ACOUSTIC INDICES

Aim

Given mel-filtered spectrograms, the aim is to predict, with reasonable certainty, if the recording -

- Contains bird sounds - biophony
- Contains only other environmental sounds - geophony/anthrophony

Recordings

Each recording is exactly 2 seconds in length. There are recordings of four classes -

- Good - definitely contains a distinct signal of interest
- Bad - definitely does not contain a distinct signal of interest
- Human - has a human voice in the recording
- Maybe - may have a signal of interest

Purpose of using Acoustic Indices

Acoustic indices aim to characterize soundscapes by quantifying a temporal or spectral feature of a spectrogram. In previous work[1], various acoustic indices were tested to predict:

- Richness: number of species detected in the recording
- Total number of avian species vocalizations in recordings
- Shannon diversity of avian species

First acoustic index: Acoustic Complexity Index[2]

- ACI: aims to quantify complex biotic songs by computing the variability of the intensities registered in audio-recordings, despite the presence of constant human-generated-noise
- Spearman's rho showed high correlation between the ACI values and the number of bird vocalizations

---

[1] "Acoustic indices as rapid indicators of avian diversity in different land ...." 9 May. 2018, https://www.veruscript.com/api/files/be13ad8b-527c-11e8-b3ad-0242ac110002/download. Accessed 15 Oct. 2018.

[2] "A new methodology to infer the singing activity of an avian community ...." http://www.iinsteco.org/people/pubblications/almo/2011/2011_A_new%20methodology_to_infer_the_singing_activity.pdf. Accessed 15 Oct. 2018.

- Soundscape can be broken down into
    - biophony (non-human biological sounds such as the vocalizations of birds, amphibians and other animals),
    - geophony (physical features of the environment such as the wind blowing through a forest or the burbling of water in a stream)
    - anthrophony (human-induced noise from whatever source)
- ACI is based on this - biotic sounds, such as bird songs, are characterized by an intrinsic variability of intensities, while some types of human generated noise (such as car passing or airplane transit) present very constant intensity values
- Calculation: for one time step and one freq bin, add absolute difference between all neighbouring pixels (horizontally adjacent) - this is D
- Divide D by sum of all intensity values for that row to normalize - this is ACI for a freq bin and a time step
- ACI for one frequency bin is the sum of ACIs over all time steps for that bin
- ACI for the whole recording is the sum of ACIs for each freq bin
- Flat-like sounds, such as insects buzz or other anthropogenic noise (i.e. cars, airplanes, etc.), are characterized by approximately constant levels of intensity, which produce very small ACI value

Results from initial ACI implementation

<u>Note</u>: Signals have not been normalized. We have taken the first 2 seconds of each signal and used the 10*log(spectrogram_values + 10^{-6}). No filtering was applied.

File name	**File type**	**ACI value**	**Other notes**
Aeroplane sound	Environmental sound	-18.59	Most intensity is in 0-1 kHz
Heavy rain	Environmental sound	-21.18	Most intensity is in 0-1 kHz
Brown	Noise	-17.45	Most intensity is in 0-200 Hz
Pink	Noise	-21.69	Most intensity is in 0-500 Hz
White	Noise	-23.38	Most intensity is in 0-7.5 kHz

Rufous_Antpitta	Bird chirps with some background noise	-9.47	Background noise limited to around 700 Hz. Clear chirps between 2 and around 2.5 kHz
Grey_headed_woodpecker	Bird chirps with some background noise	-10.09	Chirps + some harmonics visible from 1.5 kHz - 7.5 kHz. Background noise between 1 and 4 kHz
Italian_Sparrow	Bird chirps with some background noise	-10.06	Chirps + faint harmonics visible from 2 -5 kHz. Background noise between 0 and 2 kHz
Hawk scream	Very clear bird sounds	-17.34	Looks more like noise. Strong intensity bands between 2-3 kHz and at 5 kHz
Dove	Very clear bird sounds	-6.37	Most intensity at 500 Hz, and clear harmonics up to 3.5 kHz

Thoughts

- No normalization w.r.t time seems odd in the paper
- Instead of intensity, there should be something on periodicity or harmonic nature of the signal

Sources of sounds downloaded

- http://soundbible.com
- Xeno-canto
- http://whitenoisemp3s.com/free-white-noise

Notes from earlier work comparing 36 acoustic indices[3]

