza 2022)—and are highly robust to noise in these applications. In bioacoustics, convolutional learning models (CNNs and CAEs) have been predominantly used for feature extraction or supervised classification of individual targets or aggregates within echograms (Allken et al. 2021; Brautaset et al. 2020; Pena and Calise 2016; Proud et al. 2020), or for the classification of sounds produced by marine mammals. Despite these applications, convolutional models have not been applied to study the rich visual complexity that is apparent in diel echograms, much of which is lost or simplified in current feature-specific (e.g., (Proud et al. 2017)) or 1-d profile-based representations (e.g., (Lee and Staneva 2020)) of the acoustic data.

Here, we develop an unsupervised CAE workflow to extract structural features from complete diel echograms collected using ADCPs during >1,500 days of research cruises in the Pacific Ocean, cluster the echograms based on the extracted feature sets, and associate the clusters with environmental covariates to derive a biogeography of pelagic seascapes (50–850 m) across the Pacific basin (hereafter biogeography). As ADCP backscatter is not calibrated and therefore unsuitable for biomass estimation (Receveur et al. 2020c), we normalize all echograms to remove spurious amplitude variability. This ensures that clustering is based on consistent structural patterns—including vertical organization and diel dynamics—rather than raw signal intensity. This intensity-free approach enables robust feature extraction from lower-quality data, illustrating a path toward increased interoperability among instruments, while highlighting the capabilities of contemporary AI methods in "big data" bioacoustics workflows. This approach allows us to effectively capture the visual complexity present in echograms. It enables the construction of representative, two-dimensional 'fingerprints' of midwater communities that are otherwise hidden in the ocean interior. Additionally, it supports the scaling of bioacoustics workflows to handle the growing volume and diversity of datasets required to study these enigmatic deep-ocean ecosystems. This is reflected in the present study by the use of uncalibrated ADCP data, which comprise one of the largest acoustic databases worldwide that has so far had only residual use in the study of pelagic

fauna at large scales.

3.2 Methods

This study used diel echogram data collected with uncalibrated 38 kHz Acoustic Doppler Current Profilers (ADCPs) to develop a CAE workflow for classifying pelagic ecosystems (Fig. 3.1). Each echogram was standardized to cover the upper 50–850 m of the water column and a 24-hour time series of profiles (pings) centered on peak solar angle (Sec. 3.2.1). Because uncalibrated ADCP data cannot reliably be used to estimate absolute biomass, we normalized all echograms to focus on relative scattering structure (Sec. 3.2.2). The CAE was used to derive feature vectors from the image dataset (Fig. 3.1a–b), which were clustered into ecoregions based on the organization and coherence of visual components—such as layers and migrations—independent of scattering intensity (Fig. 3.1c–d). Cluster centroids were reconstructed by the decoder into 2D echogram "fingerprints," providing interpretable visual summaries of ecosystem structure. Finally, clusters were paired with environmental covariates in a regression framework to predict a pelagic biogeography across the Pacific (Fig. 3.1e–h). Elementary properties of echograms—e.g. layer depths, migration timing and day-night differences—were extracted to describe and corroborate the CAE-based regions (Sec. 3.2.4).

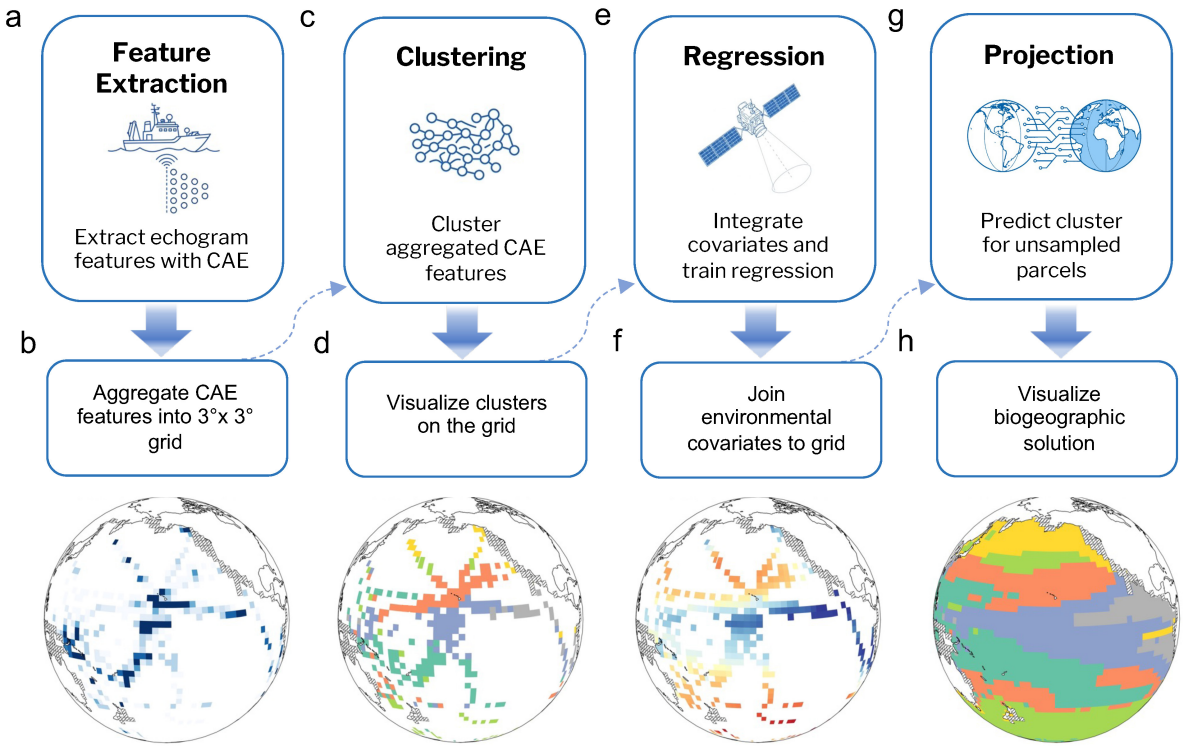

Figure 3.1: Overview of the convolutional autoencoder (CAE) workflow for echogram-based biogeography. (a) Feature extraction: echogram images are compressed into a low-dimensional latent space using the CAE encoder. (b) Aggregation: latent feature vectors are averaged into $3° \times 3°$ grid cells. (c) Clustering: aggregated feature vectors are clustered using k-means. (d) Visualization of cluster assignments on the spatial grid. (e) Regression: environmental covariates are joined to cluster labels. (f) Covariate maps for the Pacific. (g) Projection: a random forest classifier predicts cluster labels for unsampled grid cells. (h) Final biogeographic map of Pacific pelagic seascapes.

3.2.1 Data

All ADCP data were obtained from the Joint Archive for Shipboard ADCP data (JASADCP; available at https://uhslc.soest.hawaii.edu/sadcp/). Each cruise in the JASADCP is stored in a netCDF file generated by the Common Ocean Data Access System (CODAS) ADCP processing software, containing metadata and the raw received signal strength indicator (RSSI; ADCP specific units) (Firing 1995). RSSI values represent bin-averaged ensembles of pings (5 minutes in duration). The data contain quality control metrics generated by a suite of automated 'editing' algorithms for the removal of acoustic interference, identification of the bottom, the identification of low quality pings and the proportion of good quality pings per bin ensemble (pg; (Firing 1995)). Details on CODAS ADCP processing workflows accompany the data archive here (hereafter CODAS). No

correction is applied to RSSI for beam spreading or attenuation. These raw RSSI values are later converted to volume backscattering strength (S_v; dB re 1 m^{-1}), which accounts for beam spreading and absorption (Sec. 3.2.2).

Cruise transects were converted into an image database for model training. Time and position metadata were used to compute the solar angle and split transects into diel echograms centered at the peak solar angle for each 24-hour period. We restricted cruises to only those that used 38 kHz instruments in the Pacific region (latitude range from 70°S to 60°N). We did not attempt a global classification because over 95% of complete 38 kHz echograms in the JASADCP dataset came from the Pacific, and merging these with sparse data from other basins could bias results, compromise cluster stability, and hinder robust validation. Echograms were time-depth standardized to a regular depth (1 meter) and time (5 minutes) grid over the 50–850 m range and interpolated using optimal interpolation (Wunsch 2006). Thus, the image pixels and feature maps from the CAE correspond to physical locations in depth/time space that are standardized across all images.

Several steps were taken to ensure the quality of echograms in the database. First, the quality control flag in the CODAS metadata (pflag; CODAS) was used to mask poor quality profiles. Pflag is a categorical variable where nonzero values indicate 'bad' bin ensembles where the proportion of good pings in the ensemble is less than 50%. We computed the percentage of 'bad' values in each profile and masked all profiles where the proportion of flagged values exceeded 15%. Second, after masking the data, all 24-hour segments spanning less than 23-hours (from the first to the last profiles), and containing less than 21-hours of total coverage were excluded from the dataset. Third, seabed depths associated with each echogram were computed using the ETOPO global relief model. All echograms associated with seabed depths of less than or equal to 1,000 meters were excluded from the database, and echograms (n=14) associated with seabed depths of less than 1,500 meters were manually checked for seabed detections and second-bottom echoes (Proud et al. 2017; Simmonds and MacLennan 2008). While second bottom echoes could occur in deeper echograms, nearly all echograms in the database were visually inspected at some stage of the analysis, including a careful review of 25% to validate derived features (Sec. 3.2.4), and multiple large-subsets of the database to visualize cluster members.

Finally, cruises span variable depth ranges—ranging from the first to last bin midpoint ± half a bin length—yielding heterogeneous water column depths due to non-standard transducer placements and bin lengths of 16–24 m. Thus, we focused on the depth range of 50–850 m to eliminate variability in water column coverage across all echograms, promoting uniform coverage over a standard input domain. Consequently some information on the upper water column was eliminated from this analysis (Sec. 3.4.3). Summary metrics for the final echogram database are reported in Fig. 3.2b; Table 3.1.

3.2.2 From RSSI to relative scattering density

ADCP raw received signal strength indicator (RSSI) was converted into volume backscattering strength (S_v (dB re 1 m^{-1}) = 10 log$_{10}$ s_v, (Maclennan et al. 2001); Sec. 3.2.2) using corrections for the influence of environmental factors (e.g., sound speed), beam spreading and absorption (Deines 1999; Leung 2020; Mullison 2017; Receveur et al. 2020c; Song et al. 2022). The resulting S_v echograms were then normalized, converting S_v into relative scattering density (Sec. 3.2.2).

Computing S_v from RSSI

We computed S_v using Python utilities developed by the corresponding author and Dr. Shirley Leung (Leung 2020) and the equation for Teledyne Research Instruments (TRDI) ADCP instruments as specified by Mullison (2017) (Mullison 2017) (Equation 3.1), following previous studies (e.g., (Leung 2020; Receveur et al. 2020c; Song et al. 2022)).

$$S_v = C + 10\log\Big((T_x + 273.16)R^2\Big) - L_{DBM} - P_{DBW} + 2\alpha R + 10\log\Big(10^{k_c(E-E_r)/10} - 1\Big) \quad (3.1)$$

where:

- C = constant combining several parameters specific to each ADCP instrument [dB]
- T_x = temperature measured at the transducer [°C]
- R = along-beam range to the measurement [m]
- L_{DBM} = 10 log (transmit pulse length [m])
- P_{DBW} = 10 log (transmit power [Watts])
- α = absorption coefficient of the water [dB m^{-1}]
- k_c = RSSI slope [dB count^{-1}]
- E = RSSI [counts]
- E_r = noise floor [counts]

Values for T_x (°C), R (m) (computed as distance to scatterer divided by cosine (transducer beam angle)), transmit pulse length (needed to compute L_{DBM} (m)), and E (counts) are all included in

the metadata or data within the JASADCP archive, so these did not require further processing. Transducer depths are likewise included in the metadata and were used to compute range. Other parameters required to compute components of Equation 3.1 are often missing from the cruise metadata given that they are not required for velocity estimation, which is the primary purpose of ADCPs. Therefore, we made instrument-specific estimates for these values for the two Ocean Surveyor instruments (Ocean Surveyor I and Ocean Surveyor II) that comprise the 38 kHz subset of the JASADCP data used in this analysis (parameter estimates collated in Table A.1; (Leung 2020; Mullison 2017)). Values for C (-172.19 dB) and P_{DBW} (24 W) were taken from Mullison (2017) (Mullison 2017), while k_c (0.37 ± 0.004 dB count^{-1}) and E_r (14 ± 4.83 counts) values were obtained for Ocean Surveyor instruments via email correspondence with the manufacturer (J. Mullison & S. Idle, Teledyne RD Instruments). We assumed that all parameters are applicable for both generations of these instruments (OS-I and OS-II). Normalization of the data (Sec. 3.2.2) removes the dependence of S_v on constant terms in Equation 3.1 (e.g., L_{DBM}) though we reproduce the full equation here with all terms.

Parameters that do not have instrument specific routines include α and E_r. E_r was taken to be the lowest recorded RSSI within each cruise, following (Lee et al. 2004) and (Gostiaux and Haren 2010). The absorption coefficient (α) was computed using the widely used formulation given in (Francois and Garrison 1982), which requires estimates of c (speed of sound through fluid) obtained using the formulation from (Mackenzie 1981). Total absorption $2\alpha R$ was computed using the routine from (Deines 1999). Subsurface temperature, salinity and pH were also required, and were estimated using the NASA ECCO V4r4 climate model as climatological monthly-mean values at 0.5 degree spatial resolution (ECCO Consortium et al. 2021).

Normalizing S_v

After converting RSSI to S_v each 24-hour echogram was min–max normalized to [0,1] using $(S_v - S_{v,\min})/(S_{v,\max} - S_{v,\min})$. Normalization of S_v ensures that, even if S_v measurements from the same instrument across different cruises differed in mean intensity (which has been observed in previous studies, sometimes with large dB offsets), they can be compared by considering their relative scattering (Receveur et al. 2020c). Normalization also ensures that errors in S_v estimates from depth-invariant calibration parameters (e.g., constants L_{DBM} and P_{DBW}) do not impact density representation and induce spurious variability in numerical scattering intensity that could impact classification. We normalized S_v (at the decibel scale) rather than the linear volume backscattering coefficient s_v, because the former compresses dynamic range and better emphasizes structural patterns. Normalization was implemented at the echogram level—as opposed to the cruise or instrument level—to ensure that echograms were not distorted by sampling variability across cruises

or instruments in the archive (Sec. 3.4.3).

Despite normalization, incorrect estimation of calibration parameters (Sec. 3.2.2) that vary based on intensity or depth can affect vertical profile shapes and thus relative scattering density. To assess the consequence of potential mis-estimation of this kind, we varied these parameters within their distributional ranges and generated alternative input image datasets. Our results show high stability in clustering confirmed using Adjusted Mutual Information (AMI; (Hubert and Arabie 1989)) and Adjusted Rand Index (ARI; (Steinley 2004)). That said, we acknowledge that normalization can alter the relative prominence of features when high-intensity structures—such as dense schools or resonant scattering layers—are present. These features, which may vary in intensity due to physical factors like depth-dependent resonance rather than changes in biomass, can influence how surrounding lower-intensity features are represented. While our approach emphasizes spatial structure over magnitude, we revisit the implications of this context dependence in the Discussion (Sec. 3.4.3).

3.2.3 CAE workflow

We used the resulting database of quality-controlled and normalized echogram images to train a CAE and build low-dimensional representations of echograms that summarize complex visual patterns in a few variables (embeddings). We then used the model to embed the image database, cluster the embedding vectors, and train a regression model based on environmental covariates to interpolate the clusters to a biogeographical map of the Pacific. This workflow (Fig. 3.1) of dimensionality reduction, clustering, and prediction, is a common paradigm in biogeographical analyses using high-dimensional (Sonnewald et al. 2020) and bioacoustics datasets (Ariza et al. 2023, 2022; Izard et al. 2024, 2025b; Proud et al. 2018; Receveur et al. 2020b).

Model architecture and training

We used a CAE model architecture consisting of an encoder and a decoder with symmetric layer structures. The CAE used convolutional layers with 32, 64 and 128 filters associated with output feature maps of $128 \times 128 \times 32$, $64 \times 64 \times 64$ and $32 \times 32 \times 128$ respectively. Filters were applied using 3×3 kernels, followed by ReLU activation functions which introduce non-linearity, and 2×2 max-pooling operations to reduce feature map dimensions (Glorot et al. 2011; LeCun et al. 1998). The final layer in the encoder was flattened and mapped to a 20-dimensional latent space using a fully connected dense layer. The latent space dimension was selected based on visual inspection of image reconstructions, and the ability of the model to reproduce the key visual patterns in the echograms.

The decoder reconstructed inputs from the latent features using a transposed convolution with a sigmoid activation function to ensure the reconstructed image pixel values range between 0 and 1 (Rumelhart et al. 1986).

Our final model was trained using a custom loss penalty based on the structural similarity index (SSIM) (Rumelhart et al. 1986; Wang et al. 2004). The SSIM (Eq. 3.2) is a perceptual metric designed to assess the similarity between two images by considering factors that are crucial to human visual perception including luminance, contrast, and structure (Wang et al. 2004) and is defined as:

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{3.2}$$

where x and y are the two images being compared, μ_x and μ_y are the mean contrast intensities, σ_x^2 and σ_y^2 are the variances, and σ_{xy} is the covariance of the images. We implemented a loss penalty by converting Eq. 3.2 into a distance metric, such that SSIM_loss = 1 − SSIM. This encouraged the model to produce image reconstructions that retained structural and perceptual quality, and showed improved performance relative to pixel-wise approaches such as mean squared error or peak signal-to-noise ratio (Wang et al. 2004). Image reconstructions built using the CAE with SSIM loss are shown in Fig. A.6.

