Chapter 6

Conclusion

Summary

This thesis developed methods and tools for large-scale acoustic observation of biological structure in the ocean interior. The work spans uncertainty quantification for opportunistic acoustic data (Chapter 2), machine learning methods for extracting ecological information from archive-scale imagery (Chapter 3), basin-scale ecological characterization of an incompletely characterized scattering feature (Chapter 4), and the design of open-source instrumentation for broadband acoustic sensing (Chapter 5).

A recurring finding across chapters is that structural properties of acoustic scattering—layer depth, migration timing, echogram morphology—are robust to calibration uncertainty and provide ecologically valuable information without requiring precise knowledge of instrument parameters or scattering physics. In Chapter 2, normalized structural metrics (WMD MAE 6–22 m; migration timing MAE ~1 h) were invariant to parametric errors that produced NASC biases ranging from +430% to –78% per depth layer. In Chapter 3, calibration-independent structural analysis recovered basin-scale oceanographic provinces from uncalibrated ADCP data. In Chapter 4, morphometric features that proxy community behavior—including layer depth, vertical extent, and diel migration amplitude—played a central role in partitioning SSL subtypes across the Indo-Pacific. This pattern has practical consequences: it means that significant ecological information is accessible from the large, imperfectly calibrated acoustic archives that already exist across global research fleets.

Several chapters connect mesopelagic biological structure to surface-ocean properties. Surface temperature and salinity affect acoustic estimates of deep biological structure through the sonar equation (Chapter 2). Satellite-derived environmental variables predict water column acoustic structure across the Pacific (Chapter 3). Surface dissolved oxygen, light penetration, and oxygen minimum zone depth control the incidence and depth of intermediate scattering layers across the Indo-Pacific (Chapter 4). These connections are consistent with the premise motivating this work: that mesopelagic communities are structured by upper-ocean physics and biogeochemistry, and that this structuring can be leveraged to extend satellite surveillance into the ocean interior.

The methods developed here enabled analysis at scales not previously feasible. The systematic documentation of parameter error in ADCP-derived S_v provides both quantitative guidance for operators and data users and a practical framework for incorporating ADCP backscatter into integrated data workflows—a step toward unlocking the global ADCP archive for quantitative biological analysis. Convolutional autoencoders and U-Net architectures extracted ecological structure from thousands of echograms across two ocean basins, a volume of data that would not be tractable with manual or traditional numerical approaches. EchoBot demonstrated proof of concept for community-developed broadband acoustic sensing, achieving spectral target strength correlation

of $r = 0.977$ against a commercial Simrad EK80 in tank validation. By releasing all designs under open licenses, EchoBot enables community-driven development and adaptation for platforms and deployment scales that commercial instrumentation does not currently serve.

Limitations and Future Directions

Each chapter addresses a specific gap in the observational pipeline, but the methods developed here are extensible, and several productive directions for subsequent work follow directly from the present results.

The regression analysis in Chapter 2 focused on the OS-75 instrument class, primarily because of the difficulty of assembling a sufficiently large archive of paired ADCP–echosounder cruises for other configurations. Expanding the uncertainty analysis to additional instrument classes would produce more operationally significant guidance for leveraging these data. Further work could also evaluate whether regression outputs from paired cruises can be generalized to unpaired cruises through transfer functions or instrument-class-specific calibration constants, which would substantially increase the volume of quantitatively usable ADCP backscatter data.

A key finding of Chapter 2 is that structural metrics—weighted mean depth and migration timing—are preserved across instrument classes and across frequency pairings. Given the premise in Chapter 3 that correlation exists between water column structure and biological content, a premise implicit in calibrated echosounder studies that use DSL depth as an explanatory variable for water column NASC, improved global-scale coverage could be achieved through structural analysis workflows that integrate ADCP and calibrated echosounder data jointly. Such integration may permit time-resolved biogeographic solutions that capture seasonal and interannual variability in community structure. The potential value of this approach is exemplified by the correspondence between the SSL distribution mapped from calibrated echograms in Chapter 4 and the intermediate-depth features independently recovered from uncalibrated ADCP data in Chapter 3. The coincidence implies that SSL characterization could be extended across the broader ADCP archive, which would be particularly valuable for SSLs given evidence that intermediate-layer incidence exhibits pronounced phenological variability tied to seasonal water column structure.

Chapter 5 presents EchoBot as a proof of concept for open-source broadband acoustic sensing, characterized against a single calibration target in a controlled environment. Establishing its utility for field science requires multi-target calibration to assess the stability of the empirical gain offset, and deployment under realistic conditions—platform motion, ambient noise, biofouling, and long-range propagation. Reducing hardware cost and increasing transmit power for deep-water

operation are additional near-term priorities.

Finally, none of the work presented here directly addressed the potential impacts of climate change on interior ocean biological structure. As surface ocean regimes shift under warming—including the projected expansion of subtropical gyre regions, shoaling of oxygen minimum zones, and changes in mixed layer depth and primary productivity—these changes will propagate into the mesopelagic zone and reshape the communities that inhabit it. The depth, coherence, and composition of scattering layers respond to surface forcing, and shifts in layer depth have direct consequences for their accessibility as prey for surface-dwelling predators and their role in active carbon export.

Closing Remarks

Collectively, these methods and tools move the field closer to treating mesopelagic biological structure as a routinely observable quantity—a prerequisite for integrating these communities into Earth system models and for informed management of the deep ocean. The gap between what can be observed in the ocean interior and what must be known to constrain its ecological and biogeochemical function remains substantial, but the work presented here narrows it by expanding the usable data archive, introducing scalable analytical approaches, and providing open instrumentation for future deployment. Anthropogenic climate change and the emergence of commercial and industrial activity in the ocean interior provide a clear imperative for continued research aimed at understanding and monitoring this critical component of the Earth system.

Appendix A

Supplementary Material for Chapter 3

A.1 Supplementary acoustics and image data processing

This supplement describes the pre-processing steps applied to JASADCP cruise archives of raw received signal strength (RSSI) to generate training images for the CAE: 1) Cruises were segmented by the solar angle into day-length, time-depth standardized echograms (Sec. A.1.1); 2) echogram RSSI was converted into volumetric backscattering (S_v), and then normalized (Sec. A.1.2).

A.1.1 From cruises to echograms

Extracting echograms with the solar angle

Diel echograms were extracted from cruise-length time series by segmenting the 24-hour region around all peaks of the solar angle during the cruise duration (solar angle maxima in Fig. A.1), as has been done in previous approaches (e.g. Ariza et al., 2022). This process was implemented using the pysolar python library, which applies the solar position algorithm (Rhodes, https://github.com/pingswept/pysolar; e.g. Reda & Andreas, 2003).

Time-depth standardization related exclusions

Images were fixed within a standard frame of reference so that the feature maps from the CAE correspond to specific physical locations (depth and time of day; Section 3.2.1). A standard depth-time frame was built with 1-meter by 5-minute resolution, from 0–24 hours and 50 to 850 meters. After masking noisy bin ensembles using quality control flags in the JASADCP metadata

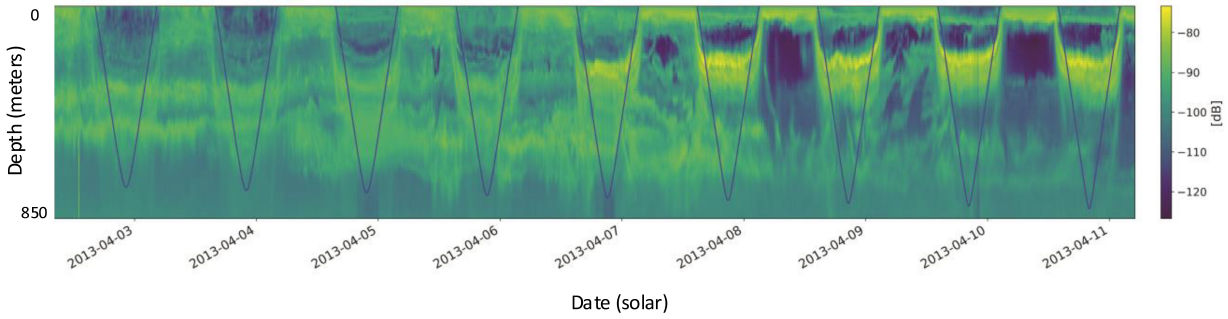

Figure A.1: Echogram extraction using the solar angle. Diel echograms are segmented from cruise-length time series by identifying the 24-hour region around each peak of the solar angle (solar angle maxima).

(pflag: CODAS), profiles from echograms were temporally joined to the grid and interpolated. We excluded all echograms from the dataset that spanned less than 23 hours of total coverage (from the first to the last profile) and or where the total amount of dropped profiles was greater than 3 hours (Fig. A.2).

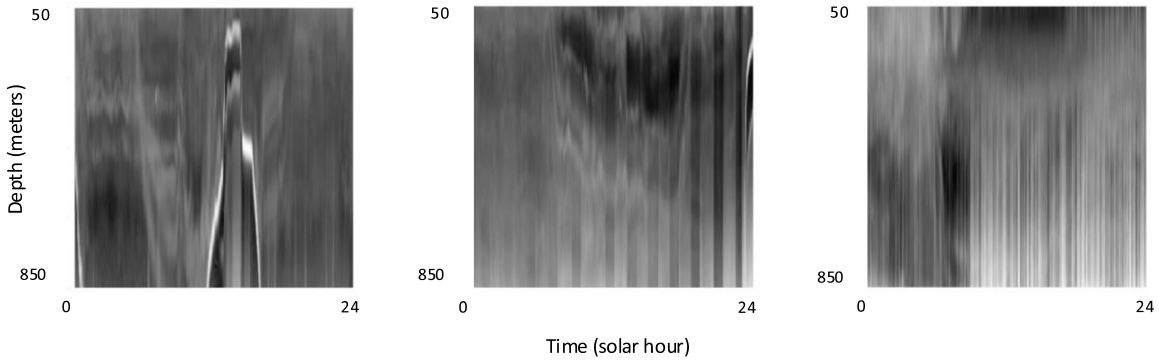

Figure A.2: Time-depth standardization exclusion criteria. Echograms spanning less than 23 hours of total coverage or with more than 3 hours of dropped profiles were excluded from the dataset.

A.1.2 From raw received signal strength indicator (RSSI) to relative scattering density

Previous work using ADCPs to estimate quantitative biological properties like biomass from scattering intensity have relied on comparisons to calibrated scientific echosounders to develop 'corrective' expressions. Receveur et al., (2020) suggests that device effects and numerical error in ADCP computations require periodic, and instrument specific corrections to make reliable quantitative estimates. The dataset used here contains observations from at least five different

instruments, some of which have operating lifetimes of over 15 years and more than 100 cruises (Table A.3). There are insufficient corresponding observations by shipboard echosounders to calibrate corrective functions for these instruments over these deployment lengths, and therefore it was not possible to use these data to quantitatively estimate physical properties from the scattering intensity. Therefore, given the unreliability of combined estimates of numerical S_v from these data, the useful information is contained in the relative density of scattering. Here, we describe how the steps of computing S_v, estimating calibration parameters and normalizing the echograms were used to promote the most reliable representations of relative scattering density.

