BioSound Exploratory Analysis
Summary Report

BioSound Exploratory Analysis Summary Report
Prepared by: Liz Ferguson & Michelle Weirathmueller
Date: February 21, 2025

Goal for Summary

This report provides a comprehensive summary of the recent developments, exploratory data analysis, and proposed next steps for the BioSound-MBON project. After completing the exploratory study and sharing the initial results, the team conducted a deeper analysis to evaluate trends and identify patterns of commonality across all datasets. The primary focus of this analysis was to assess acoustic indices, analyze diel trends, evaluate water class relationships, and apply multivariate analyses to gain valuable insights into similarities and differences of marine soundscapes.

Recent Developments to the BioSound-MBON Dashboard

The BioSound-MBON Dashboard now includes enhanced functionality, allowing users to download plots and data directly from the interface (**Figure 1**). This improvement supports greater accessibility and facilitates deeper analysis by stakeholders and researchers. The dashboard can be accessed at the following link: [BioSound Dashboard]

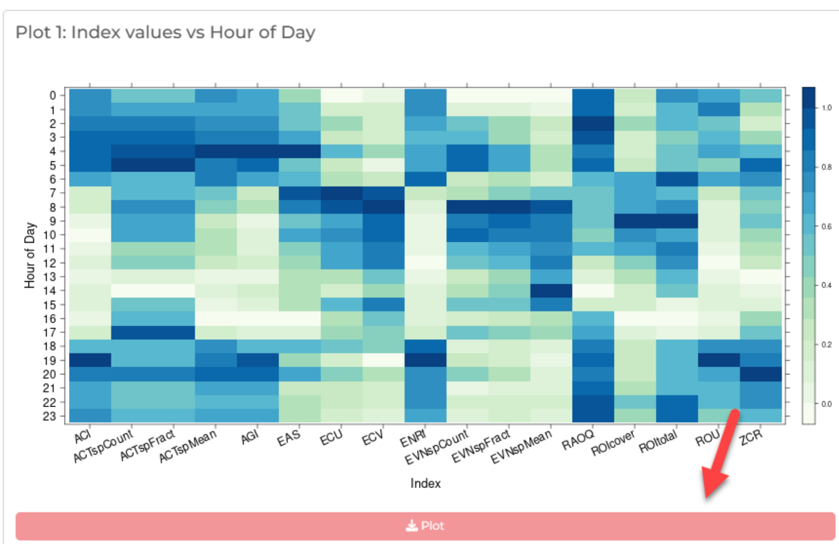


Figure 1: Download options for plots and select data available on all data explorer tabs

Evaluating Exploratory Data

The exploratory data analysis aimed to identify trends and interpret relationships within the collected datasets. Key analytical components included:

- **Diel Trends:** Analyzing acoustic indices to uncover temporal patterns across various sites.
- **Water Class Relationships:** Evaluating correlations between water class data and acoustic indices.
- **Recording Duration Effects:** Determining how different recording durations impact the calculation of acoustic indices.
- **Multivariate Analysis:** Utilizing Principal Component Analysis (PCA) and K-Means clustering to explore complex patterns within the data.

Diel & Diurnal Trends Analysis

We conducted a qualitative evaluation of trends and differences across all datasets by reviewing heatmaps generated by normalizing the value of each acoustic index. These heatmaps were produced for either the full bandwidth or a reduced 16 kHz sampling rate and plotted by the hour of the day. Our objective was to uncover diel (24-hour cycle) and diurnal (daytime-focused) trends by visualizing how acoustic indices fluctuated throughout the day. Among the three available plots in the BioSound-MBON analytical tool, Plot 3 proved to be the most impactful. This plot displayed the range of normalized, user-selected index values by date on the x-axis and hour of the day on the y-axis, providing a clear temporal pattern for the month of February. We selected February 2019 as the common month for analysis to enable consistent comparisons across most datasets, with the exception of Key West, where data was collected in February 2020.

