Supplementary material

S1 Spectrogram computation parameters

The raw displacement data are pre-processed by bandpass filtering between $1/200$–100 Hz (corners: $[1/200, 1/100, 80, 100]$ Hz), applying a 5% taper, removing the instrument response (all with `obspy` defaults), and then resampling to 100 Hz. The data are then divided into 15 min windows, demeaned, and transformed into spectrograms using short-time Fourier transforms (`matplotlib.mlab.specgram`, defaulting to Hann windowing) from 1–50 Hz using 50 s windows with 75% overlap. Aside from window and overlap, we use the default settings of `matplotlib.mlab.specgram`, which returns in units of power / Hz. The resulting spectrograms are linearly interpolated to a target grid of dimension $(60, 399)$ and converted to decibel scale ($10 \log_{10}$ of power spectral density in m^2/Hz, i.e. dB [m^2/Hz]). Computation and processing of each 15 min spectrogram took around 2 s per core on a desktop class machine.

To suppress instrument noise floor and enhance visible signal bands, spectrogram amplitudes are then globally clipped to the range –150 dB to –250 dB [m^2/Hz] for training. The exact same pre-processing settings were used for all instruments analysed in the manuscript.

S2 Synthetic feature generation

To perform synthetic data pre-training we created two simple, interpretable stochastic generators that produce training targets and a matching binary mask for each synthetic feature. This procedure is demonstrated in Fig. S2. Both generators follow the same broad pattern: they use a template that captures the local morphology of the signal, and then they randomly vary and place copies of the template across the image. To produce these copies, the generators stochastically sample counts, positions and basic size/feature amplitude parameters. We present examples of individual sampled features following these templates in Fig. S2 i). The generators were tuned qualitatively (by visual comparison to annotated examples) so the goal in these descriptions is transparency about the modelling choices rather than an attempt at exact physical realism.

Resonance generator

Resonances are modelled as smoothly varying horizontal bands: each resonance is a one-dimensional, stochastically varying signal (modelled as a "random-walk") that is scaled and offset so it is visible above the background noise level. The procedure repeatedly samples how many resonances to draw and where to put them vertically (with a bias toward a preferred band between 5 to 22 Hz, which we noticed empirically), and for each chosen row it either places a full-width band or a local segment of variable horizontal extent. Small stacked clusters of adjacent rows are sometimes generated to mimic grouped resonances. In short: we produce a smoothly varying 1-D template, stochastically vary their amplitude and place them around the target image, before summing with the background noise spectrogram.

Blue whale call generator

Blue-whale calls are represented as small 2-D teardrop templates that are skewed and injected with noise to increase variability. The generator first decides whether to place a call pattern at all; when it does, a stochastically sampled number of small patches are tiled left-to-right with random horizontal spacing and small vertical jitter within the 17 Hz call band. Each patch's size, tail length, distortion and amplitude are randomly varied so that calls differ in shape and strength. To create more natural variability, we add a small chances of skipping a call or terminating the sequence early, and the whole pattern may be flipped horizontally. These sampled patches are then summed with the background noise spectrogram.

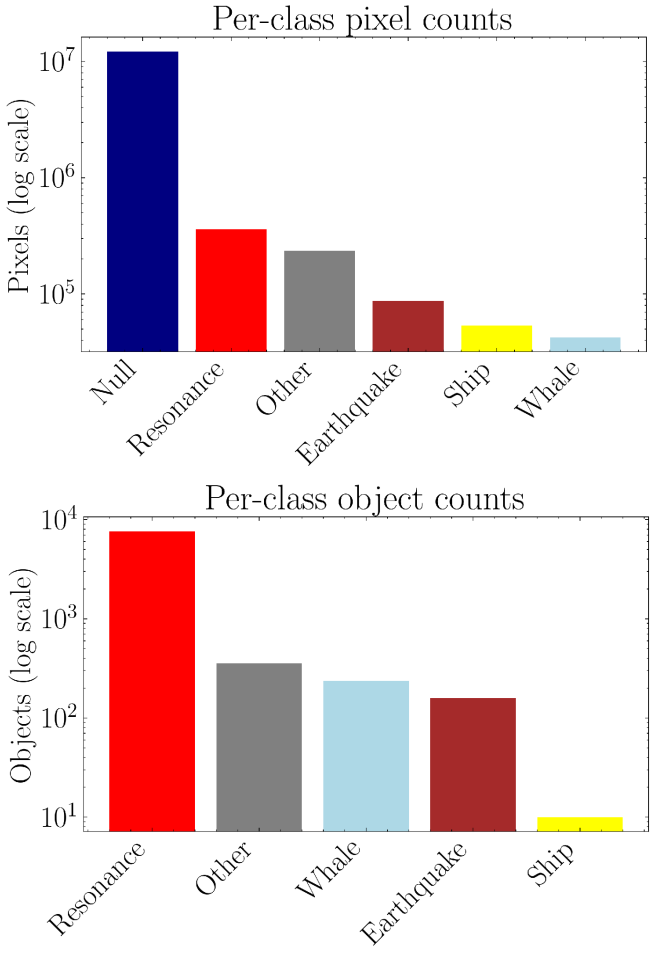

Figure S1 Per-class statistics from the manually annotated UP05 dataset. Top row: total number of pixels for each class, including the null class (pixels with no annotated features). Bottom row: number of bounding boxes annotated per class. This is not a perfect proxy for number of events (e.g. number of earthquakes), as occasionally multiple bounding boxes were used to annotate a single event with non-rectangular shape.