- Results: Limited evidence found that acoustic indices predict the richness (number of species detected in the recording) and total number of avian species vocalizations in recordings ($R2 < 0.51$)
- However, acoustic indices predicted the Shannon diversity of avian species vocalizations with high accuracy ($R2 = 0.64$, mean squared error = 0.17)
- For each recording, the Shannon diversity was calculated as - sum (p * log p) where this term is summed for every species found in that recording
- Many types of acoustic indices have been tested and have been found to reflect vocal species diversity and abundance, community composition, vegetation structure, habitat type, human perception of a soundscape, human activity, and ecological condition in terrestrial and aquatic habitats
- 36 acoustic indices tested totally
- Later RF was used with non-redundant acoustic indices as predictor variables
- Model with lowest MSE and highest R2 was selected
- The top RF model predicting Shannon diversity of vocalizations had as predictors: (1) **acoustic complexity index** (2) **roughness,** the difference between the 10th and 90th exceedance levels (3) **temporal entropy,** spectral entropy; biophony and mean sound levels in the avian frequency bands; acoustic activity, duration of acoustic events in the avian and noise bands, entropy of spectral maxima, and spectral persistence , acoustic evenness index and acoustic diversity index; spectral kurtosis
- Some studies found a significant relationship between avian species richness and acoustic indices (ACI!) - but they didn't have as much acoustic diversity
- "combining groups of indices in a flexible modelling approach is most effective to predict the variation in the acoustic environment" - Towsey, 2014
- Indices that most affected the predictive ability of models were those that reflected both the temporal and spectral distribution of acoustic energy.
- Removing ACI resulted in max increase in MSE
- "We found that acoustic descriptors, or acoustic indices summarizing sound energy (e.g., mean and median sound levels), were less important than variation+evenness in sound over time and frequency bins to predict diversity

---

[3] "Acoustic indices as rapid indicators of avian diversity in different land ...." 9 May. 2018, https://www.veruscript.com/api/files/be13ad8b-527c-11e8-b3ad-0242ac110002/download. Accessed 15 Oct. 2018.

Acoustic diversity index[4] ADI_even and a variation, ADI

- **ADI_even:** Acoustic Diversity Index by grouping into even frequency bins
  - The recordings were separated into eight frequency bands each representing a 1,000 Hz range such that band 1: 0–1 kHz, band 2: 1 kHz–2 kHz,, and so on up till band 8 = 7 kHz–8kHz.
  - We calculated the diversity across frequency bands at each site for each day using an acoustic diversity index (ADI) based on the Shannon index following Pijanowski et al. (2011b) and Villanueva-Rivera et al. (2011) with this equation:
    ADI = sum (p_i * ln p_i) for all S frequency bands where p_i is the fraction of sound in each ith frequency band in S number of frequency bands.
  - ADI thus measures how full each of the 1,000 Hz bands are overall, which indicates the extent different acoustic niches are occupied in the recording. Only sounds above -50 dBFS were used in order to get rid of background noise. The cutoff of -50 dBFS was determined after listening to the sound recordings to ensure local fauna vocalizations were not being inadvertently excluded. The acoustic diversity of each site was then obtained by averaging the ADI values at each site over the six-day period.
  - https://cran.r-project.org/web/packages/soundecology/soundecology.pdf - Refer page 4 for ADI documentation.
  - Code in R for ADI by Villanueva: https://github.com/ljvillanueva/bits-and-pieces/blob/master/R/soundscapes-primer/soundscape_band_diversity.R
- **ADI:**
  - Acoustic Diversity Index by grouping bins on the mel scale
  - Instead of bands of 1 kHz between 0 and 8 kHz, we group the bins according to the mel scale
  - There are 70 frequency values in the given mel spectrograms. We group them into 7 bins, each with 10 of the frequency values
  - Therefore, every bin encompasses a larger frequency range than the previous bin

[4] "Modeling acoustic diversity using soundscape recordings and LIDAR ...." 9 Oct. 2012, https://link.springer.com/article/10.1007/s10980-012-9806-4. Accessed 15 Oct. 2018.