Train, test and validation splits of the data were generated as random 0.80, 0.10 and 0.10 subsets of the full image dataset, respectively. Models were trained to convergence using early stopping criteria and a batch size of 64 images. We used a data augmentation protocol of random image brightening and darkening from the skimage python package and RandomAdjustSharpness from pytorch with $p = 0.5$ (augmentation randomly applied to inputs with 50% probability) and sharpness parameter of 2.5 (with values >1 indicating image sharpening) *in lieu* of explicit regularization. Learning was optimized using an Adam optimizer, which dynamically updates weights during training by incorporating past gradients and adaptive learning rates (Kingma and Ba 2015). Finally, we used a ReduceLROnPlateau scheduler to dynamically adjust the learning rate by a factor of 0.1 when the validation loss stagnated for a specified number of epochs (e.g., "patience"; (Bottou et al. 1994)). Hyperparameters were optimized based on final training loss, generalization to testing and validation data, and the visual quality of echogram reconstructions. Final and alternative hyperparameter sets, their performance impacts, and the selection criteria applied are summarized in Table A.6.

The final CAE model was used to embed and reconstruct the entire dataset, facilitating direct comparison to other dimensionality reduction algorithms where the embeddings are invertible (i.e., the images can be reconstructed from the embeddings). Comparisons of the mean squared error (MSE) and the SSIM for the full dataset were compared to linear PCA, non-linear kernel-PCA (kPCA; (Mika et al. 1998)). A wider survey of dimensionality reduction algorithms was limited by

the fact that non-invertible approaches do not facilitate 1:1 performance comparisons independent of subsequent analysis steps (clustering and prediction). As an experimental heuristic, embeddings built using Isomap and Spectral embedding were used in place of the CAE in the clustering workflow, and cluster metrics were compared (Choi and Choi 2004; Luo et al. 2003).

Clustering

For the clustering analysis, we aggregated the latent feature vectors into $3^\circ \times 3^\circ$ grid squares over the Pacific region (Fig. 3.1a; Fig. 3.2b) and computed the mean vectors. Clustering was performed on the mean vectors for each grid square (as in (Receveur et al. 2020b)) using the scipy implementation of the k-means algorithm (MacQueen 1967). We selected the k-means algorithm given that it broadly (though not dramatically) outperformed other approaches (e.g., DBSCAN, agglomerative clustering, spectral clustering) on standard cluster metrics including the silhouette score (Rousseeuw 1987), Davies-Bouldin index (Davies and Bouldin 1979) and the Calinski-Harabasz metric (Calinski and Harabasz 1974) (Appendix A.2; (Proud et al. 2017)).

Performance on clustering metrics indicated plausible solutions (peaks and 'elbows') in the k=4–10 range across different algorithms and k values. Based on k-means performance we adopted $k = 6$ where the Silhouette score was maximised, the Davies–Bouldin index was minimised, and consistent with the elbow point in the Calinski–Harabasz curve. This process does not yield an unequivocally optimal k, and we therefore used k values in the indicated range to produce alternate solutions (as in (Proud et al. 2017)). The k value (k=6) obtained through this process was used in performance comparisons to alternate dimensionality reduction algorithms (Sec. 3.2.3) and predictive models (Sec. 3.2.3). Given that clustering is naïve to geographic space, the coherence of clusters projected in latitude and longitude space (Fig. 3.3a) also provided a check on cluster performance. We formalized this for the final model using the spatial-autocorrelation metrics Moran's-I (Moran 1948) and Geary's-C (Geary 1954) to quantify the correspondence of the cluster labels to their geographic distribution (Sec. 3.3.2).

Each cluster in the final k-means solution was represented by a centroid vector summarizing the mean latent representation of all echograms within the cluster. These centroid vectors were reconstructed using the CAE decoder to produce 2D echogram images that visually represent the prototypical structural features of each region (Fig. 3.5). These synthetic reconstructions serve as a 2D analogue to profile-based representations from previous models and offer an interpretable view of dominant scattering patterns across clusters.

Predicting biogeographic regions from environmental covariates

To project cluster labels across unsampled areas, we trained a random forest model using key surface and subsurface variables known to influence mesopelagic structure including subsurface dissolved oxygen, sea surface temperature, near-surface chlorophyll (Chl-a), and mixed layer depth (Ariza et al. 2022; Bianchi et al. 2013; Bianchi and Mislan 2016; Proud et al. 2017, 2018; Song et al. 2022) (Fig. 3.1e–h). We obtained climatological monthly mean sea surface temperature and dissolved oxygen at 1-degree spatial resolution from the World Ocean Atlas 2023 (WOA), from 2000 to present (Garcia et al. 2024; Locarnini et al. 2024). We included subsurface oxygen in the model as the sum from 150–500 meters, following previous studies (Ariza et al. 2022). Near surface chlorophyll-a (Chl-a) data was obtained from NASA MODIS, using the climatological monthly mean from 2002 to present (0.1° resolution). Mixed layer depth was obtained via the NASA ECCO V4r4 climate model (0.5° resolution, (ECCO Consortium et al. 2021)).

Climatological monthly mean values for all covariates were averaged at the level of each $3° \times 3°$ grid cell and joined to the aggregated CAE-derived features. The covariate values used in model training and prediction were the averaged, climatological monthly means associated with the collection of echogram embeddings within each grid square. This approach is effectively the same seasonal weighted averaging procedure used in previous work (Ariza et al. 2022).

Our final regression model used a random forest classifier optimized using parameter search over different hyperparameter combinations for 50–1000 of estimators and 1–20 individual decision trees. This was done using the RandomizedSearchCV functionality from scikit-learn within the model_selection utility, which uses 5-fold cross-validation in iterations (Bergstra and Bengio 2012). Regression performance was compared across a number of different models and evaluated based on predictive accuracy of cluster assignments using the environmental variables. A model configuration with a max tree depth of 10 and 408 estimators produced the most accurate predictions and was used in the final analysis.

3.2.4 Echogram derived properties

Primary and secondary layer depths

To characterize the scattering distribution of the resulting clusters, we calculated the mean depths, ranges, and distributions of primary and secondary scattering layers, and the distance between them (hereafter inter-layer distance) for each region. Here, primary and secondary refer to the most prominent local maxima in the vertical profile of relative scattering intensity. Although echograms can exhibit more than two peaks, we focused on the two most prominent layers, as these typically

contain the bulk of the backscatter signal, and additional layers—when present—tend to be less distinct or biologically ambiguous. For each echogram, we determined if there were one or two layers and computed their depths using a data-driven approach that extends the commonly used weighted mean depth (WMD) or center of mass metric (e.g., (Irigoien et al. 2014; Klevjer et al. 2016, 2020; Proud et al. 2017; Song et al. 2022; Urmy et al. 2012)) to accommodate double-layer modes. Mean daily and nightly profiles were computed as ±3 hours around maximum and minimum solar angle, respectively. Mean profiles were convolved with a gaussian window using a fast Fourier transform (scipy.fftconvolve) to smooth out fine scale variability, and peak detection was performed to identify individual maxima (layers) using the scipy implementation of the topographic (peak) prominence algorithm (scipy.signal.peak_prominences). The existence of two layers was determined from the relative prominences of peaks. Specifically, the set of detected peaks in the profile $P = \{p_1, p_2, \ldots, p_n\}$ were ordered by prominence such that $\pi_1 \geq \pi_2 \geq \ldots \geq \pi_n$, where π_i is the prominence of peak p_i. The presence of two sufficiently large peaks was determined by computing the prominence ratio, $R = \pi_1/\pi_2$, which increases with the difference in the relative prominence between π_1 and π_2. If $R < T$, where T is a threshold value manually set by the user, the data was considered to contain two prominent peaks; otherwise, it was assumed to contain only one dominant peak. T was determined through supervised experimentation with the dataset. For each identified peak p_i, we defined its center of mass (COM), weighted by prominence, as $\mathrm{COM}_i = (\Sigma\, x_j \cdot w_j)/(\Sigma\, w_j)$ where: N_i is the set of indices surrounding peak p_i (typically within a local window), x_j are the positions of data points in the vicinity of p_i, w_j is the prominence-weighted contribution at x_j. Double layer classifications and layer depth estimates based on this process were manually confirmed for more than 25% of the individual data examples.

Migration attributes

We extracted DVM start (leave the surface) and end (arrive back at the surface) times from each echogram. S_v was subset to 0–100 m and depth-averaged into a 1-D signal (as in (Bianchi and Mislan 2016)). We applied a short rolling median to smooth the signal and fit a two-level "top-hat" model to determine the interval by least-squares fit (like the procedure used for DVM detection in Chun et al., 2021). Indicated change points were extracted as start and end times (analogous to the bulk arrival and departure metrics in (Bianchi et al. 2013)) (Table 3.2, Fig. 3.5). We computed two-sided standard deviations for start and end times and their coupling to sunrise and sunset respectively (as in (Bianchi et al. 2013)).

The migration fraction (MF)—the proportion of the animal biomass in each layer that participates in DVM—is a key parameter in carbon flux estimates (Archibald et al. 2019) with substantial ecosystem variability (Gorgues et al. 2019; Klevjer et al. 2020). Estimating the 'real' MF was not

possible here given normalization, the use of uncalibrated data, and the lack of observations from the upper 50 meters (Sec. 3.4.3). That said, studies using calibrated echosounder data have also used relative day-night intensities to proxy the MF given persistent uncertainty in nominal physics-based estimates (Klevjer et al. 2016). Here, we computed the ratio of the normalized scattering intensity ±50 m from the depth of each layer (primary and secondary, Sec. 3.2.4) in the 3 hours around the maximum and minimum solar angle. S_v values were clipped at 2.5 standard deviations to mitigate the impacts of extreme values. We refer to this as the pseudo migration fraction, and it was computed for all primary and secondary layers observed in the dataset.

Echogram structure index

We developed a metric to describe the coherence of echogram scattering structure. The metric combined the spectral anisotropy of the daytime subset of the image and the extent of concentration in the pixel intensities. Spectral anisotropy shows how strongly an image favours one direction, computed by taking a 2D FFT, summing energy by angle, and dividing the highest angle's energy by the average across angles. Variants of this approach have been used for image structure analysis in other disciplines (e.g. bio-medical imagery in (Brunet-Imbault et al. 2005)), and was here meant to describe the horizontal consistency of the scattering pattern during the daytime non-migrant period. We also computed the extent of concentration as the proportion of pixels (clipped at 2.5 standard deviations) that account for 70% of the total intensity (smaller values indicate higher concentration)—using the daytime subset. The echogram structure index was computed as the anisotropy score minus the concentration proportion, normalized between 0 and 1.

Basin scale property estimates

Derived metrics were estimated for the entire Pacific region to visualize how variability in key structural elements are captured by the CAE-based clustering. We predict property layers (Sec. 3.3.4, Fig. 3.6) using the same regression procedure and covariates described above for cluster prediction. For properties relating to secondary layer depths, a logistic regression for the presence or absence of secondary layers was used to mask predictions (i.e. values are only provided in a given parcel if secondary layers are predicted to occur, Fig. 3.6d).

3.3 Results

3.3.1 Quality controlled echogram image dataset

After quality control (Sec. 3.2.1), the resulting dataset contained 306 cruises from 4 different vessels (Fig. 3.2b), with 1,596 complete diel echograms spanning 50–850 m from 2001 to 2018. Normalized echogram images were used to train the CAE model, and reflected the key visual structures of interest such as scattering layers and migrations (as depicted in Fig. 3.2a). Seasonal coverage over the full archive was broadly distributed across the fall (28%), spring (34%), summer (24%) and winter (13%), though this varied with latitude.

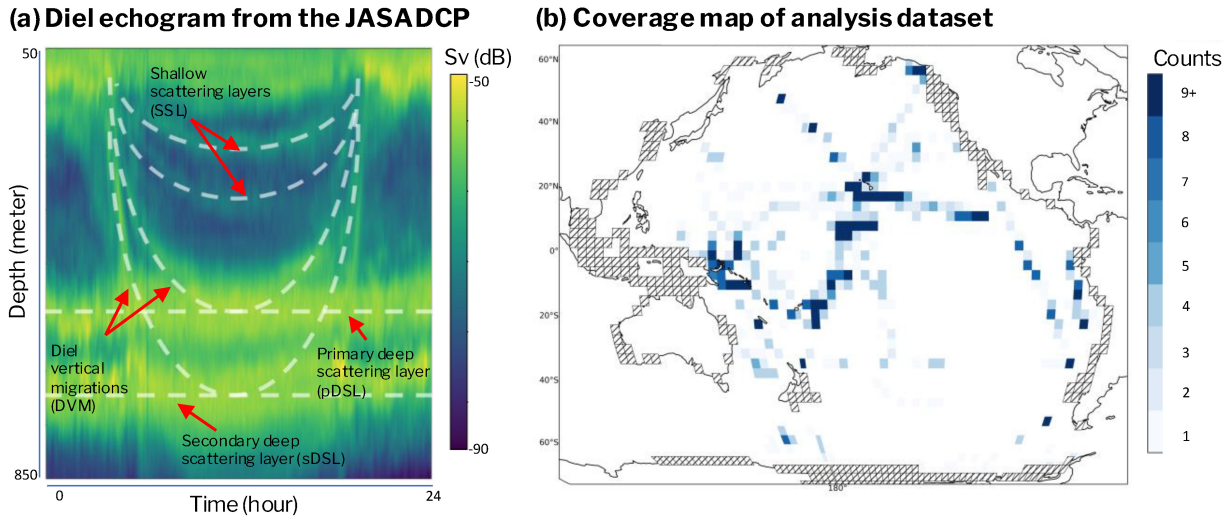

Figure 3.2: (a) Example diel echogram from the JASADCP archive, illustrating key structural features including shallow scattering layers (sSL), diel vertical migrations (DVM), the primary deep scattering layer (pDSL), and the secondary deep scattering layer (sDSL). Dashed white lines indicate approximate layer boundaries. (b) Coverage map of the analysis dataset showing the number of echograms per $3° \times 3°$ grid cell across the Pacific basin.

3.3.2 Final Model and Regionalization

CAE model and clustering

The best-performing CAE-SSIM model achieved SSIM-loss (1-SSIM; Sec. 3.2.3) of 0.164, 0.173 and 0.171 across training, testing and validation subsets respectively. The mean SSIM score for the reconstruction of the full dataset was 0.84. The CAE model trained with the SSIM loss penalty demonstrated superior performance on reconstruction metrics (SSIM & MSE) when benchmarked

against alternate reconstructions built using linear–PCA (SSIM 0.78)–and nonlinear–kernel PCA (SSIM 0.76) (Mika et al. 1998). Clustering scores based on the CAE feature set (silhouette of 0.25) exhibited better performance than scores computed by substituting embeddings built with Isomap (silhouette of 0.20), Spectral embedding (silhouette of 0.14), PCA (silhouette of 0.17), and kPCA (silhouette of 0.16). The performance of linear PCA was noteworthy given that it contains far fewer free parameters, it is highly legible, simple to implement, and PCA-variants have been used effectively in previous large-scale acoustics studies (Ariza et al. 2022). The comparison between PCA and CAE is discussed further in Appendix A.2.1, including direct comparisons between model reconstructions (Fig. A.6) and the number of learnable parameters (Table A.5).

The final 6-way k-means clustering achieved Silhouette and Davies-Bouldin scores of 0.25 and 1.36, respectively. Generally speaking, both of these scores indicate only moderate separability, though they align with expectations for inherently continuous property distributions like those expected for marine ecosystems (Duda et al. 2000). Further, while the clustering approach was naive to geographic space, the spatial projection of the clustering indicated visually apparent spatial fidelity (Fig. 3.3a). This visual pattern is supported quantitatively by a Moran's I of 0.65 and Geary's C of 0.20 (both $p = 0.001$), which indicated strong and statistically significant positive spatial autocorrelation in the clustering assignments and their distribution in latitude and longitude space. Clustering performance was also supported by the visual coherence of underlying images within each cluster, the predictability of the clusters using environmental covariates (Sec. 3.3.2, Fig. 3.4), the coincidence of cluster regions and derived properties (Sec. 3.3.4, Fig. 3.6) and the visual organization of the clusters visualized in 2-dimensions.

Predicting Pacific seascapes

Our final random forest model predicted cluster labels with a mean accuracy of 80.9% (±4.8%) over the 5 folds in cross validation, linking diel scattering structure to oceanographic conditions. The final regression model was used to predict the cluster assignment of all the parcels without data coverage in the Pacific region (Fig. 3.3b). The resulting model prediction was spatially coherent and reflected known large-scale oceanographic patterns in the Pacific, such as the Gyres and the zonal gradient of physical properties along the equator. Feature importances and an example confusion matrix are shown in Fig. A.14.

