Coral reef soundscapes analysis benefits from combining multiple ecoacoustic indices using machine learning

Authors

Ben Williams[a], Timothy A. C. Gordon[a,b], Lucille Chapuis[a], Harry R. Harding[c], Eleanor B. May[a], Mochyudho E. Prasetya[b], Andrew N. Radford[c], Stephen D. Simpson[a]

Affiliations

[a] Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4PS, United Kingdom;

[b] Mars, Inc., 6885 Elm St., McLean, VA 22101, USA

[c] School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, United Kingdom

[d] School of Health and Life Sciences, University of the West of Scotland, PA1 2BE, UK

[e] Graduate School, Hasanuddin University, 90245 Makassar, Indonesia

[f] Mars and Coral Reef Research Unit, School of Life Sciences, University of Essex, Colchester, Essex, CO3 4SQ, UK

Corresponding author

Ben Williams

bw339@exeter.ac.uk

B10 Hatherly Labs, Prince of Wales Rd, Exeter, Devon, EX4 6PU.

Abstract

Passive acoustic monitoring (PAM) is a promising new tool used to monitor tropical reef habitats. Ecoacoustic indices are an increasingly popular approach used to rapidly analyse soundscape data. However, previous investigations have primarily used individual indices in isolation to assess coral reefs, with mixed success. This investigation combines ecoacoustic indices using machine learning and demonstrates a much improved ability to generate meaningful predictions about coral reef health. PAM data collected at one of the world's largest tropical reef restoration programmes in South Sulawesi, Indonesia, was used. Multiple one-minute recordings were taken on healthy and degraded sites with 90–95% and 0–20% of measured coral cover respectively. Twelve ecoacoustic indices were calculated for each recording, in up to three different frequency bandwidths (low: 0.05–0.8 kHz, medium: 2–7 kHz and full: 0.05–20 kHz) for each recording, totalling 33 values. Fifteen of these reported a significant difference between healthy and degraded habitats. However, high variability in the distribution of results was observed, offering a limited ability for any single index to discriminate between these two habitats without extensive sampling. These indices also exhibited little to no correlation with the number of audible fish vocalisations present in recordings. Regularised discriminant analysis, a machine learning approach, was then used to better discriminate between these two habitat classes using an optimised set of ecoacoustic indices in combination. This multi-index approach discriminated between healthy and degraded sites with a much-improved accuracy than any single index in isolation. The pooled classification rate of 1000 cross-validated iterations of the model had a 91.73 % ($\pm$ 0.84) success rate. Additionally, this classification was robust to changes in the diel and lunar cycle. We then report on the success of the model to classify recordings from three artificially restored sites. This investigation presents a novel approach to perform habitat assessments using short snapshot recordings. It also demonstrates the utility of PAM to monitor reef recovery over time, reducing the need for labour intensive in-water surveys.

Keywords

Soundscape, passive acoustic monitoring, ecoacoustic index, machine learning, coral reef, restoration, bioacoustics.

1. Introduction

Monitoring of tropical reef habitats primarily relies upon visual based in-situ dive and/or camera surveys which can be logistically complicated, expensive and only capture a subset of the ecological community. Passive acoustic monitoring (PAM) is an emerging practice used to monitor habitats which has the potential to overcome many of these limitations (Lindseth and Lobel, 2018; Mooney et al., 2020). Recent progress has been driven by improvements to acoustic recorder technology, which have created the capacity to capture long-term soundscape recordings using autonomous hydrophones (Sousa-Lima et al., 2013). Although, this field is still in its infancy, numerous studies have found relationships between the soundscapes of tropical coral reefs and traditional ecological metrics such as coral cover, biological communities and overall habitat quality (Bertucci et al., 2016; Butler et al., 2016; Elise et al., 2019; Freeman and Freeman, 2016; Gordon et al., 2018; Nedelec et al., 2015). Furthermore, soundscapes are known to be important components of a reefs functioning outright, especially for orientation and recruitment of reef associated organisms (Gordon et al., 2019; Lecchini et al., 2018; Simpson et al., 2005).

A number of analytical approaches have been particularly influential in recent developments in marine soundscape ecology. First, analysis can be performed using auditory or visual inspection; investigators listen to recordings or visually examine spectrograms, manually noting points of interest such as the frequency of occurrence of certain acoustic events or diversity of fish chorusing (Archer et al., 2018; Bertucci et al., 2020b; Carriço et al., 2020; McWilliam et al., 2018, 2017; Putland et al., 2017). However, these approaches can be slow and labour intensive, introducing a severe limit on the speed at which PAM data can be analysed.

Computationally generated ecoacoustic indices are a popular approach used to overcome this limitation (Bradfer-Lawrence et al., 2019; Gibb et al., 2019). These indices have primarily been developed for terrestrial soundscape recordings (Sueur et al., 2014), and are designed to quantify particular attributes of soundscapes such as their variability across time or frequency bands (Stowell and Sueur, 2020). The key advantage of these indices is their ability to rapidly process large amounts of acoustic data, enabling significantly longer periods of recordings to be analysed. A number of indices have been trialled in the marine environment where investigations have found relationships between certain indices and elements of the ecological community, habitat quality or ecological functioning of reef habitats (Elise et al., 2019b; Gordon et al., 2018; Harris et al., 2016; Lindseth and Lobel, 2018; Mooney et al., 2020). However, indices do not offer a perfect fix, often these do not correlate with specific elements of reef ecology in the same way as they did in other investigations (Bertucci et al., 2016b; Dimoff et al., 2021; Kaplan et al., 2015).

To date, studies of marine soundscape have primarily used individual index values in isolation when testing for statistical differences between experimental groups (e.g low and high habitat quality, pre and post bleaching), or relationships with other ecological parameters (e.g species diversity, abundance). However, the results of any one index have been shown to often be overdriven by individual components of the soundscape such as close by snapping shrimps, or a repetitive fish chorusing, limiting their utility to assess the full community (Bolgan et al., 2018; Dimoff et al., 2021; Staaterman et al., 2013).

Recent studies in terrestrial soundscape ecology suggest that combining multiple indices in a ‘compound index’ design is a potential solution to this problem (Bradfer-Lawrence et al., 2019; Eldridge et al., 2018; Gibb et al., 2019; Sethi et al., 2020). This allows the construction of more complex analytical models, using machine learning or artificial intelligence. Such models allow the identification of patterns in large multivariate datasets, thus providing a more holistic approach to analysing soundscapes and offering an increased ability to identify trends and relationships between soundscapes and ecological attributes of interest.

In this study, we developed a machine learning approach that generates a compound index to predict reef health from short-term soundscape recordings. We used recordings of healthy, degraded and restored reefs, all taken at or nearby one of the world’s largest reef restoration projects. To compare, we calculated a range of individual indices between healthy and degraded sites. We also searched for relationships between these indices and fish produced sound diversity to determine the degree to which this drove index results. We then applied a discriminant analysis machine learning algorithm to an optimised set of these indices. We tested whether the compound machine learning approach delivered an improved discriminatory power between habitat classes, relative to any single index. Finally, we applied this model to recordings taken from restored sites, to test the potential of this rapid analytical approach in assessing the progress of reef restoration.

2. Methods

2.1 Study site

Recordings were taken from seven sites in the Spermonde Archipelago (South Sulawesi, Central Indonesia; 4°56.9′S, 119°18.1′E; Fig. 1.1A) around Pulau Badi (Fig. 1B) and Pulau Bontosua (Fig. 1C). An ongoing restoration project has been using a novel methodology to re-establish coral cover at sites which have been degraded by coral mining and persistent destructive dynamite fishing (Williams et al., 2019). Coral fragments are attached to ‘reef stars’ which stabilise the substrate using interlinked metal frames. Between 2013 and 2017 this increased coral cover from approximately 10% to 60% on 7000m^2 reef (Williams et al., 2019). Recordings were taken at seven sites which encompassed four distinct types of habitat (Fig. 2), these sites were: Healthy A & B, Degraded A & B, Mature Restored A & B, and Newly Restored (one site only). We measured coral cover as an assessment of the sites’ health (methodology details provided in Supp. Material 1). The two healthy sites exhibited naturally high coral cover (A: 91.2% ± 2.0; B: 93.1% ± 2.6; mean ± SE) whereas the degraded sites exhibited low coral cover (A: 2.1% ± 0.9; B: 17.6% ± 4.6). The two mature restored sites were established >24 months previously and exhibited an increased coral cover (A: 79.1% ± 3.9; B: 66.5% ± 3.8) compared to the newly restored site (25.6% ± 2.6) established <12 months previously.

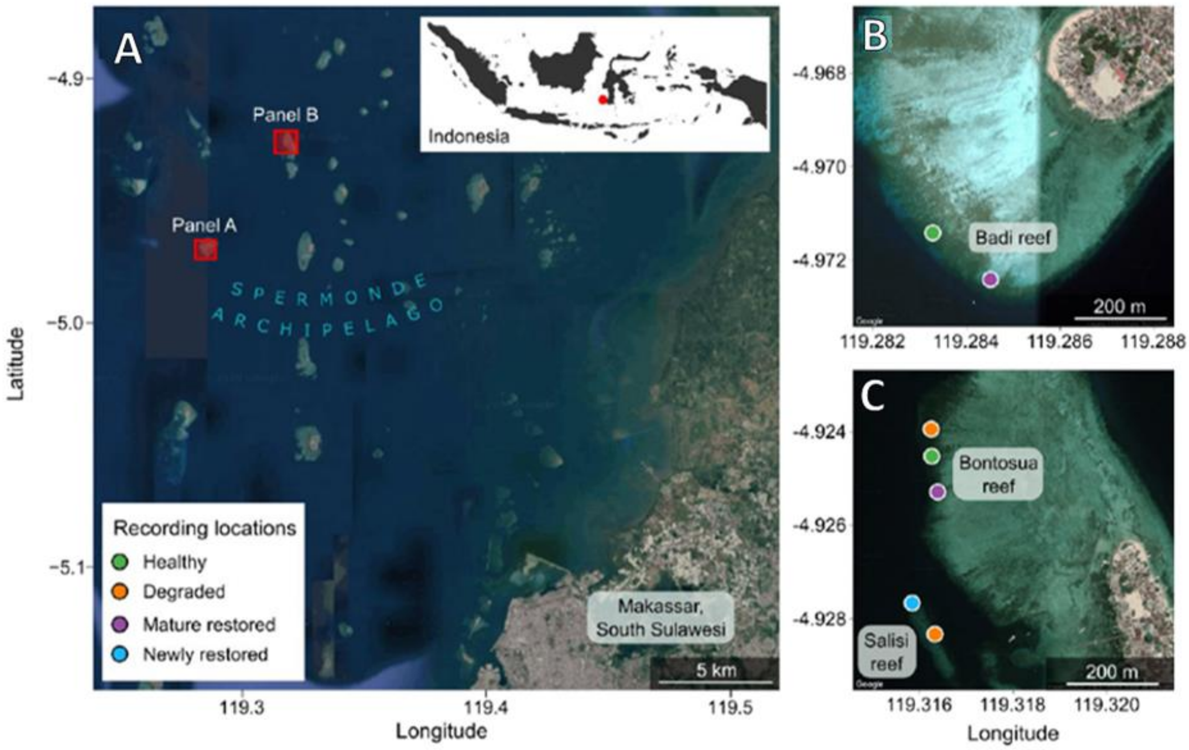


Fig. 1. Location and habitat class of the seven reef sites, present within the broader Spermonde Archipelago (A) where soundscape recordings were collected. Fringing reefs from two nearby islands: Badi (B) and Bontosua (C) were used.