Key trends observed in our analysis included:

Dataset	Type of Marine Habitat	Depth (m)	Pattern Observation
Biscayne Bay, FL	Mangrove	4.3	Notable Patterns
Chukchi Sea, Hanna Shoal	Offshore	30	Few or Unobservable Patterns
Gray's Reef, GA	Offshore	16	Notable Patterns
Key West, FL	Coral Reef	23	Notable Patterns
May River, SC	Estuary	4.5	Notable Patterns
Olowalu (Maui, HI)	Island/Nearshore	59.7	Few or Unobservable Patterns
ONC-MEF	Offshore	2,189	Few or Unobservable Patterns
OOI-HYDBBA106	Shelf	80	Few or Unobservable Patterns

Observed Patterns

Our analysis of diel and diurnal trends across multiple datasets revealed intriguing patterns that appear to correlate with tidal cycles. For several datasets and across various acoustic indices, a notable pattern emerged that aligns with tidal fluctuations, characterized by larger index measurements approximately every 6 to 7 hours (**Figure 2**). This cyclic pattern exhibited a consistent shift of about one hour forward each day throughout February, suggesting a potential link to tidal rhythms. The Biscayne Bay (a mangrove ecosystem) and May River (an estuary) datasets demonstrated this trend most prominently across multiple indices. Both sites share similar shallow depths of approximately 4.5 meters, which may amplify the influence of tidal movements on acoustic measurements.

In addition to tidal-associated patterns, we identified a distinct diurnal trend across several datasets (**Figure 3**). This pattern manifested as either lower acoustic index values during daylight hours with higher values in the evening and nighttime, or the inverse, with heightened activity during the day. This diurnal fluctuation was consistently observed across multiple indices and appeared most frequently in the Gray's Reef and Key West datasets. Unlike the shallow environments of Biscayne Bay and May River, these sites feature recorder depths of 16 meters and 23 meters, respectively. The increased depth at these locations might contribute to the observed light-driven acoustic patterns, possibly influenced by the behaviors of marine organisms, variations in abiotic noise sources, or changes in water column properties throughout the day.

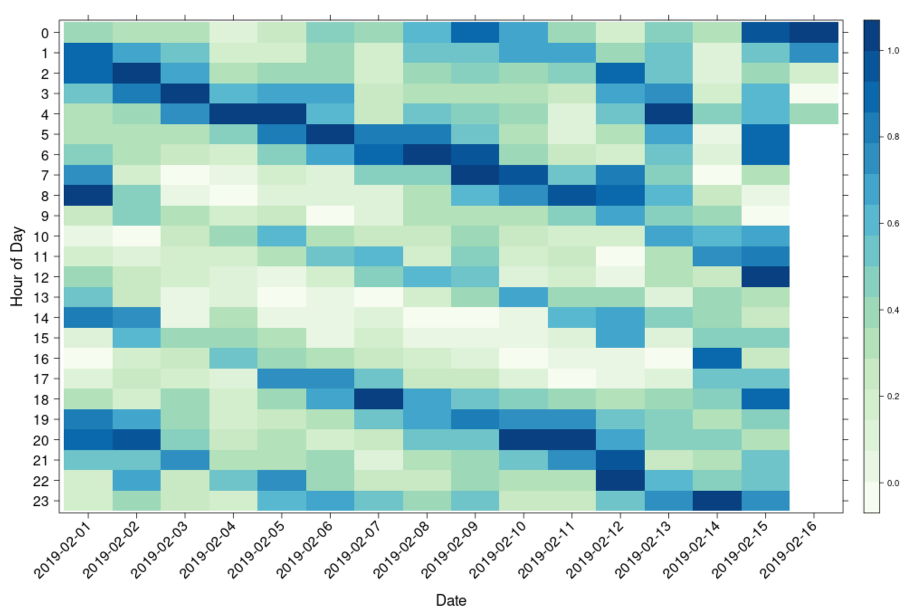


Figure 2: Example of apparent tidal pattern represented in multiple indices. This example from Biscayne Bay in February is the ACTspMean (Temporal Index), 32 kHz sampling rate.