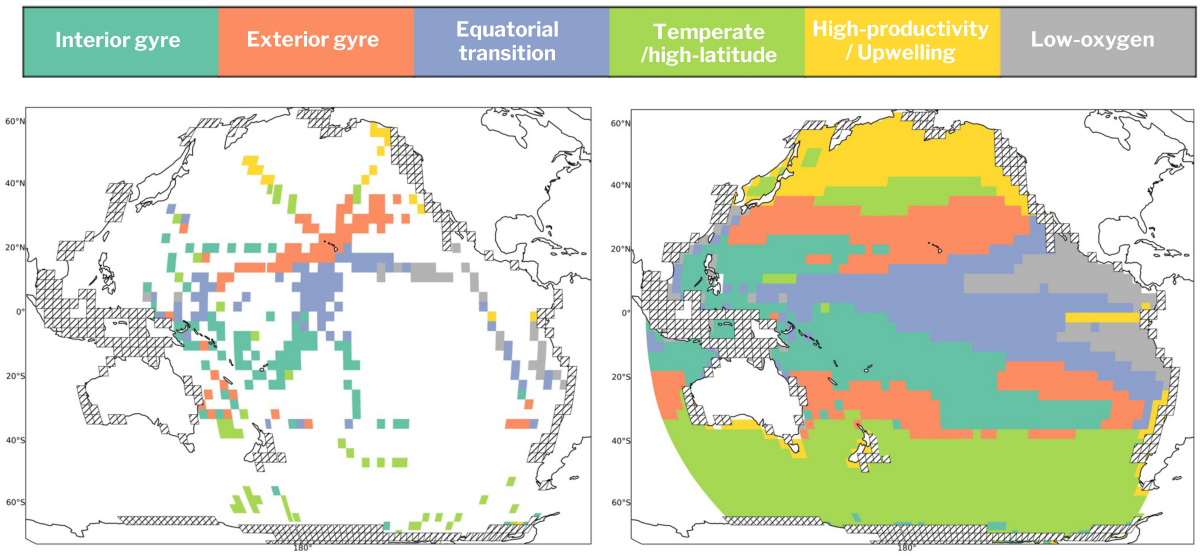

Figure 3.3: Biogeographic regionalization of the Pacific Ocean based on CAE-derived echogram features. (a) Cluster assignments for sampled $3° \times 3°$ grid cells, colored by region: Interior (tropical) Gyre, Exterior (subtropical) Gyre, Equatorial Transition, Temperate/High-latitude, High-Productivity/Upwelling, and Low-Oxygen. (b) Full biogeographic map after random forest prediction to unsampled parcels across the Pacific basin.

We assigned labels to the six clusters based on their spatial alignment with previous biogeographic solutions (Ariza et al. 2022; Proud et al. 2017) and/or biophysical properties associated with their geographic distributions: Low-oxygen, High-productivity/Upwelling, Temperate/High-latitude, Interior (tropical) Gyre, Exterior (subtropical) Gyre, and Equatorial Transition. The importance of covariates in random forest models indicated dissolved oxygen was the most influential variable (39% importance), followed by sea surface temperature (26%), near-surface chlorophyll (20%), and mixed-layer depth (13%). Mean representations of the four environmental covariates and contours of biogeographic regions are shown in Fig. 3.4 (radar plots of covariates and derived features by cluster are shown in Fig. A.15). Classification accuracy varied by region, with the Equatorial Transition being least accurate (70%) and the Exterior Gyre most accurate (92%) (Table 3.1; Fig. A.14).

Table 3.1: Summary metrics for the six biogeographic clusters, including the number of echograms (N echograms) and parcels (N 3×3 parcels), regression performance (predictive accuracy) and distributions of environmental properties. Computed means and deviation statistics are computed based on the aggregated climatological monthly mean properties of the underlying parcels.

Region	**N echo.**	**N parcels** (3×3)	**Pred. acc.**	**Mean CHL** ($mg\,m^{-3}$)	**Mean SST** ($°C$)	**Mean DO** ($mmol\,m^{-3}$)
Low-oxygen	145	29	0.86	0.17 ± 0.17	26.13 ± 2.5	171.79 ± 129.06
High-prod./Upwelling	88	22	0.80	0.70 ± 0.50	9.23 ± 3.81	1121 ± 405.71
Temperate/High-lat.	284	56	0.92	0.19 ± 0.07	10.34 ± 8.71	3251.42 ± 520.69
Equatorial transition	358	76	0.69	0.09 ± 0.05	27.49 ± 2.33	1117.84 ± 493.65
Exterior (subtrop.) Gyre	340	69	0.92	0.07 ± 0.02	24.52 ± 1.81	2573.28 ± 348.04
Interior (tropical) Gyre	381	84	0.71	0.09 ± 0.06	27.09 ± 3.14	2432.84 ± 482.97

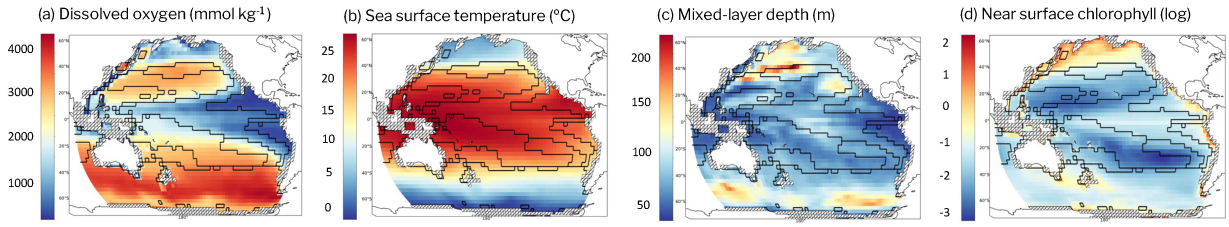

Figure 3.4: Mean representations of the four environmental covariates used in the random forest regression (sea surface temperature, near-surface chlorophyll-a, dissolved oxygen, and mixed layer depth), with contours of the six biogeographic regions overlaid. Radar plots summarize the relative magnitude of covariates for each region.

3.3.3 Fingerprints of Pacific seascapes

The centroid reconstructions (Fig. 3.5) revealed distinct structural scattering layer features in each region, including differences in layer depth and number, coherence, and diel migration patterns. These visual features were consistent with corresponding regional distributions of derived metrics (Sec. 3.2.4; Sec. 3.3.2) extracted from the underlying echograms (Table 3.2, Fig. 3.5), supporting the interpretability of the centroid representations, and underscoring the CAE model's ability to capture biologically meaningful features. Cluster centroids are shown in Figure 3.5, along with 'isolines' indicating the underlying means and standard deviations of layer depths and migration times.

Table 3.2: Derived features of echograms summarized by region. Statistics are computed based on all underlying echograms using the processes described in Sec. 3.2.4. Cells with NA* are incomplete because there are insufficient examples (<15% secondary layer incidence) to provide a robust depiction of the feature. pMFDSL and pMFSDSL indicate pseudo migration fractions for primary and secondary layers respectively.

Region	**2nd layer** (% obs)	**Prim. depth**	**Sec. depth**	**Inter-layer dist.**	**pMF DSL (%)**	**Transit time (h)**	**Structure index**	**Mig. start time**
Low-oxy.	15.2	357 ± 17	653 ± 161	321 ± 113	60.5	12.3 ± 1.0	0.77	05:44
High-prod.	12.5	444 ± 28	—	—	10.6	15.2 ± 2.0	0.53	04:16
Temp./H-lat.	7.4	518 ± 69	—	—	11.0	14.1 ± 3.2	0.36	04:28
Eq. trans.	86.0	447 ± 46	669 ± 57	224 ± 57	24.1	12.2 ± 0.4	0.48	05:56
Ext. gyre	91.8	510 ± 30	661 ± 34	160 ± 10	15.8	13.0 ± 1.1	0.46	05:37
Int. gyre	70.9	532 ± 31	753 ± 41	220 ± 30	23.4	12.3 ± 0.5	0.50	05:50

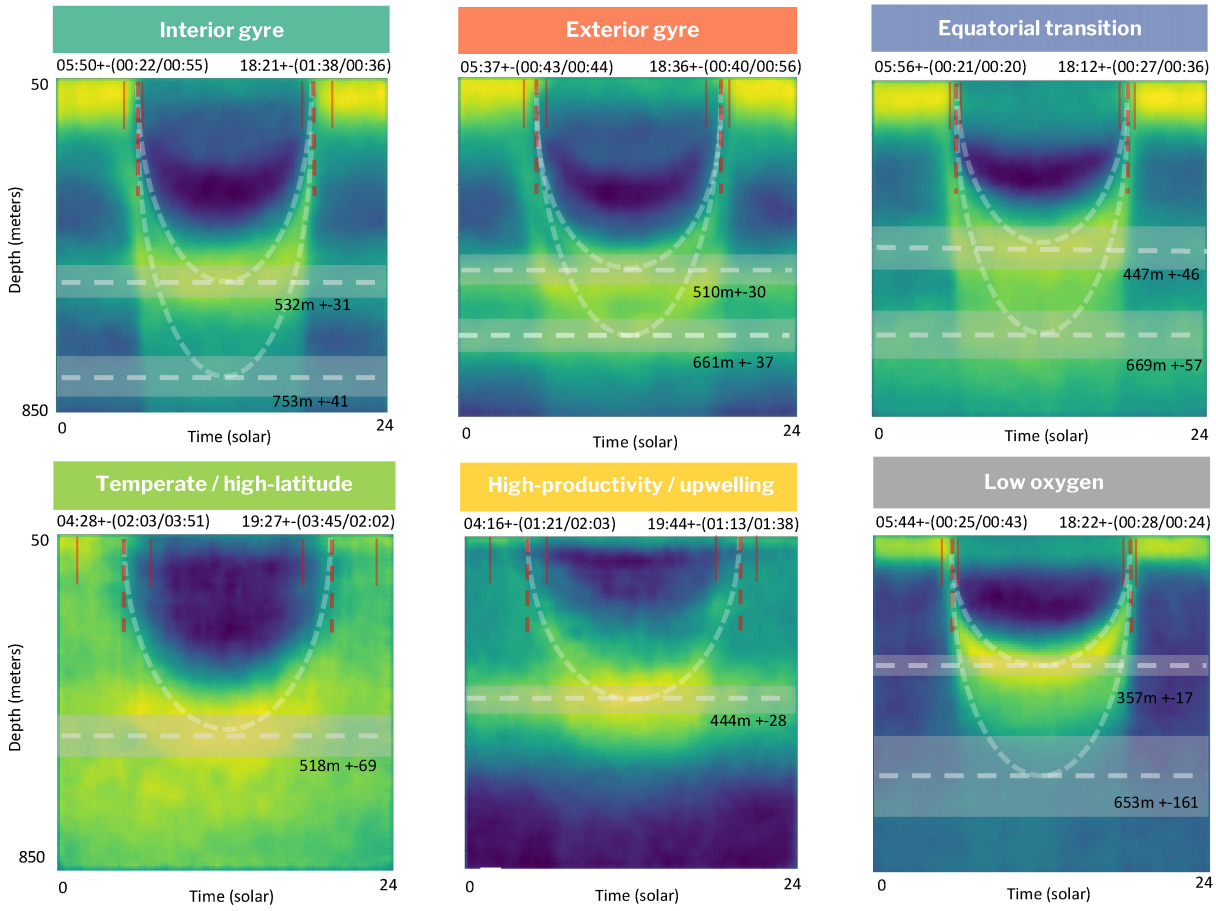

Figure 3.5: Centroid reconstructions ("fingerprints") for each of the six biogeographic regions, generated by decoding the cluster centroid vectors through the CAE decoder. Dashed horizontal lines indicate mean primary and secondary layer depths, with shading representing ±1 standard deviation. Short vertical bars indicate mean DVM start and end times. Migration start and end times with standard deviations are annotated above each panel. Primary and secondary layer depths with standard deviations are annotated within each panel. Each centroid captures the prototypical vertical structure and diel dynamics of its region.

3.3.4 Variability in layers and migrations

The correspondence of centroid reconstructions and derived properties (Fig. 3.5) was reproduced spatially in the correspondence of our biogeographic regions and basin-scale predictions of derived metrics (Sec. 3.2.4; Fig. 3.6)—including the depth and number of layers, the timing of and participation in migrations, and the level of structural coherence (Fig. 3.6).

Pacific basin. Basin-wide, migration transit time reflected broad latitudinal trends, with regional means approaching 12 h—the global mean photoperiod—while variability and maxima increased towards higher latitudes. Elevated means in upwelling and temperate zones likely resulted from seasonal and spatial sampling biases respectively. More complex regional differentiation is evident in other structural properties, which resist simple, basin-scale relationships.

Interior (tropical) Gyre. The primary (~532 m ±31 m) and secondary (~753 m ±41 m) layers were the deepest of all regions, with 71% of echograms indicating secondary layers. Both layers showed diel variability, with more intense scattering in the primary layer. The proportion of secondary deep scattering layers (present in 71% of total echograms of this region) observed was the lowest among regions where they were the dominant mode (i.e. where secondary layers occurred in >15% of echograms). Secondary layers in this region were often observed close to the 850 m depth horizon, suggesting that such layers may exist outside of the depth range of our analysis, which could explain the low proportion relative to the Exterior (sub-tropical) gyre. Coincident with the centroid reconstruction (Fig. 3.5, top), this is the only region where the pseudo migration fraction is higher, on average, for the secondary layer (28.4%) relative to the primary layer (23.4%). Arrival times of bulk migrations showed 69% greater variability (2 h 14 m) than departure times (1 h 19 m).

Exterior (subtropical) Gyre. Primary layers (~510 m ±30 m) were slightly shallower on average than in the interior gyre, and this region had the highest occurrence (91%) of secondary layers (~661 m ±37 m). The deep layers in this region showed the most similar intensities and the narrowest inter-layer depths (~160 m ±34 m; Sec. 3.2.4). Both layers were associated with clear migration trajectories and especially persistent daytime scattering. This visual aspect reflected the low pseudo migration proportion (~16%) in the region. Migration timing was essentially symmetric, though slightly more variable around sunset (1 h 36 m) than sunrise (1 h 27 m).

Equatorial Transition. Primary layers were shallower than the gyre regions (~447 m ±46 m) with secondary layers observed in 86% of cases at mean depths slightly deeper than those in the exterior gyre (~669 m ±57 m). Both layers exhibited diel migrations, with highly variable, and often stratified, secondary layers. Primary layers were generally more intense during the day, and less intense at night compared to secondary layers, reversing the trend from the interior gyre. Computed pseudo migration fractions for primary and secondary layers across the two gyre regions reproduced this visual comparison (Table 3.2). Migration timing is less variable than either of the previous regions (~42% and ~50% less than exterior and interior gyres respectively) though follows the same relative pattern as the interior gyre with 31% more variability at sunset. The mean structure index

was broadly consistent across this region and the two gyres, though here exhibited roughly twice the variability (standard deviation; Table 3.2).

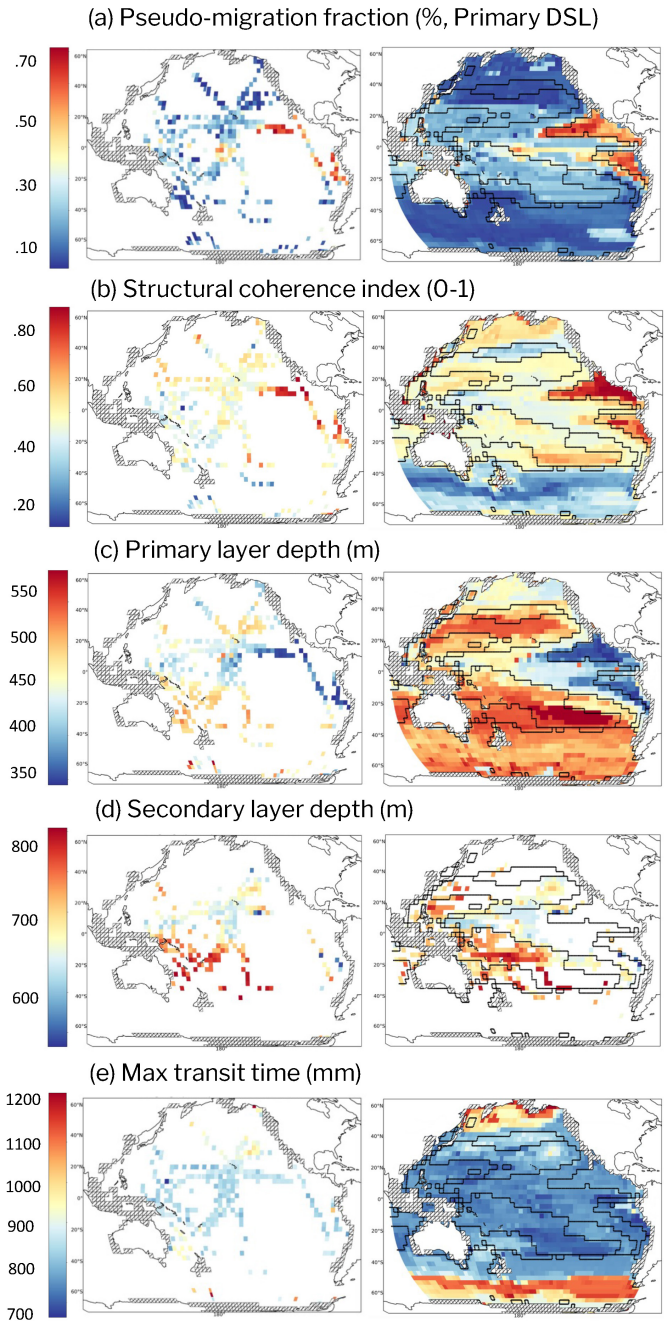

Figure 3.6: Basin-scale maps of derived structural properties predicted using the random forest regression framework and environmental covariates. Panels show (a) primary layer depth, (b) echogram structure index, (c) pseudo migration fraction for primary layers, and (d) secondary layer depth (masked where secondary layers are not predicted to occur). Biogeographic region contours are overlaid on each panel.