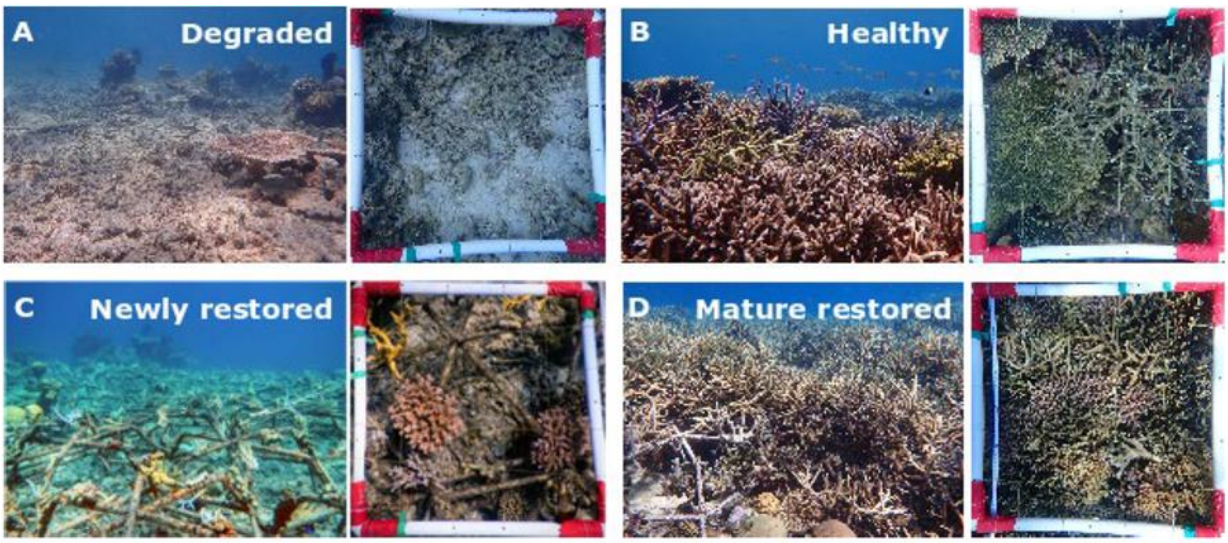


Fig. 2. Representative habitat and coral cover images from the four habitat classes at which soundscape recordings were taken. (A) Degraded, (B) healthy, (C) newly restored and (D) mature restored.

2.2 Data collection

Across the seven sites, 262 one-minute soundscape recordings were taken using SoundTrap hydrophones (SoundTrap 300 STD, Ocean Instruments, Auckland, NZ) in August-September 2018. SoundTraps were suspended 0.5 m above the seabed and set to record at a sampling rate of 48 kHz. The recordings were collected using a regime which sampled sites for five days either side of the full moon (August 26th) and three days either side of the following new moon (September 10th) in 2018 during daylight (09:00–15:00), twilight (half an hour either side of sunrise and sunset) and night time (half an hour either side of midnight) periods. These recordings were taken as part of the monitoring programme for the Mars Coral Reef Restoration Project at Badi and Bontosua Islands. We sub-sampled five non-overlapping one-minute segments from each of the hour-long periods at random, resulting in 262 samples. Only samples which were recorded under calm conditions (wind speed <20 km h^{-1}) and which contained no anthropogenic noise were included in the sample set. Three SoundTraps were rotated between sites, in an approximately even spread between each period.

2.3 Processing recordings

Each of the 262 one-minute recordings was band-pass filtered using a short-term Fourier transform filter into three frequency bands: a low-frequency band (0.05–0.8 kHz), a medium-frequency band (2 kHz–7 kHz) and a full-range band (0.05–20 kHz). The low-frequency band was selected to cover the frequencies of a range of known fish vocalisations, and the medium-frequency band was selected to encompass invertebrate sound (Elise et al., 2019a). The additional full-range frequency band encompassed the full spectrum of potentially relevant frequencies, as previously used in coral reef soundscape investigations (Kaplan *et al.*, 2015; Lyon, 2018). Frequencies below 0.05 kHz were excluded from low- and full-frequency band recordings to reduce the presence of shipping noise and geophonic noise from waves (Curtis *et al*., 1999). A new audio file for every recording in each frequency band was written to produce tracks filtered using a uniform method for subsequent analysis. All processing was performed in *R* (v3.4.2. R Development Core Team, 2020): audio files were read and written using the *tuneR* (v.1.3.3) package (Ligges *et al*., 2018) and the filter was implemented using *Seewave* (v2.1.6) (Sueur et al., 2008).

2.4 Calculating ecoacoustic indices

Twelve ecoacoustic indices chosen from a range of soundscape studies in the literature were used (Table 1). Each index was calculated for all three frequency bands, with two exceptions; Snap rate was only calculated for the middle and full frequency bands, because snapping shrimp cavitation bubbles are not audible at lower frequencies (Bohnenstiehl *et al*., 2016), and the normalised difference soundscape index (NDSI) which is typically used to quantify discrepancies in amplitude between an anthropogenic noise band up to 1 kHz and a biophonic noise band at selected higher frequencies (Kasten et al., 2012). For the first time, this index was implemented in the marine environment to instead quantify differences in the 1 kHz band where fish noise dominates, and higher frequencies where snapping shrimp sound is at its highest intensity (Au and Banks, 1998). This was therefore implemented on the full band recordings alone, to capture both the fish and shrimp bands. This results in the creation of a feature set of 33 index values across twelve indices and three frequency bands for each of the 262 one minute recordings. All indices were calculated using the *R* package *Seewave* (Sueur et al., 2008) where possible and *Soundecology* (v.1.3.3) (Villanueva-Rivera et al., 2018) for remaining indices.

Table 1. The twelve ecoacoustic indices calculated from recordings with summary description of the mechanistic principle, software used and respective settings employed.

Index	Mechanism	Software	Settings	Origin
Acoustic Complexity Index (ACI)	Measures variability in intensity of frequencies across time	*Seewave* in *R*	Window size = 512; type = Hamming; overlap = 0	(Pieretti, 2011)
Acoustic Entropy (H)	Measures randomness across temporal and spectral domains	*Seewave* in R	Window size = 512; envelope = Hilbert	(Sueur, 2008)
Acoustic Eveness Index (AEI)	Measures diversity across frequency bands	*Soundecology* in *R*	Max freq = audio tracks maximum; freq step = max freq/10; threshold = -50 dB	(Villanueva-Rivera, 2011)
Amplitude Index (M)	Measures median of amplitude envelope	*Seewave* in *R*	Envelope = Hilbert	(Sueur, 2008)
Acoustic Richness (AR)	Ranks recordings based on amplitude multiplied by randomness across the temporal domain	*Seewave* in *R*	Envelope = Hilbert	(Depraetere, 2012)
Bioacoustic Index (BI)	Measures cumulative intensity across frequency bands	*Soundecology* in *R*	Min and max frequency matched to track as appropriate; window size = 512	(Boelman, 2007)
Normalised mean difference index (NDSI)	Measures amplitude difference between two selected frequency bands	*Seewave* in *R*	Min and max frequency matched to track as appropriate; window size = 512	(Kasten, 2012)
Number of peaks	Number of major frequency peaks on obtained from a mean spectrum	*Seewave* in *R*	Window size = 512; type = Hanning; overlap = 0	(Sueur, 2008)
Spectral entropy (sh)	Measures randomness across the frequency domain	*Seewave* in *R*	No settings required	(Sueur, 2008)
Temporal Entropy (th)	Measures randomness across the temporal domain	*Seewave* in *R*	No settings required	(Sueur, 2008)

Snap Rate	Measures rate of snapping shrimp snaps	*MATLAB*	Custom script	Widely used
Sound Pressure Level (SPL)	Calibrated measure of root mean squared sound pressure level	*paPAM* in *MATLAB*	Window length = 1024; type = Hamming; Overlap = 50%	Widely used

2.5 Comparison of indices to frequency of fish vocalisations

In order to compare whether indices correlated with fish vocalisation activity, manual counts of different fish call types were obtained for each recording. The number of different fish calls present in each recording was quantified by an experienced experimentally-blind observer (T.A.C.G.). To ensure consistency in reporting, a subset of 20 tracks were listened to twice (experimentally blind to the first scoring when listening for the second time), with the same results each time. The number of unique fish calls observed in each recording was called 'phonic richness'. We tested the relationships between phonic richness and the full set of index results from all 262 recordings using Pearson's correlation tests.

2.6 Comparison of indices between healthy and degraded sites

The results for each index from the healthy and degraded habitat classes were compared against each other. The presence of a difference between the values of each of the 33 indices from recordings of healthy habitats (n = 81) and degraded habitats (n = 71) were tested for using a Mann-Whitney U test. Violin plots were also used to visualise the level of overlap between the distribution of these results for any index which reported a significant difference. If minimal overlap was observed between the two classes for any index, then the respective index would likely provide a promising measure with which to differentiate between healthy and degraded habitats.

2.6 Applying machine learning to create a compound index

Following individual index analysis, we developed a supervised machine learning model which could be used to accurately assign recordings to either healthy or degraded habitat classes. A regularized discriminant analysis (RDA) algorithm was selected to account for the high level of collinearity reported between indices (Supp. 1).

An optimised set of indices was selected in a ‘feature selection’ stage, using recursive feature elimination (RFE) and a multivariate adaptive regression spline (MAR) (Kuhn and Johnson, 2019) (Supp. 1). The RFE revealed the increases in model accuracy when using a multi-index approach as additional indices were sequentially added (Fig. 3). Starting with the most informative indices, predictive accuracy increased until a peak at eight indices was reached. This was followed by a decline as further addition of indices introduced noise to the data and/or overtraining occurred. The list of suggested features from the RFE included the following index/frequency band combinations: full-frequency band ACI, H, NDSI and th; and medium-frequency band ACI, BI, H and th. This was highly congruent with rankings obtained from the relative importance scores using the MAR (Fig. 4).

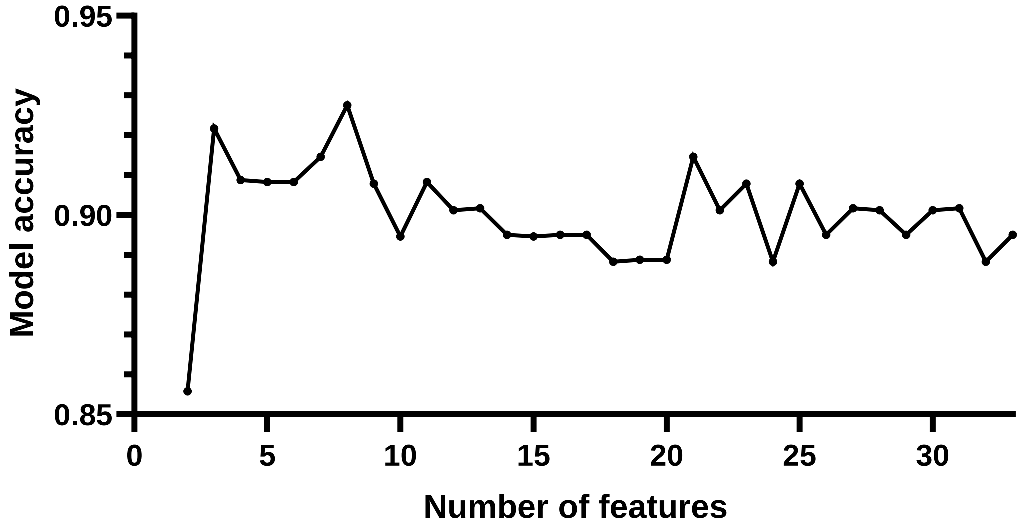


Fig. 3. Results from the recursive feature elimination performed using 1000 repeats of the discriminant analysis algorithm using k-fold cross validation with 10 folds (see methods 2.5). As additional indices are added to the model, the accuracy of the model is indicated on the y-axis until all 33 indices have been included.