Computing S_v to correct relative density

Computing S_v corrects for the influence of environmental factors, and for depth attenuation of the signal, which impact the shape of the vertical profiles. The raw received signal strength (RSSI), before correcting for these factors, represents a highly distorted spatial representation of the scattering as sound gets progressively more attenuated in the water column (Fig. A.3). Therefore, S_v conversion is an important step despite normalization, and was used to correct distortions in relative density. The impact of S_v conversion on the vertical scattering structure is represented by the example in Fig. A.3.

Computing S_v and estimating calibration parameters

We computed Sv from the archival ADCP data using the most recent form of the ADCP backscattering equation from Mullison (2017) as in previous studies (Section 3.2.2, Equation 3.1; e.g., Receveur et al. 2020; Leung et al., 2020; Song et al., 2022). Environmental variables and calibration parameters required to compute some of the terms in Equation 3.1 are not included in the cruise archive, and so need to be estimated. Several of the parameters needed to compute Sv are general to all instruments, and some are specific to individual instrument frequencies and classes.

Parameters common to all instruments include R—the slant range to depth cell—2αR—the total absorption—E_r—real time reference for echo intensity—c—the speed of sound through fluid—and α—the absorption coefficient. We computed these parameters as described in Section 3.2.2. The other parameters needed to compute the terms in Equation 3.1 are instrument and frequency specific, and include C, P_{DBW}, and Rayleigh distance—which we obtained from Mullison (2017)—and K_c—conversion factor for echo-intensity—which we obtained from field service emails with Teledyne (TRDI) engineers. There are only two instruments in our dataset—TRDI OS-I and OS-II 38kHz instrument classes. Collated parameters for these and other instruments in the OS class are shown in

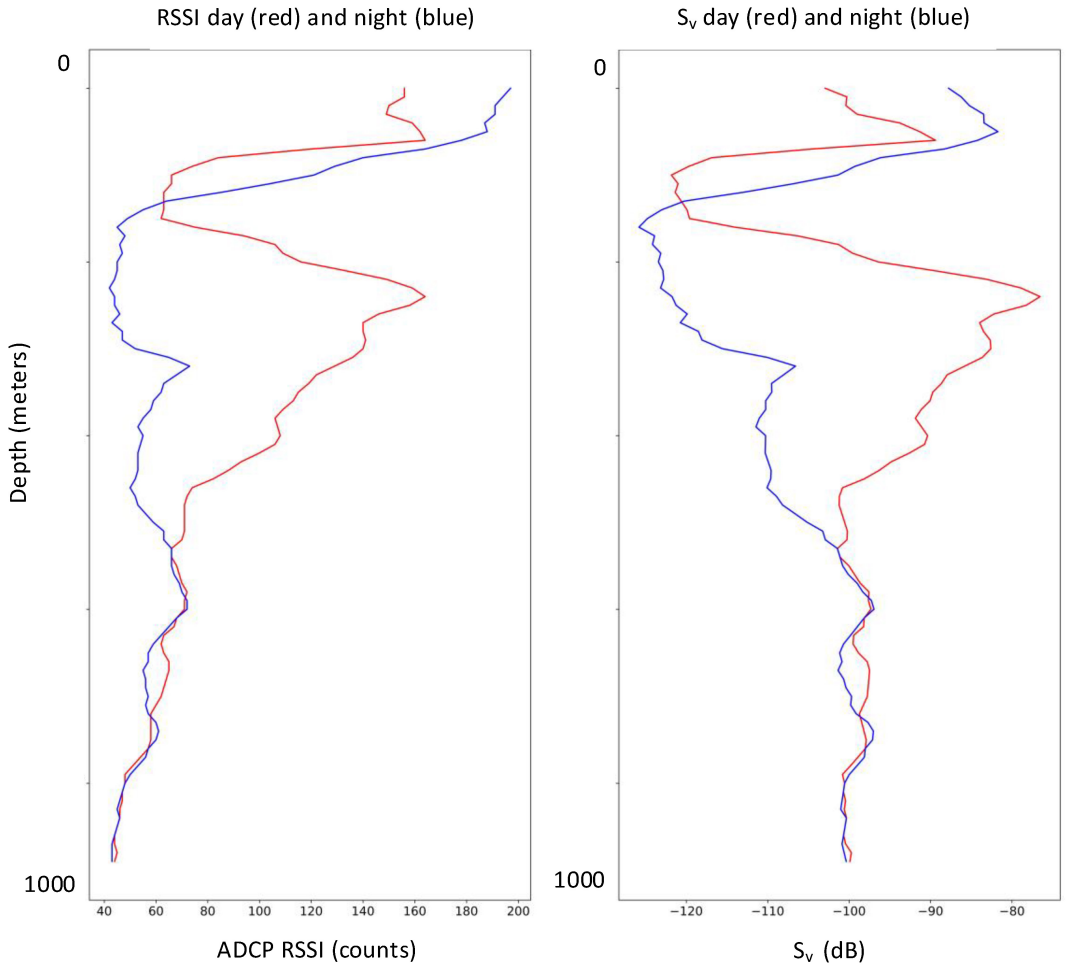

Figure A.3: Impact of S_v conversion on vertical scattering structure. The raw received signal strength (RSSI) represents a highly distorted spatial representation of scattering due to progressive attenuation with depth. S_v conversion corrects for environmental factors and depth attenuation.

Table A.1 below. For Sv computations used in the final analysis we used the mean values in the table on lines 2 and 5.

Table A.1: Collated parameter values for Ocean Surveyor-I and Ocean Surveyor-II ADCP instruments for 38, 75 and 150 kHz. Values are from Mullison (2017) and email correspondence with Teledyne (TRDI) engineers Jerry Mullison and Scott Idle in January 2020.

Instrument	**Freq (kHz)**	**C 25%**	**C 6%**	P_{DBW} **Battery**	P_{DBW} **Power**	**Rayleigh Dist.**	K_c	E_r
OS	150	−156.01	−156.01	None	21	1.62	0.42 ± 0.012	22 ± 2.08
OS	75	−164.26	−164.26	None	24	3.24	0.39 ± 0.019	19 ± 2.93
OS	38	−172.19	−172.19	None	24	8.19	0.37 ± 0.004	14 ± 4.83
OS-II	150	−156.01	−156.01	None	21	1.62	0.42 ± 0.012	22 ± 2.08
OS-II	75	−164.26	−164.26	None	24	3.24	0.39 ± 0.019	19 ± 2.93
OS-II	38	−172.19	−172.19	None	24	8.19	0.37 ± 0.004	14 ± 4.83

Normalization to eliminate spurious variability

Evidence from previous research shows that for a single ADCP instrument operated across multiple deployments, the associated corrective functions to calibrated echosounders can maintain identical correlations but show large scalar offsets in dB, such that deployment-specific intercepts for these functions are warranted even for individual devices (Receveur et al., 2020). This means that S_v measurements from a single instrument taken across different cruises—all things equal—may have distinct mean numerical scattering intensities but nearly identical relative scattering intensity. This is also true for components of the error in Sv estimates related to calibration parameters that are not depth variable—for example in L_{DBM} and P_{DBW}—such that relative density representation is conserved despite spurious variability in intensity. As noted above, the useful information in the acoustics data used here is the relative density of scattering. Normalizing the data was intended to eliminate any spurious signals related to scalar offsets in scattering intensity across instruments or within the operating life of a single instrument, or due to errors in depth-invariant calibration components, and does not change the representation of relative density.

Evaluating the impact of calibration parameter mis-estimation

To the extent that error is variable with depth or intensity, it is possible that the visual features of the echograms—such as layer depth—could be distorted. Given the lack of corresponding observations from calibrated echosounders, it was not possible to evaluate device-specific measurement errors. That said, in our analysis the numerical error in S_v is potentially increased through the need to estimate calibration parameters (Sec. A.1.2; Mullison et al., 2017; Leung, 2020). However, this component of the error can be diagnosed within the data using the means and standard errors for

the parameters (Table A.1). Specifically, we computed the clustering solution using the mean parameters, and compared the result to 500 alternate parameterizations simulated by varying the parameters within their interquartile ranges (Table A.2). The resulting clustering is highly stable across these parameter estimates based on Adjusted Mutual Information (AMI; Hubert & Arabie, 1989) and Adjusted Rand Index (ARI; Steinley., 2004) stability metrics. The mean AMI and ARI score of all comparisons and associated standard deviations are shown in Table A.2.

Table A.2: Adjusted Mutual Information (AMI) and Adjusted Rand Index (ARI) scores from iterated comparisons of clustering outputs based on the mean Sv echograms, and echograms generated using different calibration parameters from within the provided ranges.

Stability metric	**Mean and standard deviation**
Adjusted mutual information (AMI)	0.919 ± 0.08
Adjusted Rand index (ARI)	0.917 ± 0.10

A.1.3 Summary data metrics

Summary metrics for the final 1,596 echogram subset of the JASADCP used in this analysis are reported in the graphs below. Section 3.3.2, Table 3.1 reports nominal figures for the number of echograms in each region and the number of parcels ($3° \times 3°$ grid squares). The spatial coverage of observations is shown on the map in Section 3.3, Fig. 3.2b. Supplementary graphs show the within-region distribution of observations across seasons (Fig. A.4) and years (Fig. A.5).

The echogram dataset is comprised of 306 separate cruises across 4 different vessels (Table A.3). The population of unique instruments could not be determined given the absence of serial numbers in these text fields (the location where serial numbers would be provided was identified and in all cases found to be unpopulated) and the fact that the primary key of the archive (SACID; CODAS) describes unique combinations of cruise, vessel and instrument.

Table A.3: Summary statistics describing the contribution of individual research vessels to the JASADCP subset used in this analysis. Cruise metadata do not include instrument serial numbers, so we cannot report the count of unique instruments.

Vessel name	**N cruises**	**Bin length (m)**	**Transducer depth (m)**	**Instrument type**	**Frequency**	**First cruise**	**Last cruise**	**Manufacturer**
R/V Kaiyo (49XX)	5	16.0	6.0	Ocean Surveyor II	38 kHz	2001-11-29	2003-10-30	TRDI
R/V Kilo Moana	142	24.0	8.0	Ocean Surveyor	38 kHz	2003-12-18	2018-11-15	TRDI
R/V Lawrence M. Gould	128	24.0	6.0	Ocean Surveyor	38 kHz	2004-11-29	2018-11-08	TRDI
R/V N.B Palmer	30	24.0	7.0	Ocean Surveyor	38 kHz	2004-10-12	2018-03-10	TRDI

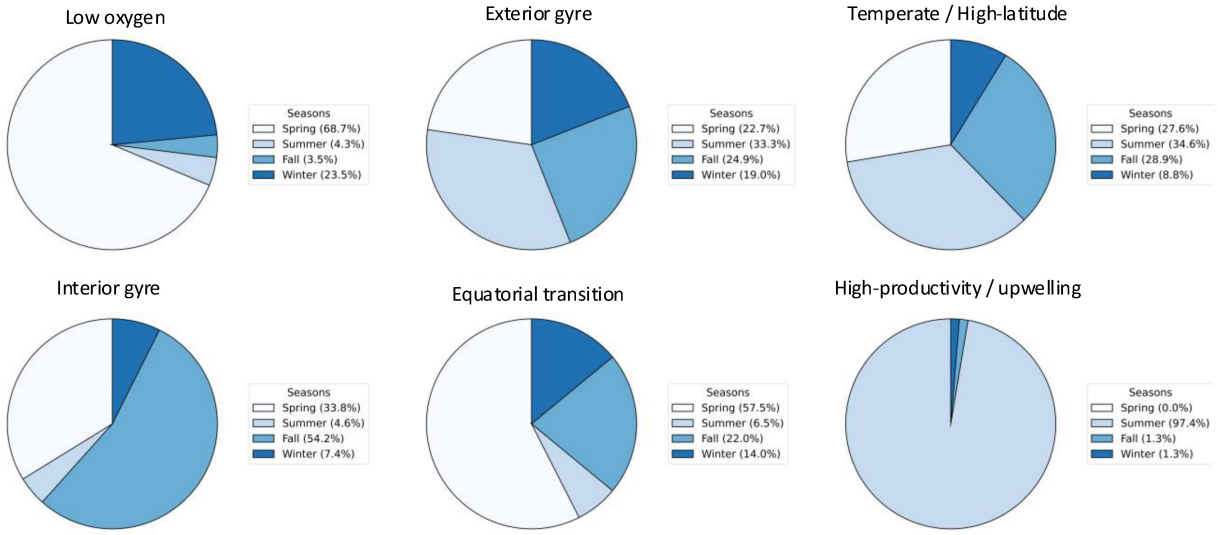

Figure A.4: Within-region distribution of observations across seasons for the final 1,596 echogram subset of the JASADCP.