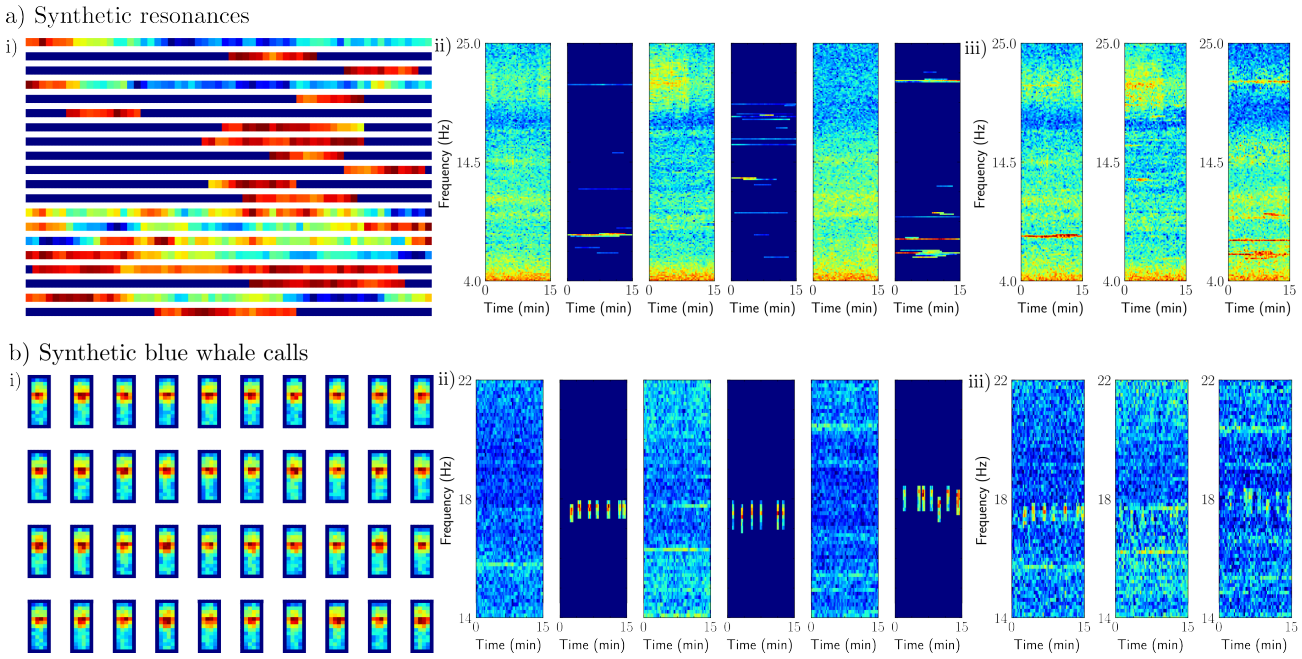

Figure S2 A visual demonstration of the synthetic feature generation procedure, broken down by instrument resonances in a) and blue whale vocalisations in b). For each feature class, we define a heuristic signal template. Panel i) shows several randomly generated examples following this template. In panel ii), we sample signal-free background spectrograms, as well as sampling from the templates according to some tuned distribution. The resulting spectrograms in iii) combine the signals and background spectrograms, mimicking realistic feature and noise distributions.

