METHODOLOGY FOR ECOSYSTEM CHANGE
ASSESSING USING ECOACOUSTICS ANALYSIS

Diana Carolina Duque Montoya

Universidad de Antioquia UDEA
Facultad de Ingeniería
Medellín, Colombia
2019
METHODOLOGY FOR ECOSYSTEM CHANGE ASSESSING USING ECOACOUSTICS ANALYSIS

Diana Carolina Duque Montoya

In partial fulfillment of the requirements for the degree of Master’s in engineering

Advisor
Claudia Victoria Isaza Narváez

Research Group
Sistemas Embebidos e Inteligencia Computacional
-SISTEMIC-

Universidad de Antioquia UDEA
Facultad de Ingeniería
Medellín, Colombia
2019
This thesis is dedicated to my family

Thank you for supporting me in all the decisions I made up to this point
Acknowledgments

First, I would like to thank God for guiding me through His will, I would have never dare to take all these steps without His help. I would also like to thank my advisor, Claudia Isaza, who patiently leaded me to overcome my failures, advised me for every decision I made on this journey and helped me to improve the quality of my work. Thanks also to the people in SISTEMIC for providing me with everything I needed to accomplish all the goals during my master, specially to the professor Sebastián Isaza, who mentored me for developing my teaching skills and accepted me as his assistant in the informatics I class; in this job I learned new tools for the development of computer algorithms and could have a financial support during my studies.

Special thanks to the interdisciplinary group formed by members of Grupo Herpetológico de Antioquia and SISTEMIC: Juan Manuel Daza, Claudia Isaza, Estefany Cano, William Gómez, Camilo Sánchez and Juan Parra, who helped me to understand the main definitions for ecology, ecoacoustics, artificial intelligence and the importance of an accurate ecosystem state assessment. They were also responsible for setting and executing the guidelines for recording the locations around ISAGEN facilities; company that I would also like to thank. Thanks also to the Alexander von Humboldt institute, for providing another set of recordings for the validation of my work.

Finally, I would like to thank the friends I made during the past couple of years, who made this experience much more pleasant and to my boyfriend and family, who supported me with their unconditional love during this time of my life and will continue to do so for the rest of it.
Abstract

Ecoacoustics has become a field of growing interest for ecosystem monitoring. Its main advantages over traditional methods include cost effectiveness, non-invasiveness and simplicity; besides the distribution of many recorder units makes possible the recollection of more information. However, for long term studies, the quantity of collected data makes the manual inspection of recordings a cumbersome task, leading to reduced analysis. As an alternative to manual inspection, a series of indices have been proposed to summarize the acoustical information in recordings. Nonetheless, these indices have been applied mainly to biodiversity studies and their connection to ecosystem state is still not clear.

In this work we trusted ANOVA robustness for non-normal data for proposing a methodology that selected the best acoustical indices or features for a specific application and used them to model the ecosystem soundscape patterns with hidden Markov models and Gaussian mixture emissions (GMMHMM). Additionally, the set of input features included by default a biodiversity indicator per 1kHz band.

This methodology was applied to two Colombian cases with defined ecosystem types. In the first case, a series of forest, stubble and pasture recordings were collected for over a year in the east of Antioquia. The second application aimed to find the soundscape patterns of dry forests transformations in two regions of the Colombian Caribbean. The model identified six and three soundscape patterns for the first and second dataset respectively. In the first application, continuous sounds, high biophonic intensity and multiple occupied frequency bands were found in the patterns associated to forest sites; on the other hand, stubble sites presented more general entropy, which we related to high geophonic presence, preventing biophonic activity. Lastly, pasture soundscapes alternated between periods of high geophony and high frequency complexity, making it an intermediate ecosystem in the acoustical sense. The adaptation of the model for classification resulted in the identification of 81% of the forest samples, 96.6% of the stubble samples and 51.2% of the pasture samples.

The classification results for the second application were not as high, with 68% for the low transformation samples, 58.8% for the medium transformation and 31.8% for the high transformation. Nonetheless, the confusion matrices indicated that the training samples were not enough, and more sampling should be provided for attaining better results.

Given that GMMHMM is a sequential model, it also presented the temporal configuration of the acoustical patterns by their transition probabilities. This feature allowed us to emphasize the importance of conservation, when we found that the most stable and inaccessible states were associated to the most acoustically diverse ecosystems.
Resumen

La ecoacústica se ha convertido en un área de creciente interés para el monitoreo de ecosistemas. Entre las principales ventajas que presenta sobre las técnicas tradicionales se encuentran su bajo costo, poca afectación al entorno y simplicidad; además de que la distribución de varias grabadoras hace posible la recolección de más información. Sin embargo, para estudios de largo plazo, la cantidad de datos hace que la inspección manual de las grabaciones sea una tarea tediosa y por consiguiente el análisis sea limitado. Como alternativa a la inspección manual, una serie de índices han sido propuestos para resumir la información acústica de las grabaciones. No obstante, estos índices han sido aplicados principalmente a estudios de biodiversidad y su relación con el estado del ecosistema no es claro aún.

En este trabajo se confió en la robustez del ANOVA frente a datos que no se distribuyen normalmente para proponer una metodología de selección de los mejores índices o descriptores acústicos para una aplicación específica y usarlos para modelar los patrones del paisaje acústico del ecosistema con modelos ocultos de Markov y emisiones por mezclas Gaussiánas (GMMHMM). Además, el conjunto de descriptores que entran al modelo incluye por defecto un indicador de biodiversidad para cada banda de 1kHz.

Esta metodología fue aplicada a dos casos colombianos con tipos de ecosistema definidos. En el primer caso, una serie de grabaciones de bosque, rastrojo y pastizal fueron colectados por más de un año en el este de Antioquia. La segunda aplicación buscaba encontrar patrones de paisaje acústico de las transformaciones de bosque seco en dos regiones del caribe colombiano. El modelo identificó seis y tres patrones acústicos para la primera y segunda base de datos respectivamente. En la primera aplicación, se encontraron sonidos continuos, alta intensidad biofónica y ocupación de varias bandas en los patrones asociados a bosque, mientras que en los rastrojos se presentó más entropía, que se relaciona con alta presencia geofónica, lo que limita la actividad biofónica. Finalmente, los paisajes acústicos de pastizal alternaron entre periodos de alta geofonía y alta complejidad frecuencial, haciéndolo un ecosistema intermedio en el sentido acústico. La adaptación del modelo para clasificación resultó en la identificación del 81% de las muestras de bosque, 96,6 % de las muestras de rastrojo y 51,2 % de las muestras de pastizal.

Los resultados de clasificación para la segunda aplicación no fueron altos, con 68% para las muestras de baja transformación, 58,9% para la transformación media y 31,8% para la transformación alta. No obstante, las matrices de confusión indicaron que las muestras de entrenamiento no fueron suficientes, y que debería proporcionarse mayor muestreo para obtener mejores resultados.

Dado que GMMHMM es un modelo secuencial, también presentó la configuración temporal de los patrones acústicos dadas sus probabilidades de transición. Esta característica nos permitió destacar la importancia de la conservación, cuando encontramos que los estados más estables e inaccesibles fueron asociados a los ecosistemas más diversos acústicamente.
3.2.5. MODEL SET UP RESULTS ................................................................................................................................. 37
3.2.6. MODEL IMPLEMENTATION RESULTS .............................................................................................................. 39

CHAPTER 4. CONCLUSIONS ........................................................................................................................................... 42
4.1. SUMMARY ............................................................................................................................................................. 42
4.2. CONTRIBUTIONS OF THIS WORK ........................................................................................................................ 43
   4.2.1. AUTOMATIC REMOVAL OF NOISY RECORDINGS DUE TO GEOPHONY ..................................................... 43
   4.2.2. ANALYSIS OF ECOACOUSTIC INDICES SIMILARITY ................................................................................... 43
   4.2.3. AUTOMATIC RECOGNITION OF DAILY SOUNDSCAPE PATTERNS .............................................................. 44
   4.2.4. TEMPORAL RELATION OF SOUNDSCAPE PATTERNS .................................................................................. 44
4.3. PUBLICATIONS ....................................................................................................................................................... 44
4.4. FUTURE WORK ......................................................................................................................................................... 45

REFERENCES ............................................................................................................................................................... 47
List of Figures

Figure 2.1 Block Diagram of the Proposed Methodology to Detect Ecosystem Acoustical Patterns.................. 5
Figure 2.2 Two spectrogram segments of discarded recordings in the Humboldt case. The upper segment
  presented rain noise, while the lower segment presented wind noise. The y axis represents frequency, the x
  axis represents time and the coloring represents the intensity of the signal........................................... 6
Figure 2.3 HMM Diagram. The Markov chain is represented in the upper half of the figure, with its hidden states
  (denoted by the blue circles) and transition probabilities (denoted by the curved arrows). In the lower part
  of the figure, the observed sequence is produced by the Markov chain given the emission probabilities (the
  observed samples are denoted by purple squares and the emission probabilities are denoted by the arrows
  connecting the states to the observations). .................................................................................................. 15
Figure 2.4 Gaussian mixture model example with three components. It shows that three different distributions of
  the observed samples could be assigned to the same hidden state............................................................. 16
Figure 2.5 Model set up diagram. This flowchart describes the process for the estimation of the model
  parameters. The purple items describe the first part of the process, where \( N_r \) and \( N_{\text{CC}} \) are selected. The green
  items describe the second part of the process where the remaining model parameters (\( A, B \) and \( \pi \)) are
  obtained. Note that \( C \) is the number of tested values for \( N_{\text{CC}} \), \( G \) is the number of tested values for \( N_r \) and \( R_g \)
  is the number of times that the Baum-Welch algorithm is executed, and the SSE is computed. ..................... 17
Figure 3.1 Google Earth screenshot of the monitored area. Yellow pin color indicates a pasture site, green pin
  color indicates a forest site and red pin color indicates a stubble site. The closest weather station is also
  marked. ...................................................................................................................................................... 22
Figure 3.2 Spectrogram segment of a discarded recording. Observe the vertical black lines towards the end of the
  segment indicating a microphone disconnection................................................................. 24
Figure 3.3 F-value for the selected features in the Jaguas application. ............................................................. 26
Figure 3.4 Grouped boxplots by ecosystem type for the three soundscape features. ................................. 26
Figure 3.5 Mean vectors for the resulting clusters and the clusters’ prototypes in the Jaguas application. ...... 29
Figure 3.6 Training distribution of ecosystem types in the resulting clusters.................................................. 31
Figure 3.7 Temporal Classification of samples per site in the Jaguas case........................................................ 32
Figure 3.8 F-value for the selected features in the Humboldt application....................................................... 36
Figure 3.9 Grouped boxplots by ecosystem type for the three soundscape features in the Humboldt application. 37
Figure 3.10 Mean vectors for the resulting clusters in the Humboldt application. ....................................... 38
Figure 3.11 Training distribution of ecosystem transformations in the resulting clusters. ............................. 40
Lists of Tables

Table 2.1 Calculation, description and reference of the extracted acoustical features. .................................................. 8
Table 3.1 Description of the recording sessions for the Jaguas application. The rows are colored according to their ecosystem type. ........................................................................................................... 21
Table 3.2 Results for the preprocessing stage in the Jaguas application. .............................................................................. 23
Table 3.3 PCC matrix for the recorders and weather station (IDEAM). The cells are colored according to their value, greener cells indicate high PCC values; redder cells indicate low PCC values. ........................................ 23
Table 3.4 Discarded features with their correspondent correlated feature, their MAPCs and their PPCs ............... 25
Table 3.5 Final feature vector for characterizing a daily sample in the Jaguas application. ........................................... 27
Table 3.6 MSE values for all the possible combinations of N_r and N_m in the Jaguas application. The table cells are colored according to their values. Greener cells indicate lower MSE and redder cells indicate higher MSE. ........................................................................................................... 28
Table 3.7 Transition matrix (A) of the resulting model in the Jaguas case. The initial probability for each state (π) is represented by row 0. Each cell represents the transition probability of the row state (i) to the column state (j). The greener the cell color, the higher probability for that transition, and the red cells indicate low transition probability. Notice that the row’s sum equals one ................................................................. 28
Table 3.8 Cluster’s label according to their training data distribution in the Jaguas case. ...................................................... 31
Table 3.9 Normalized confusion matrix for the training and testing Jaguas datasets. Notice that the sum of the rows equals one. Greener cells indicate higher values, redder cells indicate lower values. ................................. 31
Table 3.10 Description of the selected recording sessions for the Humboldt Application. The rows are colored according to their transformation value. ........................................................................................................... 34
Table 3.11 Final feature vector for characterizing a daily sample in the Humboldt application. .................................... 37
Table 3.12 MSE values for all the possible combinations of N_r and N_m in the Humboldt application. The table cells are colored according to their values. Greener cells indicate lower MSE and redder cells indicate higher MSE. ........................................................................................................... 38
Table 3.13 Transition matrix (A) of the resulting model in the forest transformation case. The initial probability for each state (π) is represented by row 0. Each cell represents the transition probability of the row state (i) to the column state (j). The greener the cell color, the higher probability for that transition, and the red cells indicate low transition probability. Notice that the row’s sum equals one ...................................................................................... 39
Table 3.14 Cluster’s label according to their training data distribution in the Humboldt case. .............................................. 39
Table 3.15 Normalized confusion matrix for the training and testing Humboldt datasets. Notice that the sum of the rows equals one. Greener cells indicate higher values, redder cells indicate lower values. ................. 40
Conventions Index