Temperate/high-latitude. Primary layers (~518 m ±69 m) were the most variable in this region, and less than 10% of examples showed well-defined secondary layers. This region was characterized by high ambient scattering in the mesopelagic and weak continuity among layers. Persistent diurnal background scattering below 400 m, and high-ambient scattering throughout the water column during the night, created a diffuse scattering pattern throughout the sampled portion of the water column. Diel migrations were less concentrated, with weakly resolved trajectories and high intra-day variability. Correspondingly, the structure index was substantially lower than all other regions. The pseudo migration fraction (primary layer) was near its lowest at 11% on average, and migration times were extremely variable—though essentially symmetric at sunrise (5 h 55 m) and sunset (5 h 48 m).

High-Productivity/Upwelling. This region was characterized by relatively shallow primary layers (~444 m ±28 m). Distinct secondary layers were rare (12.5%), though 57% of cases exhibited unique, double-layer-like migrations. These migrations targeted depths above and just below the primary layer, creating stratified daytime patterns, though these layers fully migrated at night and so were distinct from the other secondary layers in the dataset. Migration timing was highly variable as in the temperate/high-latitude region, though showed much greater variability at sunrise (4 h 05 m) than sunset (2 h 50 m). High background scattering was observed during the night and was concentrated above 450 m with limited structure below this depth, creating a diffuse nocturnal scattering pattern with a clear near-surface concentration. This lack of structure was not reflected by our structure index, likely due to the use of daytime subsets to compute anisotropy and concentration (Sec. 3.2.4). That said, the consistency of the depth-limited scattering mode amongst the underlying echograms produced regionally coherent structure index values (Fig. 3.6b). Pseudo migration fractions (primary layer) were the lowest here of all regions, though showed significant variability (Table 3.2).

Low-Oxygen. This region had the shallowest and least variable primary layers (~357 m ±17 m). Most examples lacked a visible resident deep scattering layer during the night, indicating that most of the resolved community participated in diel vertical migration. Correspondingly, the pseudo migration fraction was the highest of all regions, with values over 60% on average. Migration timing was highly stable—as in the equatorial transition region—though showed more variability (~30%) around sunrise then sunset. Secondary layers were observed in a fraction of examples (15.2%) at depths far below primary layers (~653 m ±161 m; largest regional inter-layer distance ~321 ± 121 m) and were largely non-migratory (lowest secondary layer pseudo migration fraction of ~7%).

3.4 Discussion

This study presents the first large-scale application of unsupervised deep learning to uncalibrated ADCP echograms for pelagic ecosystem classification. The ADCP archive used in this study spans nearly 500,000 km of surveyed transects throughout the Pacific basin, representing nearly 40% more data than the global 38 kHz archive used in prior work (350,000 km; (Ariza et al. 2023); Fig. 3.2b). By extracting low-dimensional structural features from >1,500 diel echograms, we identify six distinct regional "fingerprints" of mesopelagic structure that align with known environmental gradients and existing biogeographic models. Our findings show that regional patterns of mesopelagic scattering structure can be inferred from normalized acoustic imagery alone, enabling standardized analysis across platforms and offering a scalable framework for integrating structural insights into ocean observation networks.

3.4.1 Spatial Structure and organization

Our biogeographic model closely matches previous basin-scale acoustic classifications (Ariza et al. 2022; Proud et al. 2017), which are also structured around key environmental gradients such as oxygen, chlorophyll, temperature, and mixed layer depth (Fig. 3.4). Like those models, ours delineates the major gyres, temperate bands, and equatorial transition zones, with close correspondence to previous solutions (Ariza et al. 2022; Proud et al. 2017). Despite the general similarities, two key differences emerge: (1) our model extends the temperate region across the Southern Ocean at ~40°S, whereas previous models indicate an incursion of oligotrophic modes, and (2) our model does not distinguish a separate polar mode.

These differences result in part from the combination of imbalanced dataset coverage, clustering approach and the number of regions (k) in our model. The majority of echograms from the temperate/high-latitude region are from 'polar' latitudes south of 50°S (~85%) and show a high level of scattering consistency. In contrast, the subset of temperate/high-latitude echograms from north of 50°S is small (n=22 in the southern hemisphere) and inconsistent. The small size and heterogeneity of the lower-latitude subset does not computationally justify a separate mode based on cluster performance. Thus, our temperate/high-latitude region can be understood as an extended polar mode that captures an under sampled, heterogenous temperate region. Some distinction between temperate and polar modes does begin to emerge at higher clustering complexity (e.g. for $k = 8$), but it is weakly defined and contributes to decreased performance on cluster metrics. We retained $k = 6$ in the main analysis, though provide solutions for alternate k values (k={4–8}) and example echograms illustrating the substructure in the temperate/high-latitude region.

Alternatively, spatial disagreement between our biogeography and previous models could represent meaningful differences related to the method of feature extraction. For example, rather than an artifact of sampling bias, the grouping of temperate and polar modes in our solution could reflect an internally valid representation of the data based on how variability in the signal is encoded in the CAE features. While distinct from previous acoustics-based biogeographies, the broad extent of the temperate/high-latitude regions south of 40°S in our solution maps almost exactly to global models of vertical migration using satellite-based LiDar (Behrenfeld et al. 2019). These solutions find that outside of the permanently stratified ocean (e.g. outside of 40°), multi-scale spatiotemporal variability dominates the migration signal. Scattering variability at high-latitudes is caused by intra-day changes to cloud cover (Omand et al. 2021), sub-seasonal variability related to upwelling, mesoscale activity (Della Penna et al. 2022) and or fine-scale shifts in in plankton communities (Behrenfeld et al. 2019), and longer period variability due to extreme seasonality. Our approach to feature extraction explicitly captures this multi-scale variability in a way that previous acoustics-based approaches have not: intra-day variability in scattering patterns is captured through the use of diel echograms, while larger scale spatiotemporal variability is captured through parcel level aggregation of individual CAE feature vectors. High-frequency structural variability (Sec. 3.4.2) is characteristic of *individual* echograms from the polar subset of the temperate/high-latitude region (i.e., highest latitudes) but can also be encoded into the parcel level representations through the aggregation of variable scattering modes. Here, the substructure in the temperate/high-latitude region is de-emphasized by the fact that intra-day variability and spatiotemporal variability both encode a lack of structural coherence (Sec. 3.4.2) at the parcel level. Ultimately, differing representations produced by differing feature-extraction choices will emphasize different facets of regional scattering; disagreements are expected, not necessarily mutually exclusive. Future work combining high-dimensional structural embeddings (e.g., CAE-derived) with calibrated intensity data could facilitate ecosystem partitions that reflect more wholistic, multi-faceted gradients in scattering.

3.4.2 Vertical Structure

Regional modalities revealed by our clusters vary in terms of the structural *components* (layers and migrations, Fig. 3.2a) present in echograms and the visual *coherence* of the scattering patterns (the clarity and consistency of components and their organization in space and time).

Components. There is substantial variability across regions in terms of the number, presence and location scattering components (visualized in Fig. 3.2a). The range and distribution of mean primary scattering layer depths (357 to 532 m) corresponds to previous studies ((Bianchi et al. 2013) (350–600 m); (Proud et al. 2017) (460–550 m); (Ariza et al. 2022) (~350–550 m)). Secondary layer

depths are anecdotally supported by the few estimates of secondary layers (e.g. (Proud et al. 2018)). At the basin scale, layer depths and the distance between layers (inter-layer distance) coincide with insights from existing research describing the tendency of mesopelagic organisms to seek out resident depths that balance the desire for distinct light comfort zones (Aksnes et al. 2017) for predator avoidance, alongside tolerance for abiotic properties of the at-depth environment—such as the concentration of dissolved oxygen (Bianchi and Mislan 2016). For example, the zonal gradient of primary layer depths across the basin is well described by our resulting regions, and coincides with insights regarding the combined influence of deepening light penetration under low-productivity regimes and the increased availability of oxygen at depth (e.g. (Bianchi and Mislan 2016; Urmy et al. 2012)). Further, larger inter-layer depths in the oligotrophic interior tropical gyre relative to the Equatorial transition and exterior subtropical gyre regimes coincide with the expectation that the distance between isolumes increases at mesopelagic depths, governed by trade-offs between the intensity of incoming solar irradiance and near surface chlorophyll (Aksnes et al. 2017; Røstad et al. 2016).

The timing and participation of migrations across regions reflect general insights from the literature. The highest pseudo migration fractions are tightly concentrated in the low oxygen region, which closely contours the $\sim$1.5 ml O_2 l^{-1} breakpoint identified as the level at which oxygen exerts primary control on migrations (Klevjer et al. 2020). The uniquely high migration fraction here also reflects the distinct taxonomic composition—which has been proposed as one of the key determinants of the migration proportion (Klevjer et al. 2016)—characterized by a large proportion of endemic taxa adapted to extreme vertical property gradients (Sutton et al. 2017). Low migration fractions in productive systems (temperate/high-latitude and high-productivity/upwelling) likewise reflect previous insights regarding the decreased profitability of migration in pigment-rich, turbid waters where chlorophyll decreases both the proportion of total light available and the visual range of predators, thereby narrowing the antipredation window and the benefit of migration (Bianchi and Mislan 2016; Klevjer et al. 2020). Patterns in migration timing were more equivocal, though greater variability in departure times relative to arrival times in the low oxygen region coincide with previous research suggesting that early departure in this region may reflect tolerance for hypoxic conditions at depth. This pattern is also observed in the high-productivity/upwelling region, though here may be an artifact of extreme seasonality and a relative lack of stability in bulk migration time estimates due to the weak DVM signals. All other regions show either symmetric arrival and departure timing or greater variability around sunset, reflecting existing hypotheses that upward migrations take longer due to the influence of gravity (Bianchi and Mislan 2016).

Finally, centroid reconstructions contain additional complexity, including the existence of shallow migrant scattering layers in the equatorial transition and gyre regions and very deep secondary

migrations in the low-oxygen region (Fig. 3.5). Shallow scattering layers, which occur in the upper 250 m of the water column and do not connect with the primary deep scattering layer, are well documented and have been noted in many regional studies (e.g., (Arostegui et al. 2023; Klevjer et al. 2016; Receveur et al. 2020b; Wang et al. 2014; Wiebe et al. 2023)), though their variability and distribution are not well described. Likewise, the existence of very deep and relatively weakly scattering secondary migrations outside of the 850-m range in the low-oxygen region is consistent with previous work, which has noted nearly identical migrations in the same region (Klevjer et al. 2016). Both features are not reflected in previous biogeographical models using more coarse structural feature sets, highlighting the loss of potentially valuable information in other approaches. For example, accurately capturing deep migration target depths is important for quantifying their role in the biological carbon pump as they may accelerate export by bypassing more near-surface remineralization (McMonagle et al. 2023), while shallow scattering layers have been shown to be an important resource for predators (Arostegui et al. 2023). Ecologically meaningful variability in scattering structure has been documented at even finer, sub-daily time scales (Urmy and Benoit-Bird 2021), including short-term disruptions of scattering layers due to predator foraging events (Urmy and Benoit-Bird 2021) or mesoscale intrusions (Della Penna et al. 2022). Fine-scale phenomena such as these are not explicitly resolved in our model, though may be incorporated into feature sets indirectly as variability (Sec. 3.4.1; coherence below). Future work could integrate anomaly-detection techniques within the CAE framework to isolate such transient deviations from baseline structural patterns, providing new insight into dynamic ecosystem interactions.

Coherence. Echogram coherence refers to the visual clarity and smoothness of echogram features, shaped by fine-scale horizontal variability and background scattering, which affect feature definition and prominence (Lee and Staneva 2020; Proud et al. 2018). Coherence is also incorporated into our clustering through aggregation (Sec. 3.4.1) that encodes larger scale (e.g. seasonal or mesoscale) variability across multiple echograms. Like the more fine-scale component features, scattering coherence at the level of individual echograms is reflected poorly in profile-based solutions, as taking the means of day and night profiles explicitly eliminates intra-period variability. Our approach incorporates this variability via the CAE-SSIM features (Sec. 3.4.1). For example, the lack of edge structure in the echograms and centroids from the high-productivity/upwelling and temperate/high-latitude regions coincides with insights from global (Behrenfeld et al. 2019; Proud et al. 2018) and targeted (Della Penna et al. 2022; Omand et al. 2021; Sato et al. 2024) studies that reveal causes of low coherence and their spatial distributions.

As noted above (Sec. 3.4.1), fine-scale spatial and temporal variability in environmental and biological parameters—e.g. clouds and plankton community composition, respectively—have been shown

to cause highly-variable and unstructured scattering patterns in high latitude regions that directly coincide with the high latitude components of our upwelling and temperate regions. Mesoscale energetic variability has also been shown to modulate pelagic ecosystem structure (Della Penna and Gaube 2020; Della Penna et al. 2022; Perelman et al. 2023), and observations in our dataset from the temperate/subpolar region in the South Pacific are concentrated in regions of high eddy kinetic energy near the East Australian Current / Extension, the Tasman Sea, and the Arctic Circumpolar Current. Recent research has also noted fine-scale variability and unstructured scattering modes associated with upwelling regimes that are not in high-latitudes such as the equatorial component of our upwelling ecoregion (Sato et al. 2024).

3.4.3 Study limitations

Our unsupervised, image-based approach prioritizes visual structure, avoiding subjective thresholding and enabling the model to capture complex visual patterns in diel echogram data. While effective, this method has limitations, particularly the loss of intensity information due to normalization, which prevents direct inference of biomass or community composition. The use of normalized data eliminates the intensity signal, such that the solution presented here reflects the structural variability in scattering modes only (i.e., "fingerprints"), and neither incorporates nor directly reflects corresponding patterns in biomass. Therefore, normalization introduces the possibility that biologically similar echograms—differing only in the presence of a high-intensity school or depth of a resonant layer—could be scaled differently, thereby altering their contrast and potentially affecting their placement in the learned feature space. We also recognize that the exclusion of the upper 50 m may reduce model sensitivity to shallow migrants, and future work could evaluate this effect by incorporating datasets with full water column coverage. While our CAE is designed to prioritize consistent spatial patterns and minimize sensitivity to intensity spikes, this normalization step may cause the same scattering community to appear differently depending on context. Consequently, our clusters represent structural scattering modes—i.e., visual organization and coherence of features—rather than discrete ecological communities per se. These modes may integrate a mix of ecosystem composition, scattering features, and observational artifacts, particularly in cases where strong intensity anomalies or vertical migration extremes are present. Future work could mitigate this by explicitly modelling the impact of rare high-intensity features through data augmentation, or by testing normalization schemes that dampen outliers more selectively (e.g., percentile clipping or local contrast scaling).

Our solution is also impacted by instrument frequency and noise, which further impact the visual appearance of echograms. First, all data in this study were collected using 38 kHz ADCPs. Although normalized scattering strength was used to reduce the influence of ADCP-specific and

contextual variability, frequency affects how biological layers are represented, depending on scatterer composition (Bassett et al. 2020; Benoit-Bird et al. 2009a; Izard et al. 2025b). For example, organisms with gas-filled swim bladders may appear more prominently at 38 kHz than at higher frequencies, influencing the apparent structure of layers (Dornan et al. 2019; Izard et al. 2025b). Second, the use of a non-thresholding approach means that we do not formally quantify and remove a significant proportion of the noise in the echogram data. Noise varies over time, by instrument, by vessel, by sea state, by cruise, and a number of other factors, introducing considerable heterogeneity. However, because convolutional autoencoders compress the full data into a reduced set of latent components and reconstruct from those, they tend to exclude noise by design—especially when that noise is inconsistent or lacks spatial or temporal coherence (Gondara 2016; Pintelas et al. 2021). This compression acts as a form of implicit denoising, favoring consistent structural patterns while discarding idiosyncratic variation. Moreover, even if noise exhibited systematic signatures tied to specific cruises or instruments that could, in theory, bias classification in a supervised neural network, the autoencoder is not optimized to distinguish between such sources. Instead, it learns to represent generalizable features that support accurate reconstruction across all samples, thereby disincentivizing overfitting to cruise- or instrument-specific artifacts (Goodfellow et al. 2016).

Finally, seasonality in the underlying data is distributed unevenly among the resulting clusters, with the Exterior Gyre, Temperate/Subpolar, Equatorial Transition and Interior gyre showing broad seasonal coverage, and both the Low-Oxygen (<10% observations from summer and fall combined) and High-Productivity/Upwelling showing significant imbalance (97.4% from summer). A number of studies have shown that scattering patterns and features (e.g. DSL depth) show substantial seasonal variability (Cisewski et al. 2021; Geoffroy et al. 2019; Urmy et al. 2012; Wang et al. 2014), though no basin-scale bioacoustics models—including the one presented here—have had sufficient data to resolve temporal variability. The most obvious influence of seasonality is the changing of the photoperiod—especially pronounced from the sub-tropics to the poles—which directly influences light structured diel migrations. This is resolved at a high level in our derived metrics, and the within-region variability in migration timing is shown through the isolines in Fig. 3.5. Further, while computed mean transit times are relatively stable across the regions, the variance of transit time shows a highly consistent meridional pattern. That said, the impact of seasonality on scattering structure is largely unresolved. The solution presented here should be taken as an estimate of the climatological mean distribution of regions.