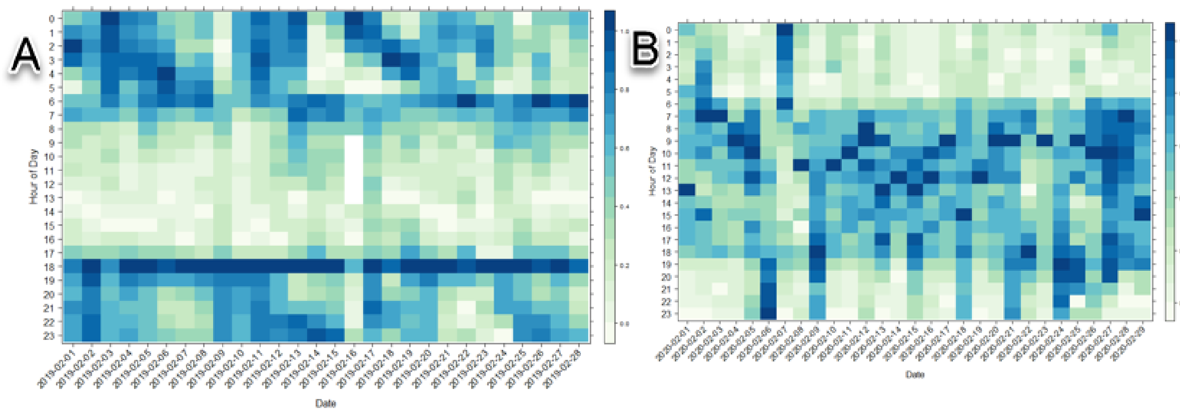


Figure 3: Diurnal trends observed in multiple datasets and indices. This example includes A) Gray's Reef, BGNf index (Amplitude Index) calculated from 48 kHz data and B) Key West, BGNf index (Amplitude Index) calculated from 48 kHz data.

We observed that the patterns present in the native bandwidth of each acoustic index were often mirrored in the low sampling rate examples, albeit with some variations in the intensity of index values (**Figure 4**). This consistency in patterns across both the full bandwidth and the reduced 16 kHz sampling rate suggests that much of the influence on these acoustic indices may be concentrated in the lower frequency ranges. Lower frequencies often capture sounds from a broad spectrum of sources, including environmental noise, certain marine mammal vocalizations, and anthropogenic activities, which might drive these observed trends. The alignment of patterns across sampling rates reinforces the potential utility of lower frequency monitoring for capturing meaningful ecological and environmental signals.

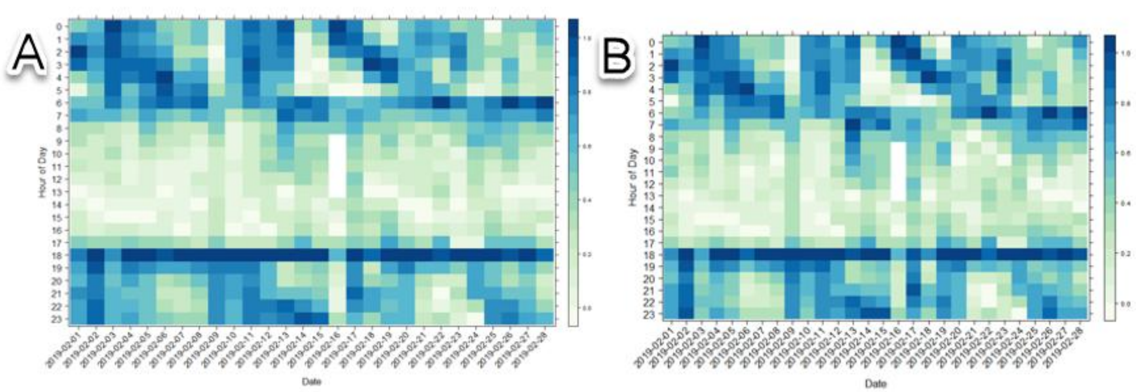


Figure 4: Similar patterns exhibited in full bandwidth and decimated data. This example includes A) Gray's Reef, BGNf index (Amplitude Index) calculated from 48 kHz data and B) Key West, BGNf index (Amplitude Index) calculated from 16 kHz data.

Recording Duration Summary

To assess the effect of recording duration on acoustic index measurements, we conducted a statistical evaluation using two datasets with varying recording intervals. The datasets included **Olowalu (Maui)** with recordings of **300 seconds** and **150 seconds**, and **Key West** with recordings of **30 seconds** and **10 seconds**. Since the acoustic index measurements were not normally distributed, we employed a non-parametric Mann-Whitney U test to compare each recording interval across all acoustic indices. Our analysis revealed that most indices exhibited statistically significant differences in measurements between recording durations for both datasets.

Table 1 summarizes the p-value by acoustic index where at least one non-significant result was observed. Indices highlighted in blue indicate non-significant results for both datasets, while all other rows demonstrate a significant difference for at least one of the two datasets. Our results showed that a significant difference was less frequently encountered in the longer 300-second versus 150-second duration comparison compared to the shorter 30-second versus 10-second intervals, albeit only for nine indices. This finding suggests that longer recording durations may provide more stable and consistent acoustic measurements, potentially reducing variability and enhancing interpretability.

Table 1: Comparison of statistical results (p-values) for Mann-Whitney U test comparisons to 300 and 150 seconds (Olowalu, Maui) and 30 and 10 seconds (Key West) datasets.