\( \sim \) New estimated value for a variable
\( \prime \) Transpose of an array variable
\( \bar{\ } \) Mean value for a variable
\([\ ]\) Nearest integer greater than a variable
\([\ ]\) Nearest integer smaller than a variable
\( \alpha \cdot \alpha_i(t) \) Technophony level - Joint probability for a partial starting sequence finishing in the state \( i \) and time \( t \)
\( a_{ij} \) Probability of transition from state \( i \) to state \( j \)
\( A \) Transition matrix of \( \lambda \)
\( ACIf_t \) Acoustic Complexity Index (frequency-time)
\( ACIf_{te} \) ACIf evenness
\( ACIt_t \) Acoustic Complexity Index (time-frequency)
\( ACIt_{te} \) ACIt evenness
\( ADI \) Acoustic Diversity Index
\( ADI_b \) Modified acoustic diversity index for band \( b \)
\( AEI \) Acoustic Evenness Index
\( ANOVA \) Analysis of Variance
\( \beta \cdot \beta_i(t) \) Bioacoustics Index, biophony level - Joint probability for the partial ending sequence starting at \( t + 1 \) and finishing at \( T \), given state \( i \) and time \( t \)
\( b_i(y_t) - b_i(y_t) \) A specific frequency band – the emission probability of state \( i \) to observation \( y_t \)
\( B \) Emission Probability Distribution of \( \lambda \)
\( BN_{f(j)} \) Frequency Background Noise (in the bin \( j \))
\( BN_t \) Temporal Background Noise
\( y_i(t) \) Probability of being in state \( i \) at time \( t \)
\( y_{im}^{e^t}(t) \) Probability for the \( e^t \) sequence that the \( m^t \) component of the \( i^t \) state generated observation \( y_t \)
\( c \) Number of temporal clusters in the recording
\( C \) Number of likely values for \( N_m \)
\( CF \) Crest Factor
\( \delta_t(i) \) Highest probability for the first \( t \) observations in the sequence ending in state \( i \)
\( d \) Audio signal duration – if preceding another variable, it represents the variable’s derivative
\( D \) Signal digitalization depth
\( e \) A specific sequence
\( \varepsilon_{(i)} \) Hilbert transform (evaluated in time step \( i \))
\( E \) Number of sequences
\( E_{peak} \) Peak energy value of the signal
\( \phi(y_t|\mu_{im}, \Sigma_{im}) \) The probability of \( y_t \) to be generated by a multivariate gaussian distribution with mean vector \( \mu_{im} \) and covariance matrix \( \Sigma_{im} \)
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_j )</td>
<td>A specific frequency value (correspondent to frequency bin ( j ))</td>
</tr>
<tr>
<td>( F_f )</td>
<td>( P_f ) for ( P_f ) values higher than a certain threshold, 0 otherwise</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>FM</td>
<td>Frequency Modulation</td>
</tr>
<tr>
<td>( F_S )</td>
<td>Frequency of sampling for an audio signal</td>
</tr>
<tr>
<td>( \zeta )</td>
<td>A specific recording session</td>
</tr>
<tr>
<td>( g )</td>
<td>A specific group (label)</td>
</tr>
<tr>
<td>( G )</td>
<td>Number of likely values for ( N_c )</td>
</tr>
<tr>
<td>GHA</td>
<td>Grupo Herpetológico de Antioquia</td>
</tr>
<tr>
<td>GMMHMM</td>
<td>Hidden Markov Models with Gaussian mixture emissions</td>
</tr>
<tr>
<td>H</td>
<td>Acoustic Entropy Index</td>
</tr>
<tr>
<td>( H_m )</td>
<td>Entropy of spectral maxima</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Model(s)</td>
</tr>
<tr>
<td>( H_s )</td>
<td>Spectral Entropy</td>
</tr>
<tr>
<td>( H_t )</td>
<td>Temporal Entropy</td>
</tr>
<tr>
<td>( H_v )</td>
<td>Entropy of spectral variance</td>
</tr>
<tr>
<td>( i )</td>
<td>A specific time step or state in the HMM context</td>
</tr>
<tr>
<td>IDEAM</td>
<td>Colombian climate authority</td>
</tr>
<tr>
<td>( l_{i,j} )</td>
<td>Normalized amplitude of each pulse in the time step ( i ) and frequency bin ( j )</td>
</tr>
<tr>
<td>( j )</td>
<td>A specific frequency bin or state in the HMM context</td>
</tr>
<tr>
<td>( k )</td>
<td>A specific temporal cluster for ACIt calculation</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>A certain HMM</td>
</tr>
<tr>
<td>( \Lambda_j )</td>
<td>Denotes major peaks in the mean spectrum</td>
</tr>
<tr>
<td>( \mu_{i(m)} )</td>
<td>Mean vector (of the ( m^{th} ) gaussian distribution) for the state ( i )</td>
</tr>
<tr>
<td>( m )</td>
<td>Frequency bins at the left of ( j ) – A certain gaussian component</td>
</tr>
<tr>
<td>( l )</td>
<td>Frequency bins surrounding ( j )</td>
</tr>
<tr>
<td>M</td>
<td>Median of amplitude envelope</td>
</tr>
<tr>
<td>( MAPC_{(Y)} )</td>
<td>Mean absolute Pearson coefficient for feature ( Y )</td>
</tr>
<tr>
<td>MD</td>
<td>Musicality Degree</td>
</tr>
<tr>
<td>MFCC</td>
<td>Mel Frequency Cepstral Coefficients</td>
</tr>
<tr>
<td>MID</td>
<td>Mid-band activity</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Squared Error</td>
</tr>
<tr>
<td>nan</td>
<td>Not a number</td>
</tr>
<tr>
<td>( N_b )</td>
<td>Total number of frequency bands to be considered</td>
</tr>
<tr>
<td>( N_c )</td>
<td>Number of states in the model</td>
</tr>
<tr>
<td>NDSI</td>
<td>Normalized Difference Soundscape Index</td>
</tr>
<tr>
<td>( nfft )</td>
<td>Number of points for the FFT</td>
</tr>
<tr>
<td>( N_f )</td>
<td>Total number of frequency bins</td>
</tr>
<tr>
<td>( N_g )</td>
<td>Number of groups</td>
</tr>
<tr>
<td>( N_m )</td>
<td>Number of gaussian components in the GMMHMM</td>
</tr>
<tr>
<td>NP</td>
<td>Number of peaks</td>
</tr>
<tr>
<td>( N_t )</td>
<td>Total number of temporal steps</td>
</tr>
</tbody>
</table>
\( N_{tr} \) \hspace{1cm} \text{Total number of recordings in the training set}
\( \xi_{ij}(t) \) \hspace{1cm} \text{Probability of being in state } i \text{ and } j \text{ in time } t \text{ and } t+1 \text{ respectively}
\( \pi_i(t) \) \hspace{1cm} \text{Initial probability distribution of } \lambda \text{ (initial probability for state } i \text{)}
\( p_b(d) - p_{im} \) \hspace{1cm} \text{Occupation of the frequency band } b \text{ (or } d \text{) – the weight for the } m\text{th} \text{ gaussian distribution in the } i\text{th state}
\( \hat{p} \) \hspace{1cm} \text{Estimated probability for the given sequence}
\( PCC_{(YZ)} \) \hspace{1cm} \text{Pearson Correlation Coefficient (for } Y \text{ and } Z \text{)}
\( P_f \) \hspace{1cm} \text{Temporal spectrogram average for frequency } f
\( PSD(n_{r\xi}) \) \hspace{1cm} \text{Power Spectral Density (averaged over the band between 600 Hz and 1200 Hz for recording } r \text{ in session } \xi \text{)}
\( \rho \) \hspace{1cm} \text{Ratio of biophony to anthrophony}
\( r \) \hspace{1cm} \text{A specific recording}
\( R_t \) \hspace{1cm} \text{Total number of recordings in the session } \xi \text{)}
\( R_{sg} \) \hspace{1cm} \text{Number of times that the Baum-Welch algorithm is run}
\( RMS \) \hspace{1cm} \text{Root mean square}
\( ROI \) \hspace{1cm} \text{Region of Interest}
\( \sigma_{eq} \) \hspace{1cm} \text{Energy-equivalent sound pressure level standard deviation}
\( \sigma_L \) \hspace{1cm} \text{Instantaneous sound pressure level standard deviation}
\( \sigma_Y(Z) \) \hspace{1cm} \text{Variance for index } Y \text{ (} Z \text{)}
\( \sigma_{Y,Z} \) \hspace{1cm} \text{Covariance for indices } Y \text{ and } Z
\( \Sigma_{im} \) \hspace{1cm} \text{Covariance matrix of the } m\text{th} \text{ gaussian distribution in the state } i
\( S_j \) \hspace{1cm} \text{Mean spectrum value for frequency bin } j
\( s_{i,w} \) \hspace{1cm} \text{Spectrogram cell in the time step } i \text{ and frequency bin } w
\( SF \) \hspace{1cm} \text{Spectral Flatness}
\( SPL \) \hspace{1cm} \text{Sound Pressure Level signal}
\( SSE \) \hspace{1cm} \text{Sum of Squared Errors}
\( STFT \) \hspace{1cm} \text{Short-Time Fourier Transform}
\( t \) \hspace{1cm} \text{A specific time in the sequence}
\( T \) \hspace{1cm} \text{Time for the last element in the sequence}
\( T_e \) \hspace{1cm} \text{Time duration for sequence } e
\( thr_{\xi} \) \hspace{1cm} \text{Threshold for heavy rain detection in the session } \xi \text{)}
\( TLSV \) \hspace{1cm} \text{Temporal Sound Level Variance Descriptor}
\( U_j \) \hspace{1cm} \text{Maximum value of the spectrogram in the frequency bin } j
\( v_t \) \hspace{1cm} \text{Assigned state to the observation } y_t
\( V_j \) \hspace{1cm} \text{Spectrogram variance along frequency bin } j
\( \omega \) \hspace{1cm} \text{Window size for the spectrogram computation}
\( w \) \hspace{1cm} \text{A specific frequency bin (different to } j \text{)}
\( W_j \) \hspace{1cm} \text{Welch’s power spectral density estimate (for frequency bin } j \text{)}
\( x \) \hspace{1cm} \text{Raw audio signal}
\( \psi_t(i) \) \hspace{1cm} \text{Temporary vector keeping track of the state sequence that maximizes } \delta_t(i)\text{)}
\( Y(Z) \) \hspace{1cm} \text{Denotes a specific index}
\( \bar{y} \) \hspace{1cm} \text{Index mean for all samples}
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_g$</td>
<td>Index mean for the group $g$</td>
</tr>
<tr>
<td>$y_{gr}$</td>
<td>Index value for the recording $r$ in the group $g$</td>
</tr>
<tr>
<td>$y_t$</td>
<td>Observation vector in time $t$</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction

1.1. Problem Description and Justification

Sound, as an important indicator for ecosystem activity, can be of great help to monitor the natural periodicity in the ecosystem and its degradation due to industrial pollution or climate change (Joo, Gage, & Kasten, 2011; B. L. Krause & Farina, 2016). Currently, there are many tools to automatize the collection of recordings and analyze the data (Aide et al., 2013; Kasten, Gage, Fox, & Joo, 2012; Merchant et al., 2015; Sueur, Farina, Gasc, Pieretti, & Pavoine, 2014) but given the diversity of applications and the different approaches to address them, there is still not a clear consensus about the best methodology for extracting the ecosystem acoustical patterns.

Summarizing the soundscape behavior for specific sites generally relies on the analysis of certain acoustic metrics or indices. These analysis have been carried on visually (Gage & Axel, 2014; Gaitán Forero, 2017; Job, Myers, Nghishineh, & Gill, 2016; Razali et al., 2015; Mangalam Sankupellay, Towsey, Truskinger, & Roe, 2015) or applying statistical tools like correlation or linear models (Fuller, Axel, Tucker, & Gage, 2015; Gage, Wimmer, Tarrant, & Grace, 2017; Jorge, Machado, da Cunha Nogueira, & Nogueira-Filho, 2018; Bernie Krause, Gage, & Joo, 2011; Pekin, Jung, Villanueva-Rivera, Pijanowski, & Ahumada, 2012; Tucker, Gage, Williamson, & Fuller, 2014). Although some of the studies have employed machine learning techniques for grouping the acoustic measurements (Mangalam Sankupellay et al., 2015; Tucker et al., 2014); these techniques have been used rather rarely for soundscape pattern recognition (Bormpoudakis, Sueur, & Pantis, 2013; Mullet, Gage, Morton, & Huettmann, 2016).

Few of the studies cited above analyze temporal variation of soundscapes beyond daily fluctuations. Those that do, seemed to describe accurately the respective seasonal soundscape patterns, but some presented analysis lacking predictive power; then, for these methodologies to detect early anomalies in the ecosystem acoustic signature is more difficult (Gage & Axel, 2014; Bernie Krause et al., 2011). Only one study proposed predictive models for each of the soundscape components, instead of building a model for the soundscape as a whole (Mullet et al., 2016). In this work, we proposed such a model by the application of predictive machine learning algorithms.

The region where the recordings are collected is another factor to consider. The acoustic richness is expected to increase as we approach to the Equator, as this region favors the conditions for life and diversity of species (Gaston, 2000). Nonetheless, few ecoacoustic studies are made on equatorial ecosystems.

In this work, we describe a new methodology for assessing the temporal and structural changes in tropical soundscapes. For this purpose, two Colombian databases were used for building the model; the first one was collected from the western mountain chain of the Andes; the second one contained recordings from the Caribbean dry forests. Firstly, the acoustic metrics were selected according to the ecosystem types of each database; then a model based on hidden Markov models (HMM) (Rabiner, 1989) was trained. HMM is a statistical model that can be fit as a machine learning model; it also groups the samples by acoustical similarity
and establishes relations between groups (components or states), in order to predict the most likely future state for a given soundscape pattern.

1.2. Background

Sound is an important indicator in ecological studies. Through sound, many species can interact with the environment, mark their territory, start mating rituals, and detect preys or predators (Mcwilliam & Hawkins, 2013). The study of animal communication by analyzing the sounds they produce has become into a science called Bioacoustics (Fletcher, 2007). Thanks to bioacoustics applications, it has been possible to discover new aspects of animal behavior, such as the differences between female and male vocalizations (Langbauer, 2000) or preferred position for making alarm calls (Collier, Blumstein, Girod, & Taylor, 2010), etc.

The development of sophisticated tools for efficient recording has allowed the enhancement of bioacoustics studies, leading to the analysis of higher organizational levels in ecology. In the past, the studies focused mainly on species behavior (Erbe, King, Yedlin, & Farmer, 1999; Leong, Ortolani, Burks K. D, Mellen, & Savage, 2003; Marcellini, 1974), now there are multiple studies at the population and community level (Baker & Logue, 2003; De Solla, Fernie, Barrett, & Bishop, 2006; Hobson, Rempel, Greenwood, Turnbull, & Van Wilgenburg, 2002; Tobias et al., 2010). Currently, the goal is to take the next step and extract information from the ecosystem level.

To study the ecosystem acoustic dynamics, it is necessary to include in the analysis sounds from non-biological sources, which are proven to modify the communication patterns in living organisms (B. Krause, 1987) and even help them choose their habitat (Slabbekoorn & Bouton, 2008). In addition, the seasonality of ecosystem soundscape must be characterized, to understand its natural cycles and recognize atypical behavior. For the first purpose, three types of sounds are to be analyzed: biological sounds, or biophony; sounds produced by physical phenomena, such as wind, rain, etc. or geophony, and sounds produced by human-made machinery, which are called technophony or anthrophony. The collection of these sounds is the definition of soundscape for ecological studies (Pijanowski, Villanueva-Rivera, et al., 2011).

The study of soundscape surpasses the bioacoustics domain, because this area is focused mainly on biological sound. On the other hand, the area of soundscape ecology is specifically addressed to landscape ecology studies (Pijanowski, Villanueva-Rivera, et al., 2011). Hence, a new term was needed to enclose the analysis of soundscape for ecosystem studies, and it has now been defined as Ecoacoustics (Sueur & Farina, 2015). Consequently, ecoacoustics concerns both bioacoustics and soundscape ecology and other spatiotemporal, large-scaled studies intended to respond questions about the ecosystem through sound.

For ecoacoustics studies to include spatiotemporal ecosystem change, a big amount of data is needed, because the perturbation gradient is only visible through long periods of time (Magurran et al., 2010). Additionally, big quantities of data demand more accuracy for acoustical measurements. Over the past few years, many acoustic metrics have been proposed to describe the ecological components in sound and summarize acoustical data (Sueur et al., 2014). Most of them were design to estimate biodiversity (Pieretti, Farina, & Morri, 2011; Sueur, Pavoine, Hamerlynck, & Duvail, 2008; Villanueva-Rivera, Pijanowski, Doucette, & Pekin, 2011), while only a few
consider other soundscape elements in the analysis (Kasten et al., 2012; Qi, Gage, Joo, Napoletano, & Biswas, 2008). Although these indices convey important information about the ecosystem dynamics, their relation to ecosystem state and communities’ distribution is not clear yet.