From here, further manual feature selection through removal and addition of indices one by one whilst executing the full model (outlined in section 2.7) was used to select a final feature set with the lowest misclassification rate that would constitute the compound index used. This led to the discarding of th in both the full and middle-frequency bands and introduction of low-frequency band ACI and middle-frequency band AR into the final set: low-frequency band ACI, medium-frequency band ACI, AR and BI, full-frequency band ACI, H and NDSI. Feature selection was performed using the *R* packages *mlbench* (v2.1.1) (Leisch and Dimitriadou, 2010) and *Caret* (v.6.0-86) (Kuhn, 2020).

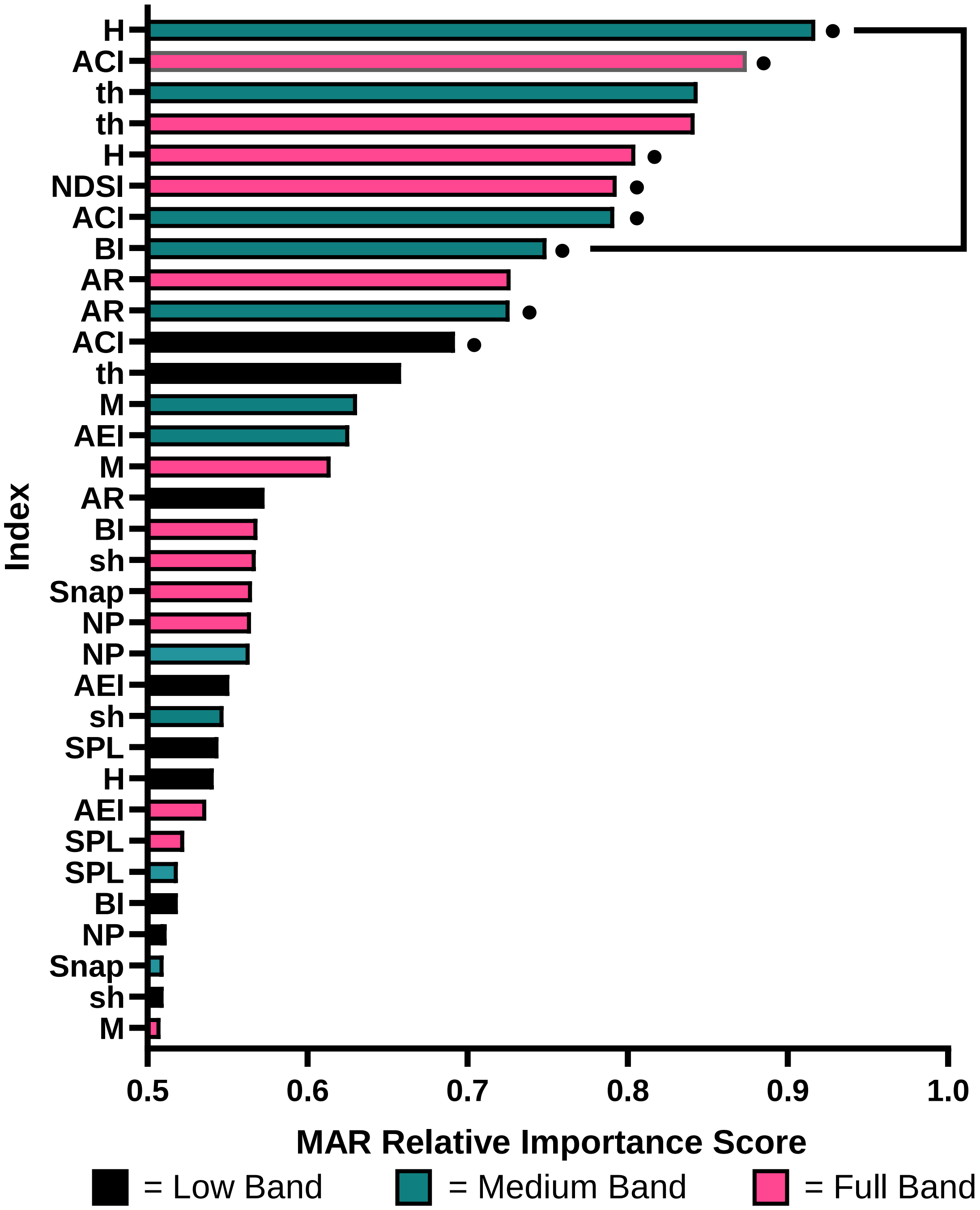


Fig. 4. Relative importance rankings of indices obtained from the multivariate adaptive regression (MAR) analysis used for feature selection. The eight recommendations obtained from the recursive feature elimination (RFE) analysis are indicated by the black line. The top eight indices of the MAR analysis were congruent with the RFE's eight recommendations, though the order was not conserved. Black dots to the right of bars indicate features which were selected for the final model after further manual feature selection.

2.7 Constructing the final model

Using the healthy and degraded datasets, an RDA model could then be constructed. Accuracy of the model was assessed using k-fold cross validation, with 10 folds. This splits the dataset into 'training' and 'test' sets to prevent overestimation of the models accuracy when presented with new data (Supp. 1). Due to random processes used in RDA, 1000 repeats of the cross-validated model construction were performed to provide a suitable level of depth for accuracy to be assessed (Rao *et al*. 2008). The RDA model was constructed using the *R* packages *MASS* (v.7.3-53) (Venables and Ripley, 200) and *KlaR* (v.0.6-15) (Weihs *et a.,* 2005).

The suitability of the data from restored reefs for entry into the model also had to be confirmed. If the restored sites exhibited soundscape properties that were highly distinct from both healthy and degraded sites, the model would be forced to attempt to fit them into a classification that was inappropriate. The presence of divergence from both classes was therefore explored using cluster analysis. This employed a principal component analysis (PCA) conducted on the feature set of the eight selected indices and a pairs plot which was also performed in *R* between every combination of two indices against one another (Supp. 3, Fig. S1).

3. Results

3.1 Comparing indices between healthy and degraded sites

Mann-Whitney U tests revealed significant differences between healthy and degraded habitat index scores for 15 of the 33 indices (Fig. 5). Violin plots of the three most significantly different index results between the healthy and degraded sites reveal a large area of overlap is present between values of these indices from both habitat classes (Fig. 6).

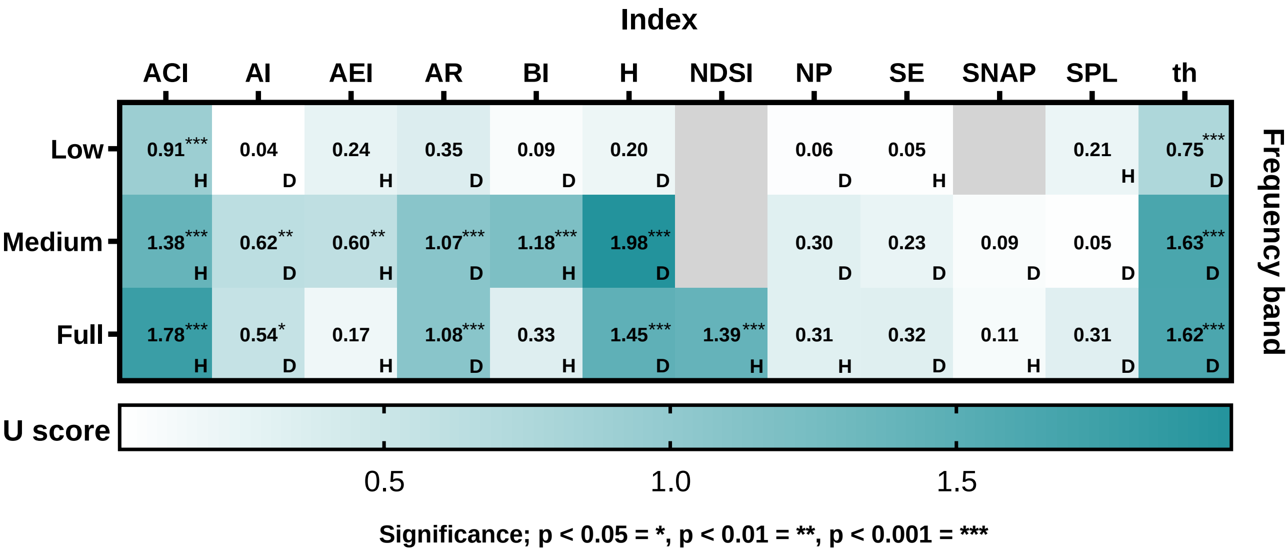


Fig. 5. Heat map displaying results from the Mann-Whitney U test between the ecoacoustic index scores calculated from recordings of healthy (n = 81) and degraded (n = 71) sites in low, medium and full-frequency bands. The habitat class with the higher mean is indicated by the letter in the bottom right corner of each cell (D = Degraded; H = Healthy). Blank cells indicate indices for which values from the corresponding frequency band were not calculated (see methods).

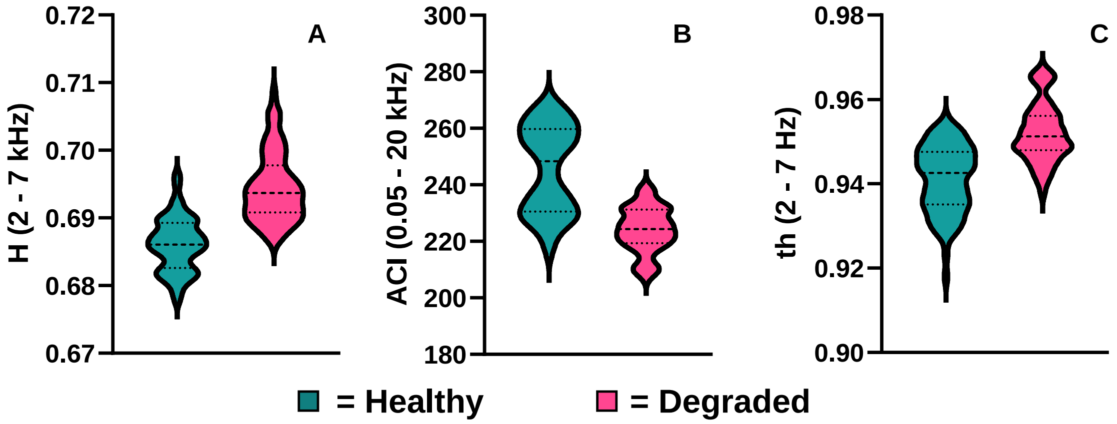


Fig. 6. Violin plots of the three indices with the most significant differences between healthy and degraded habitat (N = 152). (A) Medium-frequency band Entropy Index (H) (Mann-Whitney U; U = 1.98, p<0.001), (B) Full-frequency band Acoustic Complexity index (ACI) (U = 1.78, p<0.001), (C) Medium-frequency Temporal Entropy (th) (U = 1.63, p<0.001).