	Olowalu (Maui)	Key West, FL
BGNf	0.585	0.024
BGNt	0.902	0.806
BI	0.062	0.003
EPS_SKEW	0.162	<0.001
EVNtCount	0.311	0.828
H_gamma	0.623	<0.001
H_pairedShannon	0.097	<0.001
Hf	0.1	<0.001
MEANt	<0.001	0.466
MED	0.902	0.806
NDSI	0.062	0.003
rBA	0.062	0.003
TFSD	0.859	0.013
ZCR	0.827	0.001

Water Class Data Analysis

The exploratory study generated a series of correlation matrices that examined the relationship between acoustic indices and environmental data from the Seascapes designation of water classes. These water classes are defined using remotely sensed ocean properties, including variables such as sea surface temperature, chlorophyll concentration, salinity, and primary productivity, among others. **Figure 5** presents an example of correlation matrices for each study region, highlighting the interaction between acoustic measurements and environmental parameters. To ensure robust analysis, we calculated the water class cell counts for all regions to identify those water classes with sufficient representation in the dataset (**Figure 6**). This step was critical in avoiding skewed interpretations based on underrepresented classes. We then extracted the Pearson correlation coefficient values from each dataset to detect persistent patterns of correlation, focusing on indices that exhibited consistently high (e.g., >0.6) or low (e.g., < -0.6) values across datasets. A greater magnitude of either positive or negative correlation values would indicate a stronger relationship between the acoustic indices and specific environmental variables.

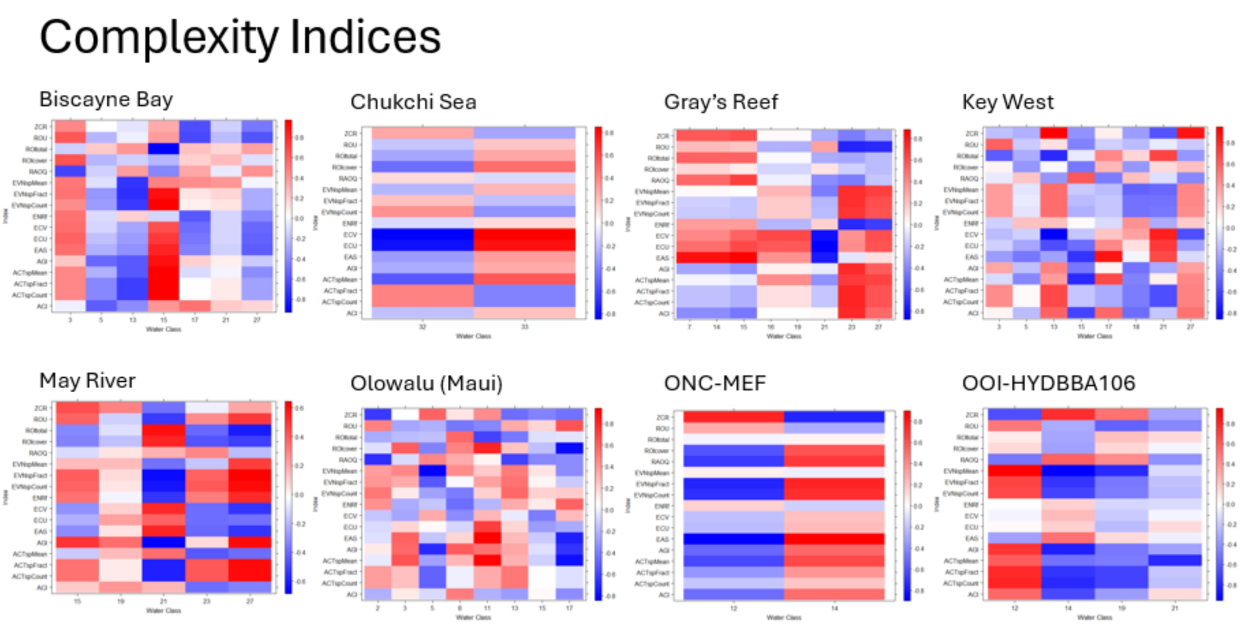


Figure 5: Series of correlation matrices from the exploratory study for the set of complexity indices, with water class on the x-axis and complexity index on the y-axis. Data from these plots were extracted for subsequent analysis.