In other acoustical sciences such as psychoacoustics and acoustic design, the relationship between the listeners and the acoustical surroundings is explored (Everest & Shaw, 2001; Truax, 1998). Particularly, in psychoacoustics, three components of sound are considered to be the most important for analyzing its effects on people: intensity, frequency changes and temporal structure (De Coensel, Botteldooren, Debacq, Nilsson, & Berglund, 2007). Although these studies focus on human perception of sound, other studies have adapted some psychoacoustical metrics to measure other species hearing (Bedoya, Isaza, Daza, & López, 2014; Blumstein et al., 2011; Delaney, Grubb, Beier, Pater, & Hildegard Reiser, 1999; Fuller et al., 2015). Given that most of ecoacoustics indices are based on spectrograms, which is a representation of sound intensity in time and frequency, the combination of these three components should be a good criterion to find the right model for soundscape characterization.

Several types of techniques have been applied to build an ecoacoustic model for ecosystem assessment. Some are based on statistical models (Fuller et al., 2015; Bernie Krause et al., 2011; Tucker et al., 2014) and others are based on machine learning techniques (De Coensel, Botteldooren, Debacq, Nilsson, & Berglund, 2008; Mullet et al., 2016; Rychtáriková & Vermeir, 2013). Machine learning techniques are being preferred recently for their versatility, practicality and adaptability to any type of problem. A subgroup of these techniques, known as unsupervised learning or clustering, are especially useful for problems where the data patterns are not clear, like soundscape analysis, because they are able to automatically identify clusters by grouping similar data (Bishop, 2006).

However, in ecoacoustic modeling studies, few machine learning approaches have taken into account the temporal dimension of data, which is key for seasonal soundscape characterizing or degradation detection. A robust model should include time series analysis for interpreting correctly the dynamism of ecosystem soundscape through time. One of the most popular method for time series analysis is hidden Markov models (HMM) (Rabiner, 1989). Hidden Markov models (HMM) are based on Markov chains, a stochastic model that establishes that the current state depends on the previous one. They have been used in many bioacoustics applications, leading to high performance in species vocalizations recognition (Agranat, 2009; Aide et al., 2013; Erbe et al., 1999; Kirschel et al., 2009; Kogan & Margoliash, 1998). In this work, it is explored for first time to assess the ecosystem change through space and time.

### 1.3. Objectives

#### 1.3.1. General Objective

Develop a methodology to analyze ecosystem changes by their soundscape, considering environmental factors and the distribution of its communities.
1.3.2. Specific Objectives

- Analyze current ecoacoustic indices to assess their main strengths and drawbacks.

- Propose a new methodology based on ecoacoustics for a structured analysis of the ecosystem soundscape.

- Test the proposed methodology on a large dataset with known ecosystem changes and compare the results to other approaches.
Chapter 2. Proposed Methodology

The proposed methodology for ecosystem soundscape characterization consists of 3 stages: in the first stage, or preprocessing stage, the noisiest recordings are discarded. Then, in the features stage, soundscape features are extracted and selected, and the acoustic diversity indices are calculated for biodiversity estimation. Finally, in the modeling stage, the parameters of the hidden Markov model with Gaussian mixture emissions (GMMHMM) are set up and the model is trained and tested with the data (Figure 2.1). In this chapter, each of the stages is introduced and justified and the details of the implementation are explained.

Figure 2.1 Block Diagram of the Proposed Methodology to Detect Ecosystem Acoustical Patterns

2.1. Preprocessing

2.1.1. Recordings Filtering

Loud geophonic and technophonic sounds can modify animal communication patterns (Slabbekoorn, 2004; Towsey, Wimmer, Williamson, & Roe, 2014). It has been proven that these sounds can prevent individuals of making calls (Lengagne & Slater, 2002) or provoke the increase of callings to overcome the acoustical obstacles (Lengagne, Aubin, Lauga, & Jouventin, 1999). Wind and rain are specially problematic, because they occupy a large range of the spectrum, masking the calls of many species (Qi et al., 2008). Therefore, in order to identify ecosystem change by acoustic analysis, the common practice is to discard the recordings with high levels of these sounds (Mangalam Sankupellay et al., 2015).

Manual inspection of recordings for this purpose could be tedious for large datasets. In this work, we estimated and implemented an automatic threshold to the algorithm proposed by Bedoya et al. (Bedoya, Isaza, Daza, &
López, 2017) for detecting the recordings containing heavy rain or wind. As mentioned in their work, we found that these recordings had the highest levels of power spectral density (PSD) in the band between 600 and 1200 Hz (not only for rain, but also for wind, see Figure 2.2). For automating this process, we tested the geometric and arithmetic mean as thresholds for filtering out the noisiest recordings per session and discovered that the best threshold was the average of the two means (see Eq. (2.1)).

\[
thr_\zeta = \frac{\sum_{r}^{R_\zeta} PSD_{r,\zeta}}{R_\zeta} + \sqrt{\prod_{r}^{R_\zeta} PSD_{r,\zeta}}
\]

Where \(thr_\zeta\) is the threshold per session, \(PSD_{r,\zeta}\) is the mean PSD in the band between 600 and 1200 Hz for each recording \(r\) in the session \(\zeta\) and \(R_\zeta\) is the number of recordings per session. A recording session is a set of consecutive recordings made by the same recorder, in the same location (during this period the recorder was not reconfigured or recharged). It was necessary to compute the threshold per session, because we noticed that, even though recorders were supposed to be set with the same specifications, there were considerable differences in intensity, which could be due to different models of recorders, microphones, or misfunctioning.

**Figure 2.2** Two spectrogram segments of discarded recordings in the Humboldt case. The upper segment presented rain noise, while the lower segment presented wind noise. The y axis represents frequency, the x axis represents time and the coloring represents the intensity of the signal.
The recordings with high level noise detected by the algorithm were discarded from the subsequent stages of the analysis.

2.2. Features

In this stage, different sets of features are extracted from the recordings. For acoustic signal characterization, there are a wide variety of proposed features which are to be considered according to the application. Maybe the most popular type of acoustical features is the Mel Frequency Cepstral Coefficients (MFCC) for automatic speech recognition. This technique has been used not only for human speech (Davis & Mermelstein, 1980; Hermansky, 1990; Milner, 2002; Wang & Paliwal, 2003), but also for birds (Kirschel et al., 2009; Kogan & Margoliash, 1998; Lellouch, Pavoine, Jiguet, Glotin, & Sueur, 2014), anurans (Bedoya et al., 2014; Duque-Montoya, Isaza, & Cano Rojas, 2017), and other species calls classification (Mazhar, Ura, & Bahl, 2007; Soltis, Leong, & Savage, 2005). However, the main use of MFCC and other automatic feature extraction techniques (like those used in deep learning) is for supervised applications, where the final clusters are commonly known. For the goal of characterizing acoustical dynamics in the ecosystem, in which we look for identifying patterns that have not been fully described, we chose to use features with physical interpretation, so the resulting clusters could be analyzed using their features profile by the final users of our model.

The users of our model will be the people responsible to study the ecosystem soundscape and set the guidelines for protecting it, as an important part of ecological conservation. These experts traditionally base their studies on identifying the species in an ecosystem and summarizing the findings using biodiversity indices (Magurran et al., 2010). To interpret the acoustical information, like in traditional methods, a series of ecoacoustic indices based on the biodiversity indices have been proposed (Sueur et al., 2014). In this work we used the alpha indices, which describe the diversity within a group by its sound intensity, acoustic complexity and soundscape elements.

Some of these indices have become key elements to many ecoacoustic applications (e.g. Depraetere et al., 2012; Harris, Shears, Radford, & Reynolds, 2016; Sankupellay et al., 2015). However, they can be easily affected by factors such as background noise, distance of sound source to the microphone, number of individuals singing or relative amplitude difference between same kind calls (Sueur et al., 2014). Furthermore, their theoretical interpretation is still unclear (Eldridge, Casey, Moscoco, & Peck, 2015). For this reason, besides these indices, we included in our work acoustic metrics for more general applications, looking to compensate indices bias and complement the soundscape description.

2.2.1. Features Extraction

For selecting the additional features to our problem, we focused on the first source of information for ecoacoustical analysis: the spectrogram. Before the indices were proposed, the spectrogram was the main tool to visualize the acoustic dynamics of a community (Blumstein & Munos, 2005; Kirschel et al., 2009; Leong et al., 2003; Urazghildiev & Clark, 2006) or an ecosystem (Job et al., 2016; Joo et al., 2011) and it is still the first step for the calculation of many indices (Sueur et al., 2014). A spectrogram is an intensity plot of the Short-Time
Fourier Transform (STFT) magnitude (Smith, 2007), it allows to see the intensity distribution of the audio signal in both frequency and time (see Figure 2.2).

If we think intuitively on the components of the spectrogram, we identify three elements: intensity, frequency and time. Thence, we looked for acoustic features that could retrieve information about these three dimensions and describe accurately the spectrogram. The selected acoustic features are mathematically explained, described and referenced in Table 2.1.

Table 2.1 Calculation, description and reference of the extracted acoustic features.

<table>
<thead>
<tr>
<th>Acoustic Indices</th>
<th>Calculation</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acoustic Complexity Index (frequency-time)</strong> (ACIfₜ)</td>
<td>[ \sum_{i=1}^{N_t} \sum_{j=1}^{N_f-1} \left</td>
<td>\frac{I_{i,j} - I_{i,j+1}}{I_{i,j} + I_{i,j+1}} \right</td>
<td>] (2.2)</td>
</tr>
<tr>
<td><strong>ACIfₜ evenness (ACIfₜₑ)</strong></td>
<td>[ \frac{1}{\sum_{i=1}^{N_t} \left( \sum_{j=1}^{N_f-1} \frac{</td>
<td>I_{i,j} - I_{i,j+1}</td>
<td>}{I_{i,j} + I_{i,j+1}} \right)^2} ] (2.3)</td>
</tr>
<tr>
<td><strong>Acoustic Complexity Index (time-frequency)</strong> (ACItₜ)</td>
<td>[ \sum_{k=1}^{c} \frac{\sum_{i=1}^{N_t}</td>
<td>I_{i,j} - I_{i+1,j}</td>
<td>}{\sum_{i=1}^{N_t} I_{i,j}} ] (2.4)</td>
</tr>
<tr>
<td><strong>ACItₜ evenness (ACItₜₑ)</strong></td>
<td>[ \frac{1}{\sum_{k=1}^{c} \left( \frac{\sum_{i=1}^{N_t}</td>
<td>I_{i,j} - I_{i+1,j}</td>
<td>}{\sum_{i=1}^{N_t} I_{i,j}} \right)^2} ] (2.5)</td>
</tr>
<tr>
<td><strong>Acoustic Diversity Index (ADI)</strong></td>
<td>[ - \sum_{b=1}^{N_b} p_b \log_2 p_b ] (2.6)</td>
<td>The Shannon index is used to estimate diversity by using 1 kHz frequency bands as acoustic categories. ( p_b ) is the occupation of the frequency band ( b ), measured as the number of cells in the band.</td>
<td>(Pekin et al., 2012)</td>
</tr>
</tbody>
</table>
| **Acoustic Evenness Index (AEI)** | \[
\frac{\sum_{b_1}^{N_b} \sum_{b_2}^{N_b} | p_{b_1} - p_{d_2} |}{2N_b^2 \bar{p}}
\] (2.7) | The Gini coefficient is applied to obtain acoustic evenness across frequency bands of 1kHz. (Villanueva-Rivera et al., 2011) |
| **Bioacoustics Index (\(\beta\))** | \[
\int_{f=2kHz}^{f=8kHz} P_f
\] (2.8) | Area below the \(P_f\) curve in the biophony frequencies (between 2kHz and 8kHz) in dBA. \(P_f\) is the mean of the spectrogram along the time dimension. This index gives an estimate of the biophonic energy in the recording. (Boelman, Asner, Hart, & Martin, 2007) |
| **Acoustic Entropy Index (H)** | \(H_t H_s\) (2.9) | It is the product of \(H_t\) and \(H_s\) (see below). It ranges between 0 and 1; 0 for pure tones and 1 for white noise. (Sueur et al., 2008) |
| **Entropy of spectral maxima (\(H_m\))** | \[U_j = \max_{i=1}^{N_f} (s_{i,w})\] (2.10) | The Shannon index is applied to the maximum values of each frequency bin in the spectrogram, but only in the band between 482 Hz and 8820 Hz (expanding the biophony band). \(s_{i,w}\) stands for a cell in the spectrogram in the time step \(i\) and frequency bin \(w\) and \(U_f\) is the maximum value in the frequency bin \(j\). (Towsey et al., 2014) |
| **Spectral Entropy (\(H_s\))** | \[
\frac{\sum_{j=1}^{N_f} S_j \log_2 S_j}{\log_2 N_f}
\] (2.12) | The Shannon evenness Index is applied to the average energy \(S_j\) in each frequency bin. (Sueur et al., 2008) |
| **Temporal Entropy (\(H_t\))** | \[
-\frac{\sum_{i=1}^{N_f} e_i \log_2 e_i}{\log_2 N_t}
\] (2.13) | Similarly, to \(H_s\), the Shannon evenness Index is applied to the amplitude envelope of the signal, obtained by applying the Hilbert transform (\(e\)). (Sueur et al., 2008) |
| **Entropy of spectral variance (\(H_v\))** | \[V_j = \frac{\text{var}_{i=1}^{N_f} (s_{i,w})}{\log_2 N_f}\] (2.14) | The Shannon index is applied to the variance value for each frequency bin. \(V_j\) stands for the variance value in the frequency bin \(j\). (Towsey et al., 2014) |

**Notes:**
- \(P_f\) denotes the power frequency.
- \(s_{i,w}\) represents a cell in the spectrogram.
- \(e_i\) denotes the amplitude at each frequency bin.
- \(V_j\) represents the variance value in the frequency bin \(j\).
<table>
<thead>
<tr>
<th>Media of amplitude envelope (M)</th>
<th>(2^{D-1} \text{median}(\epsilon))</th>
<th>The median of the amplitude envelope (obtained by using the Hilbert transform), where (D) is the signal digitalization depth. (Depraetere et al., 2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid-band activity (MID)</td>
<td>(F_f = \begin{cases} \frac{P_f}{P_f} &amp; \text{if } P_f &gt; P_f^\text{max} \ 0 &amp; \text{if } P_f &lt; P_f^\text{max} \end{cases}) (\int_{f=482\text{Hz}}^{f=3500\text{Hz}} F_f)</td>
<td>The area below (P_f) in the mid band (482 Hz – 3500 Hz) for values of (P_f) that surpass a certain threshold. (P_f^\text{max}) corresponds to the values of (P_f) over that threshold (in this case the mean of (P_f)). (Towsey et al., 2014)</td>
</tr>
<tr>
<td>Normalized Difference Soundscape Index (NDSI)</td>
<td>(\frac{\beta - \alpha}{\beta + \alpha})</td>
<td>A normalized measure of the ratio of biophony to technophony. For pure technophony NDSI = -1 and for pure biophony NDSI = 1. (\beta) is the bioacoustics index and (\alpha) is the estimated technophony, which is measured similarly to (\beta) for the frequency band between 200 Hz and 1500 Hz. (Kasten et al., 2012)</td>
</tr>
<tr>
<td>Number of peaks (NP)</td>
<td>(\Lambda_j = \begin{cases} 1 &amp; \text{if } S_j &gt; S_m \cap \frac{</td>
<td>dS_j</td>
</tr>
<tr>
<td>Ratio of biophony to anthrophony ((\rho))</td>
<td>(\frac{\beta}{\alpha})</td>
<td>The first index to estimate the effect of technophony to biophony. (Qi et al., 2008)</td>
</tr>
</tbody>
</table>