3.2 Comparing indices to phonic richness

Results from the Pearson's correlation revealed no strong relationship between phonic richness and any of the 33 indices trialled (Supp. 1, Fig. S2). The strongest relationship was a weakly negative correlation with the acoustic entropy index (H) in the full-frequency band (Pearson correlation; rho = -0.43; p<0.001), with all other indices reporting weaker correlations than this.

3.3 Regularised discriminant analysis

From the 1000 repeated constructions of the cross-validated model using the 152 recordings taken across healthy and degraded sites, the pooled mean misclassification rate was 8.27% ($\pm$ 0.84, SE). Of the 81 recording samples taken from the two healthy sites, 72.96 ($\pm$ 0.11) of these were correctly classified as healthy, with 8.04 ($\pm$ 0.11) misclassified as degraded. Of the 71 recordings taken from the two degraded sites, 67.22 ($\pm$ 0.09) of these were classified as degraded, with 3.74 ($\pm$ 0.09) misclassified as healthy. Individual results for each recording sample are also reported (Fig. 7).

Cluster analysis using the principal component analysis (Fig. 8) and pairs plot (Supp. 1, Fig. S3) were used to examine whether the 110 samples taken from recordings of the three restored sites were suitable for input into the model. Results from the plots showed these had a strong overlap with both the healthy and degraded habitat classes. For the Mature Restored and Newly Restored sites 70/81 and 70/71 samples respectively fell within one or both of the predictive ellipses for the two existing classes. This indicates that the soundscapes of the restored sites did not diverge from the soundscape present on the other two habitat types when using the properties investigated here. This supports the inputting of restored samples into the model as this is likely to generate an estimation of classification with a similar level of accuracy observed for the original two sites from which it was constructed. Additionally, the PCA showed that 61/81 samples from the Mature Restored sites fell within the ellipse that could be used predict healthy sites, whereas 24/27 samples of recordings from the Newly Restored site fell within the ellipse that can be used to predict degraded sites. However, it is important to note that there was a large region of overlap between the healthy and degraded class, with most of the ellipse of the degraded classes encompassed by that of the healthy class.

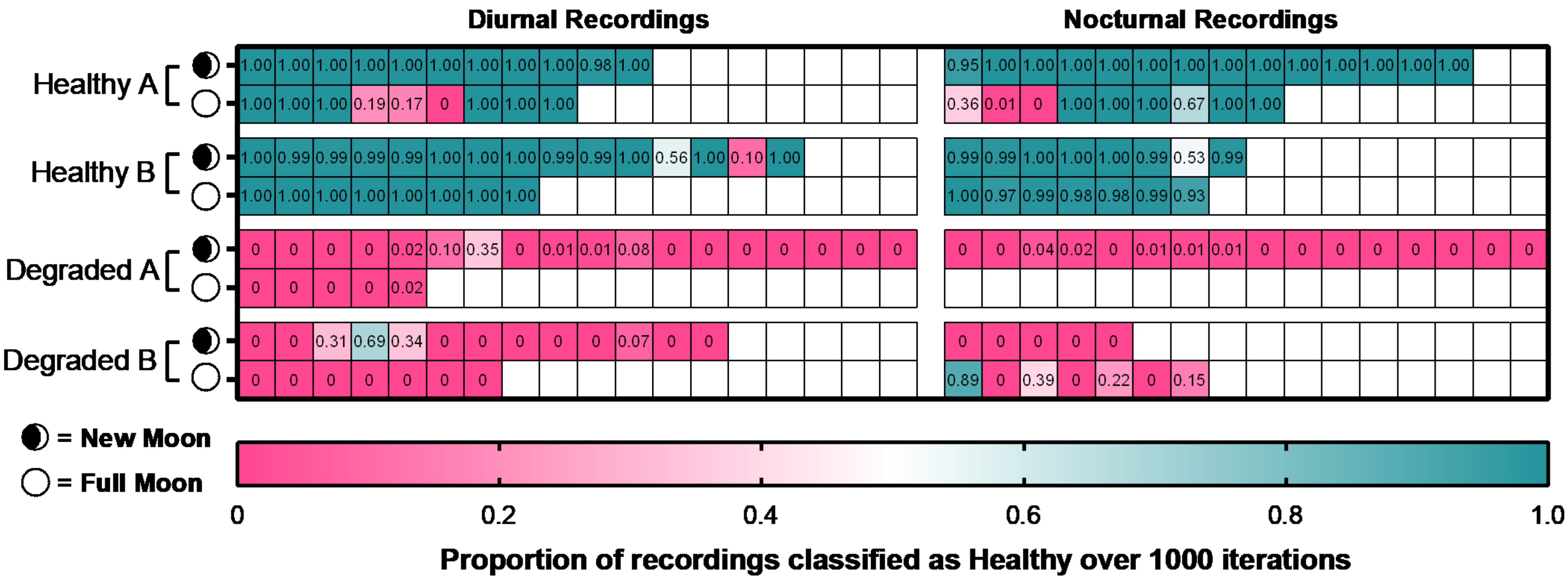


Fig. 7. Habitat classifications predicted by the machine learning model. Each cell indicates a single one minute recording from the 152 that were available across healthy and degraded habitats. The model was executed 1000 times on the dataset, generating a new habitat class prediction each time for every recording. Values within cells represent the proportion of these 1000 iterations in which the recording was predicted as originating from a healthy site, with the remaining being predicated as degraded, also represented by the colour code. Recordings taken on the left of the partition were taken during the day and recordings to the right were taken during crepuscular or night time periods. Although frequent gaps were present in the sampling regime, the order with which cells are presented within their respective blocks conserves the overall order with which they were sampled across time.

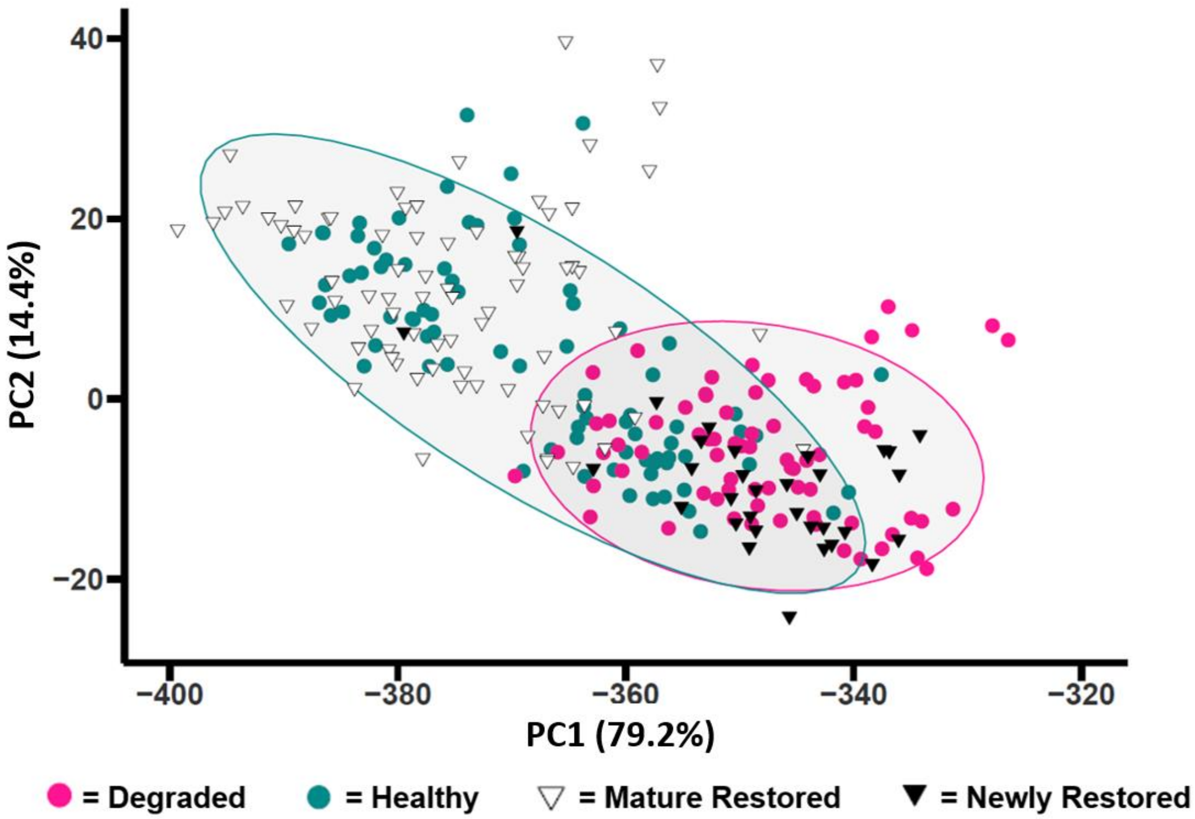


Fig. 8. Plot from the principal component analysis of PC1 and PC2 for the Healthy and Degraded site recording samples. Samples from recordings of Restored sites are overlaid on this to help determine whether these conform with either of the two existing classes or whether the properties of their soundscape are distinct. Ellipses indicate the zone within which a new sample can be assigned to a class using the two principle components presented in this figure alone. Overlapping areas indicate ambiguous results which cannot be differentiated but nonetheless fit one of the existing classes.

Execution of the model on the restored site samples was therefore performed in the same manner (Fig. 9). The majority classification of samples from mature restored sites was healthy, and samples from the newly restored site were mainly classified as degraded. A more decisive classification of Mature Restored site B was reported over Mature Restored site A, with 37/38 and 33/39 samples reporting a majority classification of healthy respectively. The six samples which reported a majority classification as degraded on Mature Restored site A occurred consecutively on the new moon at night. On the Newly Restored site, 27/33 samples reported a majority classification as degraded, all of these were during the full moon (though only four new moon samples were available) and five of these were at night.

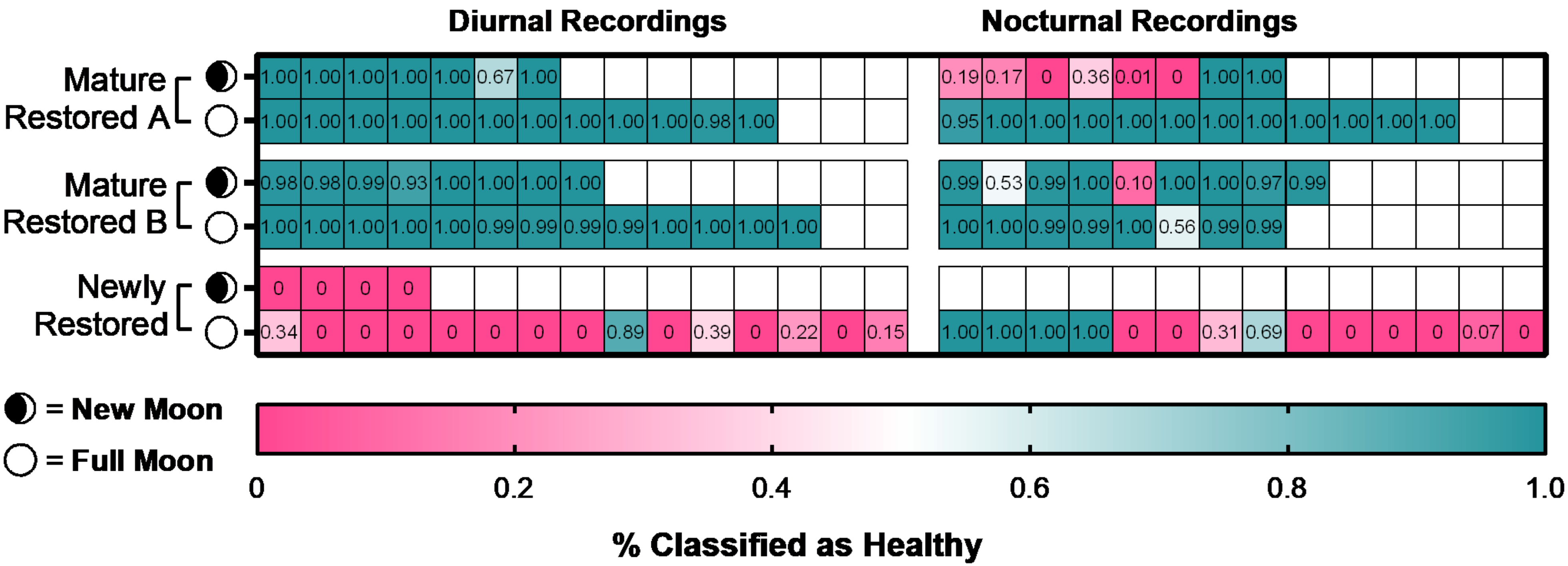


Fig. 9. Habitat classification predictions by the machine learning model for the restored site recording samples. Each cell indicates a single one-minute recording from the 110 that were taken from restored sites. The model was executed 1000 times on the dataset, generating a new habitat class prediction each time for every recording. Values within cells represent the proportion of these 1000 iterations in which the recording was predicted as originating from a healthy site, with the remaining being predicated as degraded, also represented by the colour code. Recordings taken on the left of the partition were taken during the day and recordings to the right were taken during crepuscular or night time periods. Although frequent gaps were present in the sampling regime, the order with which cells are presented within their respective blocks conserves the overall order with which they were sampled across time