3.5 Broader applications, alternate formulations and future directions

A key motivation for this research is the insight from previous acoustic-based biogeographies that vertical structure and scattering intensity are closely coupled, as evidenced by spatial patterns in earlier solutions (e.g., (Ariza et al. 2022; Proud et al. 2018)). The connection between structure and intensity is supported by regional studies (Benoit-Bird et al. 2009b; Sato et al. 2017), and by the reliance of global biomass estimates on regressions that incorporate structural features as key covariates (e.g., the use of primary layer depth to predict intensity in (Proud et al. 2017) and subsequently abundance in (Proud et al. 2019)). Through this lens, the use of CAEs reflects the intention to facilitate a more comprehensive representation of the structural signal (e.g. the rich visual detail in diel echogram imagery), and to leverage this more complete structural representation, *in lieu* of intensity, to describe regional variability across ecosystems. The success of this strategy is indicated by the congruence of our spatial model and vertical modes with previous studies despite the use of an 'intensity-agnostic' approach.

Our biogeography is based on ADCP data, which is of lower quality than the echosounder datasets used previously. All current basin scale biogeographical models face the same limitation of the lack of data coverage, which has prevented them from resolving temporal variability, which is known to be significant in certain regions (e.g., (Cisewski et al. 2021)). This is a key obstacle, as temporally varying models would facilitate the identification of more detailed relationships between scattering dynamics and surface environmental properties (e.g., through seasonal changes). Better characterizing this variability will have broad implications across ocean science research, from predator-prey interactions in commercially and ecologically important species (Braun et al. 2022, 2023) to carbon sequestration and climate (Archibald et al. 2019; Thibault et al. 2025). The lack of interoperability among instruments and frequencies is largely due to the use of scattering intensity in conventional analysis approaches that are highly sensitive to error, especially at mesopelagic depths where gas-bladdered organisms can produce resonant backscatter (Bassett et al. 2020; Cotter et al. 2021a,b; Proud et al. 2019), and so rely on precise calibration of instruments and corresponding environmental properties that influence numerical scattering. Consequently, scattering measurements made by different types of observing platforms, such as ADCPs and echosounders, have not been used in integrated datasets to facilitate increased coverage. It is not clear that this lack of interoperability extends to the structural (i.e., visual) content of cross-platform observations—for example the number, depth and coherence of scattering layers—and in fact the spatial location of aggregates in contemporaneous measurements show a high-degree of stability relative to numerical intensity (Benoit-Bird et al. 2009a; Lee et al. 2004). These insights, in combination with the close

correspondence of ecosystem structure and intensity noted above, imply the possibility that mixed input datasets could be used together in integrated models to study scattering structure across ecosystems. Given the scale of existing archives of lower quality data (e.g. JASADCP), and the proliferation of acoustic observations across a diverse range of platforms, methodological approaches that are robust to data quality and calibration status offer the potential to leverage the scale and diversity of these archives, and facilitate more sophisticated models linking spatial, vertical, and temporal variability in pelagic ecosystems.

Chapter 4

Environmental Drivers and Basin-Scale Distribution of Shallow Scattering Layers Across the Indo-Pacific Mesopelagic Zone

This chapter is an unpublished manuscript currently in revision, co-authored with Martin C. Arostegui and Camrin D. Braun.

Abstract

Sound scattering layers are among the most pervasive features of the oceanic water column, yet intermediate-depth acoustic aggregations between the near-surface epipelagic community and the canonical deep scattering layer (DSL) have received comparatively little systematic study at basin scale. These intermediate layers—which we term shallow scattering layers (SSLs)—occupy a distinct vertical niche between the DSL and euphotic zone scattering. SSLs have been noted in prior studies using shipboard acoustic echograms, yet their environmental controls, internal substructure, and geographic extent remain poorly characterized. This study presents a basin-scale description of SSL incidence, acoustic subtypes, and environmental controls across the Indo-Pacific, derived from 1,451 daily 38-kHz echograms from the Australian Integrated Marine Observing System (IMOS) Ships of Opportunity Bioacoustics (SOOP-BA) program. An XGBoost incidence model trained on environmental covariates co-located with each echogram predicts SSL-suitable habitat across 43.6% (33.1%) of a 1° Indo-Pacific climatological grid based on a modeled occurrence probability threshold of 0.50 (0.75). A convolutional U-Net was used to segment SSL regions in echograms, and acoustic features extracted from these masks were used to resolve three SSL subtypes via unsupervised clustering. These subtypes differ in scattering intensity, layer morphology and depth, and migration behavior: a cold subpolar assemblage (C0; mean SST 11.9°C), a warm mesotrophic assemblage with pronounced diel vertical migration (C1; mean SST 23.4°C), and a warm oligotrophic assemblage characterized by diffuse, largely non-migratory scattering (C2; mean SST 22.6°C). Top environmental predictors—surface dissolved oxygen (feature gain 0.193), upper oxygen minimum zone depth (0.190), and chlorophyll-*a* (0.175)—jointly reflect the light–oxygen–productivity structure of the upper mesopelagic niche. SSL mean depth (245 ± 38 m) scales with light penetration depth at a slope six-fold shallower than the DSL, indicating SSLs track a distinct, shallower isolume. Beyond their value as oceanographic indicators, SSLs may be relevant to both predator foraging ecology—as recently demonstrated for open-ocean endotherms that preferentially forage at intermediate scattering layer depths—and to mesopelagic carbon export, by adding vertical substructure to active carbon flux that may modulate sequestration efficiency in certain regions. An exploratory carbon flux analysis suggests that resolving SSL-depth migration separately from the DSL shifts standard migration proportion estimates, with the magnitude of this effect ranging from negligible in warm-water environments to potentially meaningful fractionalization in subpolar settings.

4.1 Introduction

The mesopelagic zone (200–1000 m depth) mediates a substantial fraction of the biological carbon pump through the diel vertical migration (DVM) of micronekton and macrozooplankton (Boyd et al. 2019). The deep scattering layer (DSL), typically centered at 400–600 m in the Indo-Pacific, is the primary acoustic signature of this migrant biomass and has been extensively characterized across ocean basins (Bianchi et al. 2013; Klevjer et al. 2016; Proud et al. 2017). These ecosystems exhibit complex vertical layer structure with regionally specific acoustic signatures arising from biophysical interactions between water mass properties, ambient light penetration, and the physiology and behavior of mesopelagic organisms (Ariza et al. 2022; Culhane et al. 2025; Klevjer et al. 2016; Proud et al. 2018). Among these structures, shallow scattering layers (SSLs)—discrete, acoustically coherent aggregations at 100–400 m depth, distinct from both the near-surface epipelagic community and the DSL—have been widely noted anecdotally across regional acoustic studies, yet have received comparatively little systematic attention despite their potential ecological and biogeochemical function (Arostegui et al. 2023).

A number of prior studies have noted scattering layers at intermediate depths above the DSL across multiple ocean basins and under a range of conditions. In the Indo-Pacific, such layers have been resolved across the tropical Indian Ocean, the western Pacific warm pool, and into subpolar and subtropical latitudes, commonly occupying depths between 100 and 400 m during daytime when DSL organisms are deeper (Diogoul et al. 2026; Domokos et al. 2023; Klevjer et al. 2012; Proud et al. 2018; Receveur et al. 2020a; Song et al. 2022; Wang et al. 2014). Regional studies in the Red Sea documented mesopelagic scattering structure at comparable depths in highly stratified, oligotrophic conditions, with multi-frequency evidence pointing toward distinct taxonomic assemblages including myctophids and sternoptychids (Klevjer et al. 2012). In the Atlantic, basin-wide acoustic surveys and observations along the Gulf Stream and in the Sargasso Sea have resolved layered structures above the DSL in oligotrophic waters, often co-occurring with deep light penetration and thermal stratification (Wiebe et al. 2023). Multi-frequency ship-based surveys in the subtropical western Atlantic documented distinct surface scattering layers at depths shallower than 300 m, characterized by thin, intense aggregations (~20 m thickness) at the upper edges of deeper diffuse layers, with net-validated taxonomic composition varying spatially across foraging habitats of deep-diving predators (Sato and Benoit-Bird 2017). Buoy-based and time-series acoustic records have further demonstrated that SSL-like features exhibit pronounced seasonal dynamics—appearing preferentially during thermally stratified, low-productivity periods and diminishing during winter mixing—linking their occurrence to phenological cycles in water column structure rather than fixed geographic positions (Kaartvedt et al. 2012a; Wang et al. 2014). Multi-frequency acoustic analyses and net-sampling studies have provided anecdotal evidence for taxonomic composition:

gas-bladder–bearing fish (*Vinciguerria* spp., *Maurolicus* spp.) have been associated with strong 38-kHz backscatter at intermediate depths in tropical oligotrophic settings (Klevjer et al. 2012; Proud et al. 2017), while euphausiids, siphonophores, and lipid-rich myctophids appear to dominate in cooler, more productive waters (Dornan et al. 2019; Proud et al. 2019). The frequency-dependent backscatter characteristics of these intermediate layers vary regionally in ways that reflect local assemblage composition and depth-dependent resonance effects (Izard et al. 2025a, 2021). Across these regional studies, water mass properties—particularly thermocline depth, dissolved oxygen gradients, and light penetration—have emerged consistently as co-variates of SSL occurrence and depth, suggesting that biophysical interactions structure the intermediate-depth niche that SSLs occupy.

Beyond their value as oceanographic indicators, SSLs may function as ecologically and biogeochemically consequential features of pelagic seascapes. As features that partition the water column into distinct vertical strata, they represent accessible, intermediate-depth foraging opportunities for large open-ocean predators that would otherwise face a costly descent to DSL depths. Open-ocean predators—including certain billfishes, tunas, and sharks—routinely perform dives to intermediate mesopelagic depths above the primary DSL (Braun et al. 2023). Acoustically derived micronekton abundance at mesopelagic depths has been shown to significantly improve habitat predictions for tunas and seabirds in the southwest Pacific, with prey layers at 10–200 m and 200–400 m depth contributing measurably to predator distribution models alongside oceanographic covariates (Receveur et al. 2022). Albacore tuna in the North Central Pacific have been shown to preferentially forage at intermediate scattering layer depths where warmer temperatures substantially reduce thermoregulatory metabolic costs relative to deeper DSL foraging (Arostegui et al. 2023), and pelagic thresher shark in the Red Sea consistently target similarly intermediate scattering layer depths rather than pursuing DSL prey within or below an oxygen minimum zone (Arostegui et al. 2020). SSL migrants also inject respired carbon at 200–300 m rather than the 400–600 m typical of DSL migrants, potentially intercepting a fraction of the active carbon flux before it reaches sequestration depths and reducing flux-weighted efficiency (Bianchi et al. 2013; Martin et al. 1987). Progress in understanding these ecological and biogeochemical roles has been limited by a lack of models for the spatial extent and environmental controls on SSL incidence, and of systematic characterization of SSL attributes—including migration behavior, depth, and scattering intensity—that directly influence their function in predator foraging and carbon export.

This paper presents a basin-scale assessment of SSL distribution, environmental drivers, and acoustic substructure across the Indo-Pacific. Using 1,451 daily 38-kHz echograms from the AODN SOOP-BA program (2013–2025), we apply a U-Net segmentation model and XGBoost classifiers to quantify SSL occurrence, resolve environmentally coherent subtypes, and generate climatological

habitat predictions. We explore the environmental drivers of SSL incidence and identify a multi-factor niche—shaped by light penetration, surface oxygen, and OMZ depth—where SSLs are likely to occur, and discuss the distinguishing attributes of each subtype. We additionally present an exploratory SSL-aware migration fraction analysis to assess how scattering layer substructure may modulate mesopelagic carbon export efficiency.

4.2 Materials and Methods

The analysis pipeline (Fig. 4.1) comprises five stages: (1) acquisition and quality control of daily 38-kHz echograms from the AODN SOOP-BA archive; (2) binary SSL classification via expert labeling, inter-rater validation, and XGBoost incidence modeling; (3) SSL masking via U-Net segmentation with manual audit, followed by acoustic feature extraction and environmental co-location; (4) unsupervised clustering of SSL subtypes and spatial prediction across an Indo-Pacific climatological grid; and (5) descriptive analyses including frequency differencing and exploratory migration fraction estimation.

4.2.1 Acoustic dataset

Daily 38 kHz echograms were obtained from the AODN SOOP-BA archive (Haris et al. 2021), accessed via the Australian Ocean Data Network S3 API. The full dataset comprises 1,451 echograms spanning 2013–2025, covering the region approximately 60°S–40°N, 20°–290°E. Each echogram represents a 24-hour period centered at peak solar angle, integrated at 1 nmi horizontal and 5 m vertical resolution.

Each echogram was regridded onto a standardized 1440-minute × 160-depth-bin array (5 m vertical resolution, 0–800 m) with two-tier temporal forward-fill (20-minute limit during daytime, 120-minute limit during nighttime) to account for variable ping rates across vessels. Data were accessed via the AODN S3 datastore as calibrated volume backscattering strength (S_v, dB re 1 m^{-1}), processed by the SOOP-BA facility, which applies instrument-level quality control including calibration verification and noise removal during processing (Haris et al. 2021). Echograms were retained if diel completeness—the fraction of the 1440-minute daily grid containing valid acoustic data after forward-fill—exceeded 90%, ensuring sufficient temporal coverage to resolve diel vertical migration patterns. After this exclusion, 1,451 echograms were retained for analysis.

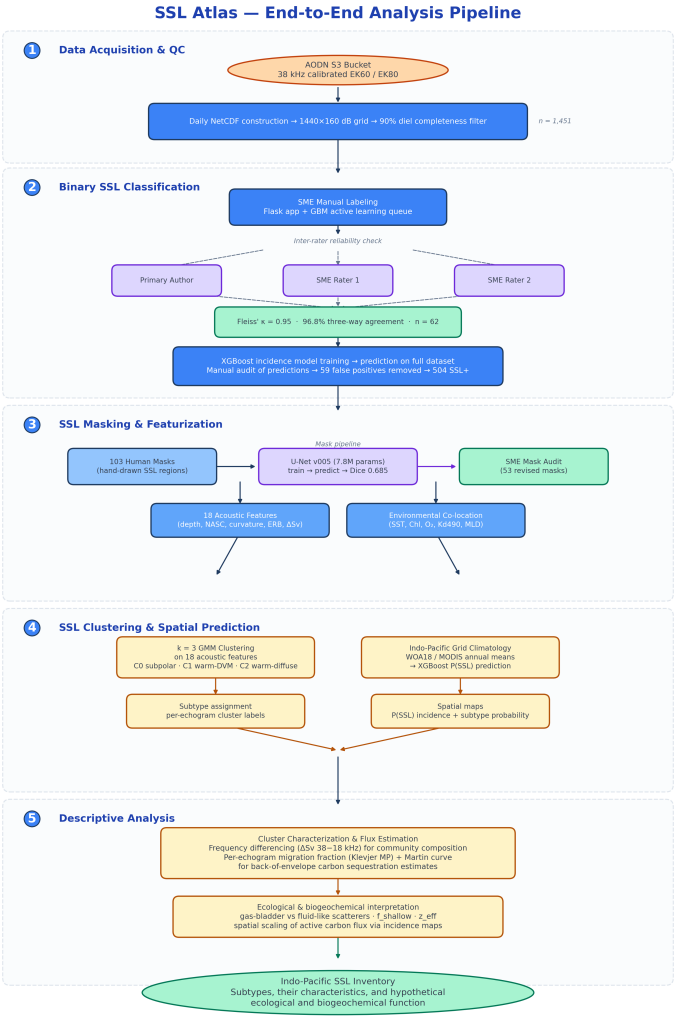

Figure 4.1: End-to-end analysis pipeline. Numbered sections correspond to the five methodological stages described in Sections 4.2.1–4.2.5. Key intermediate outputs include 504 SSL-positive echograms with consolidated masks, 11-feature acoustic characterization, and per-echogram migration fraction estimates.

4.2.2 SSL detection, masking, and validation

All manual labeling, mask drawing, and quality audit tasks described below were performed using a bespoke web-based annotation application built for this study (Fig. B.3). The application consolidates binary labeling (with model-assisted active learning queue prioritization), pixel-wise mask drawing, and batch mosaic audit into a single Flask-based interface, ensuring consistent label persistence and mask versioning across all annotation stages.

SSL detection. SSL presence/absence labels were assigned to all 1,451 echograms through a two-stage process. First, the corresponding author manually labeled a training subset of echograms using the binary labeling mode of the annotation application (Fig. B.4), identifying SSLs as distinct, acoustically coherent layers at 100–400 m depth during daytime, clearly separated from the near-surface and DSL communities (Fig. 4.2). A gradient boosting classifier trained on acoustic features extracted from the labeled subset was then used to predict SSL presence for all remaining echograms, with an active learning queue that iteratively prioritized uncertain and model-disagreeing cases for manual review. All model-predicted labels were subsequently reviewed via the batch mosaic audit mode, in which echograms were displayed in 8×8 grids and false-positive detections were flagged and corrected, yielding a final set of 504 SSL-positive echograms (34.7%).