**Complementary Acoustic Features**

<table>
<thead>
<tr>
<th>Frequency Background Noise (BN(_f))</th>
<th>(\text{mode}\left(S_{LW}\right))</th>
<th>Mean background noise of the spectrogram frequency bins, based on the mode value for each frequency bin (f). (Towsey, 2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal Background Noise (BN(_t))</td>
<td>(\text{mode}(SPL))</td>
<td>Temporal background noise in dB. Based on the mode value for the immediate sound pressure level signal (SPL). (Towsey, 2013)</td>
</tr>
<tr>
<td>Feature</td>
<td>Formula</td>
<td>Explanation</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Crest Factor (CF)</td>
<td>$\frac{E_{\text{peak}}}{RMS}$</td>
<td>The ratio of the peak value of the energy signal ($E_{\text{peak}}$) to its root mean square value ($RMS$), explained below. The higher the value, the more expected peaks in the energy signal.</td>
</tr>
<tr>
<td>Frequency Modulation (FM)</td>
<td>$\sum_{i=1}^{N_f-1} \sum_{j=1}^{N_t-1} \frac{180^\circ}{\pi} \arctan \frac{l_{i,j+1} - l_{i,j}}{l_{i+1,j} - l_{i,j}} (N_f - 1)(N_t - 1)$</td>
<td>The mean angle of the directional derivatives. The higher the value, the more abrupt changes in intensity.</td>
</tr>
<tr>
<td>Musicality Degree (MD)</td>
<td>$\frac{\sum_{j}^{N_f-1} \log(W_j^2 - W_j^2)}{\log(f_{i+1} - f_i)} N_f - 1$</td>
<td>The mean slope of the $1/f$ curve. $W$ is the Welch’s power spectral density estimate. MD is a measure of the temporal complexity of the signal.</td>
</tr>
<tr>
<td>Root mean square ($RMS$)</td>
<td>$\sqrt{\frac{x^2}{N_f}}$</td>
<td>The root mean square of the raw signal ($x$). It is a measure of the amplitude of the signal.</td>
</tr>
<tr>
<td>Spectral Flatness (SF)</td>
<td>$\frac{N_f}{\sqrt{\prod_{j=1}^{N_f} W_j}} W$</td>
<td>The ratio of the geometric mean of the power spectrum to its arithmetic mean. The higher the value, the more frequency complexity.</td>
</tr>
<tr>
<td>Temporal Sound Level Variance Descriptor (TLSV)</td>
<td>$\sigma_L \sigma_{eq}$</td>
<td>The product of the instantaneous sound pressure level standard deviation ($\sigma_L$) and the energy-equivalent sound pressure level standard deviation ($\sigma_{eq}$). TLSV is a measure of temporal complexity.</td>
</tr>
</tbody>
</table>

All the acoustic indices were computed on the selected recordings, then each site daily averages were calculated and standardized using (2.31).

$$\hat{y} = \frac{y - \bar{y}}{\sigma_Y}$$  \hspace{1cm} (2.31)

Where $\hat{y}$ is the new standardized value, $y$ is the original value, $\bar{y}$ and $\sigma_Y$ are the index mean and the index standard deviation calculated over all samples respectively.

### 2.2.2. Features Selection

In this stage, the most significant indices for our application are selected as features.
2.2.2.1. Data Division

At this point, the data was divided in two subsets: the training data, containing 80 percent of the total data; and the testing data, with the remaining 20 percent. Given the requirements of the clustering method and the purpose of this study, the data division was made in such a way that the chronological order was maintained, and each subset contained samples from each recording sequence as possible. A recording sequence is a set of recordings in the same session with possible gaps not larger than 6 days. The features extraction and selection methods and the model set up (section 2.3.1) were applied only to the training data, and after the model is configured, it is implemented to both the training and testing data. In other words, the complete methodology is executed using only the training data, except for the model implementation stage (section 2.3.2), in which both training and testing data were classified by the resulting model. The data division is suggested to avoid overfitting and obtain a model capable of generalization.

2.2.2.2. Correlation Analysis

A first stage of feature selection was based on correlation analysis. The Pearson Correlation coefficient ($PCC$) was computed among every pair of indices to identify those which were strong correlated, and therefore redundant for the analysis (2.32). For each pair of strongly correlated indices (Absolute value of $PCC$ was 0.9 or higher), only one of the indices were selected for the following stage.

$$PCC_{YZ} = \frac{\sigma_{YZ}}{\sigma_Y \sigma_Z}$$  (2.32)

Where $\sigma_{YZ}$ is the covariance for indices $Y$ and $Z$, $\sigma_Y$ is the variance for index $Y$ and $\sigma_Z$ is the variance for index $Z$.

2.2.2.3. F-value

For selecting the most representative indices for each case, the Analysis of Variance (ANOVA) F-value was calculated for each index. Computing this value requires knowing the prior classification of the data; which for our case, corresponded to the ecosystem type of the site where the audio files were recorded. Then, the F-value is computed as shown in (2.33).

$$\frac{N_{tr}(N_{tr} - N_g) \sum_g (y_g - \bar{y})^2}{(N_g - 1) \sum_g \sum_{r} (y_{gr} - y_g)^2}$$  (2.33)

Where $N_{tr}$ is the total number of recordings in the training set, $N_g$ is the number of groups, $y_g$ is the index mean for the group $g$, $\bar{y}$ is the index mean for all samples and $y_{gr}$ is the index value for the recording $r$ in the group $g$. 
The F-value assesses the discrimination power of the index, i.e. the higher the F-value, the more uniform index values for same group samples, and the more different index values for different group samples. Consequently, the indices with highest F-values were selected as soundscape features, for our hypothesis poses that soundscape differs with ecosystem type.

For this selection we excluded the ADI and $BN_f$ indices, which we used for computing the biodiversity features in the following stage. Then, based on the assumption that each soundscape feature would differentiate at least two groups, the number of soundscape features was determined by (2.34)

$$\binom{N_g}{2}$$

(2.34)

Which is the number of possible pairs between groups, or ecosystem types.

Although the ANOVA poses the normality of the data as a requirement, it is deemed to be robust for non-normal data too (Schmider, Ziegler, Danay, Beyer, & Bühner, 2010), therefore the data normality test was omitted.

### 2.2.3. Biodiversity Acoustic Features

Biophony remains still the most important soundscape element (Gage & Axel, 2014). Whereas these sounds have been traditionally demarcated in the frequency band between 2 kHz and 8 kHz (Boelman et al., 2007; Kasten et al., 2012), this convention has been less applicable to tropical habitats (Sueur et al., 2014) and has already been extended to the 2-11 kHz band for some other studies (Eldridge, Guyot, Moscoso, Johnston, & Eyre-walker, 2018; Gage & Axel, 2014). The band under 2 kHz is by no means lacking biological sounds, especially in the tropic, but it is often dominated by technophony (Fuller et al., 2015; Joo et al., 2011). To encompass biodiversity in the soundscape level completely, in this work we included as features a modified ADI vector for the 0-11 kHz band as descriptors of acoustic diversity.

As other acoustic indices (AEI, $H_g$, $H_v$, $H_m$), ADI is based on the Shannon index, which is used frequently in ecology for richness estimation (Spellerberg & Fedor, 2003; Sueur et al., 2014). This index was proposed originally for measuring the information content in a signal by summing the information of its components (Shannon, 1948). In ADI, these components correspond to 1 kHz frequency bands, which we proposed to analyze separately because it is easier to interpret (An unique value would not say much about the calls distribution in the signal, on the other hand, a value per band would help identify which are the bands with higher information or complexity, hence the most active acoustical species by mapping their calls in the spectrum).

The main advantage for ADI over other Shannon based indices is that it considers background noise as a part of the calculation (Pekin et al., 2012). Therefore, this index should be more robust to technophony effect in the corresponding band. Nonetheless, the main drawback of ADI is that it uses a fixed background noise value for
the whole spectrum. Assuming that this value differs between frequency bins, we calculated a different background noise value per bin, using the original proposed $BN_f$ vector (Towsey, 2013) \(^1\) (2.35).

$$BN_{f_j} = \text{mode}(s_{i\omega}) \tag{2.35}$$

Where $BN_{f_j}$ is the background noise value in the frequency bin $j$, and $s_{i\omega}$ is the spectrogram value for time step $i$ and frequency bin $\omega$.

Using $BN_{f_j}$ the modified ADI is calculated per 1 kHz band as indicated in (2.36)

$$ADI_b = -p_b \log p_b \tag{2.36}$$

Where $ADI_b$ is the modified ADI for the band $b$ and $p_b$ is the occupation in the band $b$, which was measured as the number of spectrogram cells in the band higher than their respective $BN_{f_j}$.

As a result, we obtained 11 biodiversity metrics, which join the soundscape features for the complete feature set per day.

2.3. Modeling

In this stage, the selected features served as inputs to a GMMHMM for modeling ecoacoustical patterns. HMM is a stochastical model in which a sequence of random observations is generated by a set of hidden states (or clusters) given a Markov chain and a emission probability distribution (Bilmes, 1998); this process is illustrated in Figure 2.3. This model assumes Markov property, i.e. the probability of a given state depends uniquely on the previous one (Bilmes, 1998). For denoting this model, (2.37) is used

$$\lambda = (A,B,\pi) \tag{2.37}$$

Where $\lambda$ is the HMM, $A$ is the transition matrix, i.e. the element $a_{ij}$ in $A$ denotes the probability of passing to state $j$ being previously in the state $i$; $B$ is the emission probability distribution, or the distribution that determines the probability that a given state produces a given observation; and $\pi$ is the initial state distribution, or the distribution that sets the probability for each state to be the first in the sequence (Rabiner, 1989).

GMMHMM is a type of HMM in which $B$ follows a gaussian mixture distribution. Therefore, the probability for the state $i$ to emit the observation vector $y_t$ at the time $t$ ($b_i(y_t)$) is given by a weighted sum of normal functions (see (2.38)).

$$b_i(y_t) = \sum_{m=1}^{N_m} p_{im} \phi(y_t | \mu_{im}, \Sigma_{im}), \quad \sum_{m=1}^{N_m} p_{im} = 1 \tag{2.38}$$

---

\(^1\) This is a simplification to the algorithm, to see the complete algorithm, please refer to (Towsey, 2013)
Figure 2.3 HMM Diagram. The Markov chain is represented in the upper half of the figure, with its hidden states (denoted by the blue circles) and transition probabilities (denoted by the curved arrows). In the lower part of the figure, the observed sequence is produced by the Markov chain given the emission probabilities (the observed samples are denoted by purple squares and the emission probabilities are denoted by the arrows connecting the states to the observations).

Where $p_{im}$ is the weight for the $m$th distribution in the state $i$, $\phi(y_t | \mu_{im}, \Sigma_{im})$ is the probability for the observation $y_t$ to be generated by the multivariate gaussian distribution described by the mean vector $\mu_{im}$ and covariance matrix $\Sigma_{im}$, and $N_m$ is the number of components in the mixture. This model has the advantage of fitting complex data but requires the calculation of more parameters for the model. In Figure 2.4, an example with a univariate 3-component-gaussian mixture model is plotted showing that the distribution is shaped by its components and, unlike the normal distribution, in which the data is grouped around a unique value, this model presents three different data groups, allowing each state to be associated to different values of the observed data.

Lastly, this model was chosen because it does not only serve to characterize the hidden states (patterns or clusters) which describe the ecosystem soundscape, but it also recognizes the relation between them, making it a predictable model, suitable for both spatial and temporal analysis.

2.3.1. Model Set Up

In this stage, the optimal GMMHMM parameters $A$, $B$ and $\pi$ are estimated using the Baum-Welch Algorithm and the Mean Squared Error (MSE) applied to the training dataset. Firstly, a set of likely values for $N_c$ (number of states) and $N_m$ (number of mixture components) are established, then the Baum-Welch Algorithm is executed and the sum of squared errors (SSE) is calculated several times for each combination of these values. Using the SSE resultant values, the MSE is computed per combination and the optimal values for $N_c$ and $N_m$ are chosen according to the lowest MSE. Finally, the Baum-Welch algorithm is executed again several times using the selected $N_c$ and $N_m$ values and the final selected model is that with the lowest SSE. The flowchart for the described process is shown in Figure 2.5. The resulting model performance is displayed in Chapter 3.
Figure 2.4 Gaussian mixture model example with three components. It shows that three different distributions of the observed samples could be assigned to the same hidden state.

In the following subsections the model set up process is detailed.

2.3.1.1. Baum-Welch Algorithm

For finding the GMMHMM parameters that best fit to the problem, we applied the commonly used Baum-Welch algorithm to the training data (Bilmes, 1998). This algorithm finds the HMM parameters that maximize the probability for the observed sequence of feature vectors.

Firstly, the parameters $A$, $B$ and $\pi$ are initialized with random values. Then, let $\pi_i$ be the initial probability for state $i$, $b_i(y_t)$ the probability for state $i$ to emit the observation $y_t$, $N_c$ the number of states, $a_{ij}$ the transition probability from state $i$ to state $j$, and $T$ the time for the last element of the sequence. Next, the algorithm is recursively executed using forward, backward and update procedures:

For the forward procedure, we denote $\alpha_i(t)$ as the joint probability for a partial starting sequence finishing in the state $i$ and time $t$. $\alpha_i(t)$ can be obtained as shown in (2.39) and (2.40).
Figure 2.5 Model set up diagram. This flowchart describes the process for the estimation of the model parameters. The purple items describe the first part of the process, where $N_c$ and $N_m$ are selected. The green items describe the second part of the process where the remaining model parameters ($A$, $B$ and $\pi$) are obtained. Note that $C$ is the number of tested values for $N_m$, $G$ is the number of tested values for $N_c$ and $R_G$ is the number of times that the Baum-Welch algorithm is executed, and the SSE is computed.

\[ \alpha_i(1) = \pi_i b_i(y_1) \]  \hspace{1cm} (2.39)

\[ \alpha_j(t+1) = \sum_{i=1}^{N_c} \alpha_i(t) a_{ij} b_j(y_{t+1}) \]  \hspace{1cm} (2.40)

Whereas for the backward procedure we denote $\beta_i(t)$ as the joint probability for the partial ending sequence starting at $t+1$ and finishing at $T$ given state $i$ and time $t$. $\beta_i(t)$ is calculated using (2.41) and (2.42).

\[ \beta_i(T) = 1 \]  \hspace{1cm} (2.41)

\[ \beta_i(t) = \sum_{j=1}^{N_c} \beta_j(t+1) a_{ij} b_j(y_{t+1}) \]  \hspace{1cm} (2.42)

Lastly, in the update step, we denote $\gamma_i(t)$ as the probability of being in state $i$ at time $t$, and it is estimated by (2.43).

\[ \gamma_i(t) = \frac{\alpha_i(t) \beta_i(t)}{\sum_{j=1}^{N_c} \alpha_j(t) \beta_j(t)} \]  \hspace{1cm} (2.43)

In this step it is also defined $\xi_{ij}(t)$ as the probability of being in state $i$ and $j$ in time $t$ and $t+1$ respectively. This probability is estimated by (2.44).
\[
\xi_{ij}(t) = \frac{a_i(t)a_j\beta_j(t+1)b_j(y_{t+1})}{\sum_{i=1}^{N_c}\sum_{j=1}^{N_c}a_i(t)a_j\beta_j(t+1)b_j(y_{t+1})} 
\]

(2.44)

For several sequences, \(a_i(t), \beta_i(t), y_i(t)\) and \(\xi_{ij}(t)\) are calculated for each sequence with different initialization parameters.