Spectral Entropy

- Implemented as a modified version of spectral entropy as described in Sueur (2008)[5].
- According to the paper: A mean spectrum s(f) is first computed using STFT with a non-overlapping sliding window
- s(f) is transformed into a probability mass function S(f) of length N, which is used to compute the spectral entropy

$$H_f = -\sum_{f=1}^{N} S(f) \times \log_2 S(f) \times \log_2(N)^{-1}, \text{ with } H_f \in [0,1].$$

- Modified implementation:
  - Take one time window in the spectrogram
  - Normalize the spectrogram by dividing each column in the spectrogram by the absolute sum of the column values (one time window, across all frequencies)
  - Take the absolute value of this computed spectrogram - this is spec_norm
  - Again, for each time window, element-wise multiply the column with $\log_2$(column values + 10^{-6})
  - Add these products, so now we have one value for each time window
  - Divide these values by $\log_2$(no. of frequency values)
  - Sum all these values to get one value of spectral entropy
  - Divide by (no. of frequency values) again

NDSI: Normalized Difference Soundscape Index[6]

- Used intensity values instead of PSD (with Welch estimation) as in the paper.
- We define anthrophony bin: 1-2 kHz and biophony bin: 2-8 kHz, and find the corresponding closest indices in the mel filtered spectrogram
- For each bin, we take absolute sum across all time windows in that freq range (like summing over a rectangle)
- NDSI is defined as:

---

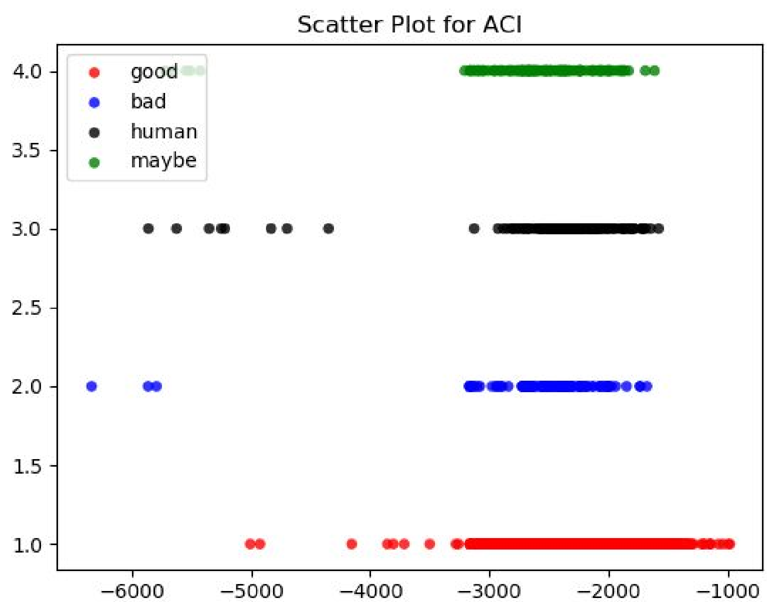
[5] "Rapid Acoustic Survey for Biodiversity Appraisal - PLOS." 30 Dec. 2008, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0004065. Accessed 15 Oct. 2018.

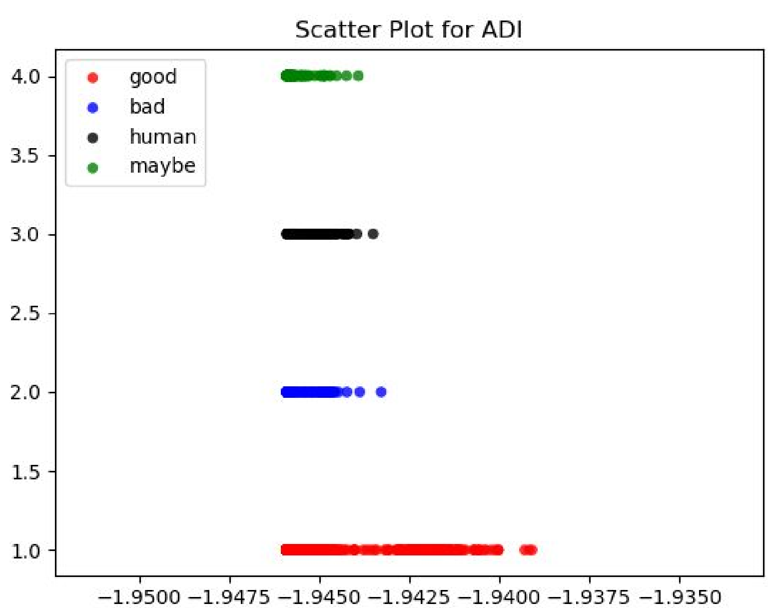
[6] "The remote environmental assessment laboratory's acoustic library ...." https://www.sciencedirect.com/science/article/pii/S157495411200088X. Accessed 15 Oct. 2018.