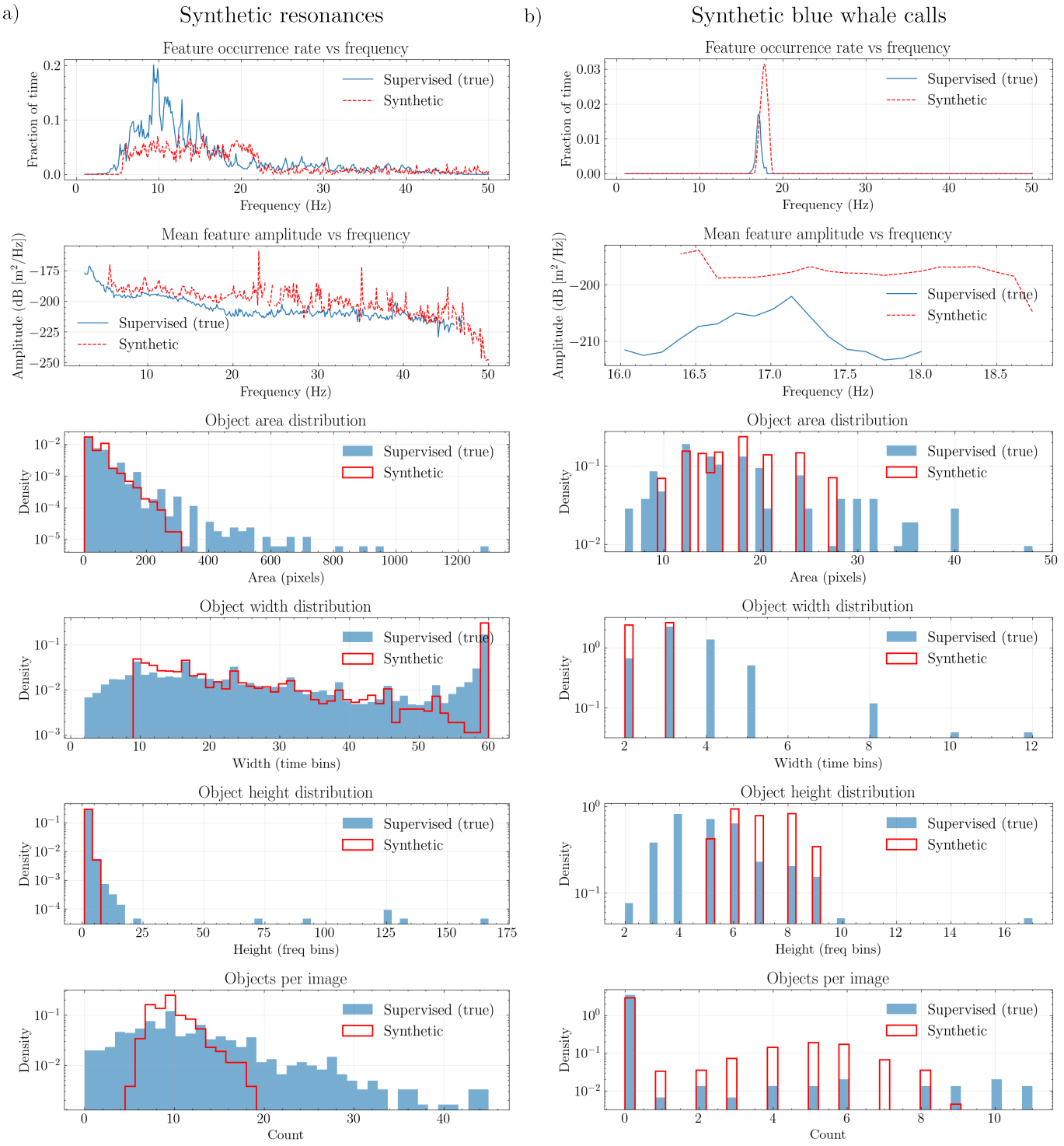

Figure S3 A comparison of the distribution of properties of manual annotations against synthetic features for the two signal classes: a) instrument resonances and b) blue whale vocalisations. The synthetic generators were designed heuristically and tuned qualitatively by eye, without rigorous optimisation. Despite noticeable differences between the distributions of synthetic and real features, note that pre-training on the synthetic features still yielded substantial performance gains in the segmentation models. This demonstrates that simple, heuristic generation of the features is sufficient to provide pre-training benefits.

Whale Distribution Animation

Figure S4 Animation of weekly blue whale call counts across the entire deployment. Upper panel: the total number of calls detected across the UPFLOW array per week. Lower panel: the total number of counts at each station, visualised by the bubble size at each station.

S3 Projected current velocity

The relationship between tidal currents and instrument resonances is likely to be a complex one. Here, we seek a simple, heuristic relationship between the modelled barotropic current from the data product, and the instrument resonances of the OBS. We constructed a projected velocity quantity v_{proj} for ocean current $\mathbf{u}$:

$$v_{\text{proj}} = \alpha\|\mathbf{u}\| + \mathbf{u} \cdot \hat{\mathbf{a}}(\varphi + \Delta\varphi), \tag{1}$$

where $\hat{\mathbf{a}}$ is the OBS station orientation unit vector for a given station azimuth φ and optimal driving direction offset $\Delta\varphi$. The optimal choices for the velocity magnitude scaling factor α and optimal driving direction offset $\Delta\varphi$ are unknown, so we perform a quick manual search that maximises correlation between the instrument resonances and v_{proj} across a subset of stations. The results shown in the main text used $\{\alpha = 0.2, \Delta\varphi = 67.5\,\text{deg}\}$.

This is a idealised model of the relationship between ocean currents and instrument resonances. The actual relationship is likely to be non-linear (Godin et al., 2024; Tan et al., 2025), with different instrument resonances activating at different current thresholds (Stähler et al., 2016; Corela, 2014). The effect of the angle of the current $\mathbf{u}$ will also vary for each vibrating component on the OBS given their differing positions on the instrument. Finally, in practice the barotropic current data product may be a very poor proxy for seafloor current, especially at stations with more complicated seafloor topography effects.

[1]The number of synthetically generated spectrograms seen during pre-training, compared to 40000 in the main text.

[2]Pre-training with a larger batch size of 128 as opposed to 10 used in the main text.

S4 Inter-annotator agreement

In order to validate the quality and consistency of the manual annotations during the annotation process, we annotated a shared subset of spectrograms from the UP05 dataset. We present the statistics of these overlapping annotations for the resonance class in Table ST4. We find good consistency between one pair of annotators (1 vs. 2; IoU $= 0.622$), and moderate consistency between the other (2 vs. 3; IoU $= 0.452$).

These results give an indication of an upper bound on possible ML model performance. However, the large difference between these IoU values demonstrate inconsistencies between the manual annotators. There are several factors driving this: we did not set out a set of rigorous standards or guidelines for annotation; there is some inherent ambiguity in what is classed as a resonance; and the fine-grained nature of the resonances, which made annotation quality dependent on the time and care invested by each annotator.

Note the best performing model performance in Table 1, with an IoU of 0.441. This approaches the lower bound of inter-annotator agreement (Annotator 2 vs. 3) but remains below the higher bound (Annotator 1 vs. 2). Thus, while the model does not yet reach the consistency of the best human pair, its performance is on par with the lower range of human annotator variability.

S5 Whale call detection accuracy

We performed further quantitative validation of the blue whale call detection accuracy in order to assess model performance. We focused on evaluating the false positive rate blue whale vocalisations. False negatives were common but highly dependent on call SNR, so we did not attempt a comprehensive analysis. Note that this evaluation is distinct from the approach in the main text, which focused on the accuracy of individual pixel segmentation accuracy as opposed to detecting whale call objects.

We manually evaluated the model performance (using the call detection approach detailed in the main text) for three separate datasets across the UPFLOW array. The datasets were: random sampling during the active vocalisation months, random sampling during the inactive months, and biased sampling for periods with two hours of continuous model call detections. For the random sampling datasets, two sets of 2 hour periods were evaluated per station. For the biased sampling, one 2 hour period per station was evaluated.

We assign false positives to four distinct classes: earthquakes, broadband transients, resonances, and unidentified signals. We introduce this broadband transient class here since it was a common source of false positives. The physical origin of these signals is unclear; they could be caused by very local microseismicity (where an earthquake-like signature could not be resolved at our time resolution), outgassing, fish bumps, or potentially high frequency, or potentially even aliased bioacoustic signals (e.g. sperm whale clicks) that could not be resolved at our sampling rate.