Now, if there are \(E\) sequences, and the \(e^{th}\) sequence has a length of \(T_e\), the parameters \(\pi_i\) and \(a_{ij}\) are updated using (2.45) and (2.46)

\[
\hat{\pi}_i = \frac{\sum_{e=1}^{E} \gamma^e_i(1)}{E}
\]

(2.45)

\[
\hat{a}_{ij} = \frac{\sum_{e=1}^{E} \sum_{t=1}^{T_e} \xi^e_{ij}(t)}{\sum_{e=1}^{E} \sum_{t=1}^{T_e} \gamma^e_i(t)}
\]

(2.46)

To estimate \(b_i(y_t)\), let \(\gamma^e_{im}(t)\) be the probability for the \(e^{th}\) sequence that the \(m^{th}\) component of the \(i^{th}\) state generated observation \(y_t\), and it is defined by (2.47)

\[
\gamma^e_{im}(t) = \frac{p_{im}\phi(y_t|\mu_{im},\Sigma_{im})}{b_i(y_t)}
\]

(2.47)

Finally, \(B\) is then updated by replacing (2.48), (2.49) and (2.50) in (2.38)

\[
\hat{\beta}_{im} = \frac{\sum_{e=1}^{E} \sum_{t=1}^{T_e} \gamma^e_{im}(t)}{\sum_{e=1}^{E} \sum_{t=1}^{T_e} \gamma^e_i(t)}
\]

(2.48)

\[
\hat{\mu}_{im} = \frac{\sum_{e=1}^{E} \sum_{t=1}^{T_e} \gamma^e_{im}(t)\gamma^e_i(t)}{\sum_{e=1}^{E} \sum_{t=1}^{T_e} \gamma^e_i(t)}
\]

(2.49)

\[
\hat{\Sigma}_{im} = \frac{\sum_{e=1}^{E} \sum_{t=1}^{T_e} \gamma^e_{im}(t)(\gamma^e_i(t) - \mu_{im})(\gamma^e_i(t) - \mu_{im})'}{\sum_{e=1}^{E} \sum_{t=1}^{T_e} \gamma^e_i(t)}
\]

(2.50)

Baum-Welch algorithm is repeated for all values of \(t\), then executed again until a desired level of convergence. Since this algorithm is gradient-based, it can stop in a local optimum. Therefore, it is suggested to be restarted and executed several times, to generate several models and choose the one with the best score.

### 2.3.1.2. Mean Squared Error (MSE)

Baum-Welch algorithm finds the optimal values for most of GMMHMM parameters, except for the number of clusters (N\(_c\)) and the number of mixture components (N\(_m\)). For finding such values, a 2D-grid with the combinations of likely quantities for these parameters was arranged, then the Baum-Welch algorithm was run for several times for each of the combinations in the grid.

Every time the Baum-Welch algorithm was executed, the SSE was computed for the resulting model. But firstly, \(u_t\) was computed for each observation using Viterbi algorithm (explained in the following section), then the mean vector \(\mu_i\) for the model state \(i\) was calculated using (2.51).
\[ \mu_i = \sum_{m=1}^{N_m} p_{im} \mu_{im} \]  

(2.51)

Let us remember that \( \mu_{im} \) is the mean vector for the \( m^{th} \) distribution that composes the emission probability for state \( i \), and \( p_{im} \) is its mixture weight. Using \( \mu_i \) and \( \nu_t \), SSE can be computed using (2.52)

\[ \sum_{t=1}^{T} (y_t - \mu_i) (y_t - \mu_i)' \quad \forall i = \nu_t \]  

(2.52)

Finally, the MSE value was obtained as an average of all SSE values within the same \( N_c \) and \( N_m \) combination (Rasmussen & Williams, 2004). MSE is a measure of cluster compactness, i.e. the higher the MSE the more variance within the same cluster samples. Accordingly, the chosen values for \( N_c \) and \( N_m \) were the ones that led to the lowest MSE value.

MSE was preferred over maximum likelihood estimation (MLE), Bayesian information criterion (BIC) and Akaike information criterion (AIC) (Bishop, 2006), because the first showed convergence with low-complexity models, whereas the other methods kept increasing their scores with model complexity (i.e. greater values for \( N_c \) and \( N_m \), which would lead to overfitting).

Once the optimal parameters for the model have been found, it is not necessary to repeat the model set up for new similar data, but only classify these data using the Viterbi algorithm (explained below). This is the reason why the data division (section 2.2.2.1) was executed in the first place, to ensure that the model responds correctly to new samples from the same study.

### 2.3.2. Model Implementation

In this stage we apply the resulting GMMHMM model to the complete dataset (both training and testing data) using the Viterbi algorithm (Forney, 1973; Rabiner, 1989).

For the implementation of the algorithm, let \( \delta_t(i) \) be the highest probability for the first \( t \) observations in the sequence ending in state \( i \), and \( \psi_t(i) \) the vector keeping track of the state sequence that maximized said probability. Now, the procedure is developed by the following steps:

1) Initialization:

\[ \delta_1(i) = \pi_i b_1(y_1) \]  

(2.53)

\[ \psi_1(i) = 0 \]  

(2.54)

2) Recursion:

\[ \delta_t(j) = \max_i [\delta_{t-1}(i)a_{ij}] b_j(y_t) \]  

(2.55)

\[ \psi_t(i) = \arg\max_i [\delta_{t-1}(i)a_{ij}] \]  

(2.56)
3) Termination:

\[ \hat{P} = \max_i \delta_T(i) \]  
\[ v_T = \arg\max_i \delta_T(i) \]  

(2.57)  

(2.58)

4) State sequence backtracking:

\[ v_t = \psi_{t+1}(u_{t+1}) \]  

(2.59)

Where \( \hat{P} \) is the estimated probability for the given sequence and \( v_T \) is the probable state that emitted observation \( y_t \). This process is repeated for each of the observed sequences.

As a result, the data is grouped in clusters and their patterns can be now described and analyzed. Notice that the obtained clusters might not correspond to the ecosystem types, but they are meant to describe the acoustical processes that configure the observed soundscape dynamics.
Chapter 3. Results

In this chapter the methodology is applied to two Colombian datasets, and the results for each of the processing stages are described. The final interpretation of the resulting clusters was reviewed by a biologist.

3.1. Jaguas Application

The first dataset was provided by Grupo Herpetológico de Antioquia (GHA)\(^2\) and ISAGEN\(^3\), an energy generation company. The recordings were collected for monitoring the ecosystems surrounding San Lorenzo dam and assess the influence of the Jaguas energy plant in the natural communities and environment.

3.1.1. Database Description

Six sites surrounding the Jaguas hydroelectric power station in Alejandría, Antioquia, were acoustically monitored for this study. A total of 124,989 recordings were collected from April 22\(^{nd}\) of 2016 to May 6\(^{th}\) of 2017 using Wildlife Acoustics\(^4\) Songmeter SM2 recorders (Wildlife Acoustics, 2011). The proprietary recording format WAC could be converted to WAV using Wildlife’s Kaleidoscope software. The resulting recordings had a digital format of 16-bit PCB \((D = 16)\), a sampling frequency of 22,050 Hz \((Fs = 22,050 \text{ Hz})\), one minute of duration \((d = 60 \text{ sec})\) and were programmed to be recorded every 20 minutes. Even though some recorders had two functioning channels during this period, only one channel per recorder was considered.

The sites were chosen and labeled by experts along a perturbation gradient. Each of the sites was assigned with a specific recorder during this monitoring period. In Table 3.1 the location of the monitored sites, their classification and the number of resulting recordings are described.

Table 3.1 Description of the recording sessions for the Jaguas application. The rows are colored according to their ecosystem type.

<table>
<thead>
<tr>
<th>Recorder</th>
<th>Location</th>
<th>Ecosystem Type</th>
<th>Number of recordings</th>
</tr>
</thead>
<tbody>
<tr>
<td>5253</td>
<td>6.356 N, 74.997 W</td>
<td>Stubble</td>
<td>20,256</td>
</tr>
<tr>
<td>5255</td>
<td>6.367 N, 74.997 W</td>
<td>Stubble</td>
<td>24,018</td>
</tr>
<tr>
<td>5256</td>
<td>6.367 N, 75.027 W</td>
<td>Forest</td>
<td>10,119</td>
</tr>
<tr>
<td>5260</td>
<td>6.361 N, 74.998 W</td>
<td>Forest</td>
<td>25,756</td>
</tr>
<tr>
<td>5257</td>
<td>6.375 N, 75 W</td>
<td>Pasture</td>
<td>19,416</td>
</tr>
<tr>
<td>7125</td>
<td>6.375 N, 74.998 W</td>
<td>Pasture</td>
<td>25,424</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td><strong>124,989</strong></td>
</tr>
</tbody>
</table>

\(^2\) [http://grupoherpetologicodeantioquia.org/](http://grupoherpetologicodeantioquia.org/)
\(^4\) [https://www.wildlifeacoustics.com/](https://www.wildlifeacoustics.com/)
In conclusion, three ecosystem types were monitored: forest, stubble and pasture, with two sites per ecosystem type (see Figure 3.1).

![Google Earth screenshot of the monitored area. Yellow pin color indicates a pasture site, green pin color indicates a forest site and red pin color indicates a stubble site. The closest weather station is also marked.](image)

**Figure 3.1** Google Earth screenshot of the monitored area. Yellow pin color indicates a pasture site, green pin color indicates a forest site and red pin color indicates a stubble site. The closest weather station is also marked.

### 3.1.2. Preprocessing Results

In this stage, a total of 25,560 recordings were automatically discarded due to probable heavy rain. This quantity corresponded to 20.4% of the total recordings. In Table 3.2, the number of original recordings, the number of discarded recordings and the percentage of excluded recordings per site are presented.

It is interesting to notice that the quantity of discarded recordings per site were very different. This seems to imply that it did not rain the same for all six sites, even though they are in the same locality. To test this hypothesis, we requested the local climate authority (IDEAM) to send the area daily rain records for the same dates (see in Fig. 3.1 the relative location of the weather station). We found that there was no strong correlation between any of the monitored sites $PSD_n$ and the weather station measurements, reinforcing our hypothesis (see Table 3.3). Notice that the longest distance between measuring points is approximately 3.9 km.
for the weather station and recorder 5256, and the shortest distance is approximately 260 m for recorder 5257 and 7125.

Table 3.2 Results for the preprocessing stage in the Jaguas application.

<table>
<thead>
<tr>
<th>Recorder</th>
<th>Number of Original Recordings</th>
<th>Number of Discarded Recordings</th>
<th>Percentage of noisy recordings</th>
</tr>
</thead>
<tbody>
<tr>
<td>5253</td>
<td>20,526</td>
<td>5,679</td>
<td>27.67 %</td>
</tr>
<tr>
<td>5255</td>
<td>24,018</td>
<td>3,590</td>
<td>14.95 %</td>
</tr>
<tr>
<td>5256</td>
<td>10,119</td>
<td>2,281</td>
<td>22.54 %</td>
</tr>
<tr>
<td>5257</td>
<td>19,416</td>
<td>1,030</td>
<td>5.3 %</td>
</tr>
<tr>
<td>5260</td>
<td>25,756</td>
<td>9,561</td>
<td>37.12 %</td>
</tr>
<tr>
<td>7125</td>
<td>25,424</td>
<td>3,419</td>
<td>13.45 %</td>
</tr>
</tbody>
</table>

Table 3.3 PCC matrix for the recorders and weather station (IDEAM). The cells are colored according to their value, Greener cells indicate high PCC values; Redder cells indicate low PCC values.

Nonetheless, IDEAM measurements seem to be the more correlated to the remaining sites measurements, slightly proving the rain algorithm reliability. On the other hand, the site 5256 shows the least correlation to the other sites, probably because it is the most remote point. Finally, relatively high PCC can be observed for same ecosystem type sites (except for the forest type), indicating that, on this band at least, there is some evidence for similar acoustical properties.

3.1.3. Feature Extraction Results

For the ecoacoustic indices calculation, MATLAB (Mathworks, 2013) home-made routines were executed. For ACIf, ACIf-te, ACIt, ACIt-e, \( \beta \), \( H_s \), \( NP \), and MID, the spectrogram was obtained using a Hann Window with no overlap \( N_{ov} = 0 \), size of 512 points \( \omega = 512 \) and 512 points for the Fast Fourier Transform (FFT) \( nfft = 512 \). This configuration produced a spectrogram of dimension 257 x 2,584 (see (3.1) and (3.2)). For \( NP \), the smoothed mean spectrum \( S_j \) was inspected for 10-point zones to find possible major peaks, and then if two peaks were found with a separation of less than 200 Hz, only the highest one was considered for the count.
\[ N_f = \left\lfloor \frac{nfft}{2} + 1 \right\rfloor \]  \hspace{1cm} (3.1)

\[ N_t = \left\lfloor \frac{d \times F_s - N_{ov}}{\omega - N_{ov}} \right\rfloor \]  \hspace{1cm} (3.2)

For ADI, AEI, \( H_m \) and \( H_v \), \( nfft \) and \( \omega \) were set as a tenth of \( F_s \) (\( nfft = \omega = 2,205 \)) leading to a spectrogram of 1,104 x 600 cells. Lastly, for NDSI and \( \rho \) the spectrogram was calculated using a Hamming window, and \( nfft \) and \( \omega \) were set to \( F_s \) (\( nfft = \omega = 22,050 \)) leading to a spectrogram dimension of 11,026 x 600. Additionally, for ADI and AEI, the threshold for computing the band occupation was set at -50 dBFS.

For the remaining features (complementary soundscape features and biodiversity features), the spectrogram was computed using a Hamming window with \( nfft = \omega = 512 \) and 256 points of overlapping (\( N_{ov} = 256 \)) producing a spectrogram dimension of 257 x 5,161. On the other hand, the Welch’s method for SF and MD was computed using a Hann window with \( nfft = \omega = 256 \) and 128 points of overlapping (\( N_{ov} = 128 \)), leading to a signal of 129 elements (this value is computed using (3.1)). The calculation of this last group of features was executed using Scipy library for Python (Jones, Oliphant, & Peterson, 2001).

Before the calculation of the daily averages, we noticed that some recordings presented “nan” values (not a number) for the ACIf index. Inspecting these recordings, we found that they presented clicks and moments of silence, as if the microphone stopped working for those instants (see Figure 3.2). We opted for discarding such recordings to avoid introducing noise to the analysis. In total, only 34 recordings were discarded for this reason.