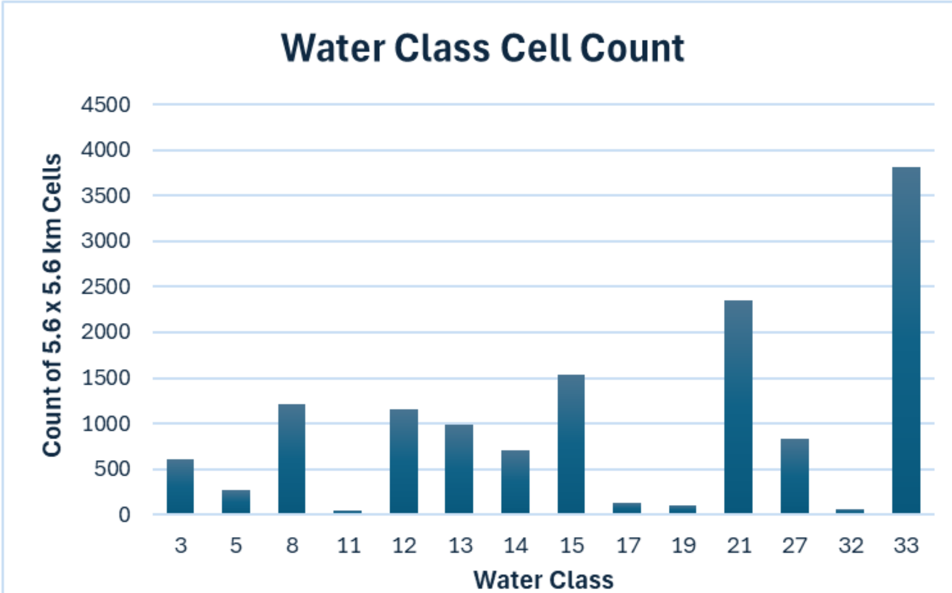


Figure 6: Water class cell count inclusive of all datasets. Water class number is indicated on the x-axis, while counts of 5.6 km by 5.6 km cells is along the y-axis.

We examined categories of acoustic indices to assess whether any consistent trends emerged by region, particularly focusing on whether similar habitats exhibited comparable correlations with water classes. However, our results did not reveal a clear pattern of consistency across similar environments. For instance, given the similarity in diel patterns between Biscayne Bay and May River, we initially hypothesized that these regions would exhibit comparable correlations with Seascapes water classes. Contrary to our expectations, **Table 2** demonstrates that each region displayed distinctly different correlation profiles. Biscayne Bay showed strong positive correlations among many of the complexity indices, whereas May River presented variable, low to moderate correlations, suggesting that even within seemingly similar shallow water habitats, underlying environmental or ecological differences influence acoustic index relationships differently.

To explore this further, we shifted our focus to per-index correlation measures per water class, aiming to determine whether at least specific acoustic index measurements maintained consistent correlations within particular water classes. This analysis involved plotting the distribution of mean index values, calculated across eight-day intervals to align with the temporal resolution of the remotely sensed water class data, and comparing these to the observed correlation coefficients. For example, Water Class 12, which comprises a substantial portion of the ONC-MEF and OOI-HYDBBA106 datasets (**Figure 7**), exhibited opposite extreme correlation coefficients between these datasets. We visualized the mean index values for the SNRt, HpairedShannon, and EVNtMean indices to interpret these trends. Our analysis revealed that increased index values were associated with lower extreme correlation values for both EVNtMean and SNRt (**Figure 8**). However, this pattern did not hold for HpairedShannon, which showed reduced index measures at the ONC-MEP site, despite a similarly strong negative correlation coefficient. Importantly, while specific trends within single water classes were occasionally evident, these associations did not translate consistently across other water classes. For instance, in Water Class 15 (**Figure 9**), high values of HpairedShannon did not correlate with a drastic decrease in correlation coefficients, demonstrating the complex and potentially site-specific nature of these relationships.

Table 2: Pearson correlation coefficient values for Biscayne Bay and May River for water class 15. Extreme green cells indicates large positive correlation, and a large negative value indicates a strong negative correlation.

Index	Biscayne Bay Correlation Coeff	May River Correlation Coeff
ACI	0.406	0.108
ACTspCount	0.972	0.340
ACTspFract	0.972	0.340
ACTspMean	0.970	-0.158
AGI	0.896	0.505
EAS	0.833	-0.312
ECU	0.744	-0.226
ECV	0.669	-0.308
ENRf	-0.130	0.339
EVNspCount	0.958	0.377
EVNspFract	0.911	0.393
EVNspMean	0.476	0.151
RAOQ	-0.420	-0.112
ROIcover	-0.253	-0.350
ROItotal	-0.933	-0.337
ROU	0.386	0.385
ZCR	0.333	0.431

