

¹ **Analyzing Audio Patterns During the 2024
Total Solar Eclipse**

³ Kaitlin Heintzman Advisor: Matt Higham

⁴ 2025-04-30

5 Table of contents

| | |
|--|-----------|
| 6 1 Abstract | 4 |
| 7 2 Data Collection Methods | 4 |
| 8 3 Indices and Packages | 8 |
| 9 3.1 Acoustic Complexity Index | 8 |
| 10 3.2 Acoustic Diverstiy Index | 9 |
| 11 3.3 Acoustic Evenness Index | 10 |
| 12 3.4 Bioacoustic Index | 10 |
| 13 3.5 Biophony | 11 |
| 14 4 Initial visualization | 12 |
| 15 4.1 Creating a cleaned data frame | 14 |
| 16 4.2 Using the HPC | 18 |
| 17 4.3 Analyzing one full file from an audio recorder | 19 |
| 18 5 Full data visualization | 27 |
| 19 5.1 Acoustic Complexity | 28 |
| 20 5.2 Acoustic Diversity | 29 |
| 21 5.3 Acoustic Evenness | 30 |
| 22 5.4 Bioacoustic Index | 31 |
| 23 5.5 Biophony | 32 |

| | | |
|----|-----------------------------------|-----------|
| 24 | 6 Modeling | 32 |
| 25 | 6.1 Bioacoustic Index | 34 |
| 26 | 6.2 Acoustic Evenness | 37 |
| 27 | 6.3 Acoustic Complexity | 40 |
| 28 | 6.4 Acoustic Diversity | 41 |
| 29 | 6.5 Biophony | 43 |
| 30 | 7 Conclusion | 45 |
| 31 | References | 45 |

³² **1 Abstract**

³³ From March 30 to April 16th, Dr. Erika Barthelmess and students from the St. Lawrence
³⁴ biology department collected data from 20 audio recording devices in the Northern New York
³⁵ area. These devices were strategically located to observe potential audio changes in wildlife
³⁶ at the time of the total solar eclipse on April 8th, 2024. We processed this audio data to
³⁷ produce organized data frames, centered around 5 commonly used audiological indices. From
³⁸ this tabular data, we constructed many visualizations displaying patterns across time, day, and
³⁹ recording device. Due to the non-linearity present in our plots, we implemented generalized
⁴⁰ additive models based on date, time of day, and recording device to assess patterns that may
⁴¹ have occurred during the time of the eclipse. From these visuals and models, we conclude the
⁴² possibility that 3 of our 5 indices are associated with audio changes during the time of the
⁴³ eclipse.

⁴⁴ **2 Data Collection Methods**

⁴⁵ This protocol was written by Dr. Erika Barthelmess from St. Lawrence University. She, along
⁴⁶ with Jessica Harmen, Evelyn Albrecht, and Kelsey Simler completed this protocol and collected
⁴⁷ data for this project.

⁴⁸ Data was collected between March 30 and April 16, 2024 by deploying 20 AudioMothTM
⁴⁹ acoustic recorders throughout St. Lawrence County, New York (Figure 1). All locations were
⁵⁰ fully within the path of totality for the April 8, 2024 total solar eclipse. The partial eclipse

51 began at 14:11:38, totality began at 15:23:52, maximum eclipse was at 15:25:29, totality ended
52 at 15:27:05 and the partial eclipse ended at 16:35:38 (all times local, times from <https://www.timeanddate.com/eclipse/in/@5111484?iso=20240408>) for a total eclipse duration of 2 hours
53 24 minutes, with totality lasting 3 minutes and 13 seconds.