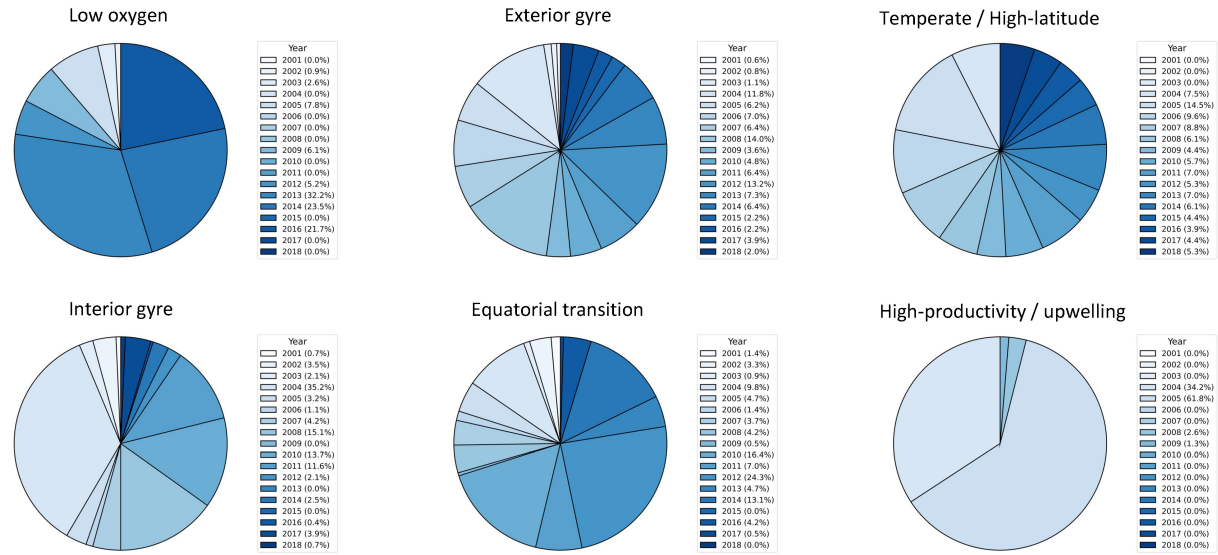

Figure A.5: Within-region distribution of observations across years for the final 1,596 echogram subset of the JASADCP.

A.2 Supplemental methods

A.2.1 CAE workflow

Comparison to alternate dimensionality reduction approaches

We compared the performance of the CAE-SSIM to principal component analysis (PCA), kernel PCA (kPCA; Mika et al., 1998), Isomap (Choi et a., 2004), spectral embedding (Luo et al., 2003) and a CAE trained using a mean-squared error loss function (CAE-MSE).

CAE and PCA-based models preserve sufficient structure to enable an approximate inverse mapping (reconstructive). Echogram images can be estimated (reconstructed) from these mappings and used to make 1:1 comparison across the approaches. We computed embeddings at 2, 10 and 20 dimensions and compared the mean SSIM and MSE scores for dataset reconstructions. The CAE-based models perform best on their respective loss penalties at all dimensions (Table A.4).

Table A.4: Mean SSIM and MSE scores for all reconstructed images in the dataset. Both metrics are computed for dataset reconstructions based on 2, 10 and 20 dimensional embeddings of the data using PCA and CAE-based approaches.

	CAE-SSIM			**PCA**			**kPCA**			**CAE-MSE**		
Dimension	2	10	20	2	10	20	2	10	20	2	10	20
SSIM	0.75	0.81	0.84	0.70	0.76	0.78	0.71	0.74	0.76	0.71	0.74	0.79
MSE	0.018	0.005	0.003	0.013	0.006	0.004	0.019	0.007	0.001	0.013	0.003	0.001

Isomap and spectral embedding are not reconstructive dimensionality reduction algorithms, and we instead made comparisons of clustering performance (Sec. A.2.1; Table A.8) using feature sets generated by the alternate dimensionality reductions.

Extended comparison of PCA and CAE

PCA performs competitively on this dataset given the complexity and computational expense of training convolutional models. It is simpler to implement, easier to interpret, and contains far fewer free parameters (Table A.5).

However, CAEs offer advantages as data volumes and complexity increase. Even on this relatively coarse ADCP dataset, the CAE achieves higher reconstruction scores (SSIM = 0.84 vs 0.78; Table A.4) and cleaner clustering (silhouette = 0.25 vs 0.17; Table A.5), particularly for disorganized echograms where linear bases underfit textures and edges (Fig. A.6). For rapid or exploratory

analyses, PCA remains appropriate and informative, but for large, heterogeneous datasets or applications targeting fine-scale structural detail, CAEs provide higher-fidelity representations and a scalable foundation for future bioacoustic workflows.

Table A.5: The number of degrees of freedom (DOF) for a PCA given the input size of our dataset, and the latent space dimensions experimented with (2, 10, 20). The corresponding number of trainable parameters in our CAE architecture for the same latent space dimensions are given for comparison.

Method	**Latent space dimension**	**DOF (PCA) / Trainable parameters (CAE)**	**CAE / PCA**
PCA	2	32,767	25.65
PCA	10	163,795	17.93
PCA	20	327,490	16.97
CAE (Sec. 3.2.3)	2	840,579	NA
CAE (Sec. 3.2.3)	10	2,937,739	NA
CAE (Sec. 3.2.3)	20	5,559,189	NA

Embedding space dimension

We selected 20 as the final CAE embedding space dimension based on the visual quality of image reconstructions. This was done to ensure that features of interest—such as layers, migrations and the structure of scattering—were effectively incorporated into the feature sets. CAE-SSIM reconstructions by models trained using 2, 10 and 20 dimensional embedding spaces are shown in Fig. A.7. Reconstructions built using 20-dimensional spaces summarize key structural modes even in echograms with low structural clarity.

Hyperparameter optimization

We employed three criteria in hyperparameter selection for the final CAE model: (1) minimize final loss (SSIM); (2) maximize generalization (convergence of training and validation loss; Fig. A.8); and (3) visually evaluate reconstructions to assess capture of scattering structures (Fig. A.9). Table A.6 summarizes final and alternate architectures, the criteria guiding each choice, and their impacts with references to supporting figures and or tables.

The latent space dimension (Sec. A.2.1; Fig. A.7) and objective function (Sec. A.2.1; Table A.4) are discussed above. Fig. A.8 (top) shows loss curves for effective learning rates that are too low (slow learning, poor generalization), too high (premature convergence to local minima), and well-matched (smooth convergence with good validation performance). Fig. A.8 (bottom) shows the

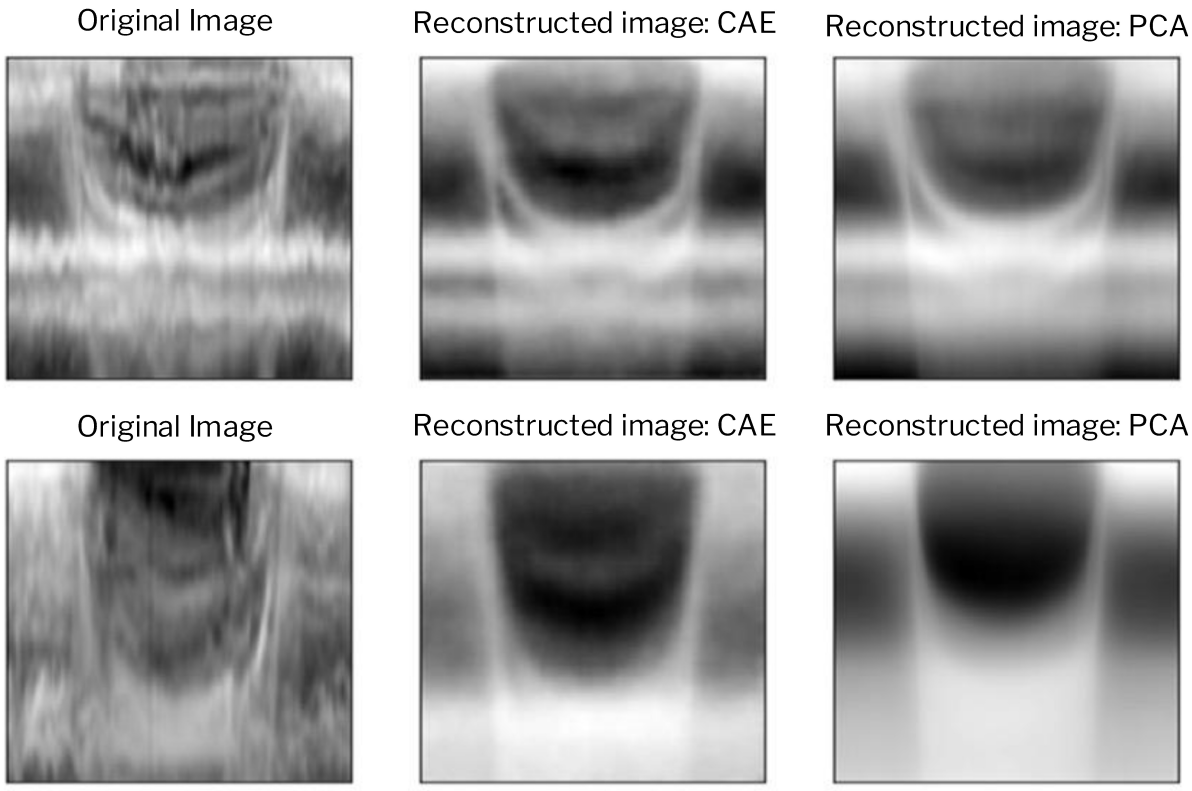

Figure A.6: Comparison of PCA and CAE reconstructions. CAE reconstructions capture finer structural detail than PCA, particularly for disorganized echograms where linear bases underfit textures and edges.