Incidence prediction. Separately, an XGBoost classifier (Model A) was trained on the final echogram-level SSL labels and co-located environmental covariates (Section 4.2.4) to predict the probability of SSL occurrence from environmental conditions alone. This model was applied to a 1° climatological grid spanning the Indo-Pacific to generate spatially continuous habitat predictions in regions without echogram coverage.

Inter-rater reliability. To assess labeling consistency, a subset of 100 echograms (50 SSL-positive, 50 SSL-absent) was independently labeled by two additional subject-matter experts using a standalone version of the labeling application.

U-Net segmentation. A subset of SSL-positive echograms (103) were manually masked by the corresponding author using the mask drawing mode of the annotation application (Fig. B.5), delineating the pixel-wise SSL boundary. Masks for the full set of SSL-positive echograms were generated using a convolutional U-Net architecture (Ronneberger et al. 2015) trained on these hand-drawn masks (representative human and U-Net mask examples shown in Fig. B.1). All U-Net predictions were subject to a manual quality audit using the batch mosaic audit mode of the

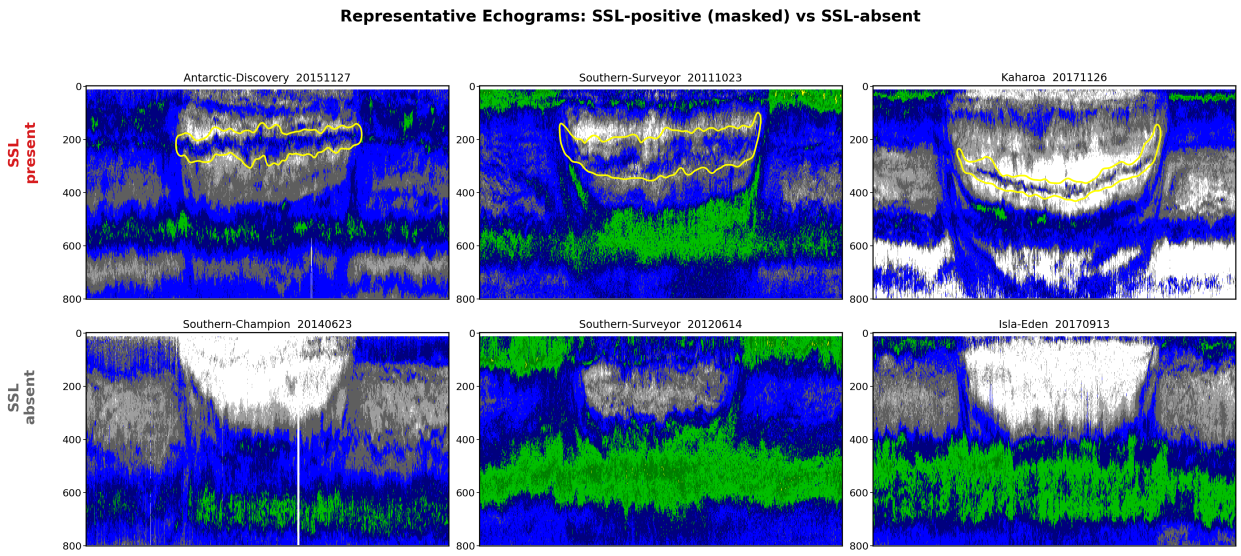

Figure 4.2: Representative echograms illustrating SSL-positive (top row, with yellow mask contours) and SSL-absent (bottom row) classifications. Echograms are displayed centered at peak solar angle over the 0–800 m depth range. SSL-positive examples show distinct intermediate-depth aggregations between the near-surface community and the deep scattering layer.

annotation application, in which predictions were reviewed in 8×8 echogram mosaics and flagged masks were corrected in an integrated revision editor.

4.2.3 Feature extraction and clustering

From each consolidated mask, a set of 11 acoustic and morphometric features was extracted, spanning four functional categories (Table 4.1). Of the 1,451 echograms, a subset of 118 had paired multi-frequency data available (18, 70, and/or 120 kHz alongside 38 kHz); for this subset, multi-frequency differencing ($\Delta S_v = S_v(38\text{ kHz}) - S_v(18\text{ kHz})$, dB) was computed within the SSL mask region, providing a proxy for dominant scatterer type: positive ΔS_v indicates enhanced 38 kHz backscatter from gas-bladder resonance, while negative values indicate fluid-like or non-resonant scatterers (Izard et al. 2025a, 2021; Proud et al. 2019; Wiebe et al. 2023). Per-echogram migration metrics were computed from day/night NASC profiles as a minimal, SSL-aware extension of the Klevjer et al. (2016) migration proportion framework (Klevjer et al. 2016). An SSL-specific migration proportion (MP_ssl) was defined as the fractional decrease in NASC within the SSL mask depth footprint between daytime and nighttime windows, with standard DNR normalization. The SSL fraction of total mesopelagic NASC (f_ssl) was computed as daytime SSL mask NASC divided by total mesopelagic NASC (200–800 m). Neither metric was included in the clustering model.

Features were standardised (zero mean, unit variance) and K-means clustering (k=3; n_init=20;

random_state=42) was applied. k=3 was selected as the optimal number of clusters (silhouette score 0.216, compared to 0.214 at k=2 and 0.161 at k=4). Cluster labels were extended to all echograms via a second XGBoost classifier (Model B) trained on the K-means labels.

Table 4.1: SSL feature set used for subtype clustering. Features are grouped by the layer property they characterise. Derived migration metrics (MP_ssl, f_ssl) are included for descriptive analysis but were not used in the clustering model.

Category	Feature	Definition	In model
Depth & position	mean_depth_m	$\bar{z} = \sum(s_v^{\text{lin}} \cdot z)/\sum s_v^{\text{lin}}$, S_v-weighted mean depth (m)	Yes
	maxsv_depth_iqr_m	IQR of per-column peak-S_v depth across time (m)	Yes
Intensity & biomass	peak_sv_db	$\max(S_v)$ within mask (dB re 1 m^{-1})	Yes
	nasc_mean	$\overline{NASC} = 4\pi \cdot 1852^2 \cdot \overline{\sum_z s_v^{\text{lin}} \cdot \Delta z}$ (m^2 nmi^{-2})	Yes
	peak_prominence_mean_db	$\overline{S_v^{\text{peak}} - S_v^{\text{mean}}}$ within mask per column (dB)	Yes
Vertical structure	erb_m	$ERB = (\sum s_v^{\text{lin}} \cdot \Delta z)^2 / \sum (s_v^{\text{lin}})^2 \cdot \Delta z$, equivalent rectangular bandwidth (m)	Yes
	energy_conc_20m	Fraction of column energy within ±10 m of peak (z_p)	Yes
DVM / curvature	parabolic_curvature	$\lvert a \rvert$ from quadratic fit $z_{\text{peak}}(t) = at^2 + bt + c$ to smoothed peak-depth trajectory	Yes
Horizontal continuity	temporal_coverage	$n_{\text{cols with mask}}/1440$, fraction of time columns containing mask pixels	Yes
	peak_depth_autocorr_lag1	Lag-1 autocorrelation of per-column peak-S_v depth series	Yes
	peak_sv_autocorr_lag1	Lag-1 autocorrelation of per-column peak S_v series	Yes
	peak_streak_max	Longest consecutive run of columns with peak depth within 15 m of predecessor	Yes
Frequency differencing	ΔS_v (38–18 kHz)	$S_v(38\text{ kHz}) - S_v(18\text{ kHz})$ within mask (dB)	No
Migration (derived)	MP_ssl	$1 - (NASC_{\text{ssl,night}} \times DNR)/NASC_{\text{ssl,day}}$; SSL-specific migration proportion (Klevjer et al. 2016)	No
	f_ssl	$NASC_{\text{ssl}}/NASC_{\text{meso,day}}$; SSL fraction of total mesopelagic backscatter	No

4.2.4 Environmental covariates and incidence prediction

The likelihood of SSL incidence across the Indo-Pacific was predicted by training an XGBoost model (Chen and Guestrin 2016) on nine environmental covariates co-located with each echogram in the dataset. Covariates were matched using monthly climatological fields at the echogram acquisition month and nearest 1° grid cell:

1. **Sea surface temperature** (SST, °C) and **mixed layer depth** (MLD, 0.5°C threshold criterion), derived from World Ocean Atlas 2018 (WOA18; (Boyer et al. 2019)).

2. **Surface dissolved oxygen** (o2_surf, μmol kg^{-1}) and **depth to the 60 μmol kg^{-1} oxygen isoline** (omz_top_60_m), from WOA18 monthly O_2 climatology.

3. **Diffuse attenuation coefficient** K_d490 (m^{-1}), derived from MODIS-Aqua monthly chlorophyll-*a* climatology (2018–2022 mean; NASA/GSFC) via the Lee et al. (2005) bio-optical model (Lee et al. 2005).

4. **Euphotic zone depth** (z_eu, m; 1% light level), computed as $z_{eu} = -\ln(0.01)/K_{d490}$.

5. **Bathymetry** (m), sourced from ETOPO1.

6. **Lunar phase** (0–1), computed analytically from echogram date and position.

The incidence model was applied to a 1° resolution climatological grid spanning the Indo-Pacific (latitude −60° to 65°, longitude 20° to 300°E; ~23,000 ocean grid points after masking non-Indo-Pacific basins), using annual WOA18 climatology and annual-mean MODIS-Aqua K_d490 to generate spatially continuous habitat predictions representing long-term mean conditions. Seasonal predictions were additionally generated using WOA18 seasonal climatologies (JFM, AMJ, JAS, OND) to assess temporal variability in SSL habitat extent. Model B (subtype classifier) was applied to assign dominant cluster probabilities at each grid point.

4.2.5 Descriptive analyses

Descriptive analyses are described in the relevant results subsections.

4.3 Results

4.3.1 Detection, validation, and segmentation

Inter-rater reliability. Three-way agreement across the 62 echograms labeled by all three raters was 96.8% (Fleiss' $\kappa = 0.95$, "almost perfect" per Landis & Koch 1977 (Landis and Koch 1977)), with only two disagreements on borderline cases. Cohen's kappa for each pairwise comparison ranged from 0.926 to 0.964.

U-Net segmentation. The production U-Net (v005, 7.77M parameters, trained on 88 / validated on 15 masks) achieved a validation Dice coefficient of 0.685. Manual quality audit of all predicted masks identified 53 requiring revision and 59 false-positive SSL labels, which were corrected before feature extraction.

Clustering. K-means clustering (k=3) on the 11-feature standardized matrix yielded a silhouette score of 0.216. Cluster sizes: C0 subpolar (n=53), C1 mesotrophic/warm-DVM (n=179), C2 oligotrophic/warm-diffuse (n=270).

4.3.2 Observed SSL incidence and geographic distribution

Of 1,451 echograms, 504 (34.7%) contained a detectable SSL. SSL occurrence was highest in the tropical and subtropical Indian Ocean and Western Pacific (Fig. 4.4). SSL-positive echograms occurred under significantly warmer SST (21.8°C vs. 15.6°C; $p \approx 0$), lower surface dissolved oxygen (225 vs. 253 μmol kg^{-1}; $p \approx 0$), and deeper euphotic zones (160 vs. 124 m) relative to SSL-absent echograms, consistent with warmer, more oligotrophic conditions typically associated with stronger thermal stratification.

4.3.3 Environmental predictors and incidence model

The XGBoost incidence model (9 environmental features; see Section 4.2.4) achieved a cross-validated AUC of 0.866 ± 0.038 (5-fold stratified). Top predictive features were surface dissolved oxygen (gain 0.193), OMZ shoaling depth (0.190), log-transformed chlorophyll-*a* (0.175), and SST (0.088). Feature importance rankings and partial dependence response curves for all nine predictors are shown in Fig. 4.3.

4.3.4 Basin-scale predicted SSL habitat

The incidence model applied to the Indo-Pacific climatological grid predicts SSL habitat across 43.6% (33.1%) of ocean grid points at P(SSL) > 0.50 (0.75) thresholds (Fig. 4.4). Highest predicted probabilities occur in the subtropical gyres (15°S–30°N), the tropical eastern Indian Ocean, and the western Pacific warm pool. Lowest probabilities occur poleward of 40°S and in the equatorial eastern Pacific upwelling zone where high K_d490 and shallow OMZ restrict intermediate-depth layer formation. The three SSL subtypes show distinct geographic distributions (Fig. 4.4): C2 (subpolar/red) concentrates poleward of ~35°S; C1 (mesotrophic/blue) occupies subtropical gyre margins and mesotrophic transition zones; and C0 (oligotrophic/green) spans broader tropical and interior gyre regions across the Indian Ocean and western Pacific.

South of 40°S, SSL-positive echograms (11.4% incidence, n = 58) appear over a region of low predicted probability. This partly reflects seasonal aliasing: the climatological grid uses annual-mean covariates that average in austral winter conditions when SSLs are absent—zero SSLs were detected

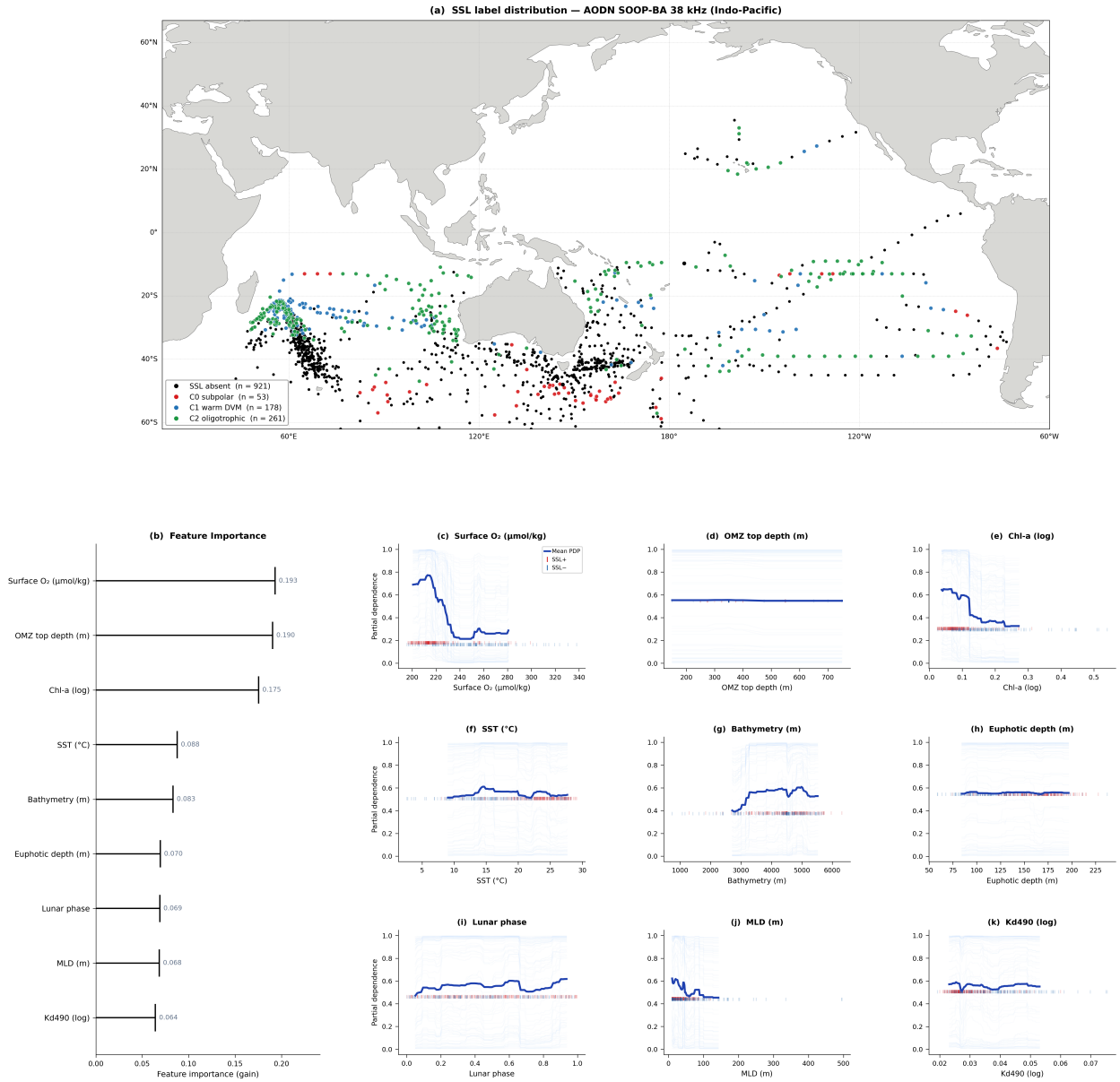

Figure 4.3: SSL incidence model overview. (a) Geographic distribution of SSL labels across the Indo-Pacific (black = SSL absent; colored by cluster: red = C0 subpolar, blue = C1 mesotrophic, green = C2 oligotrophic). (b) XGBoost feature importance (gain) for the 9-feature environmental model. (c–k) Partial dependence plots for each predictor, showing the marginal effect on predicted P(SSL). Dark blue line: mean partial dependence; light traces: individual conditional expectations (ICE). Rug marks indicate the distribution of training observations (red = SSL+, blue = SSL−).

in June–August despite 155 observations—while ships encounter summer SST averaging 10.8°C, 3.9°C warmer than the annual mean. However, even during the sampled warm months, SSL incidence in this region remains low (15.9% in austral summer vs. 47.4% north of 40°S), indicating that subpolar waters are genuinely suboptimal for SSL formation regardless of season. Seasonal predictions confirm that the P(SSL) > 0.75 boundary shifts modestly poleward in summer (Fig. B.6).