The model trained on the 2018 recordings was also tested on a smaller number of recordings taken using the same sites and methodology ten months later, in June/July 2019 . Here, the model provided similar predictions for six of the seven sites; the only site exhibited a change in prediction between 2018 and 2019 was Healthy B, which transitioned from a majority classification as healthy to a majority classification as degraded (Table 2, full results in Fig. S4)..

Table 2. Results from the application of the 2018 model when tested on recordings taken at the same sites in 2019.

	Healthy A	Healthy B	Degraded A	Degraded B	Mature Restored A	Mature Restored B	Newly Restored
Recordings classified as Healthy	9/9	2/12	0/5	5/12	9/9	12/12	8/9
Proportion classified as Healthy	1.0	0.17	0	0.42	1.0	1.0	0.89

4. Discussion

Our study compared the ability of individual ecoacoustic indices and a machine-learning based compound model to discriminate between coral reef eco-states. Our results show that while no single ecoacoustic index can reliably discriminate between healthy and degraded reefs, our supervised machine-learning approach demonstrates a strong ability to accurately predict habitat class from randomly drawn acoustic samples. This highlights the exciting potential of combining PAM with machine learning for monitoring the health of coral reef ecosystems.

Up to twelve individual ecoacoustic indices were used across three frequency bandwidths for a total of 33 features; of these 33, 15 reported a significant difference between healthy and degraded reefs (Fig. 5). Of additional interest is the lack of strong correlations between any of these indices and phonic richness (Supp. 1), indicating that fish sound diversity was not the dominant driver of these results; rather alternative aspects of the soundscape were responsible. A combined diversity and abundance metric may reveal more about the role fish vocalisations play in driving index values.

The distribution of values for the indices that reported significant differences between healthy and degraded reefs all exhibited a high degree of overlap between the two habitat classes. This means that the ability to distinguish between habitat classes from a single recording using individual indices is low, as any given value from one class is also likely to be reported from a recording of the other class. Violin plots of the three most significant results help to visualise this large overlap between the results of each class (Fig. 6). These indices therefore offer a useful tool to reveal differences between habitats if extensive sampling is achievable on all sites of interest. However, their potential to deliver reliable results from short 'snapshot' recordings is limited.

By contrast, through combining multiple indices, the regularised discriminant analysis (RDA) model reported a strong predictive ability to classify single recordings. This is observable in results from the recursive feature elimination algorithm (RFE) (Fig. 3) which highlights the increases in accuracy attainable through constructing an optimised set of multiple indices compared to individual indices (Fig. 6). The misclassification rate of the final RDA model was 8.27% ($\pm$ 0.84) when applied to recordings from the same season; this was robust to diel and lunar variation (both known to influence marine soundscapes (Staater*man et al.*, 2014)), and reliably delivered the same classification for recordings from six of the seven sites taken nine months later. The feature selection stage of this approach is specific to the data and questions considered in this study. However, indices within the final feature set may offer a useful starting place for similar investigations. To produce optimised models, investigations on alternative study systems and questions should carry out independent feature selection based on their own training data.

Following the successful classification of healthy and degraded habitats, our compound model was executed on soundscape recordings taken from nearby coral reef habitats that had been restored (Williams et al., 2019). This was used to demonstrate the ability of this approach to perform a rapid assessment of these restored sites, using one-minute soundscape recordings. The model was able to detect differences between the Mature Restored sites and the Newly Restored site. Of the recording samples from the two Mature Restored sites, 33/39 and 37/38 were given a majority classification of healthy, whereas 27/33 samples from the Newly Restored sites were classified as degraded (Fig. 9). The Mature Restored sites were more than twice as old as the Newly Restored site (restoration started >24 months prior to recordings on Mature Restored sites, compared to <12 months for the Newly Restored site), and had approximately three times higher live coral cover (79.1% ± 3.9 and 66.5% ± 3.8 for the Mature Restored sites, 25.6% ± 2.6 for the Newly Restored site; values all % live coral cover mean ± SE; full data in Supplementary 1).

Restoration progress is thus reflected in the soundscape and can be effectively detected using a machine learning driven approach. This has strong implications for marine practitioners interested in using PAM to monitor the progress of restored sites against reference habitats. More generally, it further demonstrates the potential of using machine learning on PAM data to provide a powerful level of analytical depth for coral reef monitoring programmes.

To explore how the model could have been further improved, it is worth considering the sources of the observed error rate. The presence of this error could be due to several factors in isolation or in combination. The RDA approach used operates best when the input features are of a Gaussian distribution (Wu et al., 1996), however, some of the features used exhibited a sub-Gaussian distribution. This effect was likely due to the inclusion of samples from alternative times of day and multiple sites. Diel trends are frequently observed in reef soundscapes and this is reflected in the output of ecoacoustic indices (Kaplan *et al.*, 2015; Bertucci *et al.*, 2020; Carriço *et al.*, 2020). Additionally, reef soundscapes are known to differ over small spatial scales (Putland *et al*., 2017). Considering samples were taken from spatially separated sites to provide replicates, it is to be expected that differences across the same habitat class will have occurred. Both of these factors may have skewed the distributions of the feature sets. Furthermore, the dataset used to train the model itself was likely imperfect and will have contained natural outliers through ecological randomness that cannot be resolved at the sampling resolution employed.

It is also interesting to observe six of the same seven sites recorded in 2018 reported similar results 10 months later in 2019. The outlier here was Healthy B, for which 10/12 recordings were incorrectly predicted as degraded by the model. Recordings on this site were only collected during the day in 2019, with 9/12 of these taken during the new moon period. The soundscape may thus have been inadequately sampled, or it could be an indicator of a changing state of health on this site, not yet indicated by the coral cover data which was highly similar for both years (Supp. 1). Every other recording taken from both the Mature Restored sites in 2019 were henceforth classified as healthy, potentially indicating their continued restoration progress towards becoming established ‘healthy’ habitats.

It is important to note that although this model demonstrated an impressive ability to discriminate between habitat classes, a limitation of the experimental design was spatial replication: only two example sites of each habitat type were available (and only one example for Newly Restored).. Trialling the same approach on a larger number of sites would offer a valuable contribution to elucidating the full utility of machine learning to discriminate between healthy and degraded coral reef eco-states.

Future investigations could also build on the present study by considering a more nuanced approach to classifying eco-state. For example, this study employed a binary classification of reef health. In reality reefs exist along gradients of eco-states that are not as simplified as this (Downs et al., 2005; Smith et al., 2008). A sliding gradient of eco-states could be sampled and alternative machine learning algorithms such as random forests, neural networks or logistic regression could then be trained on this data to produce models which can make predictions on a continuous scale. Additionally, although coral cover alone can be a strong indicator of overall reef health (Dietzel et al., 2020; Smith et al., 2016), other attributes of interest could be considered to better determine the eco-state of a site. Future work could attempt to fit soundscape-based machine learning models to fish abundance or diversity metrics, or other habitat attributes. Demonstrating the use of machine learning against such efoort-intensive survey methods would be valuable.

Machine learning models not driven by ecoacoustic indices have also demonstrated utility in other ecological investigations. Examples of alternative approaches includes the splitting of recordings into many smaller frequency bands and the calculation of amplitude values from these (Roca and Van Opzeeland, 2019), or, the use of a 128 strong 'universal acoustic feature set' produced from a cross-convolutional neural network applied to *AudioSet*, the world's largest manually labelled sound database (Hershey et al., 2017; Sethi et al., 2020).

5. Conclusion

This investigation demonstrates that through combining ecoacoustic indices using machine learning, improved predictions of a coral reefs eco-state (healthy or degraded) can be made from one-minute recordings. This constitutes an exciting step towards maximising the value of PAM data collected from reef habitats. It also demonstrates this concept in practice through its application on large areas of restored reef, revealing that restoration progress is detectable in the soundscape of these sites.

Funding

This work was funded by a Natural Environment Research Council–Australian Institute of Marine Science CASE GW4+ Studentship NE/L002434/1 (to Timothy

A.C. Gordon); Swiss National Science Foundation Early Postdoc Mobility fellowship P2SKP3–181384 (to Lucille Chapuis.); a University of Exeter Education Incubator Research-Inspired Learning grant (to Timothy A.C. Gordon, Lucille Chapuis and Stephen D. Simpson); the University of Exeter Global Challenges Research Fund; and MARS Sustainable Solutions, part of Mars, Inc.

CRediT authorship contribution statement

Ben Williams was primarily responsible for analysis, modelling design and writing the manuscript. **Timothy A.C. Gordon** led field collection of data with assistance from **Ben Williams**, **Lucille Chapuis, Harry Harding and Eleanor May**. **Timothy A.C. Gordon**, **Lucille Chapuis**, **Andrew Radford** and **Stephen Simpson** provided comments on analysis, modelling and the written manuscript. **Stephen Simpson** provided supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Data were collected as part of the monitoring programme for the Mars Coral Reef Restoration Project, in collaboration with Hasanuddin University; we thank Lily Damayanti, Saipul Rapi, Alicia McArdle, Freda Nicholson, Jos van Oostrum and Frank Mars for logistical support and advice regarding this restoration project. We thank the Department of Marine Affairs and Fisheries of the Province of South Sulawesi; the Government Offices of the Kabupaten of Pangkep, Pulau Bontosua and Pulau Badi; and the communities of Pulau Bontosua and Pulau Badi for their support.

Bibliography (*in progress)*

Archer, S.K., Halliday, W.D., Riera, A., Mouy, X., Pine, M.K., Chu, J.W.F., Dunham, A., Juanes, F., 2018. First description of a glass sponge reef soundscape reveals fish calls and elevated sound pressure levels. Mar. Ecol. Prog. Ser. 595, 245–252.