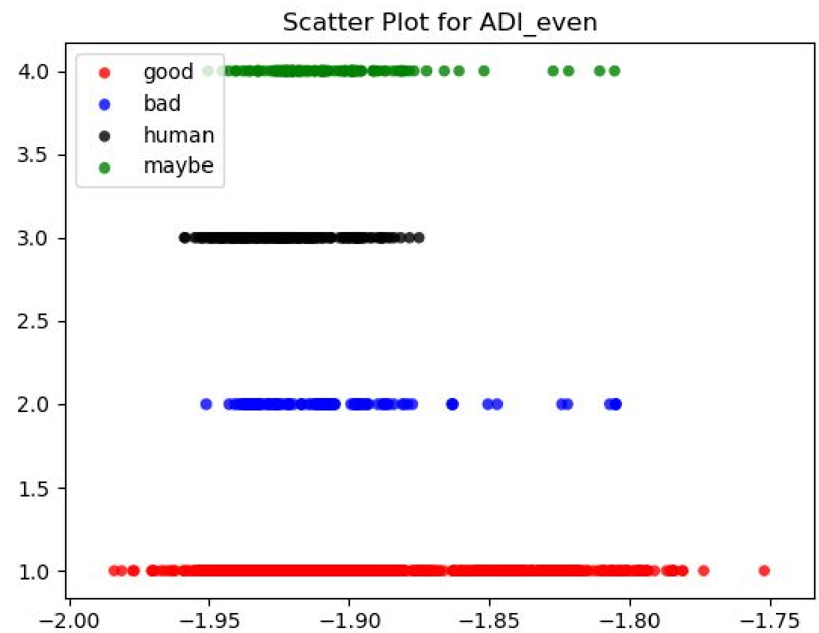
NDSI = (bio_sum-anth_sum)/(bio_sum+anth_sum)