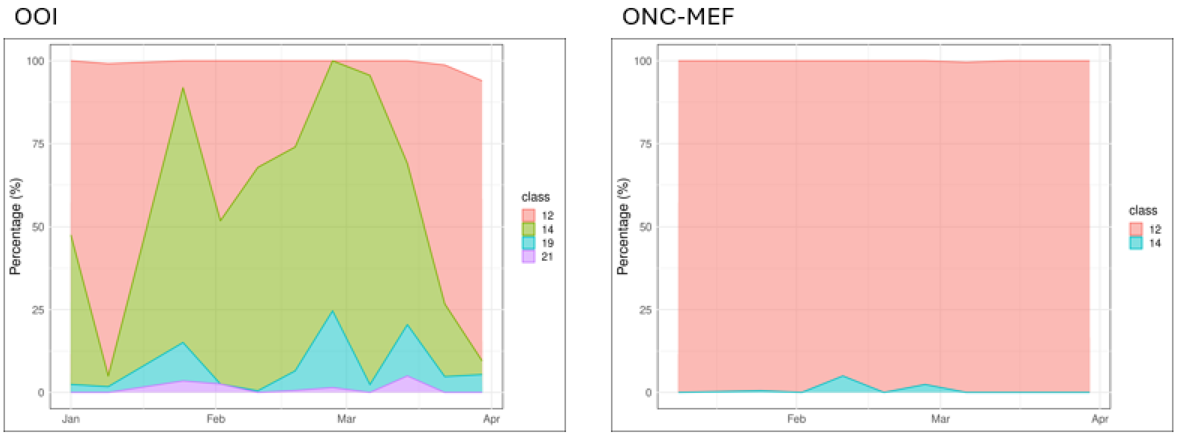
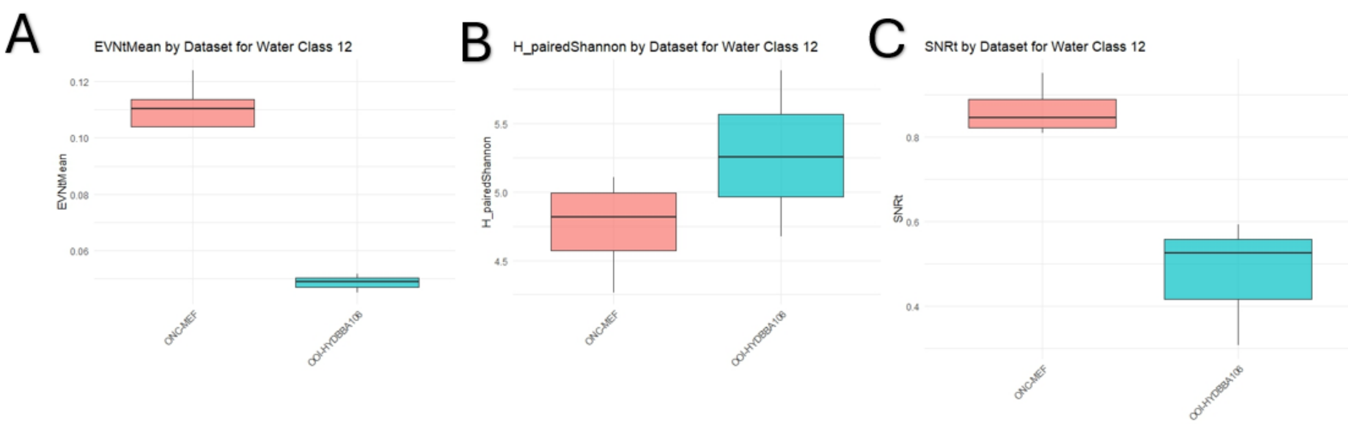


Figure 7: Composition of water class data for the period of March through April 2019 to give context in the representation of water classes in a region. Only February is incorporated in subsequent analyses. Water class 12 is only found in the OOI-HYDBBA106 and ONC-MEF datasets.