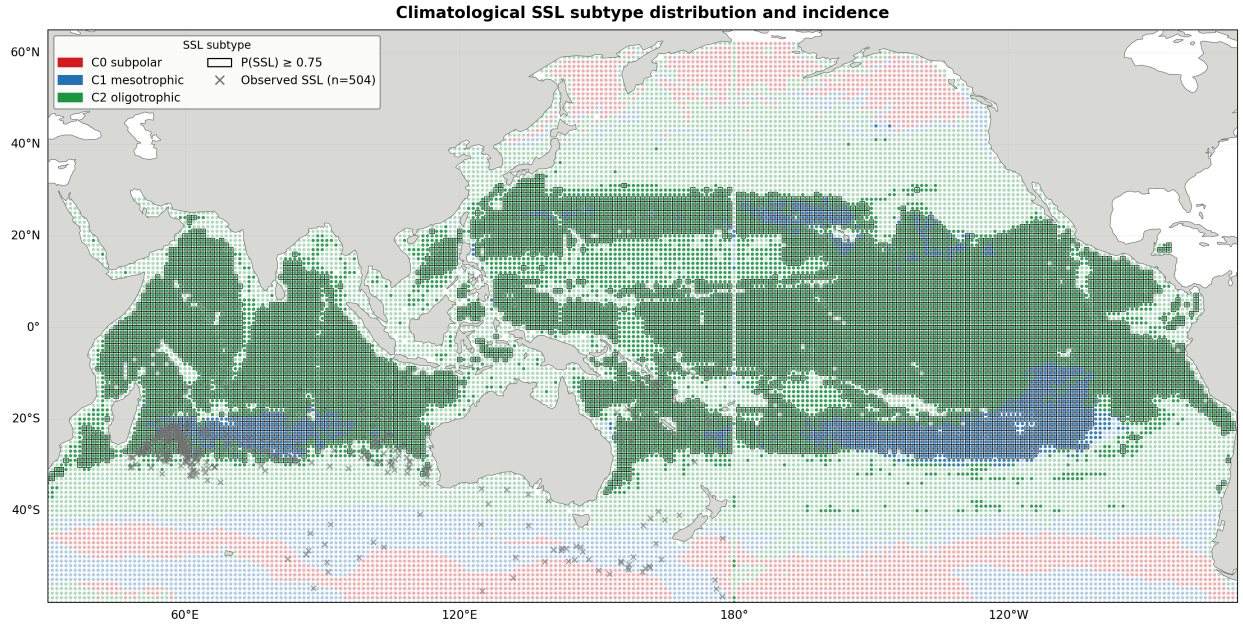

Figure 4.4: Indo-Pacific SSL habitat prediction and observed subtype distribution. Background shading shows the dominant predicted SSL subtype (green = C0 oligotrophic/warm diffuse; blue = C1 mesotrophic/warm DVM; red = C2 subpolar/cold) at each 1° climatological grid point from the V2 XGBoost incidence model (9 environmental features, AUC = 0.87). Grid cells where P(SSL) ≥ 0.75 are outlined in black. Faded shading indicates P(SSL) < 0.50. Grey crosses indicate observed SSL-positive echogram locations (n = 504).

4.3.5 SSL depth, light relationship, and vertical niche structure

SSL mean centre depth averaged 245 ± 38 m across all SSL-positive echograms. SSL depth scaled with light penetration depth (z_{depth} = 171.0 + 2.37 × ($1/K_{d490}$); r = 0.41, p < 10^{-24}), but with a slope six-fold shallower than isolume values reported for the DSL (Aksnes et al. 2017); Fig. 4.6). This indicates that SSLs track a brighter isolume and are less tightly depth-coupled to the ambient light field than the DSL. Our estimate of the SSL-preferred isolume—derived from the observed SSL depth distribution and co-located K_d490—corresponds to approximately 0.07% of surface irradiance, intermediate between the near-surface community and the DSL-preferred isolume (~0.001%; (Aksnes et al. 2017)).

SSL-positive echograms occupied a significantly wider mesopelagic niche gap—the vertical distance from the base of the euphotic zone (z_eu) to the DSL upper edge—than SSL-absent echograms. Figure 4.5 illustrates this niche context. All three SSL subtypes (panels A–C) occupy the gap between z_eu and the DSL upper edge, while two non-SSL control conditions—warm O_2-limited (panel D) and cold (panel E)—show the same vertical window unoccupied. SSL depth varies among subtypes—C0 is shallowest (209 m), C1 deepest (276 m), and C2 intermediate (232 m)—but all sit well above the DSL and below z_eu, consistent with the hypothesis that SSLs form when sufficient unoccupied vertical habitat exists between the productive epipelagic and the DSL. The overlaid light attenuation and oxygen profiles in Fig. 4.5 illustrate how the niche can be limited from above by downwelling light, from below by DSL occupancy, and laterally by oxygen depletion where the OMZ intrudes into the 200–400 m range (panel D). Notably, some regions exhibit a viable niche gap structurally but no observed SSL (panel E), suggesting additional factors such as community composition or temperature may also influence SSL formation.

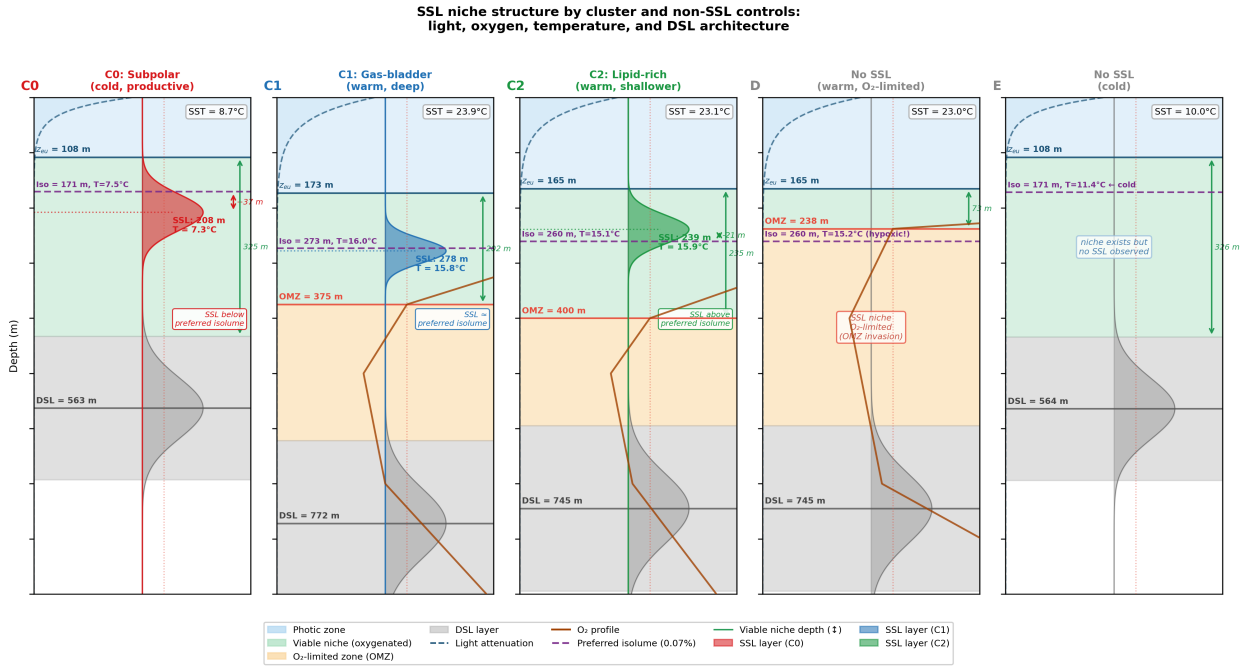

Figure 4.5: SSL niche structure by cluster (C0–C2) and two non-SSL control conditions. Each panel shows a depth profile of light attenuation (dashed blue), dissolved oxygen (brown, right axis), and Gaussian representations of the DSL (grey) and SSL (colored) layers, with key depth horizons annotated: euphotic zone base (z_eu), OMZ top (where present), DSL centre, and preferred SSL isolume (0.07% surface irradiance, purple dashed). Background shading denotes the photic zone (pale blue), viable oxygenated niche (green), O_2-limited zone (orange), and DSL layer (grey). Cluster colors: red = C0 subpolar; blue = C1 mesotrophic; green = C2 oligotrophic. Panels D–E show warm O_2-limited and cold non-SSL conditions respectively, illustrating niche closure mechanisms.

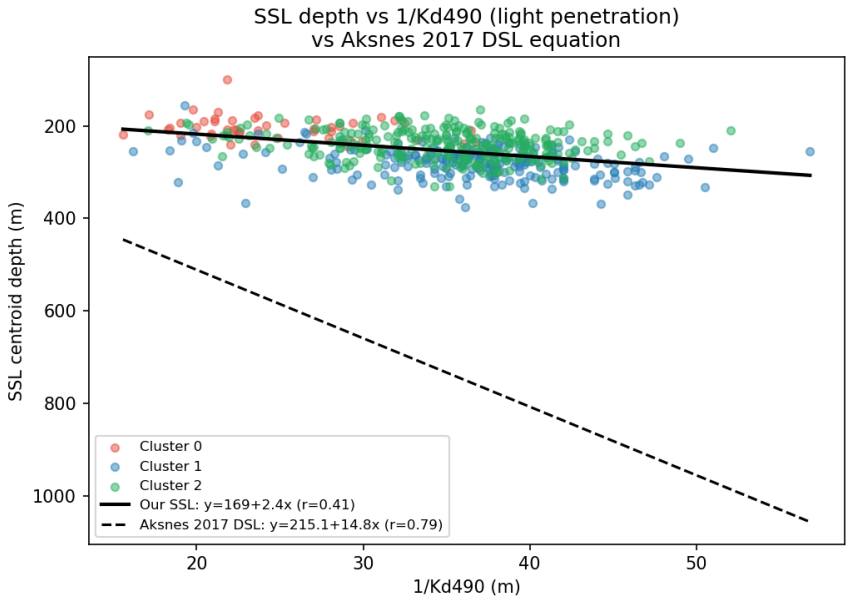

Figure 4.6: SSL mean depth vs. light penetration depth ($1/K_{d490}$). Black line: linear regression (WMD = 171.0 + 2.37/K_{d490}). Grey shading: 95% confidence interval. Inset: comparison with DSL depth relationship (Aksnes et al. 2017); slope = 14.8).

4.3.6 Acoustic subtype variability and spatial patterns

The three subtypes differ systematically in depth, acoustic intensity, scattering character, migration behavior, and environmental setting (Table 4.2; Fig. 4.7):

- **C0 (subpolar/cold; n=53)**: Mean SST 11.9°C; mean SSL depth 209 m (shallowest); highest NASC (mean 196 m^2 nmi^{-2}); weakly negative ΔS_v(38–18) (mean −1.2 dB, n=20), with uniformly positive ΔS_v(38–120) (mean +3.2 dB). Predominantly Southern Ocean and subpolar Indian Ocean. The weakly negative ΔS_v(38–18) is consistent with the subpolar resonance shift documented by Izard et al. (2025) (Izard et al. 2025a), where larger-bodied mesopelagic fish shift swim bladder resonance from 38 to 18 kHz south of the Subantarctic Front. Euphausiids, siphonophores, and wax-ester myctophids likely contribute additional non-resonant scattering (Davison et al. 2013; Dornan et al. 2019, 2022; Proud et al. 2019). This subtype shows the highest SSL migration (MP_ssl = 34.4%) and largest SSL fraction of mesopelagic backscatter (f_{ssl} = 11.6%).

- **C1 (mesotrophic/warm, strong DVM; n=179)**: Mean SST 23.4°C; mean SSL depth 276 m (deepest among subtypes); limited multi-frequency coverage (n=5 for ΔS_v(38–18), n=17 for ΔS_v(38–120)). Subtropical gyre margins and mesotrophic transition zones of the Indian Ocean and western Pacific. The strong parabolic DVM signature and deepest SSL depth suggest

vertically migrating gas-bladder fish, potentially including *Maurolicus* spp. and subtropical myctophids (Klevjer et al. 2012; Norheim et al. 2016), though robust acoustic characterization awaits additional multi-frequency coverage. Intermediate f_{ssl} (2.9%) and MP_ssl (22.6%).

- **C2 (oligotrophic/warm, diffuse; n=270)**: Mean SST 22.6°C; mean SSL depth 232 m; moderate NASC (mean 58 m^2 nmi^{-2}); significantly positive ΔS_v(38–18) (mean +3.2 dB, n=34, $p < 0.001$), indicating resonance-amplified 38 kHz backscatter from small gas-bladder fish. Tropical and interior gyre regions. The positive ΔS_v(38–18) is consistent with subtropical/tropical communities where small-bodied hatchetfish, *Cyclothone* spp., and juvenile myctophids (<30 mm) produce 38 kHz resonance at SSL depths (Izard et al. 2025a; Proud et al. 2019). Lowest SSL migration (MP_ssl = 12.2%, median 0.0%), consistent with largely non-migratory resident assemblages.

Table 4.2: SSL subtype summary statistics.

Parameter	**C0 subpolar**	**C1 mesotrophic**	**C2 oligotrophic**
n (echograms)	53	179	270
SST (°C)	11.9	23.4	22.6
SSL depth (m)	209	276	232
NASC (m^2 nmi^{-2})	196	35	58
f_ssl (%)	11.6	2.9	3.7
MP_ssl (%)	34.4	22.6	12.2
ΔS_v(38–18) (dB)	−1.2 (n=20)	−3.3 (n=5)	+3.2 (n=34)

Multi-frequency differencing within SSL masks revealed a cluster-level community signal in the ΔS_v(38–18 kHz) metric (n=59 echograms with both frequencies; Fig. 4.8). C2 (oligotrophic) echograms showed significantly positive ΔS_v(38–18) (mean +3.2 dB, n=34, $p < 0.001$), indicating enhanced 38 kHz backscatter consistent with resonant scattering from small gas-bladder fish whose swim bladder dimensions produce 38 kHz resonance at SSL depths (200–350 m) in warm subtropical waters (Izard et al. 2025a; Proud et al. 2019). C0 (subpolar) echograms showed weakly negative ΔS_v(38–18) (mean −1.2 dB, n=20), consistent with the latitudinal shift in resonance frequency documented by Izard et al. (2025) (Izard et al. 2025a) across the Subantarctic Front, where larger-bodied mesopelagic fish shift resonance toward 18 kHz. C1 had limited multi-frequency coverage (n=5 for ΔS_v(38–18)) and cannot be robustly characterized by this metric. Notably, ΔS_v(38–120 kHz) and ΔS_v(38–70 kHz) were uniformly positive across all three clusters (means +2.5 to +3.2 dB, n=90 and n=57 respectively), indicating that gas-bearing scatterers contribute to SSL backscatter in all subtypes; the cluster-level variation in ΔS_v(38–18) likely reflects regional differences in fish body size and swim bladder volume rather than a simple gas-bearing vs. fluid-like

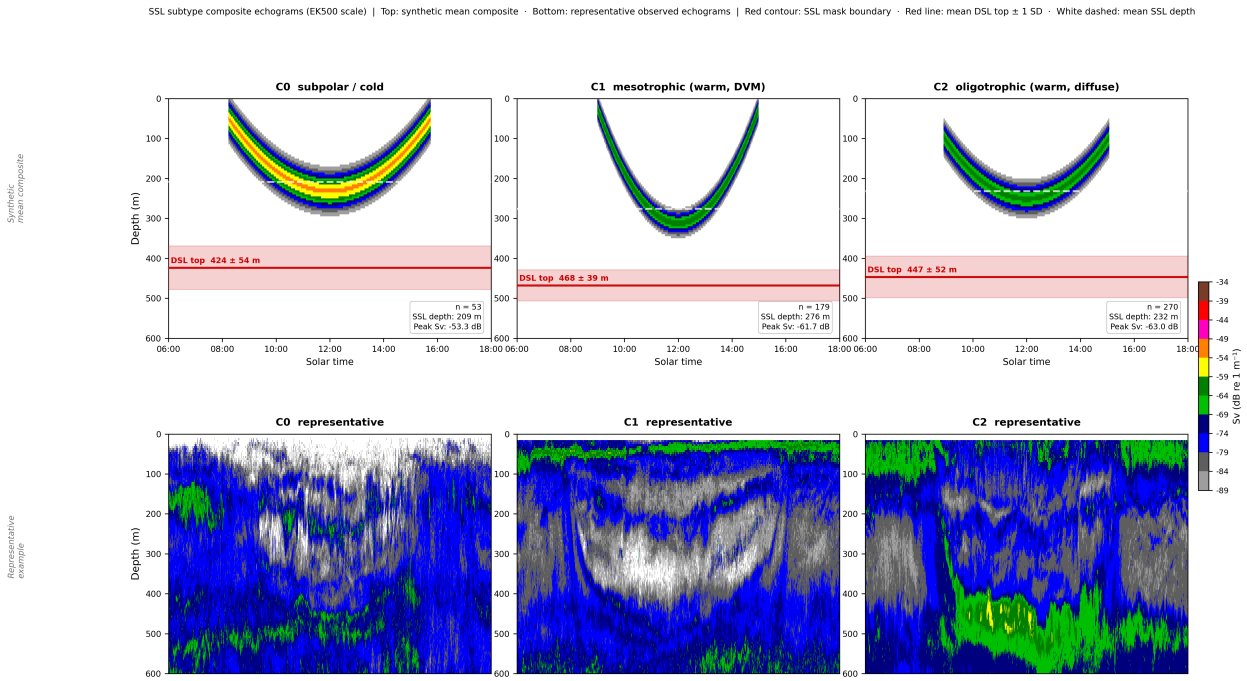

Figure 4.7: SSL subtype echogram composites. Top row: synthetic mean composite echograms for each SSL subtype displayed using the EK500 color scale (dark grey = low S_v; blue → green → yellow → orange → red = increasing S_v). Each panel reconstructs the mean layer structure from cluster-averaged SSL depth, vertical extent, and peak S_v, shown as a function of solar time. White dashed line: mean SSL depth. Red line with shading: mean observed DSL top ± 1 SD. Bottom row: representative observed echograms for each subtype with SSL mask contour overlay (white line). Panels from left to right: C0 subpolar/cold, C1 mesotrophic (warm, DVM), C2 oligotrophic (warm, diffuse).

dichotomy (Izard et al. 2025a). The full three-pair frequency differencing comparison (38–18, 38–70, 38–120 kHz) is shown in Fig. B.2.