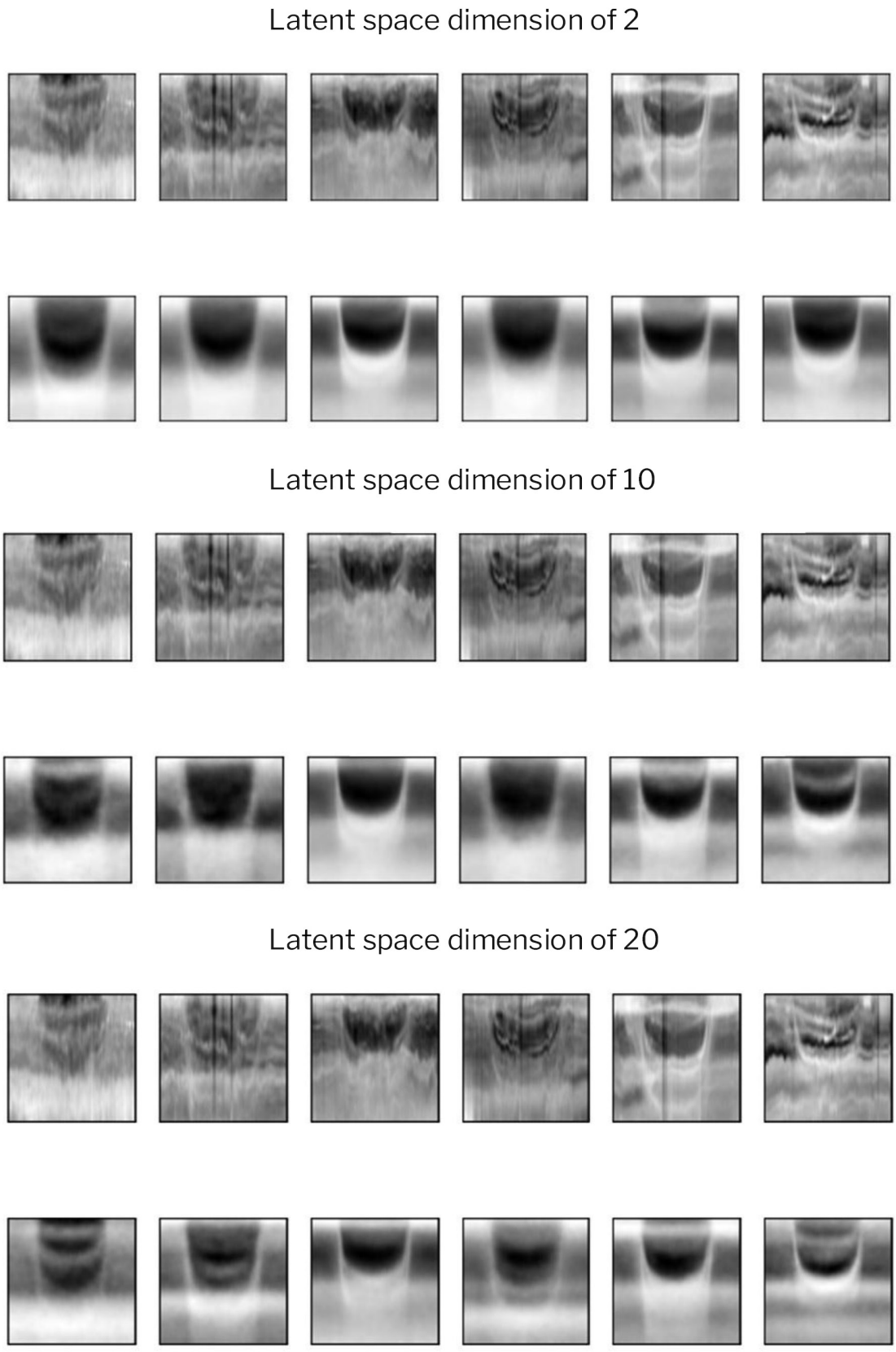

Figure A.7: CAE-SSIM reconstructions by models trained using 2, 10 and 20 dimensional embedding spaces. Reconstructions built using 20-dimensional spaces summarize key structural modes even in

Table A.6: Model architecture and training hyperparameters selected for the final CAE model are reported along with alternate configurations attempted, the criteria used to make the final selection, and the impact of the final selection on the model performance.

Architectural hyperparameters	**Final model selection**	**Software utility**	**Alternate configurations**	**Parameter selection criteria used**	**Model impact**
Latent space dimension	20	NA	2, 10	Numeric SSIM loss; Visual inspection of reconstructions	SSIM, MSE scores reported in Table A.4 (reconstructions shown in Fig. A.7)
Objective function	SSIM	NA	PSNR, MSE	Visual inspection of reconstructions	PSNR showed poor performance; MSE model performance reported in Table A.4 (reconstructions shown in Fig. A.6)
Layer architecture	3 conv layers (32, 64, 128)	pytorch.Conv2D	3 conv layers (16, 32, 64); 4 conv layers (32, 64, 128, 256)	Visual inspection of reconstructions	Shallow architecture produces insufficient detail, deeper architecture reproduces noise (reconstructions shown in Fig. A.9)
Activation function	Sigmoid	pytorch.Sigmoid	None	No experimentation done	No experimentation done
Training hyperparameters	**Final model selection**	**Software utility**	**Alternate configurations**	**Parameter selection criteria used**	**Model impact**
Regularization approach	None / Data augmentation	NA	Dropout (p=0.1; 0.2; 0.5)	Model convergence	Model convergence was not improved by Dropout regularization; Data augmentation was used in lieu of explicit regularization.
Optimizer	Adam	pytorch.optim.Adam	SGD + momentum	Numeric SSIM loss; Model convergence	Loss not significantly impacted by optimizer
Scheduler	ReduceLROnPlateau (factor=0.1, patience=2)	pytorch.optim.lr_scheduler.ReduceLROnPlateau	Constant LR; CyclicLR; patience=3	Model convergence	Adaptive learning rate improves performance (Fig. A.8, bottom)
Batch size	64	NA	32, 128, 256	Model convergence	Batch size and learning rate jointly optimized for training loss smoothness and coupling with testing and validation loss (Fig. A.8, top)
Learning rate	0.0001	NA	0.001, 0.0005, 0.00001	Model convergence	Batch size and learning rate jointly optimized for training loss smoothness and coupling with testing and validation loss (Fig. A.8, top)

impact of adding data augmentation (centre) and increasing the patience parameter (right). Selected hyperparameters were associated with the best performance on numeric SSIM loss (0.16, Table A.4), and the most tightly coupled training and validation loss curves (right).

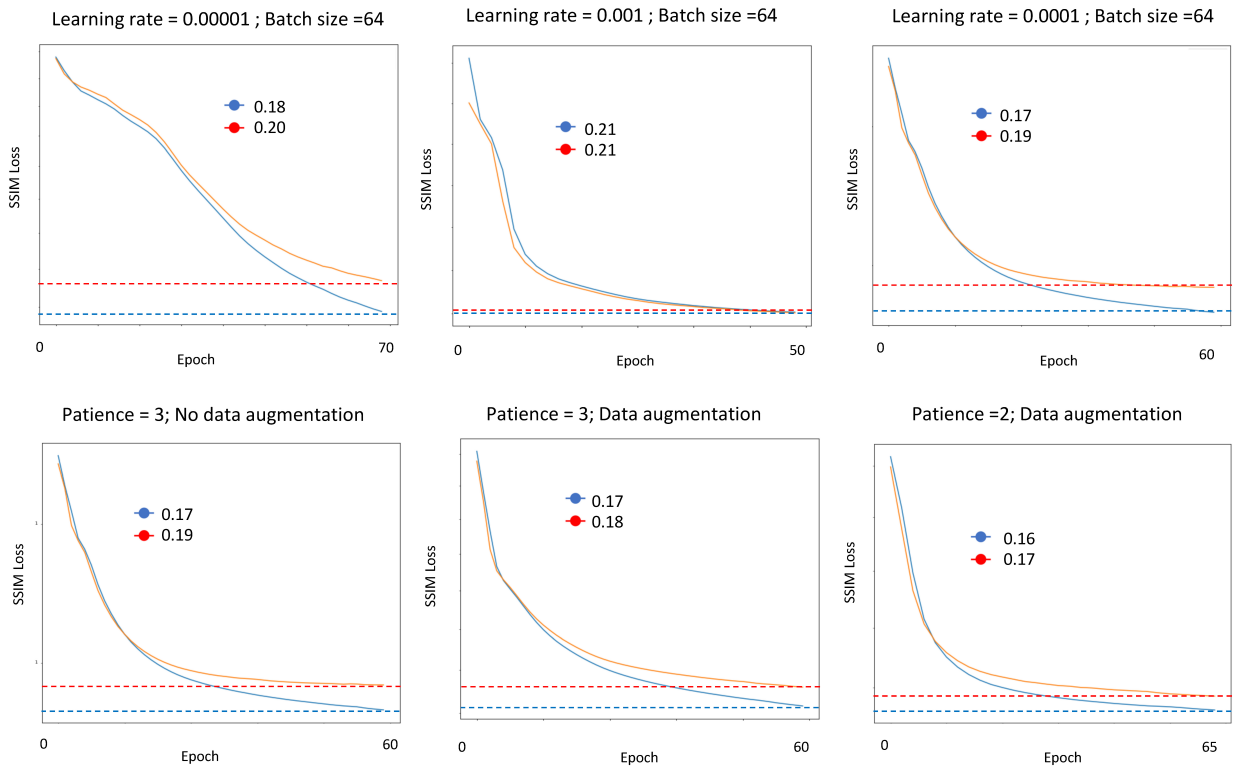

Figure A.8: (Top) Loss curves for learning rates that are too low, too high, and well-matched. (Bottom) Impact of data augmentation and increasing the patience parameter on training and validation loss convergence.

Fig. A.9 shows an example echogram with significant, structured noise as an example input (first panel). The deep architecture (second panel) achieves low reconstruction loss but overfits by replicating noise from the original image (dark vertical band). The shallow architecture (third panel) underfits, producing insufficient structural detail (much like the 2 and 10 dimensional embedding spaces; Sec. A.2.1; Fig. A.7). The architecture used in the final model (fourth panel) captures meaningful structural features while suppressing noise.

Clustering algorithm

Clustering performance was compared for k-means (MacQueen, 1967), agglomerative clustering (Ward, 1963), DBSCAN (Ester et al., 1996), and spectral clustering (Ng et al., 2001). Performance

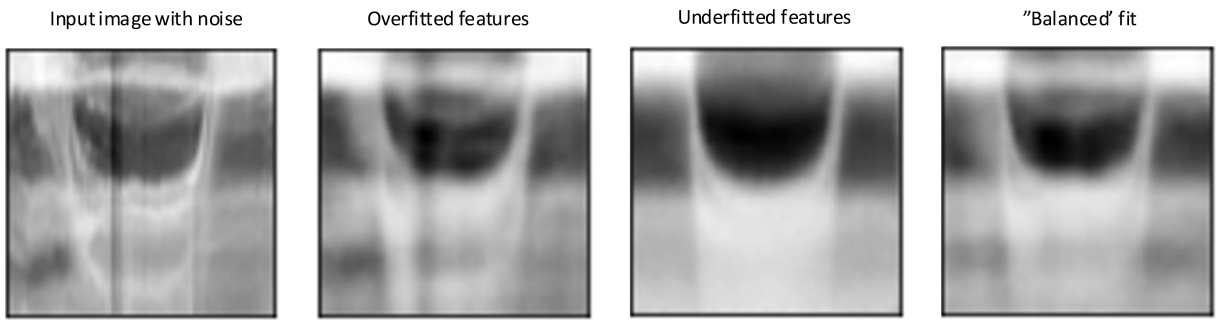

Figure A.9: Comparison of CAE reconstructions across different layer architectures for an echogram with significant structured noise. The deep architecture overfits, the shallow architecture underfits, and the final model architecture captures meaningful structural features while suppressing noise.

was evaluated with the Calinski–Harabasz (CH; Calinski & Harabasz, 1974) index, the Davies–Bouldin (DB; Davies & Bouldin, 1979) index, and the Silhouette score (SS; Rousseeuw, 1987). Results for k = 2–20 are reported in Table A.7 and Fig. A.10.