Table ST1 Performance on the UP05 validation set for both signal classes, broken down by neural network architecture. Model variations are broken down by the encoder and decoder architecture (bold row heads). Model performance for the parameters utilised in the main text are given at the top ('Main text'). The mean and standard error are computed over three independent training runs on the same train / validation split.

Signal class	Instrument resonances			Blue whale calls		
Architecture	IoU	Precision	Recall	IoU	Precision	Recall
Main text	0.420 ± 0.001	0.623 ± 0.031	0.572 ± 0.022	0.135 ± 0.032	0.229 ± 0.059	0.333 ± 0.099
Encoders						
ResNet50	0.419 ± 0.001	0.644 ± 0.008	0.545 ± 0.008	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
ResNet101	0.415 ± 0.002	0.598 ± 0.039	0.584 ± 0.038	0.072 ± 0.072	0.151 ± 0.151	0.097 ± 0.097
MIT-B1	0.411 ± 0.002	0.608 ± 0.019	0.561 ± 0.014	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
Decoders						
UNet++	0.409 ± 0.001	0.578 ± 0.015	0.584 ± 0.014	0.137 ± 0.069	0.244 ± 0.123	0.220 ± 0.116
DeepLabV3	0.055 ± 0.007	0.301 ± 0.009	0.063 ± 0.009	0.048 ± 0.048	0.055 ± 0.055	0.179 ± 0.179
Segformer	0.243 ± 0.002	0.408 ± 0.006	0.377 ± 0.010	0.075 ± 0.075	0.097 ± 0.097	0.167 ± 0.167

Table ST2 Model performance on the instrument resonance segmentation task, varying a range of different hyperparameters introduced in the main text. Hyperparameters are broken down by the training technique (bold row heads). Model performance for the parameters utilised in the main text are given at the top of each block ('Main text'). We report Intersection over Union (IoU) and recall for the three manually annotated datasets, with the mean and standard error computed over three independent training runs. Results are presented to greater precision when the standard error allows.

Dataset	UP05 Validation Set		UPFLOW UP34		RHUM-RUM RR40	
Metric	IoU	Recall	IoU	Recall	IoU	Recall
Varied hyperparameter						
Supervised						
Main text	0.420 ± 0.002	0.57 ± 0.02	0.49 ± 0.02	0.81 ± 0.02	0.52 ± 0.01	0.73 ± 0.04
Batch size 64	0.35 ± 0.01	0.47 ± 0.03	0.40 ± 0.02	0.64 ± 0.01	0.37 ± 0.06	0.58 ± 0.04
Class weight $w = 0.1$	0.32 ± 0.01	0.80 ± 0.01	0.27 ± 0.01	0.94 ± 0.01	0.36 ± 0.01	0.85 ± 0.01
Class weight $w = 1.0$	0.39 ± 0.01	0.46 ± 0.02	0.52 ± 0.01	0.65 ± 0.02	0.47 ± 0.02	0.56 ± 0.04
No ImageNet weights	0.42 ± 0.01	0.54 ± 0.02	0.52 ± 0.01	0.76 ± 0.02	0.48 ± 0.02	0.63 ± 0.06
ResNet50 encoder	0.42 ± 0.01	0.54 ± 0.03	0.53 ± 0.01	0.78 ± 0.03	0.50 ± 0.01	0.65 ± 0.04
U-net++ decoder	0.41 ± 0.01	0.61 ± 0.03	0.45 ± 0.02	0.81 ± 0.03	0.47 ± 0.01	0.65 ± 0.03
Multiclass						
Resonances & blue whales	0.410 ± 0.004	0.53 ± 0.03	0.49 ± 0.01	0.75 ± 0.03	0.48 ± 0.01	0.61 ± 0.04
Semi-supervised						
Main text	0.431 ± 0.002	0.59 ± 0.01	0.47 ± 0.01	0.85 ± 0.01	0.50 ± 0.02	0.72 ± 0.06
$\alpha = 0.9$	0.430 ± 0.002	0.56 ± 0.01	0.520 ± 0.002	0.80 ± 0.03	0.45 ± 0.01	0.58 ± 0.04
$\alpha = 0.999$	0.22 ± 0.06	0.23 ± 0.07	0.31 ± 0.06	0.33 ± 0.07	0.31 ± 0.04	0.32 ± 0.04
$\alpha = 0.995$	0.430 ± 0.001	0.61 ± 0.01	0.46 ± 0.01	0.84 ± 0.02	0.500 ± 0.002	0.72 ± 0.04
$\lambda_{\text{cons}} = 0.1$	0.43 ± 0.01	0.59 ± 0.02	0.49 ± 0.02	0.79 ± 0.03	0.45 ± 0.01	0.56 ± 0.01
$\lambda_{\text{cons}} = 2.0$	0.43 ± 0.01	0.58 ± 0.02	0.47 ± 0.01	0.83 ± 0.01	0.450 ± 0.003	0.55 ± 0.01
No λ_{cons} Ramp Up	0.430 ± 0.003	0.600 ± 0.002	0.47 ± 0.03	0.83 ± 0.01	0.46 ± 0.03	0.59 ± 0.04
Unlabelled batch size: 10	0.431 ± 0.005	0.57 ± 0.03	0.49 ± 0.01	0.78 ± 0.04	0.47 ± 0.02	0.58 ± 0.03
Unlabelled batch size: 200	0.422 ± 0.006	0.56 ± 0.02	0.48 ± 0.02	0.77 ± 0.02	0.47 ± 0.01	0.60 ± 0.03
Transfer						
Main text	0.441 ± 0.002	0.60 ± 0.01	0.53 ± 0.01	0.83 ± 0.01	0.49 ± 0.01	0.70 ± 0.02
20k spectrogram pretraining[1]	0.43 ± 0.01	0.58 ± 0.02	0.52 ± 0.01	0.80 ± 0.03	0.48 ± 0.01	0.68 ± 0.04
Pretraining batch size: 128 [2]	0.41 ± 0.01	0.57 ± 0.04	0.490 ± 0.003	0.75 ± 0.03	0.45 ± 0.01	0.59 ± 0.03
No ImageNet weights	0.440 ± 0.002	0.60 ± 0.01	0.52 ± 0.01	0.85 ± 0.01	0.490 ± 0.004	0.72 ± 0.02
No weight decay	0.430 ± 0.002	0.61 ± 0.01	0.51 ± 0.01	0.84 ± 0.01	0.460 ± 0.001	0.67 ± 0.02
Only weight decay pretraining	0.440 ± 0.001	0.61 ± 0.01	0.52 ± 0.01	0.83 ± 0.01	0.49 ± 0.01	0.70 ± 0.01