Au, W.W.L., Banks, K., 1998. The acoustics of the snapping shrimp Synalpheus parneomeris in Kaneohe Bay . J. Acoust. Soc. Am. 103, 41–47. https://doi.org/10.1121/1.423234

Bertucci, F., Guerra, A.S., Sturny, V., Blin, E., Sang, G.T., Lecchini, D., 2020a. A preliminary acoustic evaluation of three sites in the lagoon of Bora Bora, French Polynesia. Environ. Biol. Fishes 1–12. https://doi.org/10.1007/s10641-020-01000-8

Bertucci, F., Maratrat, K., Berthe, C., Besson, M., Guerra, A.S., Raick, X., Lerouvreur, F., Lecchini, D., Parmentier, E., 2020b. Local sonic activity reveals potential partitioning in a coral reef fish community. Oecologia 193, 125–134. https://doi.org/10.1007/s00442-020-04647-3

Bertucci, F., Parmentier, E., Lecellier, G., Hawkins, A.D., Lecchini, D., 2016a. Acoustic indices provide information on the status of coral reefs: An example from Moorea Island in the South Pacific. Sci. Rep. 6, 1–9. https://doi.org/10.1038/srep33326

Bertucci, F., Parmentier, E., Lecellier, G., Hawkins, A.D., Lecchini, D., 2016b. Acoustic indices provide information on the status of coral reefs: an example from Moorea Island in the South Pacific. Sci. Rep. 6, 33326. https://doi.org/10.1038/srep33326

Bohnenstiehl, D.R., Lillis, A., Eggleston, D.B., 2016. The curious acoustic behavior of estuarine snapping shrimp: Temporal patterns of snapping shrimp sound in sub-tidal oyster reef habitat. PLoS One 11, 1–21. https://doi.org/10.1371/journal.pone.0143691

Bolgan, M., Amorim, M.C.P., Fonseca, P.J., Di Iorio, L., Parmentier, E., 2018. Acoustic complexity of vocal fish communities: A field and controlled validation. Sci. Rep. 8, 1–11. https://doi.org/10.1038/s41598-018-28771-6

Bradfer-Lawrence, T., Gardner, N., Bunnefeld, L., Bunnefeld, N., Willis, S.G., Dent, D.H., 2019. Guidelines for the use of acoustic indices in environmental research. Methods Ecol. Evol. 10, 1796–1807. https://doi.org/10.1111/2041-210X.13254

Butler, J., Stanley, J.A., Butler, M.J., 2016. Underwater soundscapes in near-shore tropical habitats and the effects of environmental degradation and habitat restoration. J. Exp. Mar. Bio. Ecol. 479, 89–96. https://doi.org/10.1016/j.jembe.2016.03.006

Carriço, R., Silva, M.A., Vieira, M., Afonso, P., Menezes, G.M., Fonseca, P.J., Amorim, M.C.P., 2020. The Use of Soundscapes to Monitor Fish Communities: Meaningful Graphical Representations Differ with Acoustic Environment. Acoustics 2, 382–398. https://doi.org/10.3390/acoustics2020022

Curtis, K.R., Howe, B.M., Mercer, J.A., 1999. Low-frequency ambient sound in the North Pacific: Long time series observations. J. Acoust. Soc. Am. 106, 3189–3200. https://doi.org/10.1121/1.428173

Dietzel, A., Bode, M., Connolly, S.R., Hughes, T.P., 2020. Long-term shifts in the colony size structure of coral populations along the Great Barrier Reef. https://doi.org/10.1098/rspb.2020.1432

Dimoff, S.A., Halliday, W.D., Pine, M.K., Tietjen, K.L., Juanes, F., Baum, J.K., 2021. The utility of different acoustic indicators to describe biological sounds of a coral reef soundscape. Ecol. Indic. 124, 107435. https://doi.org/10.1016/j.ecolind.2021.107435

Downs, C.A., Woodley, C.M., Richmond, R.H., Lanning, L.L., Owen, R., 2005. Shifting the paradigm of coral-reef “health” assessment. Mar. Pollut. Bull. 51, 486–494. https://doi.org/10.1016/j.marpolbul.2005.06.028

Eldridge, A., Guyot, P., Moscoso, P., Johnston, A., Eyre-Walker, Y., Peck, M., 2018. Sounding out ecoacoustic metrics: Avian species richness is predicted by acoustic indices in temperate but not tropical habitats. Ecol. Indic. 95, 939–952. https://doi.org/10.1016/j.ecolind.2018.06.012

Elise, S., Bailly, A., Urbina-Barreto, I., Mou-Tham, G., Chiroleu, F., Vigliola, L., Robbins, W.D., Bruggemann, J.H., 2019a. An optimised passive acoustic sampling scheme to discriminate among coral reefs’ ecological states. Ecol. Indic. 107, 105627. https://doi.org/10.1016/j.ecolind.2019.105627

Elise, S., Urbina-Barreto, I., Pinel, R., Mahamadaly, V., Bureau, S., Penin, L., Adjeroud, M., Kulbicki, M., Bruggemann, J.H., 2019b. Assessing key ecosystem functions through soundscapes: A new perspective from coral reefs. Ecol. Indic. 107, 105623. https://doi.org/10.1016/j.ecolind.2019.105623

Freeman, L.A., Freeman, S.E.S., 2016. Rapidly obtained ecosystem indicators from coral reef soundscapes. Mar. Ecol. Prog. Ser. 561, 69–82. https://doi.org/10.3354/meps11938

Gibb, R., Browning, E., Glover-Kapfer, P., Jones, K.E., 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods Ecol. Evol. 10, 169–185. https://doi.org/10.1111/2041-210X.13101

Gordon, T.A.C., Harding, H.R., Wong, K.E., Merchant, N.D., Meekan, M.G., McCormick, M.I., Radford, A.N., Simpson, S.D., 2018. Habitat degradation negatively affects auditory settlement behavior of coral reef fishes. Proc. Natl. Acad. Sci. 115, 5193–5198. https://doi.org/10.1073/pnas.1719291115

Gordon, T.A.C., Radford, A.N., Davidson, I.K., Barnes, K., McCloskey, K., Nedelec, S.L., Meekan, M.G., McCormick, M.I., Simpson, S.D., 2019. Acoustic enrichment can enhance fish community development on degraded coral reef habitat. Nat. Commun. 10, 1–7. https://doi.org/10.1038/s41467-019-13186-2

Harris, S.A., Shears, N.T., Radford, C.A., 2016. Ecoacoustic indices as proxies for biodiversity on temperate reefs. Methods Ecol. Evol. 7, 713–724. https://doi.org/10.1111/2041-210X.12527

Hershey, S., Chaudhuri, S., Ellis, D.P.W., Gemmeke, J.F., Jansen, A., Moore, R.C., Plakal, M., Platt, D., Saurous, R.A., Seybold, B., Slaney, M., Weiss, R.J., Wilson, K., 2017. CNN architectures for large-scale audio classification, in: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings. Institute of Electrical and Electronics Engineers Inc., pp. 131–135. https://doi.org/10.1109/ICASSP.2017.7952132

Kaplan, M., Mooney, T., Partan, J., Solow, A., 2015. Coral reef species assemblages are associated with ambient soundscapes. Mar. Ecol. Prog. Ser. 533, 93–107. https://doi.org/10.3354/meps11382

Kasten, E.P., Gage, S.H., Fox, J., Joo, W., 2012. The remote environmental assessment laboratory’s acoustic library: An archive for studying soundscape ecology. Ecol. Inform. 12, 50–67. https://doi.org/10.1016/j.ecoinf.2012.08.001

Kuhn, M., Johnson, K., 2019. Feature engineering and selection: A practical approach for predictive models. CRC Press.

Lecchini, D., Bertucci, F., Gache, C., Khalife, A., Besson, M., Roux, N., Berthe, C., Singh, S., Parmentier, E., Nugues, M.M., Brooker, R.M., Dixson, D.L., Hédouin, L., 2018. Boat noise prevents soundscape-based habitat selection by coral planulae. Sci. Rep. 8, 9283. https://doi.org/10.1038/s41598-018-27674-w

Lindseth, A. V., Lobel, P.S., 2018. Underwater soundscape monitoring and fish bioacoustics: A review. Fishes 3. https://doi.org/10.3390/fishes3030036

Lyon, R.P., 2018. Fish biodiversity, habitat complexity, and soundscape characteristics of patch reefs in a tropical, back-reef nursery 42.

McWilliam, J.N., McCauley, R.D., Erbe, C., Parsons, M.J.G., 2017. Patterns of biophonic periodicity on coral reefs in the Great Barrier Reef. Sci. Rep. 7, 1–13. https://doi.org/10.1038/s41598-017-15838-z

McWilliam, J.N., McCauley, R.D., Erbe, C., Parsons, M.J.G.G., 2018. Soundscape diversity in the Great Barrier Reef: Lizard Island, a case study. Bioacoustics 27, 295–311. https://doi.org/10.1080/09524622.2017.1344930

Mooney, T.A., Di Iorio, L., Lammers, M., Lin, T.-H., Nedelec, S.L., Parsons, M., Radford, C., Urban, E., Stanley, J., 2020. Listening forward: approaching marine biodiversity assessments using acoustic methods. https://doi.org/10.1098/rsos.201287

Nedelec, S.L., Simpson, S.D., Holderied, M., Radford, A.N., Lecellier, G., Radford, C., Lecchini, D., 2015. Soundscapes and living communities in coral reefs: Temporal and spatial variation. Mar. Ecol. Prog. Ser. 524, 125–135. https://doi.org/10.3354/meps11175

Putland, R.L., Constantine, R., Radford, C.A., 2017. Exploring spatial and temporal trends in the soundscape of an ecologically significant embayment. Sci. Rep. 7, 1–12. https://doi.org/10.1038/s41598-017-06347-0

Rao, R.B., Fung, G., Rosales, R., 2008. On the dangers of cross-validation. An experimental evaluation, in: Proceedings of the 2008 SIAM International Conference on Data Mining. SIAM, pp. 588–596.

Roca, I.T., Van Opzeeland, I., 2019. Using acoustic metrics to characterize underwater acoustic biodiversity in the Southern Ocean. Remote Sens. Ecol.

Conserv.

Sethi, S.S., Jones, N.S., Fulcher, B.D., Picinali, L., Clink, D.J., Klinck, H., Orme, C.D.L., Wrege, P.H., Ewers, R.M., 2020. Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set. Proc. Natl. Acad. Sci. U. S. A. 117, 17049–17055. https://doi.org/10.1073/pnas.2004702117

Simpson, S.D., Meekan, M., Montgomery, J., McCauley, R., Jeffs, A., 2005. Homeward Sound. Science (80-.). 308, 221. https://doi.org/10.1126/science.1107406

Smith, J.E., Brainard, R., Carter, A., Grillo, S., Edwards, C., Harris, J., Lewis, L., Obura, D., Rohwer, F., Sala, E., Vroom, P.S., Sandin, S., 2016. Re-evaluating the health of coral reef communities: baselines and evidence for human impacts across the central Pacific. Proc. R. Soc. B Biol. Sci. 283, 20151985. https://doi.org/10.1098/rspb.2015.1985

Smith, T.B., Nemeth, R.S., Blondeau, J., Calnan, J.M., Kadison, E., Herzlieb, S., 2008. Assessing coral reef health across onshore to offshore stress gradients in the US Virgin Islands. Mar. Pollut. Bull. 56, 1983–1991. https://doi.org/10.1016/j.marpolbul.2008.08.015

Sousa-Lima, R.S., Norris, T.F., Oswald, J.N., Fernandes, D.P., 2013. A Review and Inventory of Fixed Autonomous Recorders for Passive Acoustic Monitoring of Marine Mammals. Aquat. Mamm. 39.