Selecting the number of regions (k)

Clustering metrics plot over k (Fig. A.10). Maxima and visual 'elbow' points occur in the 4–10 cluster range. Spectral clustering with $k = 4$ produces the best overall score on the DB metric (0.96), while k-means produces top scores on the CH index (276) and silhouette score (0.255). K-means was used to select the optimal k. Silhouette is maximized at 6 (left); the CH index has an 'elbow' point between 4 and 6 (centre); the DB index is minimized at 6 (right). Clustering metrics do not indicate an unequivocal choice for the optimal k. Solutions for alternate k values are shown in Fig. A.11.

Sub-structure in the temperate / high-latitude region

Comparison to alternate dimensionality reduction approaches (non-reconstructive)

The three best performing clustering algorithms were used to evaluate alternate dimensionality reduction methods that could not be compared using reconstructions. The metrics are reported in Table A.8.

Table A.7: Cluster metrics CH, DB, and SS for the three best performing clustering algorithms on k values in the range 2–20. The maximum score on each metric and algorithm, for $k > 2$, is shown in bold.

	Agglomerative			**KMeans**			**Spectral**		
k	SS	CH	DB	SS	CH	DB	SS	CH	DB
2	0.167	268.538	2.091	0.210	324.714	1.845	0.286	64.862	0.772
3	0.147	238.880	1.946	0.161	275.695	1.855	0.211	122.987	1.078
4	0.176	228.900	1.589	0.187	269.895	1.623	**0.221**	153.115	***0.966***
5	0.181	**236.493**	1.710	0.210	*276.182*	1.488	0.183	188.552	1.421
6	0.171	231.609	1.541	**0.255**	269.644	**1.365**	0.191	190.260	1.571
7	0.199	227.546	1.563	0.226	266.109	1.423	0.199	197.350	1.383
8	**0.202**	220.938	1.473	0.220	253.972	1.437	0.206	202.316	1.401
9	0.190	214.507	1.539	0.220	244.847	1.424	0.209	192.897	1.290
10	0.197	207.357	**1.429**	0.227	238.334	1.401	0.210	**203.510**	1.302
11	0.201	199.745	1.467	0.224	227.389	1.479	0.203	191.792	1.332
12	0.194	193.105	1.507	0.222	219.784	1.397	0.195	187.520	1.352
13	0.189	186.829	1.547	0.209	210.206	1.417	0.193	182.325	1.336
14	0.189	180.686	1.519	0.216	203.951	1.423	0.189	183.128	1.365
15	0.191	175.726	1.482	0.208	196.448	1.498	0.179	171.248	1.398
16	0.193	170.881	1.520	0.201	191.825	1.521	0.174	161.841	1.453
17	0.177	166.729	1.556	0.210	186.378	1.404	0.183	163.989	1.466
18	0.177	162.875	1.553	0.206	180.093	1.494	0.175	153.103	1.473
19	0.170	158.510	1.595	0.198	176.511	1.553	0.177	154.096	1.438
20	0.169	154.492	1.615	0.199	172.540	1.497	0.169	145.115	1.440

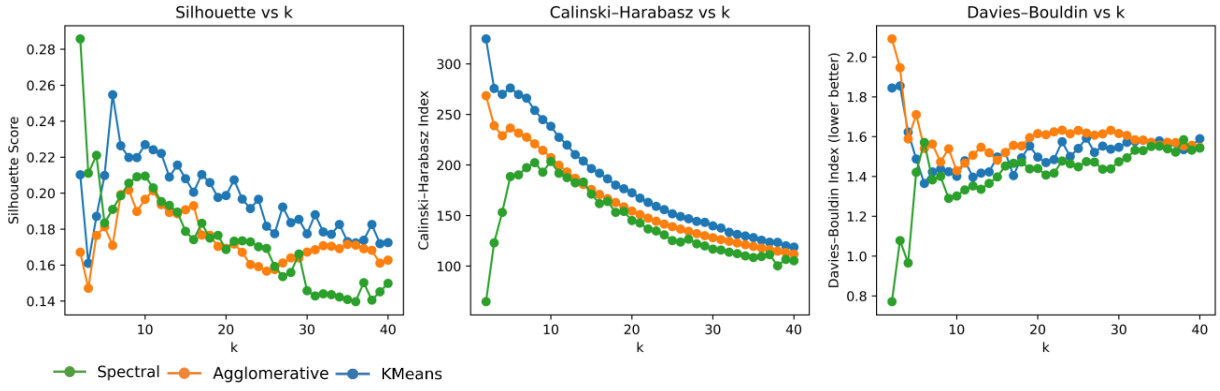

Figure A.10: Clustering metrics as a function of the number of clusters k for k-means. Silhouette is maximized at $k = 6$ (left); the CH index has an 'elbow' point between 4 and 6 (centre); the DB index is minimized at 6 (right).

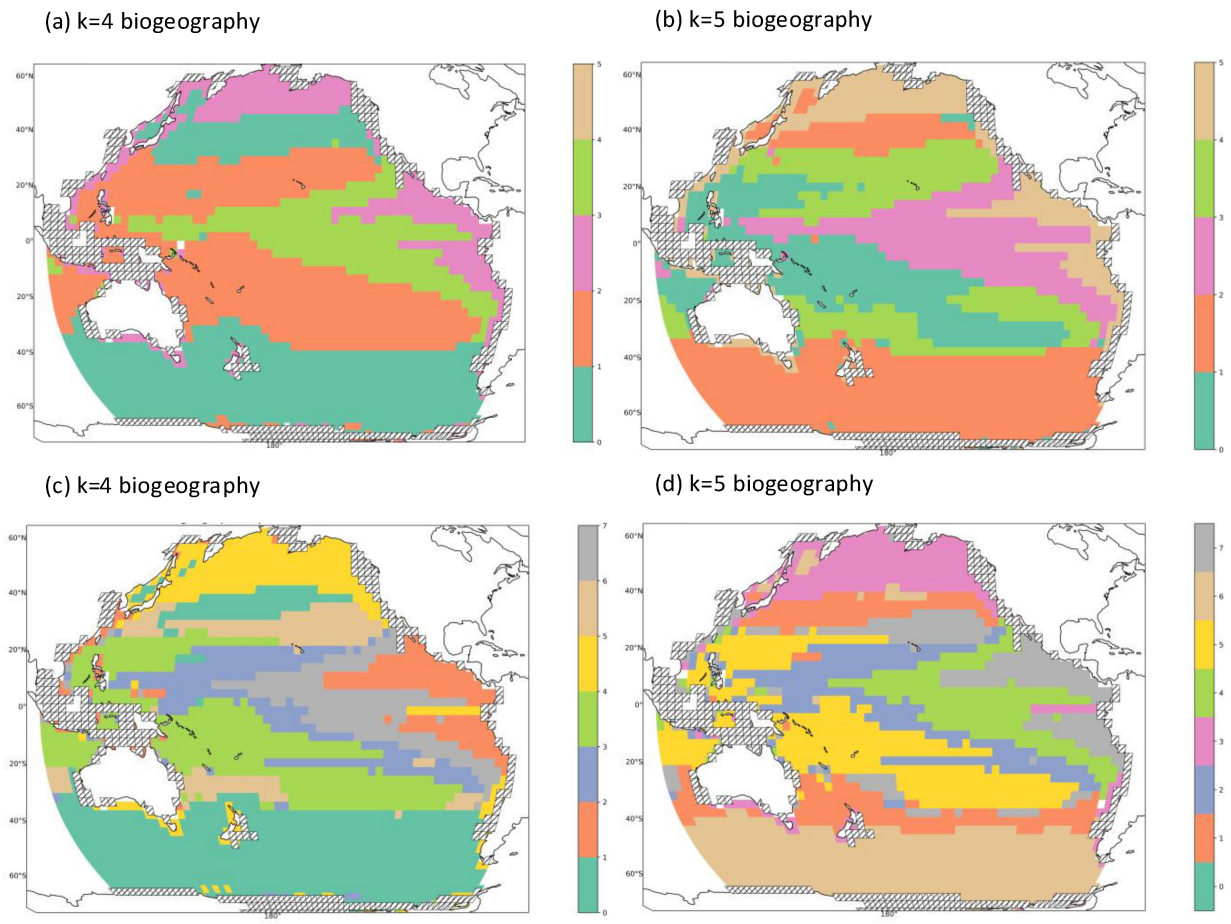

Figure A.11: Biogeographical solutions for alternate values of k, showing the spatial distribution of cluster assignments for different numbers of regions.

Table A.8: Cluster metrics SS and DB for each of the dimensionality reduction approaches compared. CH values are included for the CAE-SSIM feature set. All comparisons are based on a workflow using 6 clusters and a 20-dimensional embedding space.

	CAE-SSIM			**PCA**		**kPCA**		**Isomap**		**Spectral embedding**	
Metric	CH	SS	DB	SS	DB	SS	DB	SS	DB	SS	DB
kMeans	269.6	0.25	1.36	0.17	1.71	0.16	1.71	0.20	1.41	0.14	2.00
Agglomerative	231.6	0.17	1.54	0.12	1.87	0.13	1.61	0.14	1.58	0.12	2.00
Spectral	190.2	0.19	1.57	0.15	1.50	0.17	1.46	0.20	1.61	0.12	2.13