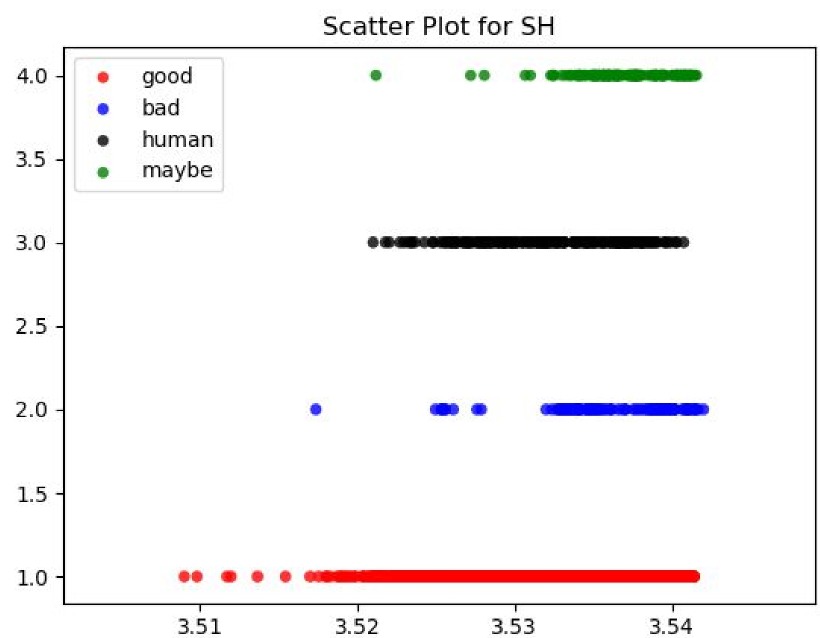
Jitter Plots

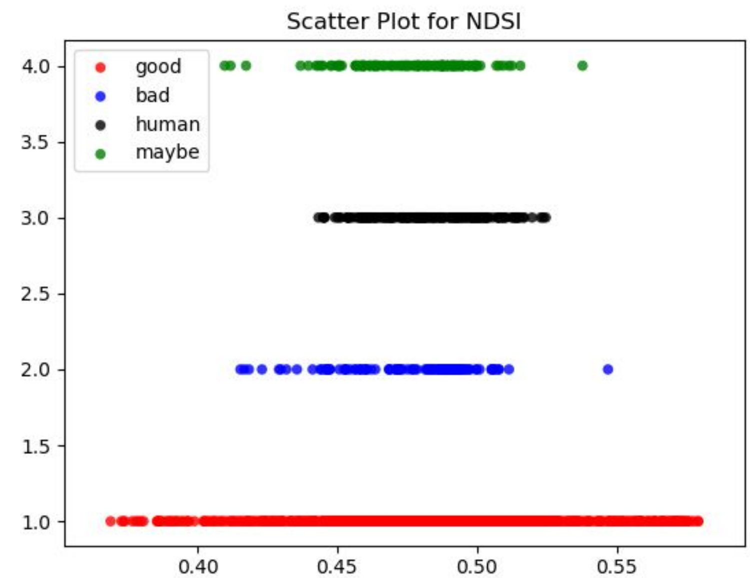
https://www.reddit.com/r/dataisbeautiful/comments/4kxtid/whats_the_best_way_to_show_a_scatter_plot_with/
http://vault.hanover.edu/~altermattw/courses/220/R/ttest/assets/jitter.png



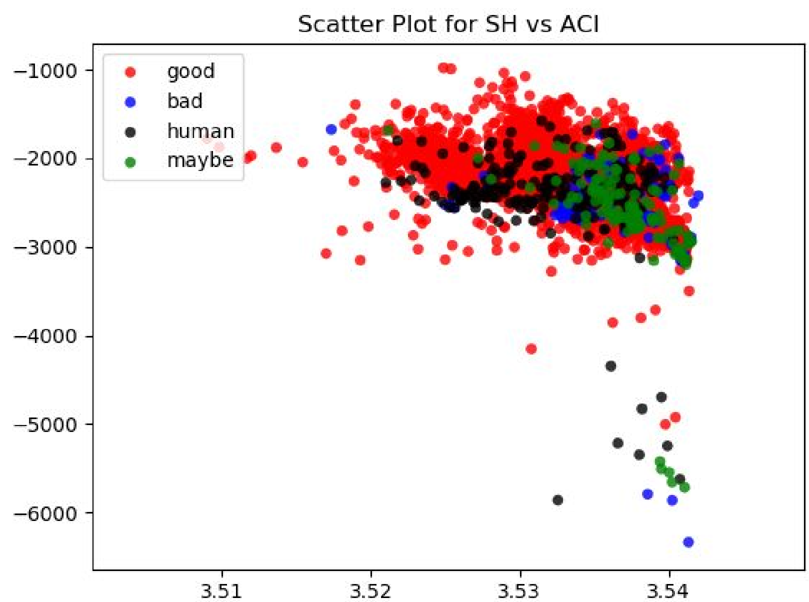



Scatter Plot for ADI_even
good
bad
human
maybe
4.0
3.5
3.0
2.5
2.0
1.5
1.0
−2.00
−1.95
−1.90
−1.85
−1.80
−1.75

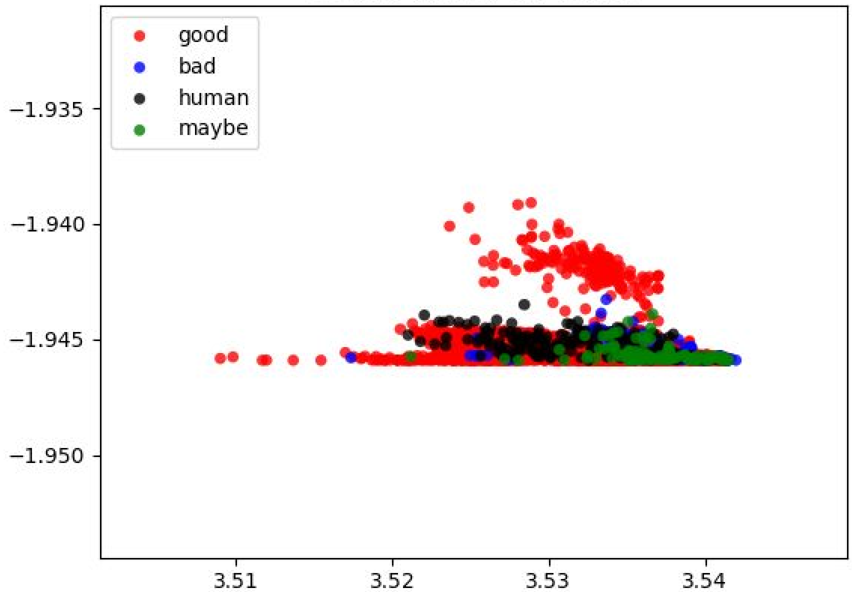

Scatter Plot for SH
good
bad
human
maybe
4.0
3.5
3.0
2.5
2.0
1.5
1.0
3.51
3.52
3.53
3.54