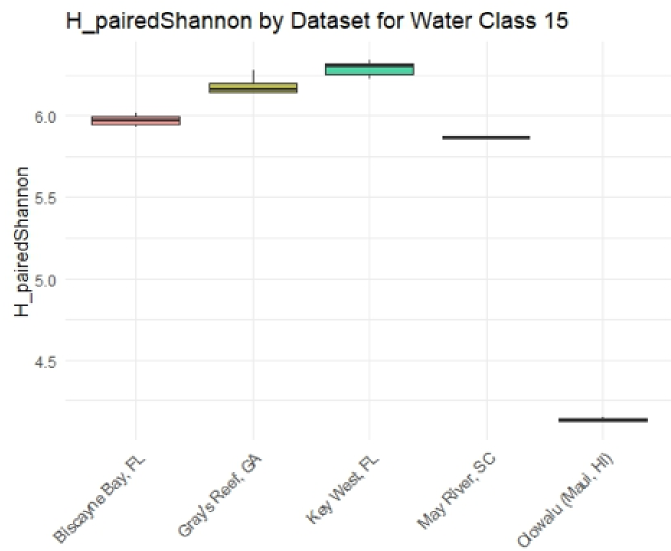


ONC-MEF: -0.81
OOI-HYDBBA106: 0.76

ONC-MEF: -0.63
OOI-HYDBBA106: -0.61

ONC-MEF: -0.96
OOI-HYDBBA106: 0.85

Figure 8: Plots of the distribution of mean index values associated with water class 12. Only datasets with profiles consistent of 5 or more cells of the specified water class are reported. Pearson correlation coefficients are reported for comparison to index measurements. Water class 12 is only represented in ONC-MEF and OOI-HYDBBA106. This figure displays measurements for EVNtMean (A), H_pairedShannon (B), and SNRt (C).



Dataset	Correlation Coefficient
Biscayne Bay	-0.26
Gray's Reef	0.34
Key West	-0.30
May River	0.009
Olowalu (Maui, HA)	-0.08

Figure 9: Distribution of mean H_pairedShannon measurements by dataset and associated correlation coefficient for water class 15. No relationship between correlation coefficients and measurements is noted across sites.

Multivariate Analysis

As a cursory investigation into the potential use of a combined index approach for evaluating the unique natures of these datasets, we applied two advanced analytical techniques to explore how a multi-index framework might enhance our understanding of marine soundscapes. This preliminary analysis aimed to determine whether combining multiple acoustic indices could reveal latent patterns and relationships not evident when indices are considered in isolation. The analytical methods employed included:

- **Principal Component Analysis (PCA):** Performed on 10 select presentative acoustic indices from each index category (e.g., per site to reduce dimensionality and highlight key patterns.
- **K-Means Clustering:** Classified sites into 8 groups based on their acoustic profiles, providing insights into site relationships and ecological significance.

We explored the natural structure of the acoustic indices by combining two complementary analytical techniques. We applied Principal Component Analysis (PCA) to the normalized indices to show the multidimensional relationships could be mapped onto two principal dimensions (**Figure 10**). The dimensional reduction allowed us to identify the main gradients of variation. Each feature's contribution to these principal components was carefully measured through loading coefficients, revealing which variables were most influential. In parallel, we introduced k-means clustering (k=8) to identify whether the natural groupings that emerge coincided with the geographical origins of the measured indices.