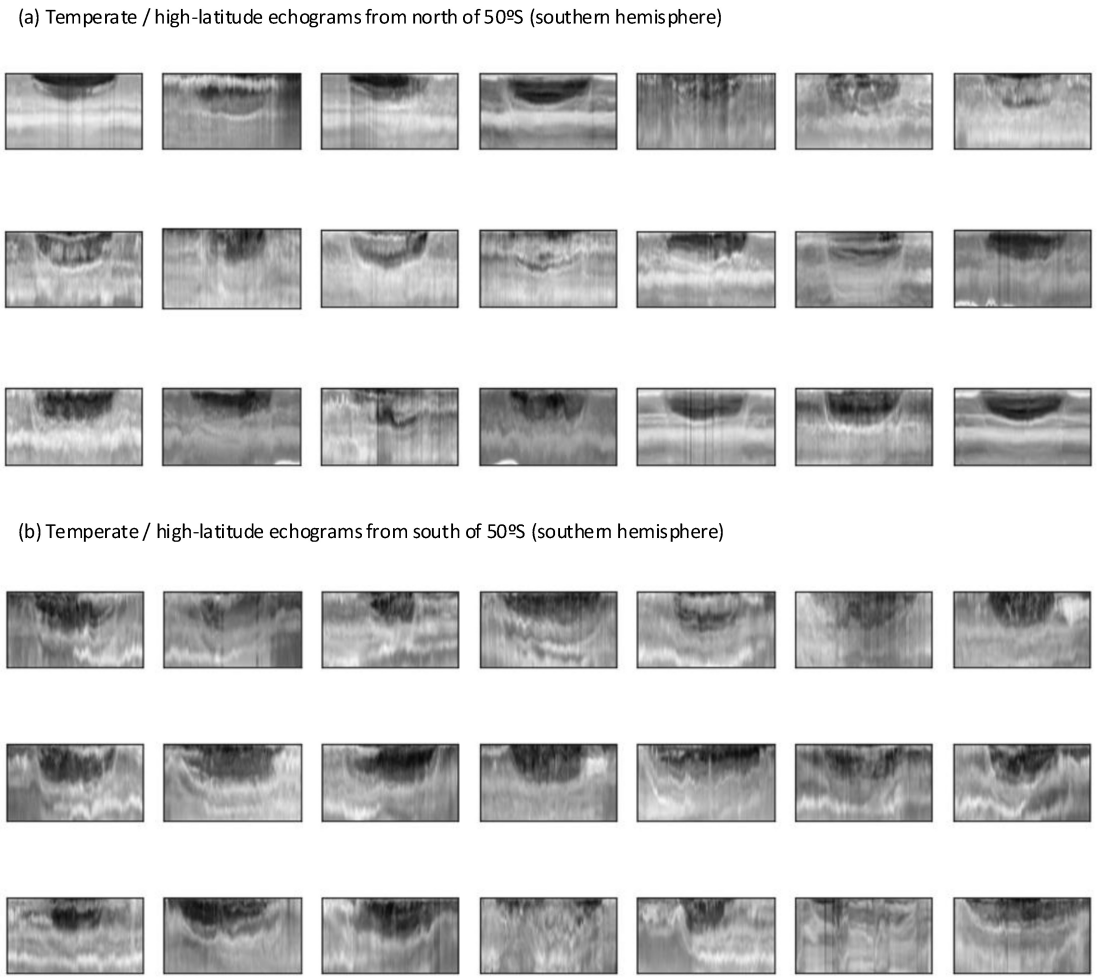

Figure A.12: Sub-structure in the temperate / high-latitude region, revealing additional patterns within this broadly defined biogeographical zone.

Regression Models

We experimented with different predictive models including XGBoost (Chen & Guestrin, 2016), decision tree (Breiman, 2017), random forest (Breiman, 2001; Ariza et al., 2022), support vector machine (Cortes & Vapnik, 1995), logistic regression (Berkson, 1944) and multi-layer perceptron (Rosenblatt, 1958). Mean predictive accuracy and standard deviations for all models are reported in Table A.9. Models were cross-validated and hyperparameters were optimized using RandomizedSearchCV functionality from scikit-learn within the model_selection utility (e.g., Bergstra & Bengio, 2012). Random forest generates the most accurate predictions, followed by XGBoost and multi-layer perceptron.

Table A.9: Mean predictive accuracy and standard deviation for all predictive models compared. Comparisons are made using the clustering output from the production workflow (k-means with 6 clusters).

Predictive model	**Accuracy**
Support vector machine	76.7 ± 6.1
Random forest	80.9 ± 4.8
Decision tree	72.3 ± 6.9
Multi-layer perceptron	77.1 ± 5.3
Logistic regression	71.0 ± 3.5
XGBoost	79.2 ± 5.0

Biogeographical predictions based on the top three performing models are shown in Fig. A.13. XGBoost closely corresponds to the CAE-SSIM solution, while the multi-layer perceptron preserves the general structure, but closely resembles the 4-cluster biogeographical solution (Fig. A.13).

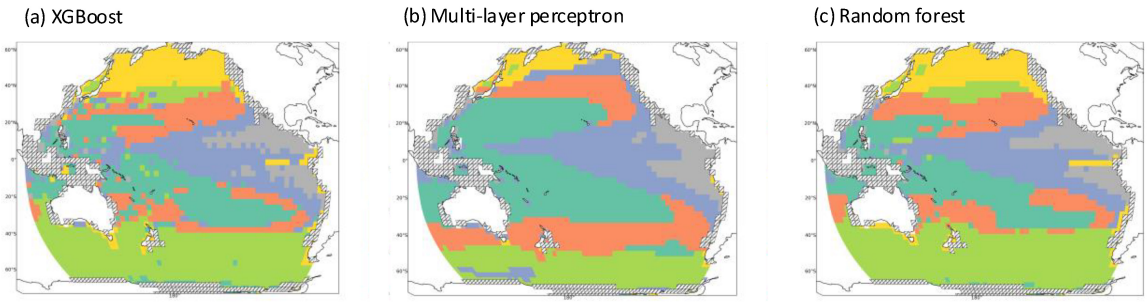

Figure A.13: Biogeographical predictions based on the top three performing predictive models: random forest, XGBoost, and multi-layer perceptron. XGBoost closely corresponds to the CAE-SSIM solution, while the multi-layer perceptron closely resembles the 4-cluster biogeographical solution.

A.3 Supplementary results

A.3.1 Final random forest model

An example confusion matrix from an iteration of the random forest model is shown in Fig. A.14 (bottom). Class specific accuracy is reported in Section 3.3.2, Table 3.1. Feature importances obtained using mean decrease in impurity (Louppe et al., 2013) for the final cross-validated model are reported in the bar graph in Fig. A.14 (top).

A.3.2 Regional distribution of covariates and derived properties

Figure A.15 provides an alternate visual representation of the correspondence of environmental covariates (Section 3.2.3) and derived properties (Section 3.2.4) to cluster regions. These relationships are visualized spatially in Fig. 3.4 and Fig. 3.6. The raw data values are included in Table 3.1 and Table 3.2.

A.3.3 Visualizing the clusters in 2 and 3 dimensions

Manifold visualizations were built by reducing the 20-dimensional CAE feature set into 2 dimensions using a diffusion-based algorithm (PHATE; Moon et al., 2017). PHATE (Potential of Heat-diffusion for Affinity-based Transition Embedding) is a dimensionality reduction algorithm that preserves both local and global structures by leveraging heat diffusion potentials, capturing intrinsic trajectory structures in high-dimensional data for smoother and more interpretable embeddings. PHATE generated more legible low dimensional representations of the CAE feature space than PCA and tSNE. 2-dimensional PHATE manifolds coloured by dissolved oxygen, primary layer depth, cluster labels and the echogram structure index (Section 3.2.4) are shown in Fig. A.16.

A.3.4 Comparison to previous biogeographical solutions

We compared our biogeographical solution to the two previous basin-scale solutions for the Pacific Ocean built using bioacoustics data (Section 3.4; Proud et al., 2017; Ariza et al., 2022). We reproduce all three solutions for the Pacific in Fig. A.17.

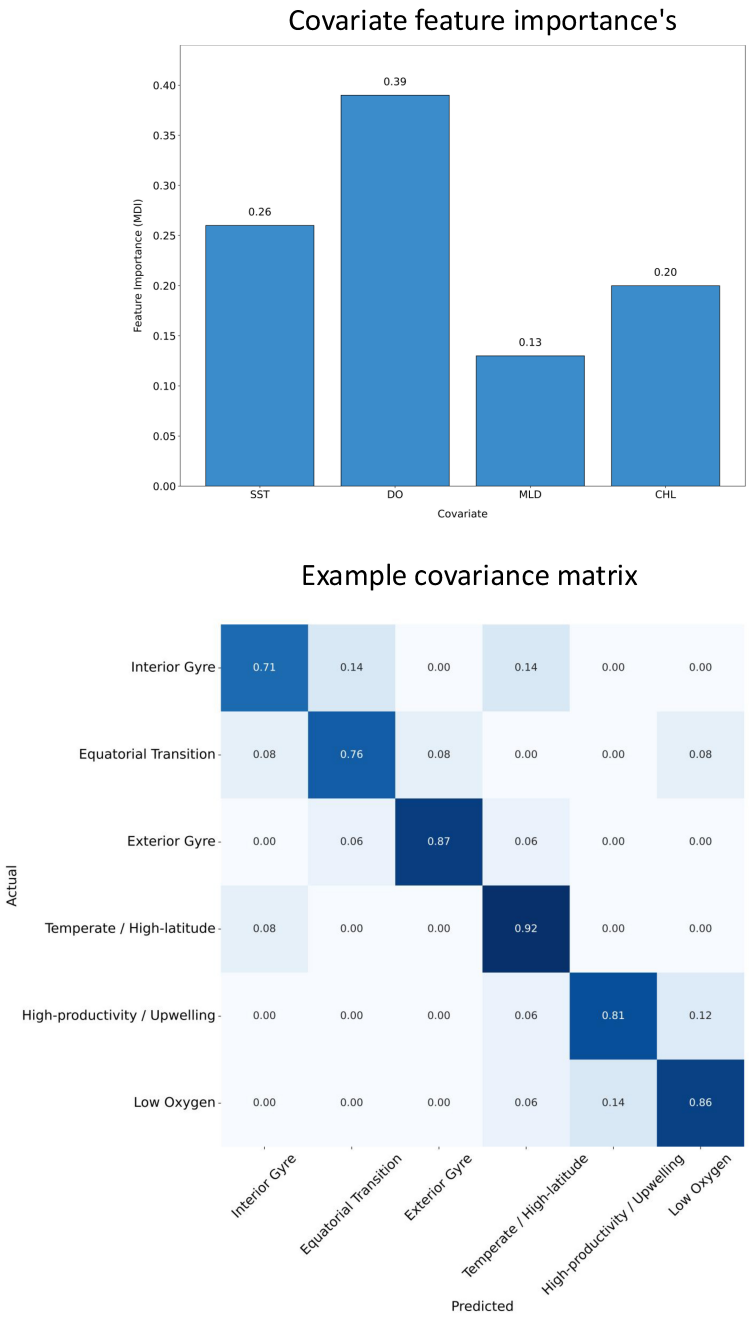

Figure A.14: (Top) Feature importances obtained using mean decrease in impurity for the final cross-validated random forest model. (Bottom) Example confusion matrix from an iteration of the random forest model.

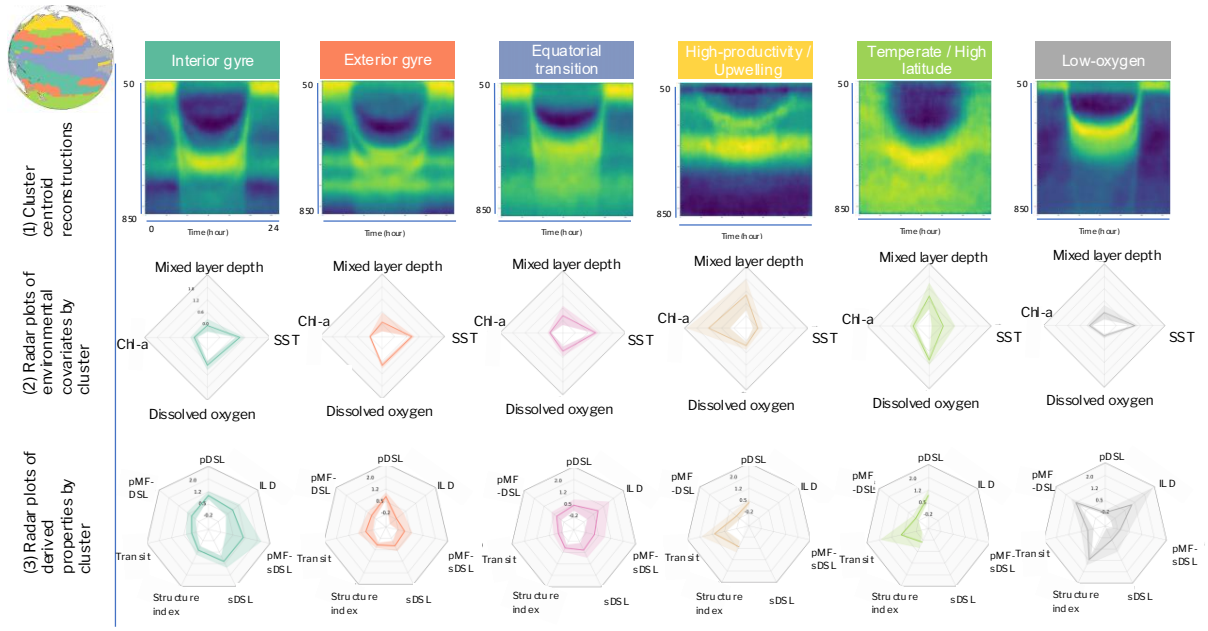

Figure A.15: Regional distribution of environmental covariates and derived properties across the six biogeographical regions. These relationships are visualized spatially in Fig. 3.4 and Fig. 3.6.