Staaterman, E., Paris, C.B., DeFerrari, H.A., Mann, D.A., Rice, A.N., D'Alessandro, E.K., 2014. Celestial patterns in marine soundscapes. Mar. Ecol. Prog. Ser. 508, 17–32. https://doi.org/10.3354/meps10911

Staaterman, E., Rice, A.N., Mann, D.A., Paris, C.B., 2013. Soundscapes from a Tropical Eastern Pacific reef and a Caribbean Sea reef. Coral Reefs 32, 553–557. https://doi.org/10.1007/s00338-012-1007-8

Stowell, D., Sueur, J., 2020. Ecoacoustics: acoustic sensing for biodiversity monitoring at scale. Remote Sens. Ecol. Conserv. 6, 217–219. https://doi.org/10.1002/rse2.174

Sueur, J., Aubin, T., Simonis, C., 2008. Equipment review: Seewave, a free modular tool for sound analysis and synthesis. Bioacoustics 18, 213–226. https://doi.org/10.1080/09524622.2008.9753600

Sueur, J., Farina, A., Gasc, A., Pieretti, N., Pavoine, S., 2014. Acoustic indices for biodiversity assessment and landscape investigation. Acta Acust. united with Acust. 100, 772–781. https://doi.org/10.3813/AAA.918757

Villanueva-Rivera, L.J., Pijanowski, B.C., Villanueva-Rivera, M.L.J., 2018. Package 'soundecology.'

Williams, S.L., Sur, C., Janetski, N., Hollarsmith, J.A., Rapi, S., Barron, L., Heatwole, S.J., Yusuf, A.M., Yusuf, S., Jompa, J., Mars, F., 2019. Large-scale coral reef rehabilitation after blast fishing in Indonesia. Restor. Ecol. 27, 447–456. https://doi.org/10.1111/rec.12866

Wu, W., Mallet, Y., Walczak, B., Penninckx, W., Massart, D.L., Heuerding, S., Erni, F., 1996. Comparison of regularized discriminant analysis, linear discriminant analysis and quadratic discriminant analysis, applied to NIR data. Anal. Chim. Acta 329, 257–265. https://doi.org/10.1016/0003-2670(96)00142-0

Supplementary information

Authors:

Ben Williams, Timothy A. C. Gordon, Lucille Chapuis, Harry R. Harding, Eleanor B. May, Mochyudho E. Prasetya, Andrew N. Radford, Stephen D. Simpson.

Coral Cover

Live coral cover percentages were measured at each site in 2018 and 2019. Three 10 m transects were laid parallel to each other, 5m apart, at every site. Quadrats were placed every metre along each transect and photographed from above using a digital camera (Olympus TG-5). Twenty-five points were overlaid on each image using Coral Point Count software (Kohler & Gill 2006); the live coral percentage cover for each quadrat was taken as the percentage of these points that overlaid live coral.

Table S1. Coral cover percentage values.

	Healthy A		Healthy B		Degraded A		Degraded B	
	2018	2019	2018	2019	2018	2019	2018	2019
Mean percentage live coral cover	91.2	91.5	93.1	94.3	2.1	3.3	17.6	11.6
Standard error	2.0	3.2	2.6	2.2	0.9	1.3	4.6	2.7

	Mature Restored A		Mature Restored B		Newly Restored	
	2018	2019	2018	2019	2018	2019
Mean percentage live coral cover	79.1	56.5	66.5	76.3	25.6	34.5
Standard error	3.9	5.7	3.8	2.7	2.6	2.8

Feature selection algorithms used for RDA model

We used two approaches to select relevant indices amongst the feature set with a regularized discriminant analysis (RDA) algorithm. The first approach used was recursive feature elimination (RFE). This operates by selecting subsets of features and adding or removing a small number of other features progressively over multiple iterations until an optimised combination is found (Kuhn and Johnson, 2019). The second approach used was a multivariate adaptive regression spline (MAR) which constructs models with different combinations of features. It then progressively adds the remaining features and scores the associated increase or decrease in parameters, such as the predictive error in the model, to determine the importance of a feature (Kuhn and Johnson, 2019). One-hundred iterations of each algorithm were performed. Both approaches use a RDA algorithm.

Cross-validation of RDA model

It is important to note that models constructed on the full dataset available typically overestimate their own accuracy. It is therefore essential to perform cross-validation of the model if a more representative estimate of its accuracy is required, as was desired here. Cross-validation involves splitting the data into two groups. The first is a ‘training set’ in which the model is provided with samples and informed of the correct classification for each, enabling it to construct its predictors which will be used to classify new data. The second is a ‘test set’, upon which the model is executed whilst blind to the true class of each sample. This yields a prediction of the class for each sample within the test set, allowing the accuracy of the model to be obtained when presented with new data that was not used in its construction (Stone, 1974). There are several varieties of cross-validation. In this instance, K-fold cross validation using 10 folds was identified as a suitable and conservative technique for estimating error (Hastie *et al*., 2009). This split the data into 10 groups, treating nine of the ten as the training set and then testing the model on the remaining fold which acted as the test set. This process was then repeated for all combinations of the initial 10 folds and the accuracy reported.

Fig. S1. Results from a Spearmans collinearity test between all 12 indices in each of three frequency bands: low (0.05 – 0.8 kHz), medium (2 – 7 kHz), full (0.05 – 20 kHz). Darker cells represent a stronger correlation. Blank cells indicate values <0.00.

	ACI Low	ACI Medium	ACI Full	AEI Low	AEI Medium	AEI Full	AI Low	AI Medium	AI Full	AR Low	AR Medium	AR Full	BI Low	BI Medium	BI Full	H Low	H Medium	H Full	NP Low	NP Medium	NP Full	SE Low	SE Medium	SE Full	TE Low	TE Medium	TE Full	SPL Low	SPL Medium	SPL Full	SNAP Medium	SNAP Full
ACI Medium	0.35																															
ACI Full	0.24	0.65																														
AEI Low	-0.07	0.02	0.02																													
AEI Medium	-0.16	-0.23	0.14	-0.07																												
AEI Full	-0.04	-0.10	0.27	0.43	0.60																											
AI Low	-0.13	-0.16	-0.28	-0.11	0.13	-0.21																										
AI Medium	0.04	0.09	-0.46	0.14	-0.46	-0.34	0.03																									
AI Full	0.09	0.06	-0.50		-0.42	-0.43	0.05	0.89																								
AR Low	-0.21	-0.29	-0.43	-0.20	0.08	-0.32	0.92	0.07	0.13																							
AR Medium	-0.05	-0.15	-0.68	0.12	-0.50	-0.39	0.16	0.87	0.83	0.24																						
AR Full	-0.02	-0.22	-0.73		-0.35	-0.42	0.21	0.77	0.88	0.32	0.89																					
BI Low	-0.13	0.06	0.09	0.16	-0.07	0.09	-0.13	-0.11	-0.15	-0.19	-0.10	-0.13																				
BI Medium	0.10	0.17	0.10	-0.21		-0.20	0.29	-0.29	-0.16	0.28	-0.24	-0.14	0.02																			
BI Full	0.10	0.14	0.18	-0.06	-0.22		-0.24	-0.13	-0.24	-0.26	-0.16	-0.38	0.37																			
H Low	0.14	0.13	-0.02	-0.48	-0.17	-0.23	-0.22	0.02		-0.18		-0.07	-0.22	0.03	0.28																	
H Medium	-0.27	-0.83	-0.75	0.08		-0.02	0.08	0.31	0.27	0.21	0.49	0.48	-0.08	-0.43	-0.18	-0.13																
H Full	-0.16	-0.54	-0.80	0.01	0.04	-0.23	0.29	0.43	0.53	0.42	0.61	0.79	-0.17	-0.09	-0.56	-0.17	0.67															
NP Low	-0.23	0.07	0.20	0.07	0.03	0.16	-0.39	-0.13	-0.14	-0.42	-0.19	-0.21	0.01	-0.22	-0.02	0.09	-0.08	-0.18														
NP Medium	0.07	-0.13	0.05	0.02	0.15	0.20	-0.10	-0.02	-0.09	-0.14	-0.05	-0.07	-0.11	-0.27	0.03	-0.01	0.17	-0.01	-0.07													
NP Full	0.25	0.11	0.12	0.03	0.08	0.09	-0.13	0.18	0.19	-0.12	0.14	0.16	-0.25	-0.27	-0.19	0.08	-0.01	0.15	-0.02	0.18												
SE Low	0.25	0.25	0.13	-0.36	-0.20	-0.10	-0.41	-0.02	-0.05	-0.42	-0.08	-0.17	-0.12		0.35	0.95	-0.22	-0.31	0.16	0.03	0.08											
SE Medium	-0.08	-0.31	0.36	0.02	0.54	0.51	-0.22	-0.53	-0.58	-0.26	-0.59	-0.55	-0.04	-0.32	0.05	-0.09	0.22	-0.22	0.17	0.32	0.11	-0.04										
SE Full	0.04	-0.18	-0.12	0.03	0.36	0.06	0.10	0.03	0.14	0.13	0.06	0.31	-0.23	-0.08	-0.70	-0.24	0.21	0.65		0.04	0.42	-0.30	0.20									
TE Low	-0.38	-0.40	-0.48	-0.29	0.10	-0.38	0.63	0.11	0.18	0.80	0.26	0.35	-0.28	0.10	-0.28	-0.01	0.34	0.50	-0.23	-0.11	-0.02	-0.32	-0.15	0.24								
TE Medium	-0.19	-0.54	-0.91	0.06	-0.36	-0.36	0.22	0.63	0.63	0.36	0.84	0.81	-0.05	-0.18	-0.19	-0.05	0.76	0.76	-0.18	-0.06	-0.08	-0.17	-0.47	0.06	0.41							
TE Full	-0.20	-0.58	-0.91		-0.07	-0.29	0.31	0.50	0.57	0.45	0.70	0.83	-0.12	-0.08	-0.44	-0.12	0.72	0.97	-0.21	-0.03	0.03	-0.27	-0.33	0.44	0.51	0.87						
SPL Low	-0.08	0.10	-0.02	0.01	-0.06	-0.16	0.39	0.06	-0.01	0.27	0.08	-0.05	-0.18		0.07		-0.13	-0.16	-0.03	-0.06	-0.10	-0.04	-0.24	-0.30	0.15	0.05	-0.09					
SPL Medium	-0.02	0.49	-0.14	0.08	-0.51	-0.39	0.08	0.79	0.70	0.05	0.65	0.50	-0.06	-0.10	-0.02	0.09	-0.16	0.07	-0.07	-0.10	0.05	0.08	-0.68	-0.22	0.01	0.31	0.15	0.30				
SPL Full	0.05	0.43	-0.15	-0.09	-0.47	-0.45	0.06	0.66	0.75	0.06	0.55	0.54	-0.05	-0.06	0.01	0.14	-0.15	0.06	-0.06	-0.09	0.01	0.13	-0.60	-0.23	0.01	0.27	0.15	0.25	0.85			
NAP Medium	0.13	0.43	-0.18	0.11	-0.44	-0.32	-0.02	0.86	0.78	-0.05	0.68	0.58	-0.07	-0.16	-0.05	0.12	-0.06	0.19	-0.08	-0.09	0.19	0.12	-0.57	-0.02	-0.03	0.33	0.23	0.10	0.85	0.74		
SNAP Full	0.23	0.46	-0.12	-0.09	-0.41	-0.39	-0.02	0.71	0.80	-0.01	0.55	0.56	-0.11	-0.03	-0.03	0.13	-0.18	0.10	-0.07	-0.13	0.15	0.13	-0.60	-0.10	-0.04	0.24	0.15	0.05	0.77	0.84	0.84	
NDSI Full	0.09	0.22	0.22	-0.21	0.05	-0.11	0.25	-0.42	-0.32	0.21	-0.40	-0.33	0.10	0.87	0.12	0.04	-0.51	-0.30	-0.21	-0.22	-0.31	0.04	-0.24	-0.24	-0.01	-0.30	-0.27	0.11	-0.16	-0.10	-0.25	-0.12