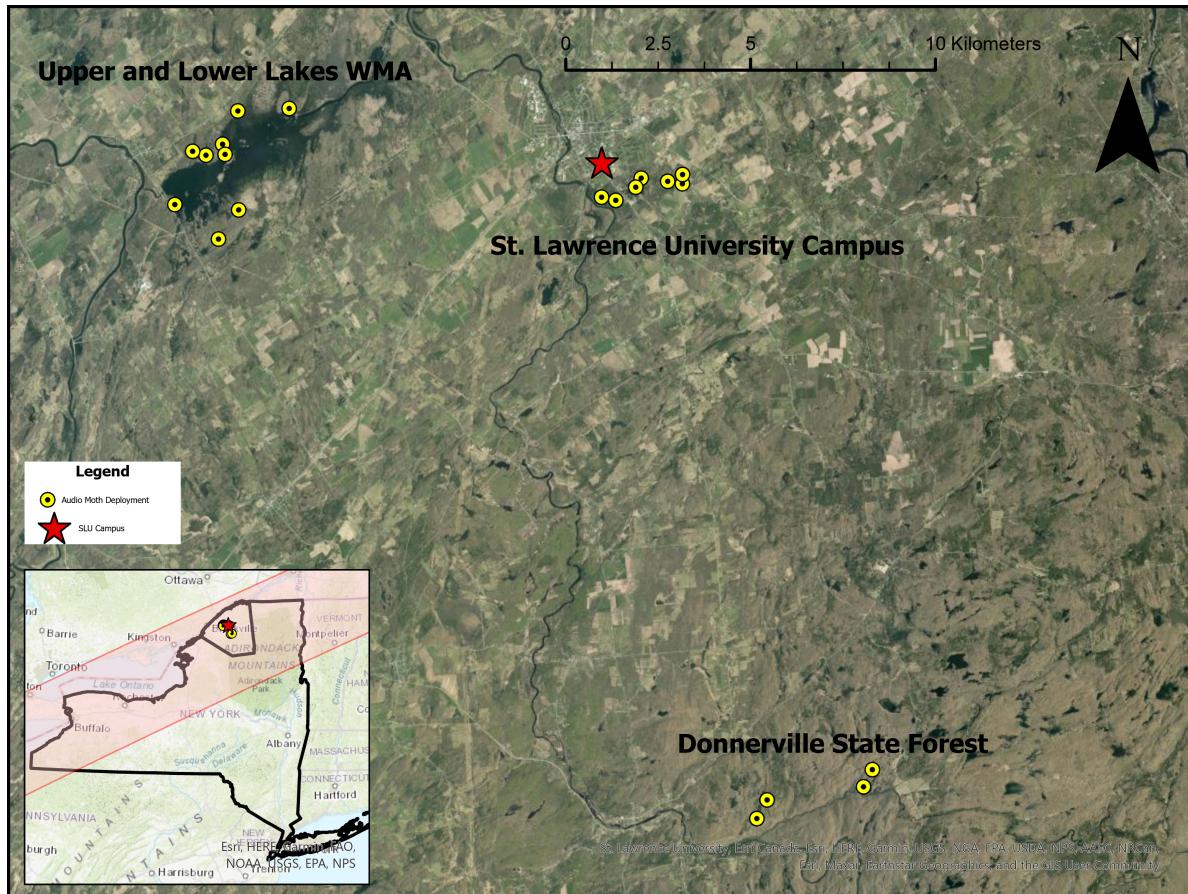


Figure 1: Audiemoth Locations

55 Each Audiemoth was configured to record within four temporal windows on each day of the
56 deployment. The first window was from 05:45 to 07:15, the second from 14:00 to 16:50, third
57 from 19:00 to 20:00 and the last from 23:00 to 23:30 (Table 1).

Table 1: Recording Times

| Times | Reasoning |
|---------------|--|
| 5:45 - 7:15 | 30-45 minutes before sunrise |
| 14:00 - 16:50 | full period of time corresponding to the eclipse |
| 19:00 - 20:00 | ~30 minutes before and after sunset |
| 23:00 - 23:30 | sample nocturnal activity |

58 Within each time window, each AudioMoth recorded in a repeated cycle with 55 seconds
 59 of recording and 5 seconds to write data for every minute of the recording window (Table
 60 2). Sample rate measures the density of recordings per unit time and therefore the range of
 61 frequencies that can be recorded. High sample rates record a higher range of frequencies but
 62 take up more space on the microSD card. We selected a sample rate of 96 kHz to capture sound
 63 frequencies up to about 48 kHz, which allowed us to capture common bird and amphibian songs
 64 and calls as well as at least some insects and bat echolocation sounds. Gain is a measure of
 65 the degree to which the microphone amplifies the sound as it is recorded. Higher gain enables
 66 detection of quieter sounds but can also result in clipping and distortion. After collecting pilot
 67 recordings near wetlands where wood frogs were calling, we determined that a gain setting of
 68 4 would help increase our detection of animal sounds.