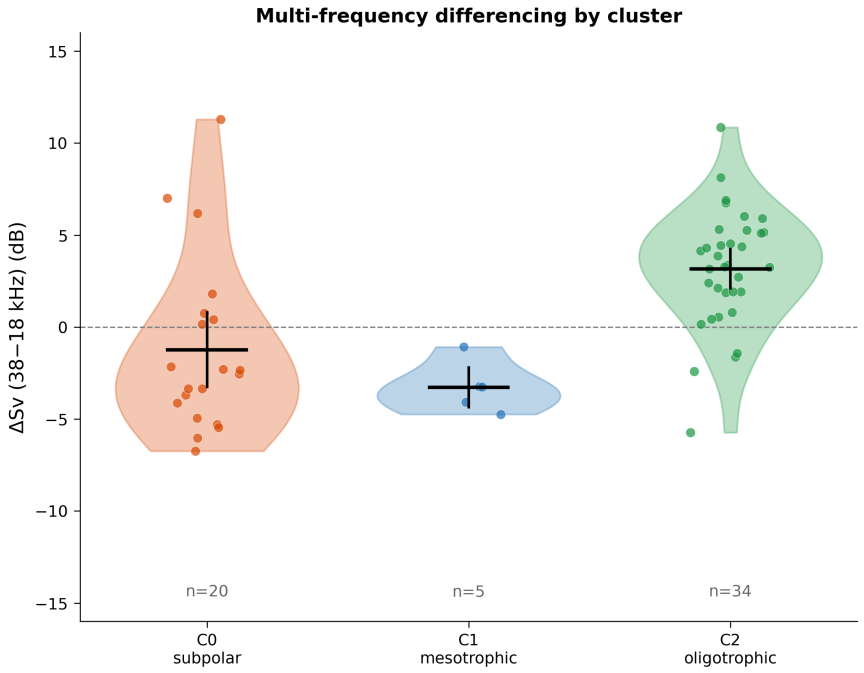

Figure 4.8: Multi-frequency differencing (ΔS_v(38–18 kHz)) within SSL masks by cluster. Violin plots show per-echogram distributions; horizontal bars indicate cluster means with 95% CI. Positive values indicate resonance-amplified 38 kHz backscatter from small gas-bladder fish; negative values indicate larger-bodied scatterers or fluid-like organisms. Note: C1 has limited coverage (n=5) for this frequency pair.

4.3.7 SSL migration and carbon flux fractionalization

The per-echogram migration metrics derived above (MP_ssl, f_ssl) permit a back-of-envelope assessment of how SSL vertical substructure may fractionate mesopelagic active carbon flux. Following the migration proportion framework of Klevjer et al. (2016) (Klevjer et al. 2016) and applying the Martin et al. (1987) (Martin et al. 1987) power-law flux attenuation at dataset-mean SSL (283 m) and DSL (731 m) injection depths, we estimate the reduction in sequestration efficiency that arises when a fraction of migrant biomass injects respired carbon at SSL depth rather than DSL depth. This analysis inherits the simplifying assumptions of the standard MP framework—including the use of NASC as a linear biomass proxy, horizontal variability in community composition and resonance properties, and the canonical Martin curve exponent—but demonstrates that even under these same assumptions, the simple addition of vertical layer structure to an otherwise single-layer MP calculation shifts the result. The resulting sequestration reduction is governed by the interaction

of the depth separation between the layers, the SSL fraction of mesopelagic NASC (f_ssl), and the SSL-specific migration proportion (MP_ssl). Across the observed parameter space (Fig. 4.9), this reduction ranges from negligible (<1%) in warm-water environments to potentially meaningful fractionalization (~5–15%) in subpolar settings.

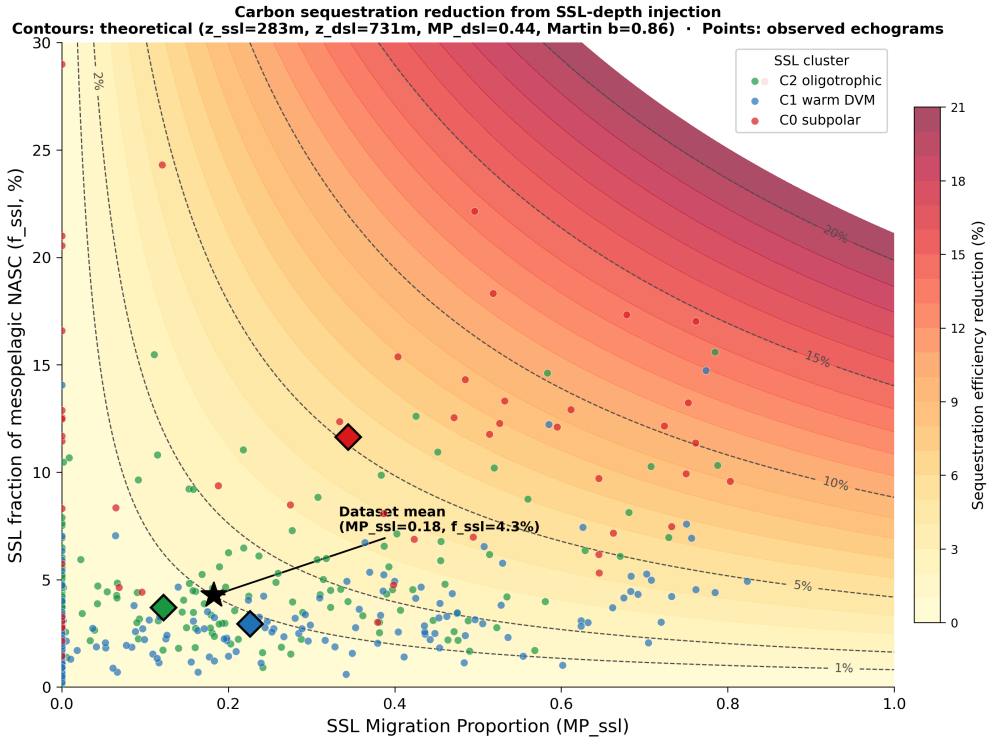

Figure 4.9: Carbon sequestration reduction as a function of SSL migration proportion (MP_ssl) and SSL fraction of mesopelagic NASC (f_ssl). Dashed contour lines show theoretical reduction at dataset-mean injection depths (SSL = 283 m, DSL = 731 m). Points show individual SSL-positive echograms; diamonds indicate cluster means; star indicates dataset mean.

4.4 Discussion

4.4.1 Environmental controls on SSL formation

The dominance of surface O_2, OMZ depth, and chlorophyll-*a* in the incidence model, combined with the significant association between SSL presence and deeper euphotic zones and wider niche gaps, points to a coherent biophysical control on SSL formation rooted in light field structure, water-mass properties, and vertical habitat availability in the upper mesopelagic.

Light penetration establishes the lower boundary of visual predation risk for mesopelagic organisms

during daytime. In clear, oligotrophic water (low K_d490), the visual refuge depth—below which epipelagic visual predators cannot effectively hunt—extends to 250–350 m, creating a wide vertical window between the surface scattering community and the DSL in which intermediate-depth aggregations can persist (Kaartvedt et al. 2012a; Langbehn and Aksnes 2018). The relationship between SSL depth and K_d490 we document (slope = 2.37 m per unit $1/K_{d490}$) is six-fold shallower than the DSL–light relationship (Aksnes et al. 2017), consistent with SSLs tracking a distinct, brighter isolume (~0.07% surface irradiance) that marks this visual refuge boundary.

Surface dissolved oxygen emerged as the top predictive feature (gain = 0.193), with SSL incidence highest at the lowest surface O_2 values. Rather than reflecting a direct physiological O_2 requirement, this association likely arises because surface O_2 acts as a water-mass proxy: warm, stratified conditions that depress surface O_2 through reduced solubility and ventilation co-occur with the deep euphotic zones, stable vertical structure, and low productivity that collectively favor SSL formation. OMZ depth operates through a distinct, complementary mechanism: when the oxygen minimum zone shoals into the 200–400 m depth range (e.g., eastern tropical Pacific), it physically compresses the habitable window for intermediate-depth organisms. Our niche framework (Fig. 4.5, panel D) illustrates this closure mechanism directly: when the OMZ top shoals above ~300 m, it compresses the viable habitat between z_eu and the OMZ to a narrow band that may be insufficient to support persistent SSL aggregations. We note that OMZ depth ranks second in feature importance despite exhibiting a relatively flat partial dependence curve (Fig. 4.3); this apparent contradiction reflects OMZ depth's role as a binary gating variable—it matters greatly whether the OMZ intrudes into the SSL depth range or not, but once the niche is either open or closed, further variation in OMZ depth contributes little additional predictive power.

Chlorophyll-*a* ranked third among predictive features (gain = 0.175), with the partial dependence relationship showing SSL incidence highest at the lowest Chl-a values and declining with increasing productivity. This pattern is consistent with the geographic concentration of SSLs in oligotrophic subtropical gyres, where clear water permits deep light penetration and extends the vertical window between the euphotic zone and DSL in which intermediate-depth aggregations can persist. In productive coastal and upwelling environments, elevated near-surface scattering, high K_d490, and shallow euphotic zones may both obscure and genuinely suppress intermediate-depth acoustic layer formation. The association between low Chl-a and high SSL incidence aligns with past studies documenting prominent shallow scattering layers in subtropical gyre interiors (Culhane et al. 2025; Klevjer et al. 2016; Proud et al. 2017) and reinforces the interpretation that the SSL niche is defined primarily by deep light penetration and vertical habitat space rather than by local productivity. Notably, some regions exhibit a viable niche gap but no observed SSL, suggesting that community composition and temperature may also play a role—regional shifts in assemblage structure and

depth-dependent resonance properties (Izard et al. 2025a, 2021) could modulate whether a given water column supports detectable SSL formation.

4.4.2 SSL subtype ecology and taxonomic interpretation

The three SSL subtypes differ not only in their acoustic character but in their inferred ecological composition, geographic distribution, and biogeochemical implications. These interpretations are necessarily tentative, as multi-frequency differencing provides a coarse compositional proxy and direct trawl or MOCNESS validation at SSL depths is absent from our dataset.

The C0 (cold/subpolar) subtype is characterized by weakly negative ΔS_v(38–18) (mean −1.2 dB), the highest NASC, and occurrence at high latitudes. The modestly negative ΔS_v(38–18), combined with uniformly positive ΔS_v(38–120), indicates a mixed scattering community that includes both gas-bearing organisms and fluid-like scatterers. The weakly negative ΔS_v(38–18) is consistent with the subpolar resonance shift documented by Izard et al. (2025) (Izard et al. 2025a): south of the Subantarctic Front, larger-bodied mesopelagic fish (>35 mm) shift swim bladder resonance from 38 to ~18 kHz, while euphausiids, siphonophores, and wax-ester myctophids contribute non-resonant scattering (Davison et al. 2013; Dornan et al. 2019; Proud et al. 2019). Southern Ocean and subpolar Indian Ocean environments support dense euphausiid aggregations that are well-documented at intermediate mesopelagic depths during daytime (Dornan et al. 2019, 2022).

The C1 (mesotrophic/warm, strong DVM) subtype is distinguished by the deepest mean SSL depth (276 m), strong parabolic DVM curvature, and occurrence at subtropical gyre margins. Multi-frequency coverage for C1 is limited (n=5 for ΔS_v(38–18), n=17 for ΔS_v(38–120)), precluding robust acoustic characterization of its scattering community. The strong DVM signature and environmental setting are consistent with vertically migrating gas-bladder fish including *Maurolicus* spp. and subtropical myctophids (Kaartvedt et al. 2019; Klevjer et al. 2012; Norheim et al. 2016). However, depth-dependent resonance physics (Izard et al. 2025a) predict that at C1's mean SSL depth of 276 m, only fish smaller than ~25–30 mm would produce 38 kHz resonance; larger adults would shift resonance to lower frequencies. This depth dependence complicates acoustic taxonomic inference at C1 depths and highlights the need for direct net sampling.

The C2 (oligotrophic/warm, diffuse) subtype shows the most robust multi-frequency signal: significantly positive ΔS_v(38–18) (mean +3.2 dB, n=34, $p < 0.001$). This positive ΔS_v(38–18) indicates resonance-amplified 38 kHz backscatter from small gas-bladder fish, consistent with subtropical/tropical mesopelagic communities where small-bodied hatchetfish, *Cyclothone* spp., and juvenile myctophids (<30 mm) dominate (Davison et al. 2013; Izard et al. 2025a; Receveur et al. 2020a). The resonance amplification means the true biomass of C2 SSLs may be lower than their

acoustic prominence at 38 kHz implies (Izard et al. 2025a; Proud et al. 2019). Despite this caveat, the broad geographic distribution and large sample size suggest C2 captures the dominant tropical SSL assemblage.

4.4.3 Implications for mesopelagic carbon flux

SSLs represent one element of a broader set of vertically stratified scattering structures—including multiple deep scattering layers, distinct migration trajectories, and depth-specific residence patterns—that reflect regionally specific interactions between biological composition and the physical-chemical environment (Ariza et al. 2022; Culhane et al. 2025; Proud et al. 2019). These structures collectively mediate active carbon transport from the euphotic zone into sequestration depths. While the direct impact of SSL-depth migration on sequestration efficiency is modest in the exploratory analysis presented here (Fig. 4.9), the cumulative effect of multiple such structural departures from a single-layer model—each routing migrant biomass to different injection depths—may be substantial. Bulk migration proportion estimates, which treat the mesopelagic as a single migrating compartment, do not capture this complexity. Incorporating vertically resolved layer structure into regional flux budgets may improve the fidelity of active transport estimates.

4.4.4 Limitations and future directions

This analysis has several limitations. The SOOP-BA dataset has spatially uneven coverage—dense in the Indian Ocean, sparser in the western and eastern Pacific—which may bias the incidence model toward Indian Ocean environmental conditions and underrepresent Pacific SSL variability. The subtype interpretations are based on coarse multi-frequency differencing rather than net sampling, and direct biological validation of the assemblage hypotheses remains a priority for future work. The carbon flux estimates carry uncertainties from mask boundary definitions, DNR assumptions, and the Martin-curve parameterization. The association between SSL incidence and specific water mass properties is consistent with earlier hypotheses identifying water-mass-specific scattering layer signatures (Klevjer et al. 2016; Proud et al. 2017), but the causal mechanisms linking water mass formation to SSL composition remain to be elucidated. Future analyses evaluating the contribution of seasonality to SSL incidence may improve the resolution of grid predictions and reveal within-region temporal variability not captured in the current annual climatological framework.

4.5 Conclusions

This study presents a basin-scale assessment of SSL incidence, environmental controls, and acoustic substructure across the Indo-Pacific. The principal findings are:

- SSL habitat is predicted across 40.5% of the Indo-Pacific, concentrated in warm, oligotrophic subtropical and tropical waters, with the environmental niche shaped primarily by surface oxygen (as a water-mass proxy), OMZ depth, and light penetration.
- Three SSL subtypes—subpolar/cold (C0), mesotrophic/warm with strong DVM (C1), and oligotrophic/warm and diffuse (C2)—are separable using acoustic morphometric clustering and show distinct geographic distributions, migration behaviors, and ecological inferences.
- An exploratory carbon flux analysis demonstrates that resolving SSL-depth migration separately from the DSL shifts standard migration proportion estimates, with sequestration reduction ranging from negligible (<1%) in warm-water environments to potentially meaningful fractionalization (~5–15%) in subpolar settings where SSL biomass and migration rates are highest.
- The niche framework established here, linking SSL presence to light penetration, surface oxygen, and OMZ depth, provides a quantitative foundation for incorporating SSL dynamics into regional active flux budgets and predator habitat models.