![Recording with Mic Disconnections](image)

**Figure 3.2** Spectrogram segment of a discarded recording. Observe the vertical black lines towards the end of the segment indicating a microphone disconnection.
Given the discarded recordings in the previous stage and in this one, a minimum number of recordings per day and recorder was established for computing the daily average. Considering that a recorder worked every 20 minutes, a total of 72 recordings should be obtained per day for that recorder. Thence, we established that at least 60 recordings per day and recorder should have passed the previous filters for computing the daily average and including that day in the subsequent analysis.

In conclusion, a total of 1172 daily samples were obtained: 142 for recorder 5253, 229 for recorder 5255, 84 for recorder 5256, 259 for recorder 5257, 195 for recorder 5260 and 262 for recorder 7125.

3.1.4. Feature Selection Results

Once the daily averages were obtained and standardized, the correlation analysis was realized. For every pair of features with absolute value of $PCC$ higher than 0.9, only one feature was selected for the following stages. The criterion for selecting the chosen feature was its mean absolute Pearson coefficient ($MAPC$) calculated as the average of the feature’s absolute $PCC$s over the rest of the features (3.3). The feature with the highest $MAPC$ within the pair, was removed from the study. The discarded features, the corresponded correlated feature, their $MAPC$s and $PCC$s are displayed in Table 3.4.

$$MAPC_Y = \left| \overline{PCC_{YZ}} \right|, \quad \forall Z \neq Y$$ (3.3)

Where $MAPC_Y$ is the $MAPC$ value for feature $Y$, $PCC_{YZ}$ is the Pearson correlation coefficient for features $Y$ and $Z$ and $Z$ is any feature different than $Y$.

<table>
<thead>
<tr>
<th>Discarded Features</th>
<th>Correlated Features</th>
<th>$MAPC$</th>
<th>$PPC$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AEI</td>
<td>ADI</td>
<td>0.418</td>
<td>-0.97</td>
</tr>
<tr>
<td>$H_V$</td>
<td>$H_m$</td>
<td>0.4071</td>
<td>0.939</td>
</tr>
<tr>
<td>ACIfte</td>
<td>ACIftr</td>
<td>0.3354</td>
<td>-0.983</td>
</tr>
<tr>
<td>ACIfte</td>
<td>ACIftr</td>
<td>0.4328</td>
<td>-0.999</td>
</tr>
<tr>
<td>$H_S$</td>
<td>ADI</td>
<td>0.3879</td>
<td>0.938</td>
</tr>
<tr>
<td>$H$</td>
<td>ADI</td>
<td>0.3887</td>
<td>0.941</td>
</tr>
<tr>
<td>TLSV</td>
<td>ADI</td>
<td>0.3902</td>
<td>-0.946</td>
</tr>
</tbody>
</table>

Notice that, though ADI is preferred over several other features, the original ADI is not used for training the model, but the modified ADI vector is used instead. Nonetheless, including the ADI for the feature selection stage helped us identify the features that would have been redundant for the analysis.

Next, the training samples are grouped according to the ecosystem type classification and the $F$-value is computed. The obtained $F$-value is depicted in Figure 3.3. Considering that there are 3 ecosystem types and applying (2.34), the selected soundscape features are the three features with the highest $F$-value: $H_t$, $\beta$ and $NP$. The grouped boxplots for these features are displayed in Figure 3.4.
Figure 3.3 F-value for the selected features in the Jaguas application.

Figure 3.4 Grouped boxplots by ecosystem type for the three soundscape features.
It is interesting to notice that each of the soundscape features are related to one of the spectrogram dimensions: time ($H_t$), frequency ($NP$) and intensity ($\beta$). Also, notice in Figure 3.4 that the three ecosystem types are not completely distinguished using the soundscape features, even though their $F$-values are very high (Figure 3.3). In general, we can infer that the forest values seem to be more heterogeneous, while the pasture values are the most homogeneous. However, since there is not a clear distinction between the classes, we expect to find common underlying acoustical patterns for the three ecosystems soundscapes.

Finally, the biodiversity acoustic features are extracted as explained in section 2.2.3 and the set of features is completed (see Table 3.5).

**Table 3.5 Final feature vector for characterizing a daily sample in the Jaguas application.**

<table>
<thead>
<tr>
<th>Soundscape feat.</th>
<th>Biodiversity features</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>$NP$</td>
</tr>
</tbody>
</table>

3.1.5. Model Set Up Results

For the model set up and implementation stage, the Hmmlearn library for Python was used (Hmmlearn developers, 2016). Firstly, $N_m$ was tested in the range between one and five ($C = 5$) and $N_c$ was tested in the range between 3 and 20 ($G = 18$). Then, the Baum-Welch algorithm was executed 20 times ($R_G = 20$) for each of the $N_c$ and $N_m$ combinations; and 20 times again when the optimal $N_c$ and $N_m$ were found. The MSE values for each $N_m \times N_c$ combination are shown in Table 3.6.

In the table we find the lowest values in the column for $N_m = 3$, and the lowest value in this column for $N_c = 16$. However, when we inspected the resulting model, we found that there were seven repeated states; therefore, we maintained the value for $N_m$, changed $N_c$ to 9 and ran the Baum-Welch algorithm 20 more times, finding this time only 3 repeated states. Lastly, we retrained the model for $N_c = 6$ and obtained our final model for Jaguas.

Figure 3.5 and Table 3.7 summarize the resulting model. The graphs in the left column of Figure 3.5 show the mean vectors for each cluster, which were computed using (2.51); the graphs at the right correspond to the clusters’ prototypes, i.e. the recordings that best fit to the corresponding mean vectors. Table 3.7 shows the transition matrix $A$ and initial probability $\pi$ of the final model.

For cluster one, the prototype recording presented four occupied frequency bands by mainly constant calls. This type of occupation presents low biodiversity features because the entropy is low. On the other hand, $\beta$ is high, because the constant calls sum to the biophonic intensity and $NP$ is also high for the number of occupied bands.
Notice that the series of temporal pulses (increasing temporal entropy) and occupied mosquitoes recording had a high value for \( H_c \) and entropy for the medium band because it presented rain, which produced a series of temporal pulses (increasing temporal entropy) and occupied most of the spectrum.

Table 3.6 MSE values for all the possible combinations of \( N_c \) and \( N_m \) in the Jaguas application. The table cells are colored according to their values. Greener cells indicate lower MSE and redder cells indicate higher MSE.

<table>
<thead>
<tr>
<th>Values for ( N_c )</th>
<th>( 1 )</th>
<th>( 2 )</th>
<th>( 3 )</th>
<th>( 4 )</th>
<th>( 5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>13,582.69</td>
<td>13,558.85</td>
<td>11,874.24</td>
<td>12,603.17</td>
<td>12,741.83</td>
</tr>
<tr>
<td>4</td>
<td>13,582.69</td>
<td>13,557.71</td>
<td>11,784.15</td>
<td>12,433.33</td>
<td>12,427.01</td>
</tr>
<tr>
<td>5</td>
<td>13,582.69</td>
<td>13,533.03</td>
<td>11,781.18</td>
<td>12,367.13</td>
<td>12,260.71</td>
</tr>
<tr>
<td>6</td>
<td>13,582.69</td>
<td>13,483.86</td>
<td>11,607.86</td>
<td>12,179.76</td>
<td>12,049.71</td>
</tr>
<tr>
<td>7</td>
<td>13,582.69</td>
<td>13,495.78</td>
<td>11,574.52</td>
<td>11,965.19</td>
<td>12,013.15</td>
</tr>
<tr>
<td>8</td>
<td>13,582.69</td>
<td>13,507.9</td>
<td>11,292.71</td>
<td>12,030.13</td>
<td>11,813.66</td>
</tr>
<tr>
<td>9</td>
<td>13,582.69</td>
<td>13,532.99</td>
<td>11,403.33</td>
<td>11,991.37</td>
<td>11,747.56</td>
</tr>
<tr>
<td>10</td>
<td>13,582.69</td>
<td>13,519.78</td>
<td>11,174.39</td>
<td>11,819.75</td>
<td>11,869.43</td>
</tr>
<tr>
<td>11</td>
<td>13,582.69</td>
<td>13,519.98</td>
<td>11,094.87</td>
<td>11,846.88</td>
<td>11,543.39</td>
</tr>
<tr>
<td>12</td>
<td>13,582.69</td>
<td>13,481.7</td>
<td>11,098.45</td>
<td>11,735.96</td>
<td>11,640.71</td>
</tr>
<tr>
<td>13</td>
<td>13,582.69</td>
<td>13,507.62</td>
<td>11,178.56</td>
<td>11,831.09</td>
<td>11,693.77</td>
</tr>
<tr>
<td>14</td>
<td>13,582.69</td>
<td>13,494.92</td>
<td>11,028.94</td>
<td>11,579.31</td>
<td>11,596.84</td>
</tr>
<tr>
<td>15</td>
<td>13,582.69</td>
<td>13,481.28</td>
<td>11,075.92</td>
<td>11,688.5</td>
<td>11,458.84</td>
</tr>
<tr>
<td>16</td>
<td>13,582.69</td>
<td>13,480.98</td>
<td>10,881.52</td>
<td>11,538.69</td>
<td>11,511.13</td>
</tr>
<tr>
<td>17</td>
<td>13,582.69</td>
<td>13,492.81</td>
<td>10,895.7</td>
<td>11,449.9</td>
<td>11,443.27</td>
</tr>
<tr>
<td>18</td>
<td>13,582.69</td>
<td>13,467.54</td>
<td>10,960.35</td>
<td>11,633.1</td>
<td>11,484.58</td>
</tr>
<tr>
<td>19</td>
<td>13,582.69</td>
<td>13,441.88</td>
<td>10,950.44</td>
<td>11,461.99</td>
<td>11,479.11</td>
</tr>
<tr>
<td>20</td>
<td>13,582.69</td>
<td>13,455.77</td>
<td>10,896.47</td>
<td>11,488.65</td>
<td>11,374.74</td>
</tr>
</tbody>
</table>

Table 3.7 Transition matrix (A) of the resulting model in the Jaguas case. The initial probability for each state (\( \pi \)) is represented by row 0. Each cell represents the transition probability of the row state (\( i \)) to the column state (\( j \)). The greener the cell color, the higher probability for that transition, and the red cells indicate low transition probability. Notice that the row’s sum equals one.

<table>
<thead>
<tr>
<th>Current State (( i ))</th>
<th>( 1 )</th>
<th>( 2 )</th>
<th>( 3 )</th>
<th>( 4 )</th>
<th>( 5 )</th>
<th>( 6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0456</td>
<td>0.1364</td>
<td>0.2723</td>
<td>0.1807</td>
<td>0.137</td>
<td>0.2273</td>
</tr>
<tr>
<td>1</td>
<td>0.8535</td>
<td>0.0275</td>
<td>0.0070</td>
<td>0.0770</td>
<td>0.0350</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>0.0292</td>
<td>0.7772</td>
<td>0.0818</td>
<td>0.0665</td>
<td>0.0453</td>
<td>0.0000</td>
</tr>
<tr>
<td>3</td>
<td>0.0062</td>
<td>0.0784</td>
<td>0.7838</td>
<td>0.0096</td>
<td>0.0858</td>
<td>0.0362</td>
</tr>
<tr>
<td>4</td>
<td>0.0845</td>
<td>0.0437</td>
<td>0.0388</td>
<td>0.7820</td>
<td>0.0361</td>
<td>0.0150</td>
</tr>
<tr>
<td>5</td>
<td>0.0296</td>
<td>0.0269</td>
<td>0.0618</td>
<td>0.0163</td>
<td>0.8654</td>
<td>0.0000</td>
</tr>
<tr>
<td>6</td>
<td>0.0000</td>
<td>0.0169</td>
<td>0.0374</td>
<td>0.0152</td>
<td>0.0000</td>
<td>0.9305</td>
</tr>
</tbody>
</table>

Cluster two prototype had low values for \( \beta \) and \( NP \) because there were not many species calls. However, this recording had a high value for \( H_c \) and entropy for the medium band because it presented rain, which produced a series of temporal pulses (increasing temporal entropy) and occupied most of the spectrum.
Figure 3.5 Mean vectors for the resulting clusters and the clusters’ prototypes in the Jaguas application.

Cluster three prototype also presented rain, but this time no biophonic sound, which explained the low value for $\beta$. 
Cluster four prototype had a couple of totally occupied frequency bands and some others partially occupied by periodical calls.

Although cluster five seemed to be the most interesting for high frequency analysis, we only found one peculiar compound call occupying the band between 7 kHz and 11 kHz. This recording also had constant, periodical and sporadic calls but with low intensity.

Finally, cluster six prototype was very similar to cluster one’s, but with higher intensity and variety of calls.

In conclusion, cluster 1 and 6 are the most acoustically complex, because they presented more occupied bands, but the found calls tended to be periodical or constant. Clusters 5 and 4 had singing species but in less intensity and continuity. Lastly, clusters 2 and 3 had the least number of singing species but presented geophonic sound which produced high temporal entropy (for cluster 2) or high values for the ADIs, given that biophony was not constant and therefore their presence meant more information.

Certainly, higher values of ADI do not mean high occupation for the respective frequency bands, but higher randomicity of sounds, which can be due to rain or to the presence of different species in the band; whereby these values must be interpreted carefully.

Finally, looking at the transition matrix, we conclude that all clusters are stable, meaning that the probability of transition to a different cluster is relatively low (around 8% for the highest values). Cluster six seems to be the most stable and the most isolated cluster, given that from the other states is practically impossible to pass to this state (see column six in Table 3.7). However, this is one of the most likely starting points with a probability of 22.73%.

The physical interpretation of the obtained clusters is made more clearly in the next section.

### 3.1.6. Model Implementation Results

In this stage, the Viterbi algorithm is applied to the dataset and the results for the cluster assignation to the observed samples are presented. Firstly, we inspected the cluster composition using the classification of the training samples, or how the ecosystem types of the training samples were distributed into the final clusters in Figure 3.6. It shows that cluster 1 and 6 were mostly assigned to forest samples, but both had a high percentage of pasture observations; cluster 2 had the highest percentage of stubble samples, but also had an important share of forest. On the other hand, cluster 3 and 4 had a majority of stubble samples but were also assigned to a big share of pasture samples. Lastly, cluster 5 had the highest percentage of assigned pasture observations and probably is the cluster that best represent the typical pasture soundscape.
Figure 3.6 Training distribution of ecosystem types in the resulting clusters.

Notice that there is a high percentage of pasture samples in most clusters (in all of them, except for cluster 2), meaning that the pasture habitat presents many soundscape patterns, probably because its acoustical components are constantly changing through time or the local physical structure allows the propagation of many sound sources.

Next, each of the clusters was labeled with the ecosystem type presenting the highest proportion in it (the cluster’s label is shown in Table 3.8). Then, using these labels and the cluster’s assignation to the complete dataset (training and testing), a compound confusion matrix was constructed for assessing the generalization capacity of the model. The matrix can be observed in Table 3.9.

Table 3.8 Cluster’s label according to their training data distribution in the Jaguas case.

<table>
<thead>
<tr>
<th>Label</th>
<th>Forest</th>
<th>Stubble</th>
<th>Pasture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clusters number</td>
<td>1, 6</td>
<td>2, 3, 4</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 3.9 Normalized confusion matrix for the training and testing Jaguas datasets. Notice that the sum of the rows equals one. Greener cells indicate higher values, redder cells indicate lower values.