Table ST3 Mean and standard error for Intersection over Union (IoU), recall, and precision across bootstrap sizes and the main (50) setting. The reshuffled dataset column presents results with a reshuffled train-validation split. We present the standard error to higher precision as this is the main quantity we wish to probe here. The full UP05 validation dataset contains 100 spectrograms, while the UP34 and RR40 datasets contain 50 spectrograms each.

Dataset	Metric	Bootstrapped dataset size				
		10	20	30	Reshuffled dataset	Full dataset
UP05 Validation Set	IoU	0.44 ± 0.02	0.42 ± 0.01	0.44 ± 0.007	0.44 ± 0.0008	0.44 ± 0.001
	recall	0.59 ± 0.03	0.59 ± 0.02	0.60 ± 0.01	0.60 ± 0.009	0.60 ± 0.005
	precision	0.65 ± 0.02	0.60 ± 0.02	0.63 ± 0.01	0.62 ± 0.006	0.63 ± 0.006
UPFLOW UP34	IoU	0.53 ± 0.02	0.51 ± 0.008	0.53 ± 0.005	0.54 ± 0.006	0.54 ± 0.004
	recall	0.83 ± 0.009	0.84 ± 0.006	0.83 ± 0.004	0.83 ± 0.006	0.83 ± 0.004
	precision	0.59 ± 0.02	0.57 ± 0.009	0.59 ± 0.007	0.60 ± 0.005	0.60 ± 0.006
RHUM-RUM RR40	IoU	0.54 ± 0.03	0.45 ± 0.05	0.49 ± 0.02	0.49 ± 0.01	0.49 ± 0.005
	recall	0.69 ± 0.05	0.63 ± 0.05	0.76 ± 0.02	0.71 ± 0.004	0.71 ± 0.008
	precision	0.75 ± 0.03	0.61 ± 0.06	0.58 ± 0.03	0.62 ± 0.01	0.62 ± 0.009

Table ST4 Inter-annotator IoU for the instrument resonance class, along with the number of overlapping spectrograms.

	IoU	# shared spectrograms
Annotator 1 vs Annotator 2	0.622	109
Annotator 2 vs Annotator 3	0.452	106

We summarise our results in Table ST5. We also include an estimate of the false positive rate from the inactive months (April, May, June and July). We find our dataset for the inactive period has a false positive rate per station per week (5.9 ± 2.5) consistent with the overall detection rate in this period (4.4). This is a very low false positive rate and gives some indication of the strengths of the model.

On the other hand, the false positive rate for the randomly sampled active periods is significantly higher than during the inactive periods. One explanation for this finding is that the model learned the repeating nature of the whale vocalisations, but this led to overconfidence, classifying both calls and other signals as whale calls incorrectly. We found that broadband signals, such as earthquakes, were more likely to be misclassified as whale calls during periods of active whale vocalisation. Some examples of these false positives and the effect discussed here can be seen in the first two rows of Fig. S7. The biased sampling approach shown by "Highly active periods" provides further quantitative evidence of this phenomenon, with significantly higher false positive rates during periods with more active whale vocalisations. We therefore conclude that earthquakes and other broadband transient signals present a source of systematic false positives for the model.

Another potential explanation for the increased false positive rate is that of "confounding" factors leading to (somewhat) spurious correlations. It is possible that the model learned to rely on the statistics of the seasonally (and day-night) varying noise background to strengthen its performance. Nonetheless, the precision during the active period is still very high (0.98), indicating a robust model with the vast majority of predictions of high trustworthiness. This analysis therefore lends weight to the conclusions drawn in the main text.

S6 Further resonance analysis

Fig. S5 presents the raw, unfiltered time series that were used to compute the correlations in Fig. 8 (main manuscript). There are clear periods of strong correlation between the current product and the resonant energy, particularly towards the start of the deployment. It is not clear what factors change that lead to the correlation breaking down. Substantial further work is needed to build predictive models of the current from the resonances.

Fig. S6 shows the accumulated energy deposited across the deployment per frequency band for several UPFLOW stations near Madeira. We note that this frequency specific analysis, which is capable of characterising resonances in a physically interpretable way, is uniquely enabled by our pixel-level segmentation approach. There are clear peaks corresponding to narrowband resonances for all the stations analysed, though the amplitude of these peaks varies significantly between station. All the stations plotted here are DEPAS "LOBSTER" instrument types (Tsekhmistrenko et al., 2025). The substantially different resonance profiles across different stations requires further investigation. Nevertheless, there are also clear, shared peaks at certain stations.

Fig. S6 then repeats the analysis of Fig. 7 (main manuscript), computing the PSD of the resonant energy a narrow band around 7 Hz, highlighted in Fig. S6a). We found that this narrow band had a very strong relationship with the tidal periods, and the resonances in this band showed a very strong tidal peak. Similar to Fig. 7, we find that the resonance segmentation model is capable of producing a significant amplification of the tidal signal over the unprocessed background noise in this narrow band.