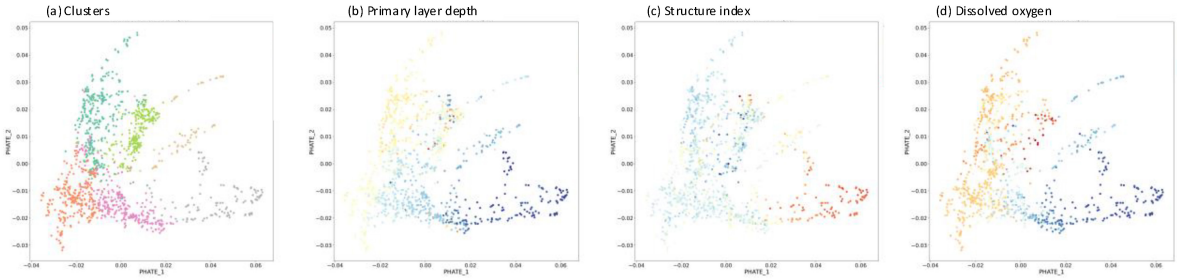

Figure A.16: 2-dimensional PHATE manifold visualizations of the 20-dimensional CAE feature set, coloured by dissolved oxygen, primary layer depth, cluster labels and the echogram structure index.

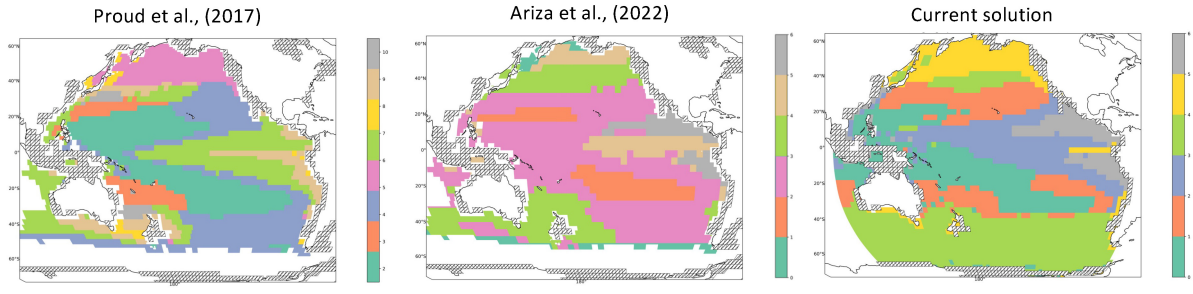

Figure A.17: Comparison of our biogeographical solution with the two previous basin-scale solutions for the Pacific Ocean built using bioacoustics data (Proud et al., 2017; Ariza et al., 2022).

A.3.5 Using the Pacific model to make global predictions

The vast majority (>95%) of the 38kHz subset of the JASADCP (Section 3.2.1) is from within the Pacific basin. We elected to restrict our biogeographic prediction to the Pacific, given the possibility that inclusion of sparse observations from other basins could lead to biased results, compromise cluster stability, and hinder robust validation. 'Experimental' global predictions based on our model (Section 3.2.3) are shown in Fig. A.18, and closely resemble previous solutions outside the Pacific basin (ex. Proud et al., 2017; Ariza et al., 2024).

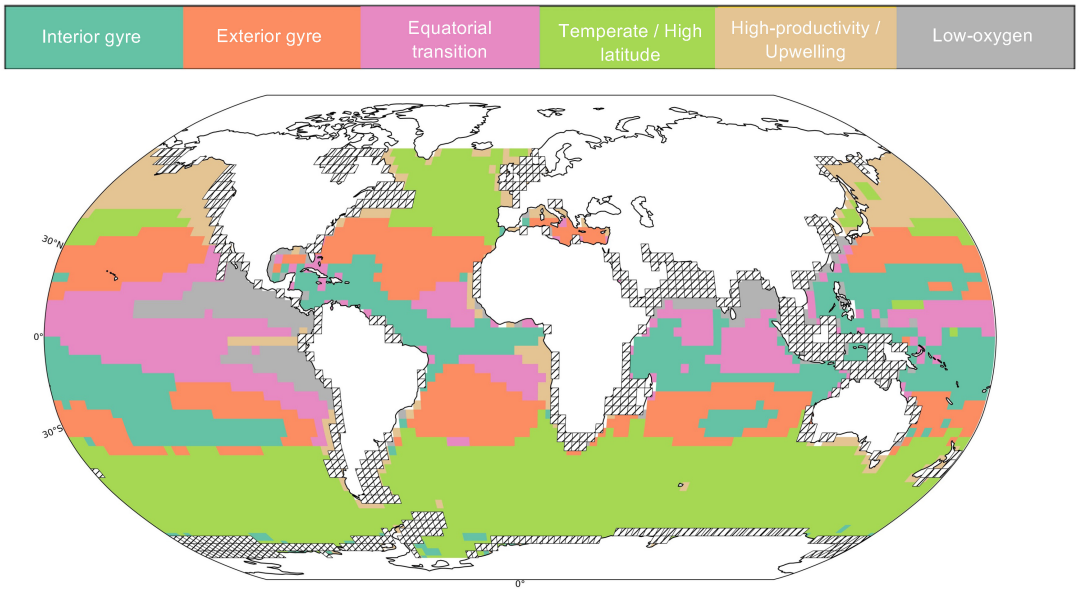

Figure A.18: Experimental global biogeographical predictions based on the Pacific-trained model. Predictions closely resemble previous solutions outside the Pacific basin.

A.3.6 DVM timing extraction with least squares 'top-hat' segmentation

Figure A.19 shows examples of estimated bulk migration times using the process described in Section 3.2.4.

A.3.7 Cluster image mosaics and reconstructions

Visual inspection of the underlying image dataset can be used as a component of performance assessment. Mosaics of regional echograms show visually consistent patterns that align with the regional centroids. Figures A.20–A.25 show image mosaics of cluster subsets and corresponding CAE reconstructions for each of the 6 regions.

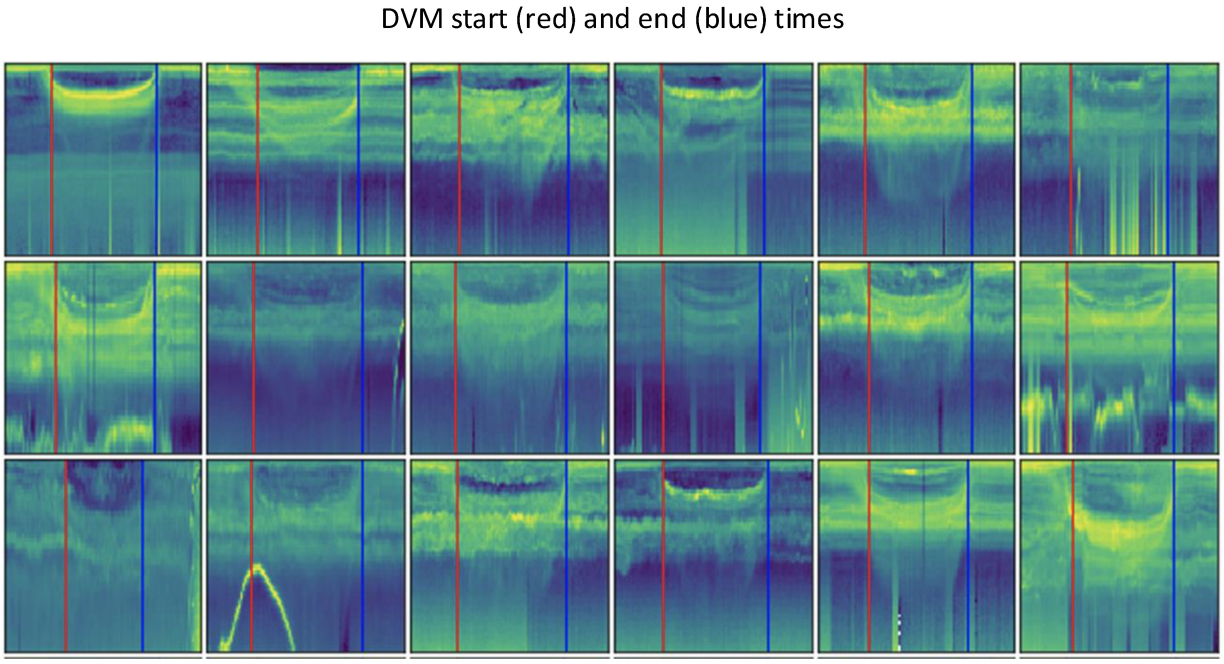

Figure A.19: Examples of estimated bulk migration times using the least squares 'top-hat' segmentation process described in Section 3.2.4.

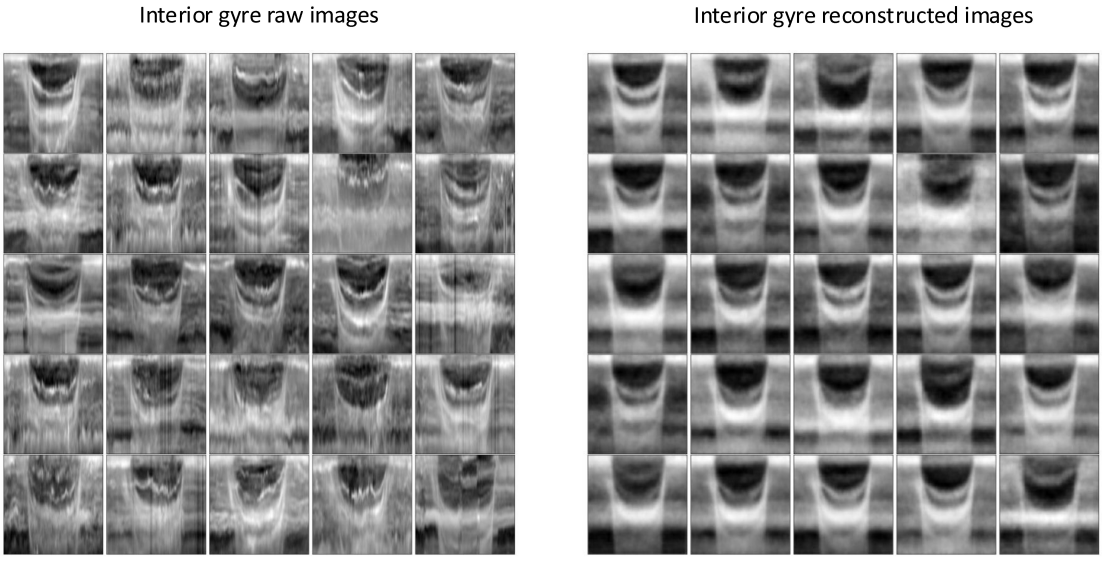

Figure A.20: Image mosaic of echograms assigned to Cluster 1 and corresponding CAE reconstructions.