<table>
<thead>
<tr>
<th>True Label</th>
<th>Dataset</th>
<th>Predicted Label</th>
<th>1 (Forest)</th>
<th>2 (Stubble)</th>
<th>3 (Stubble)</th>
<th>4 (Stubble)</th>
<th>5 (Pasture)</th>
<th>6 (Forest)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forest</td>
<td>Training</td>
<td>0.39</td>
<td>0.4</td>
<td>0.02</td>
<td>0.15</td>
<td>0.02</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>0.46</td>
<td>0.43</td>
<td>0</td>
<td>0.07</td>
<td>0.04</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Pasture</td>
<td>Training</td>
<td>0.13</td>
<td>0.07</td>
<td>0.53</td>
<td>0</td>
<td>0.16</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>0.2</td>
<td>0</td>
<td>0.44</td>
<td>0.01</td>
<td>0.3</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Stubble</td>
<td>Training</td>
<td>0.01</td>
<td>0.02</td>
<td>0</td>
<td>0.35</td>
<td>0.31</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0.28</td>
<td>0.41</td>
<td>0.3</td>
<td></td>
</tr>
</tbody>
</table>
A well generalized model would have similar values for training and testing data, which seems to be the case, although there is some variation for pasture samples. If the model is intended for classification, a diagonal confusion matrix is ideal. For our case, we found that 81% of the forest samples were classified in the forest clusters (1 and 6), 51.2% of the pasture samples were classified in the pasture cluster (5) and 96.6% of the stubble samples were classified in the stubble clusters (2, 3 and 4). In conclusion, forest and stubble were the easiest ecosystem types to be classified given the proposed methodology, but pasture was the hardest one to identify. Considering that ecosystem type classification is a complex task even for ecologists, because they can share multiple characteristics, we deemed these results very promising.

Finally, the temporal change in the sites is explained by Figure 3.7. In it, the classification for the complete dataset (training and testing) is displayed through time.

We can observe that site 5256 was the most stable site, but also had less samples than the others. This site was the furthest monitored point, and was in the depths of the forest, therefore it corresponded to the least disturbed habitat. Most of its samples were classified in cluster 6, proving that the better preserved the forest, the more different communities and more stability through time.

Cluster six was also assigned to the last samples of site 5260 during the first months of 2017. This indicates increase of calls intensity. For the past samples in this site, the associated cluster was number one, with a few exceptions in April and November, when cluster two had some apparitions, maybe due to rain.

Cluster two was mostly present in site 5255, which is a stubble site, but then disappeared in November of 2016 and made way for cluster 4 from that moment. This could indicate that for the first six months of monitoring, there was not much biophonic activity, but the following six months more species started to sing, probably because of a climatic factor.

![figure](image)

*Figure 3.7 Temporal Classification of samples per site in the Jaguas case.*
Cluster four and one alternated from November of 2016 to February of 2017 in the pasture site with recorder 5257. However, in the first and last monitored months, cluster five was more dominant. This site seemed to vary its acoustical activity by seasons, but apparently there were always some singing species in this place.

The same cannot be said about site 5253, which seemed to be a rather lacking space of biophonic activity, but with more presence of geophony. Finally, pasture site 7125 could have sometimes dominance of geophony (explained by cluster 3) or some biophonic activity (indicated by cluster 5).

Summing up, forest presented more occupied bands, indicating more species singing constant and periodical calls. Pasture seemed to be more of a passage area, where geophony dominated for some seasons, but species calls were still noticeable. Lastly, stubble sites did not present continuous and periodical sounds, but rather random sounds, probably due to geophony influence; also, did not seem a preferred space for making calls, given that not many frequency bands were occupied.

3.2. Dry Tropical Forest Transformation Application

This database was provided by the institute Alexander von Humboldt. The original project was aimed to understand the biodiversity changes due to landscape transformation and succession in the Colombian tropical dry forests (Rodríguez-Buritica, 2017). Given the dimension and the heterogeneity of the dataset, in this work we only selected a subgroup of recordings from the Cañas river basin in the Guajira and Arroyo river basin in Bolívar.

3.2.1. Database Description

From December of 2015 to March of 2017, a series of recording sessions were realized in different points of the Cañas river basin and the Arroyo river basin in the Colombian Caribbean region. The sessions were established to record continuously for five days, then stop for another five days and restart this process during approximately one month. This process was repeated for other months during this period. However, in practice we found that this protocol was not strictly followed, or some recordings were missing, which negatively affected the model performance. A complete description of the selected recording sessions is presented in Table 3.10.

The recorders were programmed to register acoustical activity every 10 minutes. The resulting audio files were monaural, presented a digital format of 16-bit PCB \(D = 16\), a sampling frequency of 48,000 Hz \(F_s = 48,000\) Hz and lasted five minutes \(d = 300\) sec. In general, the exact monitored sites were not repeated between sessions, except for sessions 3 and 7, sessions 4 and 5, sessions 10 and 12 and session 11 and 13.

For this application, we chose the recording sessions made by the same model of recorders: Wildlife Acoustics SM3. We found that SM2 recorders were also present in the original dataset and configured with the same settings as SM3 recorders, which actually produced significant differences in intensity (Eldridge et al., 2018).
The recording sessions were also selected to obtain a balanced model respecting the ecosystem transformation, which is the ecosystem type for this case.

Table 3.10 Description of the selected recording sessions for the Humboldt Application. The rows are colored according to their transformation value.

<table>
<thead>
<tr>
<th>Session Number</th>
<th>Region</th>
<th>Transf.</th>
<th>Recorder</th>
<th>Location</th>
<th>Initial Date (AAAAMMDD)</th>
<th>Final Date (AAAAMMDD)</th>
<th>Number of Recordings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Guajira</td>
<td>High</td>
<td>5067</td>
<td>11.213 N, 73.434 W</td>
<td>20151212</td>
<td>20160123</td>
<td>3,239</td>
</tr>
<tr>
<td>2</td>
<td>Guajira</td>
<td>High</td>
<td>5071</td>
<td>11.214 N, 73.44 W</td>
<td>20161221</td>
<td>20161231</td>
<td>786</td>
</tr>
<tr>
<td>3</td>
<td>Bolívar</td>
<td>High</td>
<td>5071</td>
<td>9.896 N, 75.162 W</td>
<td>20160305</td>
<td>20160326</td>
<td>1,332</td>
</tr>
<tr>
<td>4</td>
<td>Bolívar</td>
<td>High</td>
<td>5069</td>
<td>9.893 N, 75.139 W</td>
<td>20160224</td>
<td>20160228</td>
<td>574</td>
</tr>
<tr>
<td>5</td>
<td>Bolívar</td>
<td>High</td>
<td>5071</td>
<td>9.893 N, 75.139 W</td>
<td>20170123</td>
<td>20170204</td>
<td>1,087</td>
</tr>
<tr>
<td>6</td>
<td>Bolívar</td>
<td>High</td>
<td>5071</td>
<td>9.888 N, 75.141 W</td>
<td>20170213</td>
<td>20170314</td>
<td>628</td>
</tr>
<tr>
<td>7</td>
<td>Bolívar</td>
<td>High</td>
<td>302143</td>
<td>9.896 N, 75.162 W</td>
<td>20170123</td>
<td>20170224</td>
<td>2,476</td>
</tr>
<tr>
<td>8</td>
<td>Guajira</td>
<td>Low</td>
<td>JSC5069</td>
<td>11.197 N, 73.436 W</td>
<td>20151212</td>
<td>20160123</td>
<td>3,237</td>
</tr>
<tr>
<td>9</td>
<td>Guajira</td>
<td>Low</td>
<td>302143</td>
<td>11.198 N, 73.435 W</td>
<td>20161221</td>
<td>20170114</td>
<td>1,103</td>
</tr>
<tr>
<td>10</td>
<td>Bolívar</td>
<td>Low</td>
<td>5071</td>
<td>9.908 N, 75.187 W</td>
<td>20160224</td>
<td>20160227</td>
<td>576</td>
</tr>
<tr>
<td>11</td>
<td>Bolívar</td>
<td>Low</td>
<td>5071</td>
<td>9.899 N, 75.192 W</td>
<td>20160214</td>
<td>20160217</td>
<td>576</td>
</tr>
<tr>
<td>12</td>
<td>Bolívar</td>
<td>Low</td>
<td>302151</td>
<td>9.908 N, 75.187 W</td>
<td>20170202</td>
<td>20170313</td>
<td>2,597</td>
</tr>
<tr>
<td>13</td>
<td>Bolívar</td>
<td>Low</td>
<td>302298</td>
<td>9.899 N, 75.192 W</td>
<td>20170123</td>
<td>20170206</td>
<td>1,440</td>
</tr>
<tr>
<td>14</td>
<td>Bolívar</td>
<td>Low</td>
<td>302298</td>
<td>9.903 N, 75.185 W</td>
<td>20170212</td>
<td>20170306</td>
<td>1,571</td>
</tr>
<tr>
<td>15</td>
<td>Guajira</td>
<td>Medium</td>
<td>5072</td>
<td>11.167 N, 73.444 W</td>
<td>20151212</td>
<td>20160102</td>
<td>2,010</td>
</tr>
<tr>
<td>16</td>
<td>Guajira</td>
<td>Medium</td>
<td>302151</td>
<td>11.168 N, 73.431 W</td>
<td>20161221</td>
<td>20170120</td>
<td>2,219</td>
</tr>
<tr>
<td>17</td>
<td>Bolívar</td>
<td>Medium</td>
<td>5070</td>
<td>9.94 N, 75.170 W</td>
<td>20160214</td>
<td>20160316</td>
<td>1,865</td>
</tr>
<tr>
<td>18</td>
<td>Bolívar</td>
<td>Medium</td>
<td>5067</td>
<td>9.936 N, 75.159 W</td>
<td>20160214</td>
<td>20160326</td>
<td>2,535</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>29,851</strong></td>
</tr>
</tbody>
</table>
For the classification of the sites, the institute Alexander von Humboldt collected forest/non-forest layers from satellite imagery in the period between 1990 and 2012. Then, forest patches (of 30m x 30m dimension) were classified as “retained” if they were old forests, “lost” if they were younger forest than 4 years old, or “new” if they were older forest than 4 but younger than 22 years old. Consequently, high transformation zones corresponded to more proportion of lost forest than retained or new, low transformation zones presented higher proportion for retained or new forest and medium transformation zones were an approximate balance between forest and non-forest (see the third column of Table 3.10).

3.2.2. Preprocessing Results

Using the $PSD_n$ algorithm, a total of 2,861 recordings were removed from the analysis due to heavy wind or rain, for 9.58% of the total recordings.

3.2.3. Features Extraction Results

Similar parameters to Jaguas’ were chosen for computing the set of features. Given that this is a comparative study, notice that the details for feature calculation are not relevant as long as the main concept is respected, and it is uniformly applied to the data samples.

For this case, the discarded metrics from Table 3.4 were not computed to speed up the process. This omission was made on the assumption that the correlation converges as the number of samples increase. Then, since Jaguas case had a considerable number of samples, the obtained correlation represented a generalized measure that could be applied to other applications. Also, $H_v$ was replaced by $H_m$ because in a previous result, the former had presented a lower $MAPE$. This change should not affect the results, because these features were highly correlated.

In this application the ACIf index presented nan values as well. For this reason, 7 more recordings were discarded. In total 9.6% of the initial database recordings was rejected in the previous and current stage.

Finally, the averages were computed over the days with at least 120 recordings (the proportion of minimum required recordings was maintained; remind that in this study the number of recordings per day doubled Jaguas’). A total of 143 daily samples were obtained. These samples were also standardized using (2.31).

3.2.4. Feature Selection Results

An additional correlation analysis between the remaining features was executed. Although Jaguas correlation analysis showed a general relation between features, the small number of samples for this case could lead to an increase of $PCC$ for some features. Indeed, we found that the $PCC$ for $M$ and $RMS$ was higher than 0.9 as
well as the \( PCC \) for ACIf and FM (the respective Jaguas' \( PCCs \) were 0.72 and 0.81). Consequently, we removed M and ACIf; from the analysis.

Next, we calculated the F-value for the obtained features given the transformation of the training samples. Figure 3.8 shows the resulting F-values.

![F-value](image)

**Figure 3.8 F-value for the selected features in the Humboldt application.**

Observe that the features with highest F-value correspond to temporal complexity or intensity metrics; conversely, frequency complexity measures (as \( NP \) and ACIt) presented the lowest values, indicating that the number of singing species did not differ much between transformations. Note also that the highest F-values are much lower than Jaguas', which could be due to the dataset size difference or the difficulty in this case for identifying ecosystem types from acoustical measures. In Figure 3.9 we inspected the boxplots of the three highest scored features.

The boxplots confirmed that it was very difficult to separate transformation types given the soundscape features. Low and medium transformations seemed to present similar values for CF and \( H_t \) but differed a little in their \( RMS \) values. Nonetheless, transformation high and low were very similar for this last feature. Apparently, the low transformation could be easily confused with the other transformations on the acoustic level.
Lastly, we computed the $ADIs$ values maintaining the analysis up to 11 kHz. Soundscape studies usually limit to this frequency band, because higher frequencies present little information (Gage & Axel, 2014; Pijanowski, Farina, Gage, Dumyahn, & Krause, 2011; M. Sankupellay et al., 2016). The final feature vector is described in Table 3.11.

Table 3.11 Final feature vector for characterizing a daily sample in the Humboldt application.

<table>
<thead>
<tr>
<th>Soundscape feat.</th>
<th>Biodiversity features</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RMS$</td>
<td>$H_t$</td>
</tr>
</tbody>
</table>

3.2.5. Model Set Up Results

In this case the $N_m$ values were tested in the range between one and five, and the $N_c$ values were tested in the range between three and fifteen, because higher $N_c$ values were slowing down the Baum-Welch algorithm. The resulting MSEs are displayed in Table 3.12.

As shown in Table 3.12, the best $N_c \times N_m$ combination was 10 × 4. In this case, five states were replications of the others, then the model was reconfigured for having only five states and the Baum-Welch was repeated 20 more times. This process was repeated until the states were unique. At the end, only three states remained. Figure 3.10 and Table 3.13 summarize the final model. Figure 3.10 shows the mean vectors for each cluster and Table 3.13 presents the transition matrix $A$ and initial probability $\pi$ of the resulting model.

Although there were three final states, states two and three were very similar; the main difference between them was that state two had lower ADI values for the lowest frequency bands, whereas state three had lower ADI values for the highest frequency bands. These values indicate that state two presented a little more
information in the upper bands and state three had more information at the lowest bands. Besides, state two presented a lower \( RM^2 \) value, which is related to intensity, meaning that either the sound sources were further than in state three, or the soundscape was quieter. In general, these two states should present monotonous soundscapes, either for continuous sounds or silence.

Table 3.12 MSE values for all the possible combinations of \( N_c \) and \( N_m \) in the Humboldt application. The table cells are colored according to their values. Greener cells indicate lower MSE and redder cells indicate higher MSE.