Importantly, we also computed the PSD of a number of other resonance bands, and found that these did not contain a strong tidal signal. This provides

Sampling approach	Precision	False positive rate (per station per week)	False positives by class (%)			
			Earthquake	Broadband transients	Resonances	Unclassified
Random (active period)	0.98 ± 0.01	15.6 ± 5.3	1.0%	0.5%	0.3%	0.4%
Random (inactive period)	-	5.9 ± 2.5	0%	0%	0%	100%
Highly active periods	0.93 ± 0.02	174 ± 45	1.2%	3.5%	0.2%	1.8%
Total detections (inactive period)	-	4.4	-	-	-	-

Table ST5 Results of the false positive rate analysis across the three manually validated datasets. Errors were estimated by bootstrapping the annotations. The "Random (inactive period)" dataset contained no true whale calls, and all the misidentified calls resulted from unidentified signals. We also report an estimate of the false positive rate using the total detections across the array during the inactive period (bottom row) as a reference point.

Table ST6 The total number of blue whale call detections (17 Hz) extracted by the ML-based algorithm across the UPFLOW deployment.

Station	Count	Station	Count	Station	Count
UP01	1926	UP17	373	UP33	11921
UP02	10235	UP18	19720	UP34	19639
UP03	2907	UP19	10520	UP35	13186
UP04	6268	OBS17	14463	UP36	4112
UP05	5161	OBS19	22152	UP37	34
UP06	23487	UP20	19911	UP38	956
UP07	12072	UP22	15244	UP39	6271
UP08	4887	UP24	30333	UP40	5173
UP09	900	UP25	27294	UP41	22355
UP11	8654	UP27	8695	UP42	8076
UP12	9868	UP28	1252	UP43	6558
UP13	5702	UP29	11431	UP45	3354
X25	19153	UP32	23674	UP46	3825
OBS12	15941	UP31	15369		
UP16	9989	UP30	9376		

evidence that the current regime induced by the tidal currents produces a resonance signature that is physically distinguishable from other regimes. It is therefore also possible that these other resonance bands could be connected to different current regimes. In order to progress, current-meter data are required to calibrate these relationships between currents and instrument resonances (Godin et al., 2024; Tan et al., 2025).