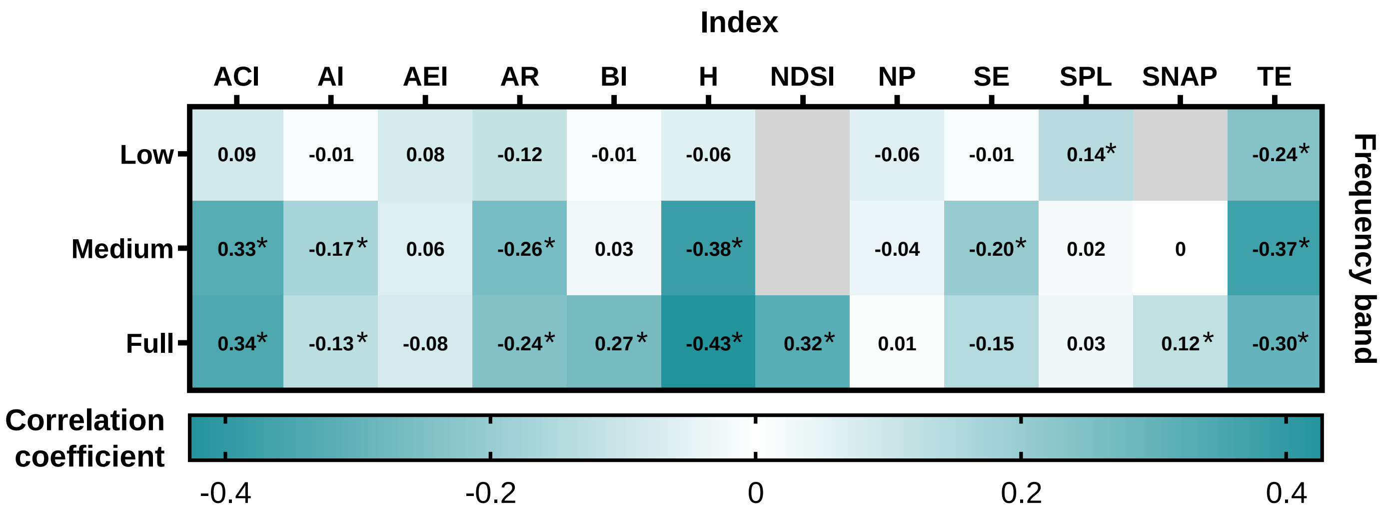


Fig. S2. Heat map displaying results from Pearson correlation tests between eco-acoustic indices and phonic richness scores in the low, medium and full frequency bands. Strength of correlation is indicated by the colour bar. Cells marked with an asterisk indicate those with a significant correlation (*p*<0.05). Blank cells indicate indices for which values from the corresponding frequency band were not calculated (see methods).

Fig. S3. Scatterplots between each of the eight indices selected for inclusion as features in the final model. Values from healthy and degraded sites alongside values from restored sites are included to enable divergence between these two groups to be observed if present.

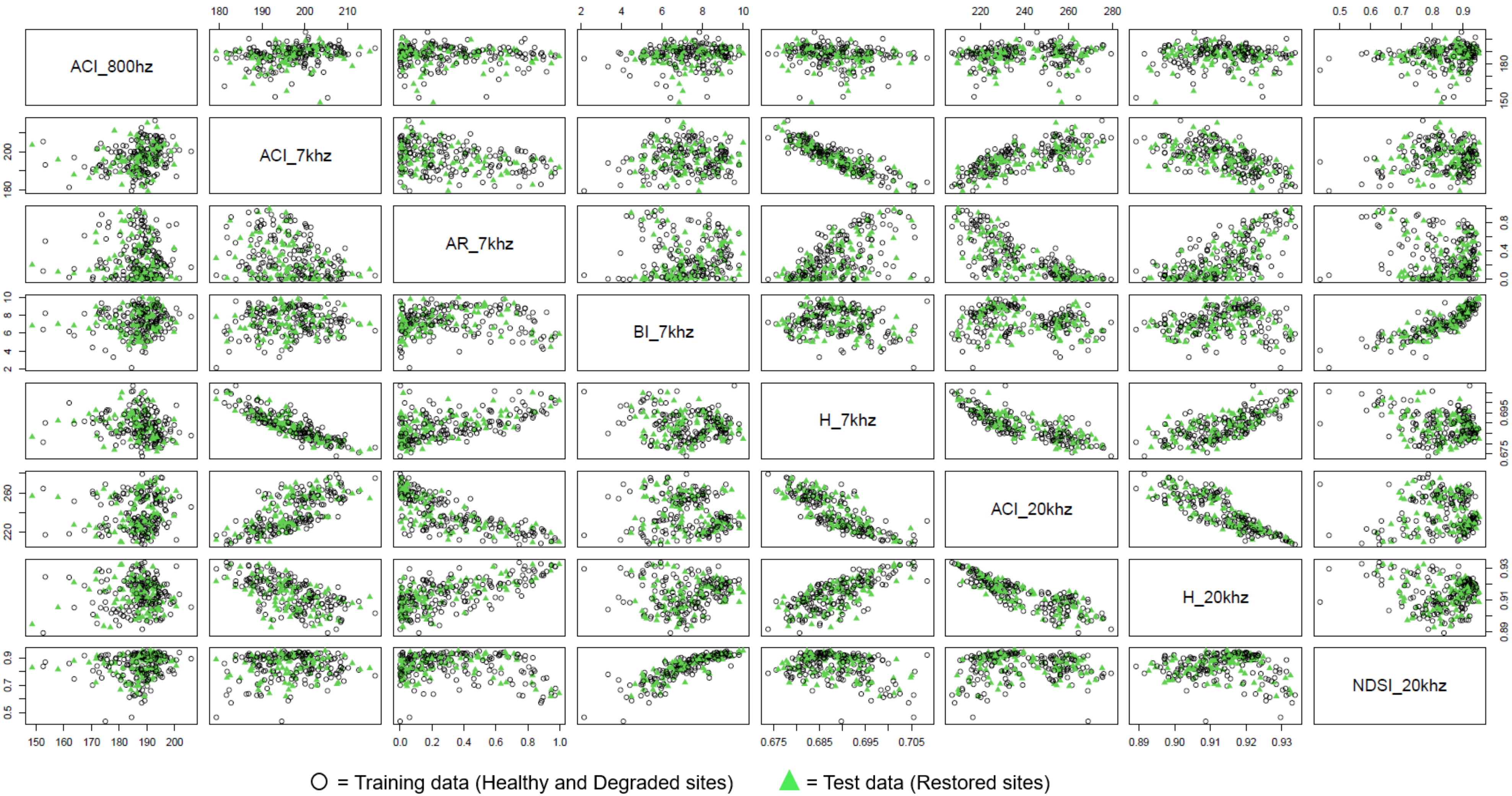
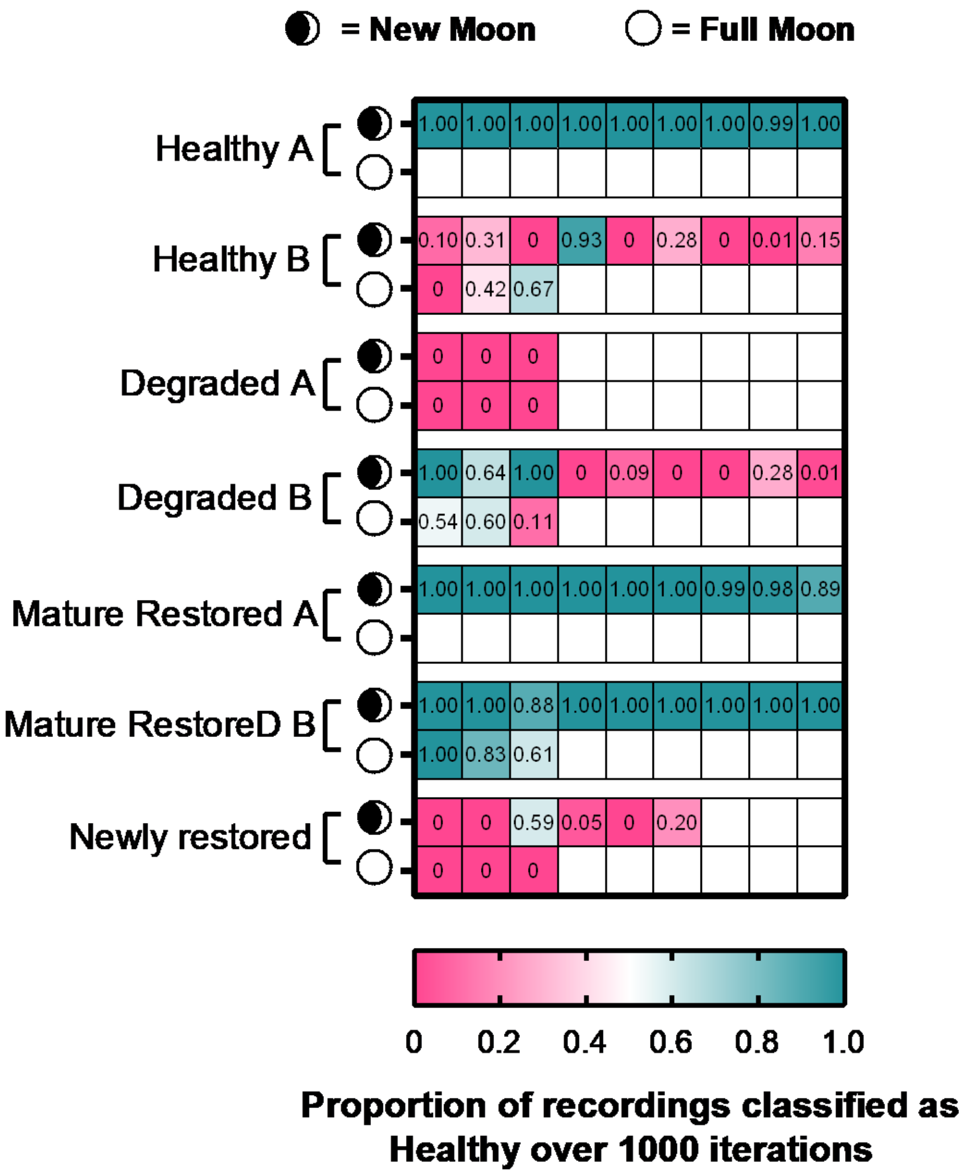


Fig. S4. Additional recordings were taken approximately nine months later using the same methodology, from the same seven sites, within three days either side of the new moon (June 17th) and full moon (July 3rd) in 2019. All recordings were taken during the daytime only. The results from 1000 iterations of the model are presented below.



References

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.

Kuhn, M., Johnson, K., 2019. Feature engineering and selection: A practical approach for predictive models. CRC Press.

Stone, M., 1974. Cross-Validatory Choice and Assessment of Statistical Predictions. J. R. Stat. Soc. Ser. B 36, 111–133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x