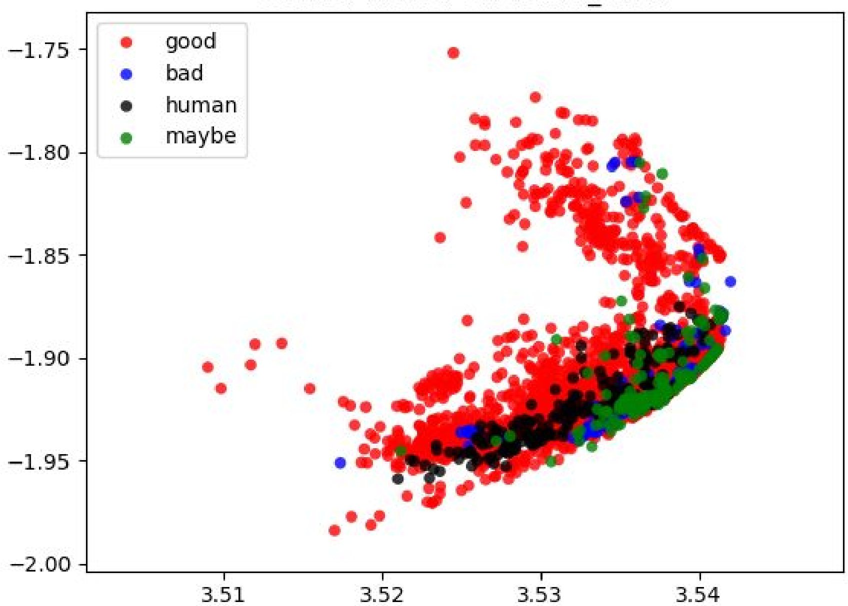
Scatter plots of SH vs all

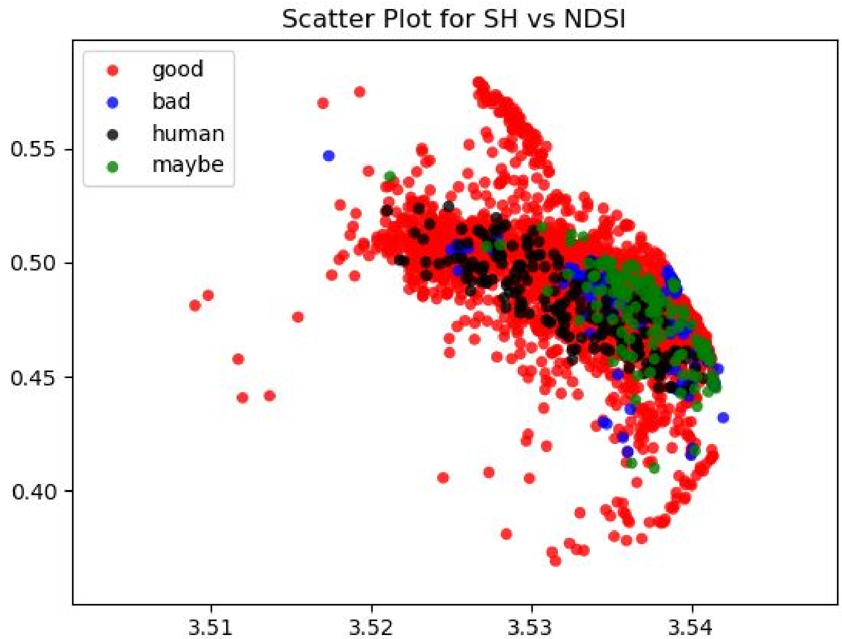


Scatter Plot for SH vs ADI



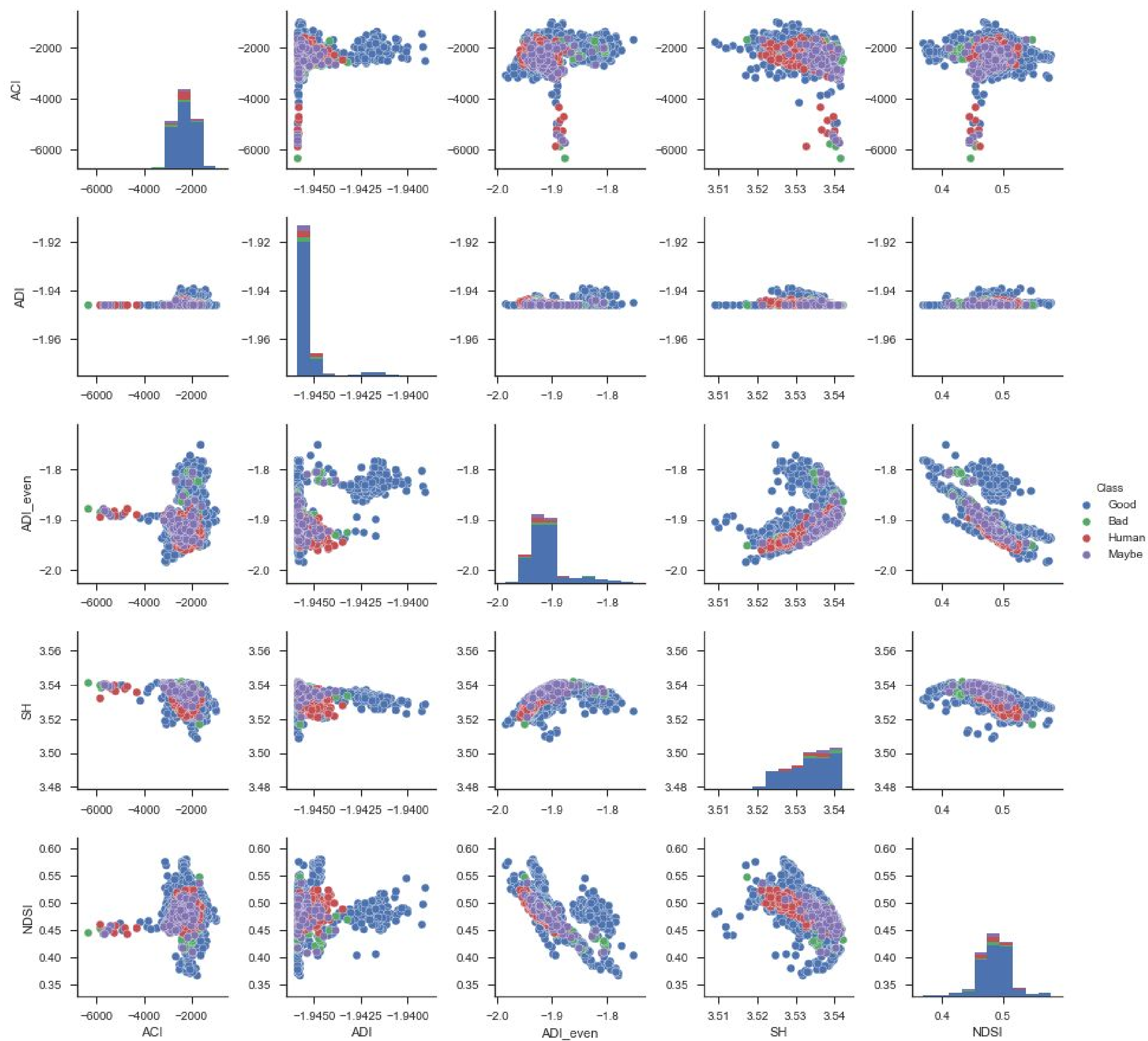
Scatter Plot for SH vs ADI_even



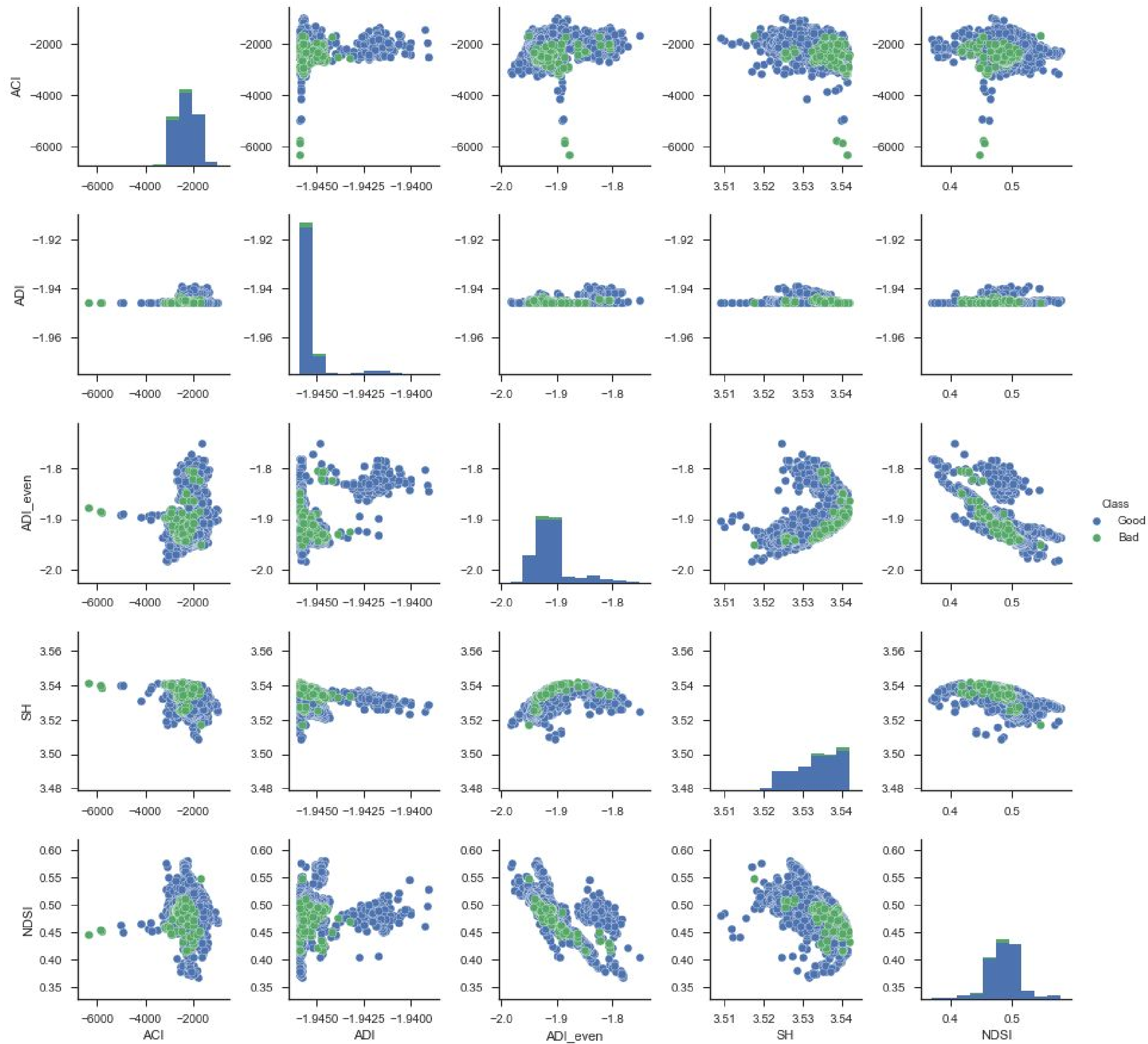


Scatterplot matrix with Pandas

https://machinelearningmastery.com/visualize-machine-learning-data-python-pandas/
http://seaborn.pydata.org/examples/scatterplot_matrix.html



Scatterplot matrix for all 4 classes with 5 acoustic indices



Scatterplot matrix for all 2 classes - Good, Bad - with 5 acoustic indices

Acoustic indices to be explored

- **Spectral kurtosis paper**[7]
  - Mentioned as an acoustic index in the review paper but check other spectral features from p112 onwards
- **Peakfreq**[8]
- **Spectral diversity**[9]
  - Spectral diversity: measured as the number of distinct spectral clusters in a one minute recording segment.
  - Expected to be a helpful indicator of spectral richness and therefore of species richness.
- Seewave documentation: specprop, p162

Interpretation of scatter matrix

- In the scatter matrix with two classes - Good and Bad - try to draw a hyperplane
- Open and manually check the plots for those spectrograms which have been misclass

---

[7] "Features for Content-Based Audio Retrieval - ScienceDirect." https://www.sciencedirect.com/science/article/pii/S0065245810780037. Accessed 15 Oct. 2018.
[8] "Rapid Acoustic Survey for Biodiversity Appraisal - PLOS." 30 Dec. 2008, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0004065. Accessed 15 Oct. 2018.
[9] "The use of acoustic indices to determine avian species richness in ...." https://www.sciencedirect.com/science/article/pii/S1574954113001209. Accessed 15 Oct. 2018.