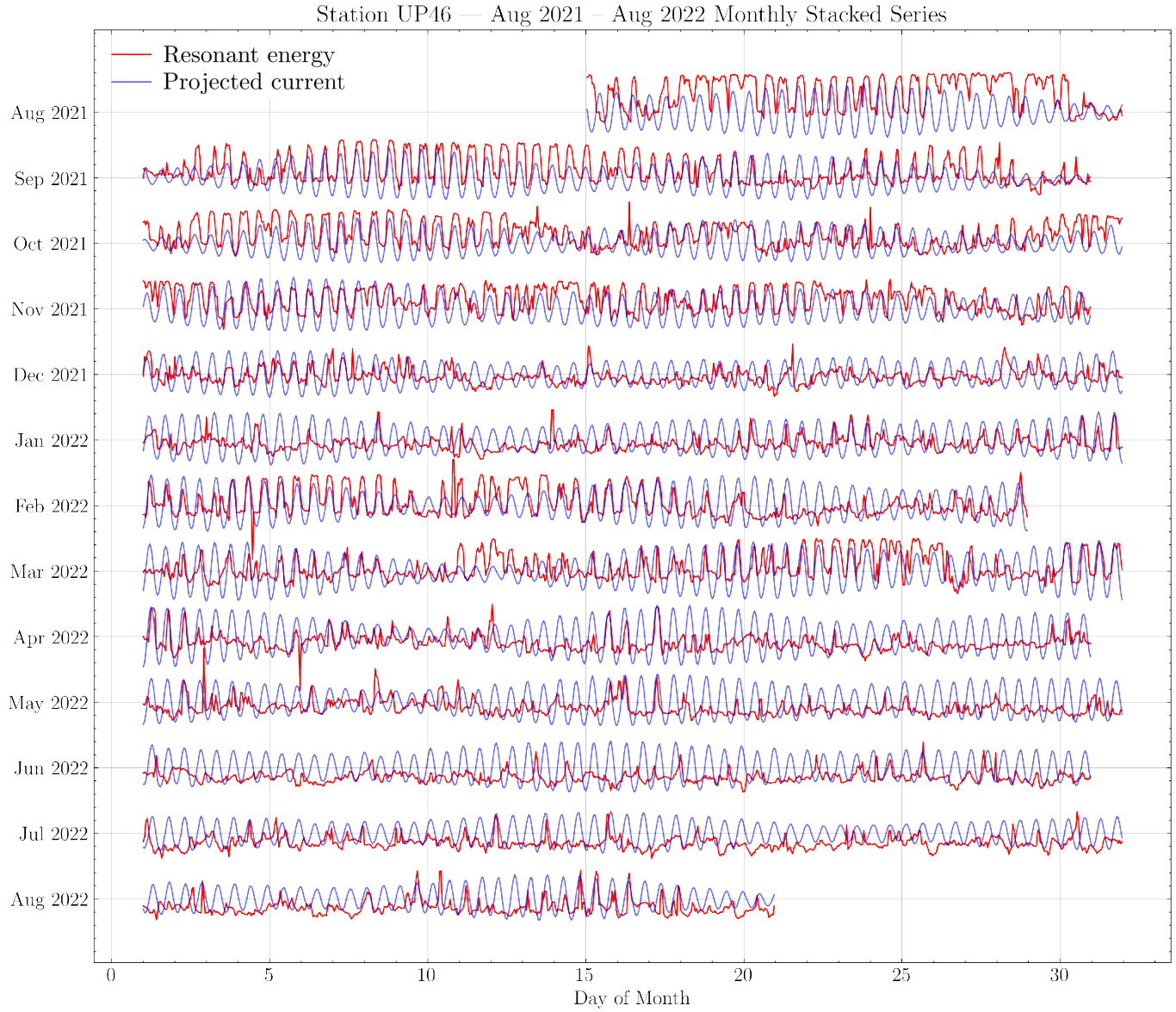

Figure S5 The full year-long time series of the extracted resonant energy (over all frequencies) against the projected current product for station UP46. Both series are sampled hourly, and neither series is filtered. The resonant energy is log-scaled (back to units of dB) and standard scaled by subtracting its global mean and dividing by its standard deviation. The currents series is locally demeaned (rolling window of 2 days) for visualisation purposes (this does not effect the overall correlation, but better aligns the series visually). Periods of strong correlation are particularly clear in Sep, Oct, and Nov 2021.

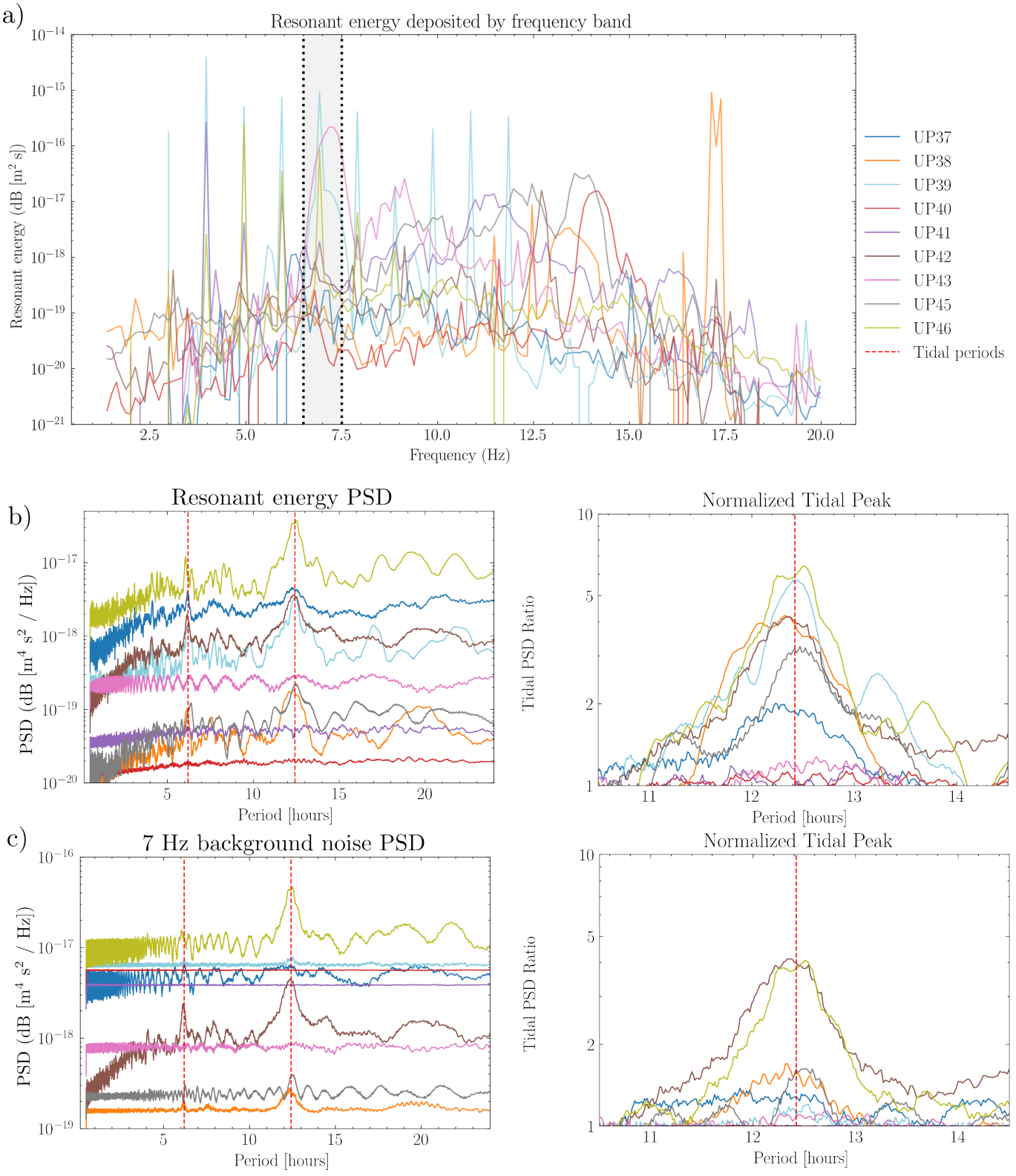

Figure S6 Further analysis of the estimated resonant energy for a subset of instruments near Madeira Island. a) shows the mean resonant energy deposited in each frequency bin between 1 - 20 Hz across the entire deployment, with clear spikes corresponding to strong, reoccurring resonances. b) shows the PSD and tidal peak as in Fig. 7 in the main text, but this time computed using resonant energy for a narrow, tidally driven resonance band around 7 Hz highlighted in a). c) repeats this procedure for the mean energy deposited in this frequency band without differentiating between resonances and background noise.