<table>
<thead>
<tr>
<th>Values for ( N_c )</th>
<th>Values for ( N_m )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1,192.40</td>
</tr>
<tr>
<td>4</td>
<td>1,192.40</td>
</tr>
<tr>
<td>5</td>
<td>1,192.40</td>
</tr>
<tr>
<td>6</td>
<td>1,192.40</td>
</tr>
<tr>
<td>7</td>
<td>1,192.40</td>
</tr>
<tr>
<td>8</td>
<td>1,192.40</td>
</tr>
<tr>
<td>9</td>
<td>1,192.40</td>
</tr>
<tr>
<td>10</td>
<td>1,192.40</td>
</tr>
<tr>
<td>11</td>
<td>1,192.40</td>
</tr>
<tr>
<td>12</td>
<td>1,192.40</td>
</tr>
<tr>
<td>13</td>
<td>1,192.40</td>
</tr>
<tr>
<td>14</td>
<td>1,192.40</td>
</tr>
<tr>
<td>15</td>
<td>1,192.40</td>
</tr>
</tbody>
</table>

Figure 3.10 Mean vectors for the resulting clusters in the Humboldt application.

On the other hand, state one presented more randomicity expressed by its high levels of \( Ht \) and \( ADIs \). Therefore, this state should represent more chaotic soundscapes.

Finally, notice that CF was approximately the same for the three states. This indicates that in general, the signal peaks did not separate as much from the effective value \( RM^2 \).

The transition probabilities indicate that state two would be the intermediate state for states one and three, given that its transition probability to either of these states was the same. Nonetheless, once the sequence reached state one, it was very difficult for it to pass to a different state. Notice that state one was the most
stable and isolated cluster but was also the most likely starting point, as happened to cluster six in Jaguas’ case. Lastly, state three had a high probability to pass to state two, which makes sense given their similarity, however it was almost impossible for this state to pass to state one.

Table 3.13 Transition matrix (A) of the resulting model in the forest transformation case. The initial probability for each state (it) is represented by row 0. Each cell represents the transition probability of the row state (i) to the column state (j). The greener the cell color, the higher probability for that transition, and the red cells indicate low transition probability. Notice that the row’s sum equals one.

<table>
<thead>
<tr>
<th>Current State (i)</th>
<th>Next State (j)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0.5556</td>
</tr>
<tr>
<td>1</td>
<td>0.9268</td>
</tr>
<tr>
<td>2</td>
<td>0.0714</td>
</tr>
<tr>
<td>3</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

3.2.6. Model Implementation Results

In this stage we implemented the Viterbi algorithm to find the final classification and we inspected the relation between the obtained clusters and the ecosystem transformations. In Figure 3.11 we used pie charts to estimate the transformation composition of each of the resulting states for the training dataset. We found that each cluster could be associated to an ecosystem transformation (see Table 3.14), but the average proportion for the assigned transformations was 56.57% for all states. In other words, we should not expect that a given cluster would produce always the same transformation, although most of the emissions would correspond to the assigned label.

Table 3.14 Cluster’s label according to their training data distribution in the Humboldt case.

<table>
<thead>
<tr>
<th>Label</th>
<th>Low Transformation</th>
<th>Medium Transformation</th>
<th>High Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster number</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Given this assignation and looking back at Figure 3.10, we might also conclude that the high and medium transformations presented similar soundscapes and the low transformation soundscape tended to be very different. However, low transformation had an important presence in the medium and high clusters, which made it the most changing soundscape of the three.

Next, with the ecosystem type assignation to each cluster and the Viterbi algorithm results for both training and testing data, we computed the confusion matrices to assess the generalization capacity and classification accuracy. The compound matrix is displayed in Table 3.15.
Figure 3.11 Training distribution of ecosystem transformations in the resulting clusters.

Table 3.15 Normalized confusion matrix for the training and testing Humboldt datasets. Notice that the sum of the rows equals one. Greener cells indicate higher values, redder cells indicate lower values.

<table>
<thead>
<tr>
<th>True Transformation</th>
<th>Dataset</th>
<th>Predicted Label</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>3 (High)</td>
</tr>
<tr>
<td>High</td>
<td>Training</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>0.11</td>
</tr>
<tr>
<td>Low</td>
<td>Training</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>0</td>
</tr>
<tr>
<td>Medium</td>
<td>Training</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>0.1</td>
</tr>
</tbody>
</table>

The matrix showed that the model was not sufficiently generalized because the dataset was too small. On the other hand, the classification results indicated that the model identified 68% of the total low transformation samples, 58.8% of the medium transformation samples and only 31.8% of the high transformation samples.

A temporal analysis was not presented because the locations were not continuously monitored, and the sequences were very short (the average sequence lasted $4.63 \pm 3.62$ days).

In conclusion, this methodology was not the best for this application. Although the original Humboldt’s study analyzed image information for approximately 30 years (Rodríguez-Buritica, 2017), the acoustical analysis period was much shorter and irregular, presenting considerable gaps. GMMHMM is a sequential model, which requires long sequences to be properly trained and the dataset was both small and too heterogeneous to have long sequences. Consider also that in theory the methodology was applied to the same ecosystem type: tropical dry forests but with different ages. From this point of view, it seemed reasonable that the frequency complexity (related to the species diversity) did not vary between transformations. Additionally, the transformation
classification was applied to large areas, whereas this methodology is rather for local problems given the scope of the recorders.

Nonetheless, we might conclude that in this application, a high general entropy (temporal and biophonic) would represent older forests, whereas younger forests would present mostly monotonous chorus. Besides, both studies indicated that the better-preserved habitats would have high transition probability from the assigned state and itself, and that their initial probability would be high as well.
Chapter 4. Conclusions

4.1. Summary

In this work, a new methodology for assessing ecosystem change through ecoacoustics analysis was proposed. Firstly, the elimination of noisy recordings was executed using an automatic threshold on the $PSD$ values in the 600 – 1200 Hz band. Secondly, the most common ecoacoustics indices and other acoustical features for soundscape analysis were analyzed using the Pearson correlation coefficient and the calculation of ANOVA F-value. The first was used to detect similar features and remove the duplicates to avoid redundancy, the second led to find the fittest features for the application and comprehend the main acoustical differences between the ecosystem types. After selecting the soundscape features, the biodiversity features were computed per band as a modification of the ADI. These values allow to recognize the frequency bands with more information or entropy, due to either random calls or geophony when generalized across the spectrum. Lastly, the model was adjusted using hidden Markov models with Gaussian mixture emissions. This model identified the main acoustical patterns that produced the observations. Each of the resulting states could be associated to a specific ecosystem type, but this was not the main purpose, but to recognize stationary and transitory acoustical patterns, which could be common to all types of habitat. The GMMHMM was also used to analyze the data as a time series and recognize the probable transitions between states on a daily basis.

The methodology was developed for the Jaguas study in the first place. The purpose of this study was to identify the acoustical signature of three ecosystem habitats by distributing two recorders per ecosystem type and registering the acoustical activity for over a year. The model identified six clusters: two were mostly assigned to forest samples, three had mostly stubble data and one presented a pattern almost unique for pasture sites. The classification led to a correct recognition of the 81% for forest, 51.2% for pasture and 96.6% for stubble. We could also recognize that the forest soundscape presented more uniformity in their sounds, more occupied frequency bands (related to the number of species) and more biophonic intensity. On the other hand, the stubble soundscape had more entropy in their sounds, which was related to geophony but less biophonic activity. Finally, the pasture sites had the most versatile soundscapes with both geophony and biophony alternating during the year.

A second case in the Colombian Caribbean was used to apply the methodology. In this case, the recordings comprised a longer time span, but there were few continuous daily records. Additionally, the ecosystem types corresponded to forest transformations, which could be best interpreted as forest longevity. The resultant model did not present an accurate generalization capacity, given the heterogeneity and small size of the dataset. However, it showed that there was not a striking acoustical difference between high and medium transformed forests, whose age is less than 30 years old; but showed that older or low transformed forests presented a different acoustical signature to the rest, with more entropy across the spectrum. The model identified three underlying states, which we associated to the transformation levels and resulted in the recognition of 68% of the low transformed forest samples, 58.8% for the medium transformation and 31.8% for the high transformation.
In conclusion, this methodology is more appropriate for long term studies with uninterrupted sampling, preferably several times per day. It also requires the correct classification of the monitored habitats, for obtaining clusters that, associating to the studied ecosystem types, would be easier to interpret.

4.2. Contributions of this work

The proposed methodology made the following contributions to the ecoacoustics field:

4.2.1. Automatic removal of noisy recordings due to geophony

Based on the algorithm proposed by Bedoya et al. for rain detection (Bedoya et al., 2017), we proposed an automatic threshold for identifying the recordings with very high levels of geophony that would introduce noise to the analysis. We noticed that this threshold did not only discard recordings in which heavy rain was present, but it also detected intense wind, which could also mask most of the acoustical activity. Notice that our interest was not to discard all recordings with geophonic presence, because these sounds are an important part of the soundscape and may explain some characteristics of biophonic behavior (such as the acoustic adaptation hypothesis (Sueur & Farina, 2015)). Therefore, this threshold is suggested for soundscape applications and not for specific species study, in which biophony needs to be isolated and all geophony should be removed.

Although noise filtering is desired for most ecoacoustic studies, we only found Towsey’s report addressing this problem in this field (Towsey, 2013). However, as himself put it, this method was aimed at removing constant sounds during a recording, that could correspond to any of the soundscape components, probably leading to an important loss of information. Since our method focuses in a frequency band typically occupied by geophony, we consider that it gives more information to the user about the type of noise that it is rejecting.

4.2.2. Analysis of ecoacoustic indices similarity

Most of the studies involving ecoacoustic indices are directed towards the identification of certain ecological qualities of the monitored habitats. These ecological qualities could include species richness (e.g. (Harris, Shears, & Radford, 2016)), ecosystem type (e.g. (Eldridge, Casey, Moscoso, & Peck, 2016)), landscape configuration (e.g. (Fuller et al., 2015)) etc. However, we did not find an actual study that compared the indices measurements to identify those who provided the same information. In this work, we computed the Pearson correlation coefficient for every pair of acoustical features to find the most representative features of all. For instance, we found that ADI is highly correlated to AEI, H and $H_s$, therefore using the four indices for a study would be redundant. Although these results, specially Jaguas’, should show a general tendency; smaller scaled studies could present considerable differences in the features similarities given the bias of limited sampling.
4.2.3. Automatic recognition of daily soundscape patterns

Automatic recognition algorithms have been used mainly on the biophonic level for species identification (Bedoya et al., 2014; Eldridge et al., 2018; Tucker et al., 2014), but rarely used for characterization of terrestrial soundscape patterns and connection to habitat structure (Bormpoudakis et al., 2013; Ulloa, Aubin, Llusia, Bouveyron, & Sueur, 2018).

When the state of the art of this research was established, only Bormpoudakis et al. work had used clustering algorithms for grouping and linking soundscape patterns to ecosystem type (Bormpoudakis et al., 2013). However, this study was made on a small dataset, for a very specific time of the day and year and avoided other elements of soundscape as geophony and technophony. Therefore, the scope of its findings was very reduced.

Towards the end of our research, we discovered Ulloa et al. work, which identified regions of interest (ROIs) in the recordings, grouped them by unsupervised algorithms and characterized two types of ecosystems by using heat maps showing the distribution of the ROIs clusters (Ulloa et al., 2018). This study was also tested on a small dataset, for diurnal recordings and focused mainly on biophonic calls, given the procedure for the detection of the ROIs.

Although our methodology highlights the importance of biophonic sounds by extracting a higher number of features for characterizing this component (the ADIs), it does not discard the remaining soundscape components (geophony and technophony) from the recordings, because they are a relevant part of the ecosystem acoustical signature and model the communication behavior of communities as well. Besides, our analysis unit, which is the daily average of the recordings, encompasses all the acoustical information in the day, detailing the totality of the acoustical processes in the monitored site.

4.2.4. Temporal relation of soundscape patterns

Many studies have addressed the temporal variation of soundscapes (Bernie Krause et al., 2011; Matsinos et al., 2008; Mullet et al., 2016) but none of them has proposed an inner model that explains the relation between soundscape patterns and their stationary or transitional probabilities.

This feature allows our method to identify the temporal stability of soundscape patterns, indicating the permanence of singing communities; and the natural sequence of acoustical states, finding the seasonality of the ecosystem in the acoustical level. Furthermore, the transition matrix could help us detect forthcoming ecosystem changes and take early conservation decisions based on that information.

4.3. Publications

During the development of this master’s project, the author published the following preliminary studies:

This work consisted on the comparison of the vocalizations of two Espadarana prosoblepon populations located in different sites, a disturbed and a conserved forest. The results showed that the degradation of this species’ habitat affected its calls structure.


In this work, psychoacoustical features and two types of machine learning algorithms were tested for forest and stubble discrimination. The results indicated that neural networks and two kernel functions of support vector machines were able to recognize the difference between forest and stubble by their psychoacoustical description.


The previous three works were presented during the 10th international conference on ecological informatics (ICEI 2018) in Jena, Germany. The first presentation showed the general results from the application of passive acoustic monitoring and machine learning techniques on the conservation of different sites in the northern Andes of south America. The second presentation highlighted the results of our research group in soundscape analysis for ecosystem change assessing. And the third presentation showed a study on the relationship between ecoacoustic indices and landscape features in the tropical Andean environment.

4.4. Future Work

Extensive work is still to be done in this field. Firstly, ecoacoustics highlights the importance of three soundscape components on the acoustical communities: biophony, geophony and technophony. However, little research has been done on the latter two components, as the studies keep focusing in the communities’ communication behavior. Additionally, the acoustical properties of each of these components is still unclear. Traditional methods have assumed each element to be limited on a certain frequency band, but we know that in practice,
that is not true, and the three elements constantly overlap on the spectrum. This overlapping is precisely one of the explanations for changes in the species communication behavior, whereby a key reason for studying all soundscape elements is being practically ignored.

Secondly, further exploration of the soundscape and ecological processes is required. The studies have focused on correlating soundscape and biodiversity or landscape configuration but are not going deeper to describe soundscape as a manifestation of the underlying ecosystem processes. For instance, an interesting study would relate soundscape characteristics to species migration, as an evidence to the acoustic adaptation hypothesis (Sueur & Farina, 2015) and its importance to ecosystem constitution.

Nonetheless, the limited scope of acoustical recording is still an obstacle to such studies. This observation leads to another necessity in soundscape studies, which is the clear definition of the intrinsic limitations of acoustical research. The reliability of soundscape findings can be severely affected by the actual spatial scope of the study, which is limited by the physical configuration of the monitored site and the nature of the acoustical signals of interest.

Lastly, the relevance of ecoacoustic indices is still uncertain. The proposal of many similar indices indicates that biodiversity abstraction from acoustical complexity is not clear yet. Furthermore, recent studies questions the validity of many acoustic indices, either for their theoretical basis (Sandoval, Barrantes, & Wilson, 2018) or for their actual applicability (Eldridge et al., 2018). For future applications a decision must be made either to continue to find a universal descriptor of ecoacoustic complexity, or to define application-specific features that maximize soundscape description.

Personally, I think that the current state of the art demands for more studies of the last type, for which machine learning algorithms present as valuable tools. Nonetheless, a better understanding of specific soundscape patterns would lead us inevitably to associate similar findings and slowly construct a formal theory from where the definitive ecoacoustical features would emerge.
References


Aval

Thesis
Methodology for ecosystem change assessing using ecoacoustics analysis

Advisor

______________________________
Claudia Victoria Isaza Narváez