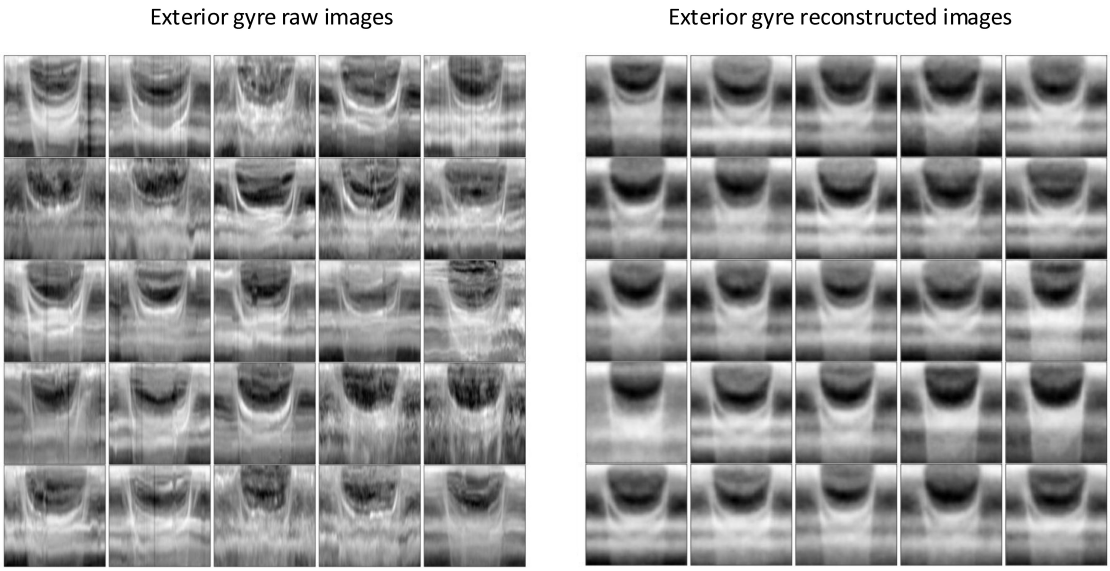

Figure A.21: Image mosaic of echograms assigned to Cluster 2 and corresponding CAE reconstructions.

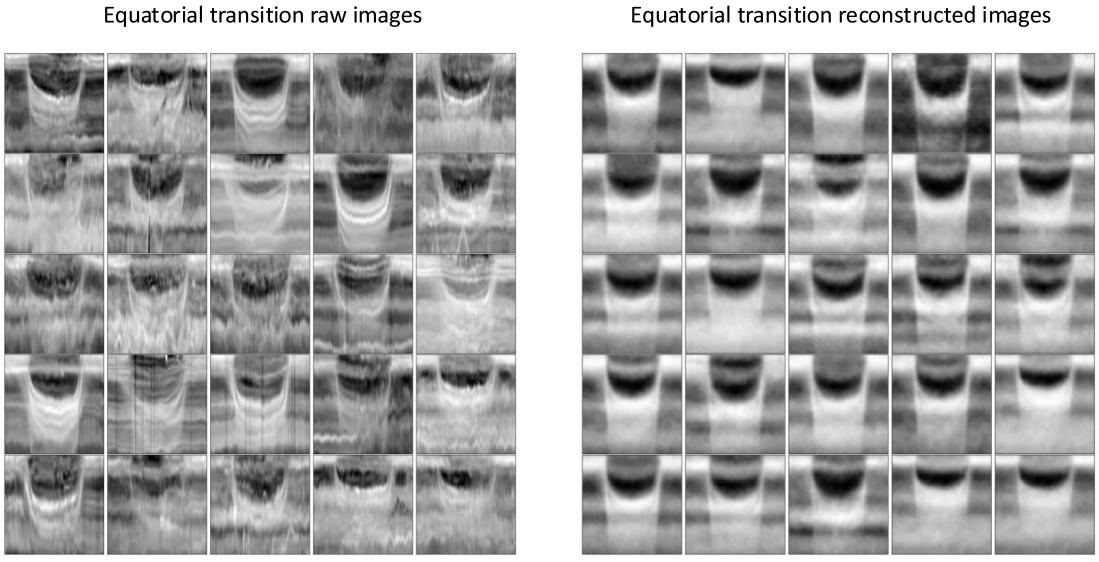

Figure A.22: Image mosaic of echograms assigned to Cluster 3 and corresponding CAE reconstructions.

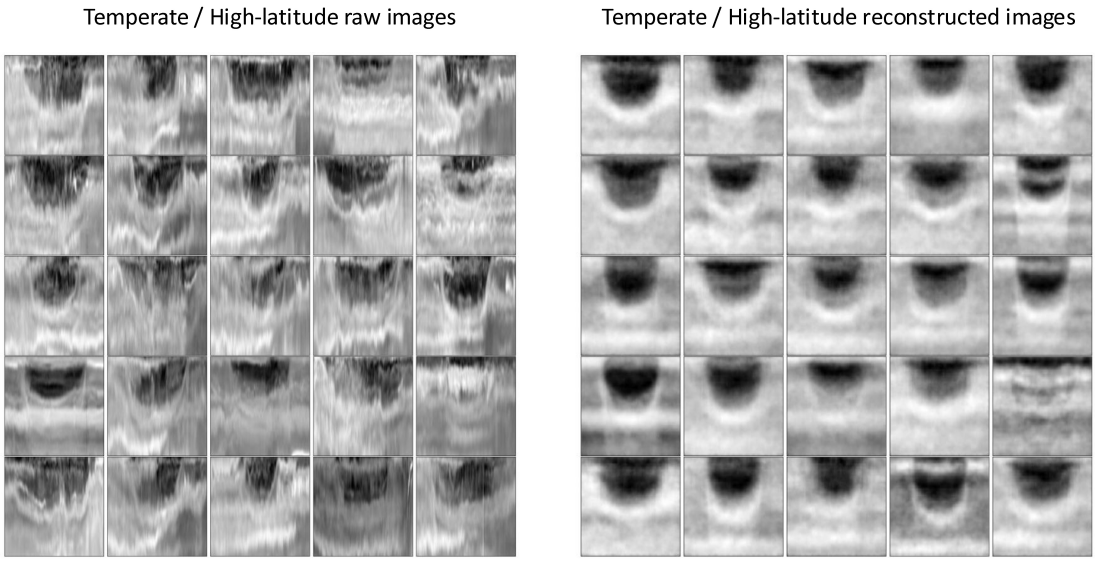

Figure A.23: Image mosaic of echograms assigned to Cluster 4 and corresponding CAE reconstructions.

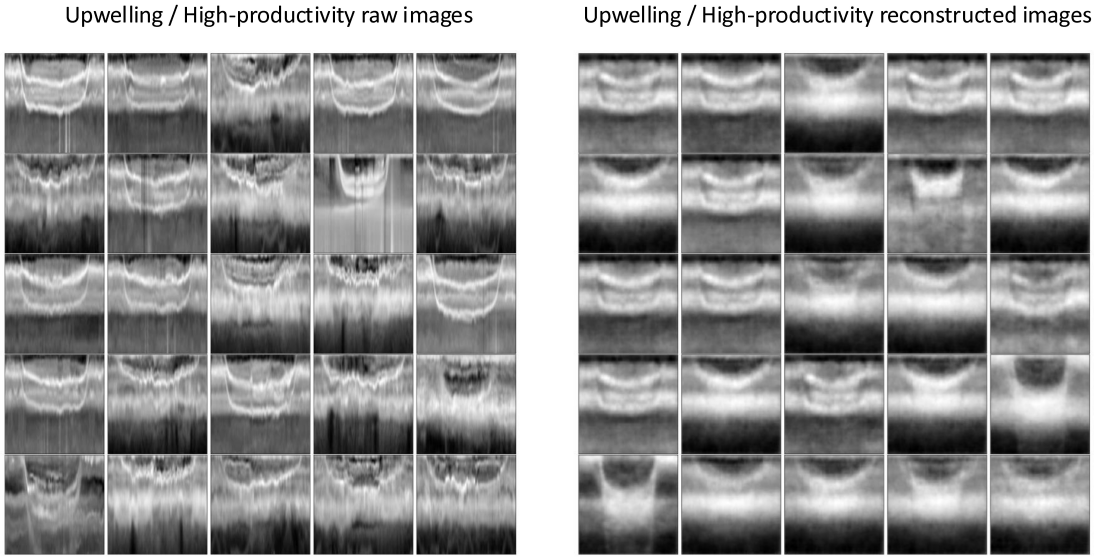

Figure A.24: Image mosaic of echograms assigned to Cluster 5 and corresponding CAE reconstructions.

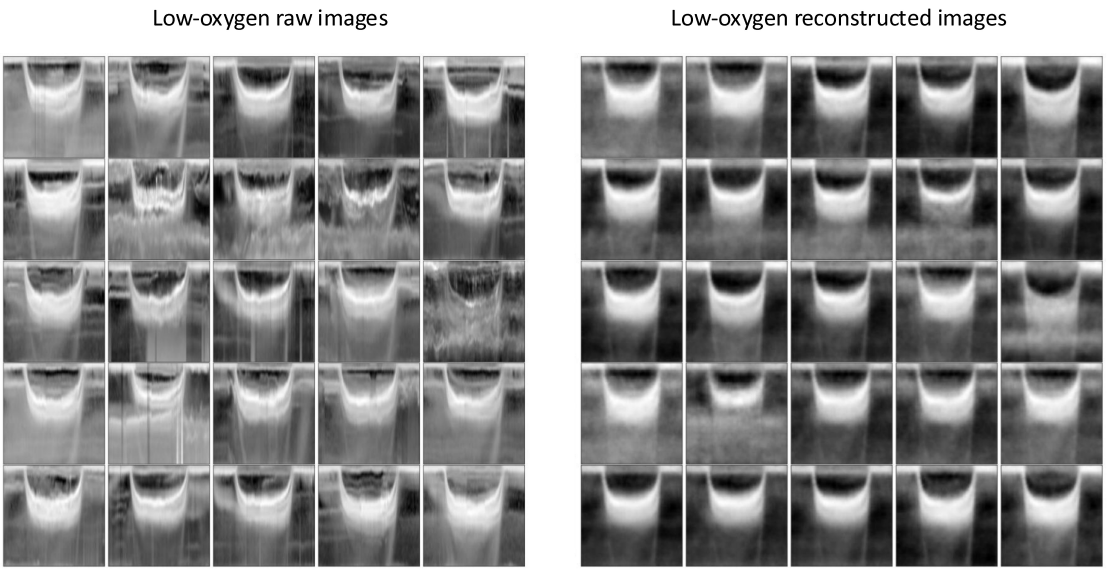

Figure A.25: Image mosaic of echograms assigned to Cluster 6 and corresponding CAE reconstructions.