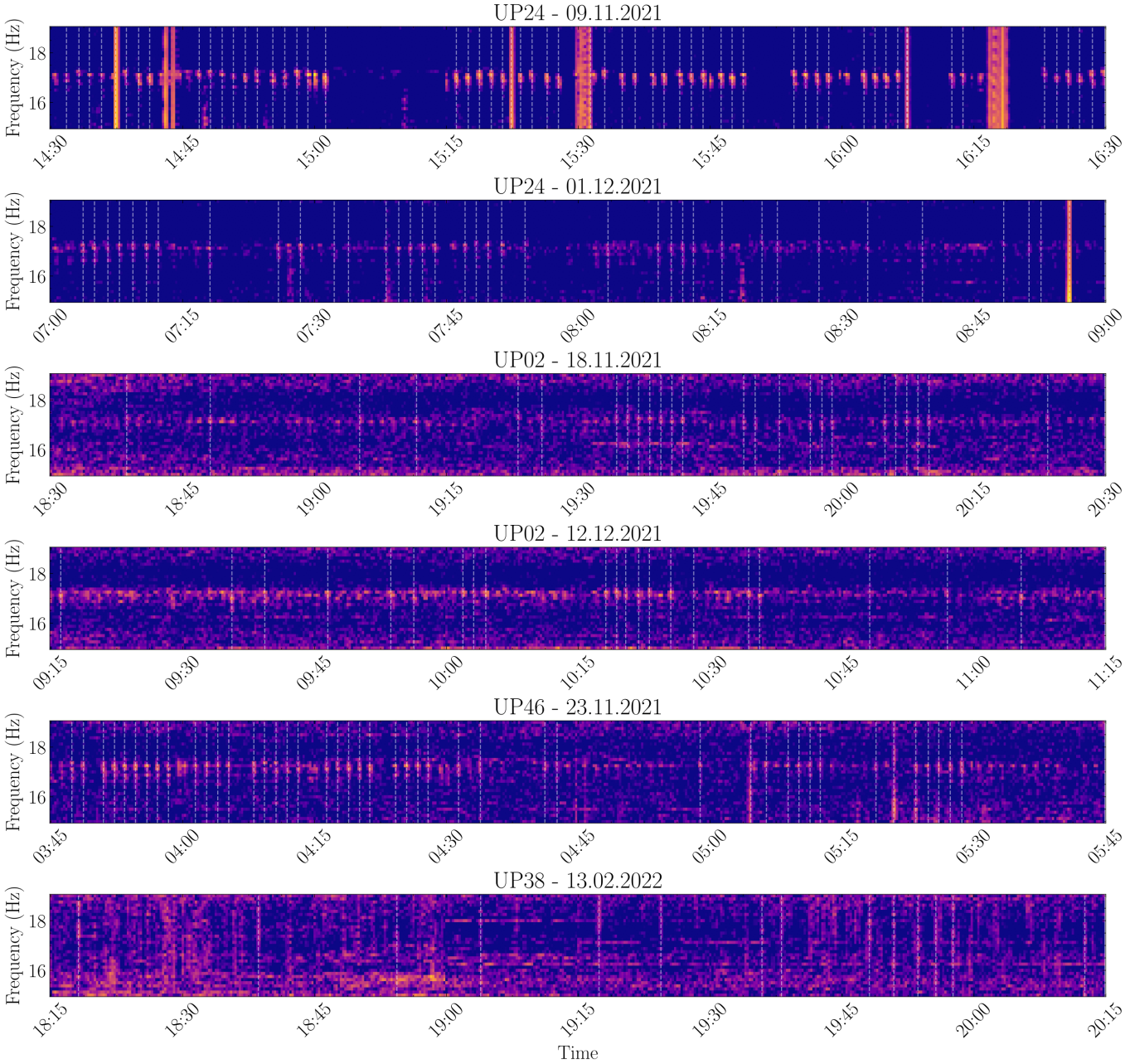

Figure S7 Highly active two hour spectrograms from across the UPFLOW array, alongside the ML annotations. The centres of the segmented whale call detection masks (exceeding the 6 pixel threshold) are shown as a vertical dashed white line. The UP38 (bottom row) example was selected to demonstrate the issues at this station, where a high false positive rate was caused by rare impulsive signals potentially related to nearby outgassing.

References

Corela, C. *Ocean bottom seismic noise : applications for the crust knowledge, interaction ocean-atmosphere and instrumental behaviour*. Phd thesis, Faculdade de Ciências da Universidade de Lisboa, 2014. http://hdl.handle.net/10451/15805.

Godin, O. A., Tan, T. W., Joseph, J. E., and Walters, M. W. Observation of exceptionally strong near-bottom flows over the Atlantis II Seamounts in the northwest Atlantic. *Scientific Reports*, 14(1): 10308, 2024. doi: 10.1038/s41598-024-60528-2.

Stähler, S. C., Sigloch, K., Hosseini, K., Crawford, W. C., Barruol, G., Schmidt-Aursch, M. C., Tsekhmistrenko, M., Scholz, J.-R., Mazzullo, A., and Deen, M. Performance report of the RHUM-RUM ocean bottom seismometer network around La Réunion, western Indian Ocean. *Advances in Geosciences*, 41:43–63, 2016.

Tan, T., Godin, O. A., Walters, M. W., and Joseph, J. E. Physics-informed and machine learning-enabled retrieval of ocean current speed from flow noise. *The Journal of the Acoustical Society of America*, 157(2):1084–1096, 2025.

Tsekhmistrenko, M., Ferreira, A. M., Miranda, M., Baranbooei, S., Cabieces Diaz, R., Carapuço, M., Corela, C., Duarte, J. L., Ferreira, H., Geissler, W. H., Harris, K., Hicks, S. P., Hosseini, K., Ke, K.-Y., Krüger, F., Lange, D., Loureiro, A., Makus, P., Marignier, A., Neres, M., Ramos, L., Rein, T., Saoulis, A., Schlaphorst, D., Schmidt-Aursch, M. C., and Tilmann, F. Performance of the 2021-2022 UPFLOW large ocean bottom seismometer array in the Azores-Madeira-Canary Islands region, Atlantic Ocean. *Seismica*, 2025.