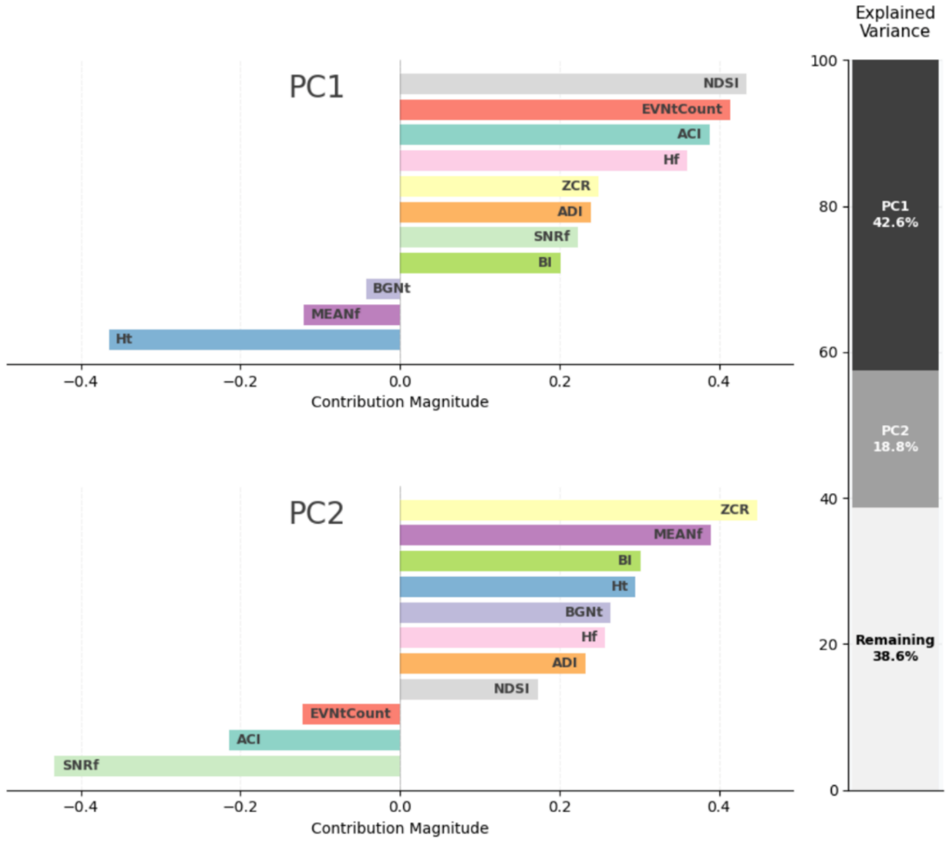


Figure 10: Principal component contributions of acoustic index to PC1 and PC2.

Our analysis revealed that k-means clustering patterns did not significantly correlate with geographical locations (**Figure 11A**). Interestingly, however, location groupings formed distinct clusters within the principal component space (**Figure 11B**). We observed that geographically proximate regions often displayed similar acoustic profiles, suggesting a spatial gradient in acoustic properties. This spatial coherence provides promising evidence that acoustic indices may effectively characterize different soundscapes, potentially offering a quantitative method for distinguishing between acoustic environments.

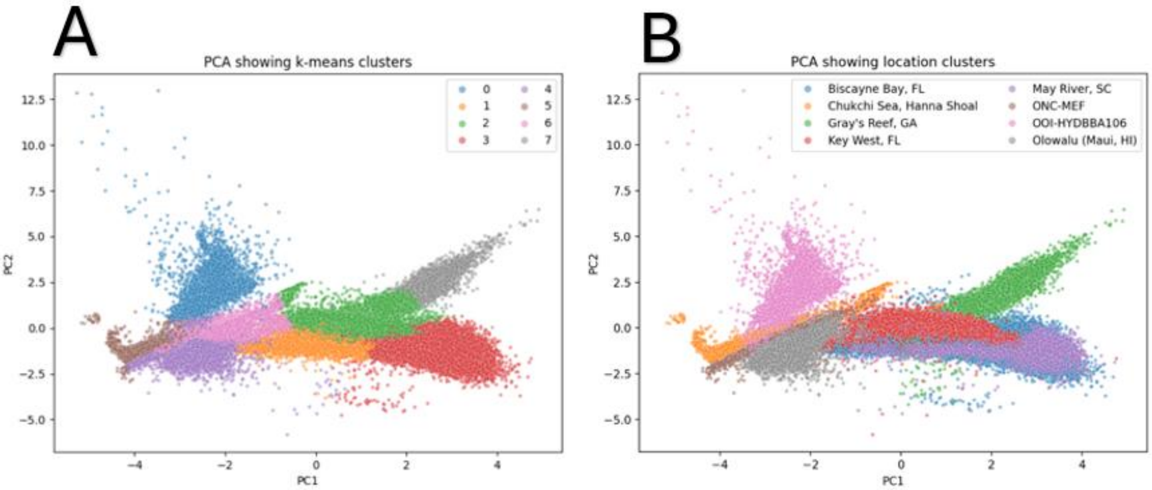


Figure 11: PCA plots by k-means clusters (A) and location clusters (B).

Multivariate Analysis Visualization

Visual representations of the Principal Component Analysis (PCA) and K-Means clustering results were generated to support the interpretation of clustering and site associations. These visualizations allow for an interactive exploration of the multivariate analysis, enabling users to compare K-Means clusters and site-specific clusters for the set of 10 indices utilized in the cursory review of multivariate methods. To provide an interactive perspective, these visualizations are available online at: Exploring BioSound Data.

Next Steps

An initial next step in this effort was the submission of a small proposal effort to the WILDLABS 2025 awards. To further enhance the analysis and understanding of acoustic indices, we proposed to evaluate the existing data using multivariate methods, including principal component analysis (PCA) and k-means clustering, and Uniform Manifold Approximation and Projection (UMAP) unsupervised learning methods. Indices will be evaluated collectively and in subsets based on their categorization (e.g., temporal indices, spectral indices) and collectively. These methods will allow us to derive quantitative pattern outputs that highlight similarities and differences in acoustic data across various underwater environments. Specifically, we will assess how factors such as recording instrument depth, diel and diurnal differences, and site-specific acoustic indices contribute to the observed variability. Additionally, as a result of subsequent conversations from the February 18th meeting, we will explore the comparison of these combined metrics in reference to a processed, longer-term dataset from May River to further understand the relationship between acoustic indices and the soundscape.