Table 2: Recording Parameters

| Parameters | Setting |
|------------------------------|--|
| Sample rate (kHz) | 96 |
| Gain | Relatively high (4 on a 5 point scale) |
| Sleep Duration (seconds) | 5 |
| Recording Duration (seconds) | 55 |

69 We used ArcGIS Pro to identify areas of forest-wetland interface or forested wetland occurring
 70 on public or University-owned land. Our intention was to place the recorders at locations where
 71 they could capture the sounds of both forest birds and pond-breeding amphibians (as well as
 72 other biotic sounds including any active insects or bats). Due to our northern location within
 73 the path of the eclipse in North America, the onset of spring was just beginning. Red-winged
 74 blackbirds (*Agelaius phoeniceus*) had returned to the area and were establishing breeding
 75 territories and both wood frogs (*Lithobates sylvaticus*) and spring peepers (*Pseudacris crucifer*)
 76 had begun to chorus at least 5 days prior to deployment of the recorders. To reduce the time
 77 required to deploy units, we located the devices near but out of view of hiking trails and within
 78 20 miles of the St. Lawrence University campus (44.58931027483651° N, -75.1613716006626 °
 79 W).

⁸⁰ **3 Indices and Packages**

⁸¹ In our analysis, we took a lot of inspiration from a paper published in 2020 titled “*Soundscape*
⁸² *shifts during the 2017 total solar eclipse: An application of dispersed automated recording*
⁸³ *units to study ephemeral acoustic events*” by Jacob E. Gerber, Dakota Howard, and John E.
⁸⁴ Quinn.

⁸⁵ Using the **soundecology** package, we selected 5 specific indices that appeared in our reference
⁸⁶ paper (Gerber, Howard, and Quinn (2020)), that we believed would be important in under-
⁸⁷ standing the changes in acoustic activity. Due to settings on the recording devices, all data
⁸⁸ was from the left channel as the right channel values were not available.

⁸⁹ **3.1 Acoustic Complexity Index**

⁹⁰ The acoustic complexity index focuses on understanding the spatial and temporal complexity
⁹¹ of sound. It does so by numerically portraying the variation in sound frequency within the
⁹² provided audio file. This index will reflect intensity and frequency shifts in an environment.

⁹³ A higher ACI value indicates a more diverse soundscape compared to a lower value, which
⁹⁴ could indicate more monotone and stable sounds.

⁹⁵ In the Gerber et. al paper in 2017, this index was not found to be significant in their model but
⁹⁶ it was apparent that ACI was greatest during totality (Gerber, Howard, and Quinn (2020)).

97 This index is obtained using the `acoustic_complexity()` function within the soundecology
98 package. The specific numerical values used are from subsetting `$acl_left_vals` (Villanueva-
99 Rivera and Pijanowski (2018)).

100 **3.2 Acoustic Diverstiy Index**

101 The acoustic diversity index (ADI) assesses both the variety and evenness in the sound dis-
102 tribution across different bands. This index is calculated by generating proportions of data
103 within a specific interval that reach above a specified threshold (default -50dBFS).

104 Similar to ACI, a higher ADI value corresponds to a increase in diversity within a habitat and
105 a low ADI refers to a less biodiverse location.

106 In the Gerber et. al paper in 2017, this index was not found to be significant in their model but
107 it was apparent that ADI was highest at the times right before and after the eclipse (Gerber,
108 Howard, and Quinn (2020)).

109 This index is obtained using the `acoustic_diversity()` function in the soundecology package.
110 The specific numerical values used are from subsetting `$left_band_values` (Villanueva-Rivera
111 and Pijanowski (2018)).

112 For both the Acoustic Diversity index and the Acoustic Complexity index, multiple values are
113 stored per .WAV audio file. In this analysis, the list of values were summed to produce the
114 total ACI or ADI value per file, which made visualization easier.

¹¹⁵ **3.3 Acoustic Evenness Index**

¹¹⁶ The acoustic evenness index (AEI) measures how even the distribution of sound is across
¹¹⁷ different frequency bands. This index assesses the equality and inequality of sound power
¹¹⁸ distribution in different ranges by calculating the Gini index over segmented portions of an
¹¹⁹ audiofile. This index does not separate anthropogenic and biological sound.

¹²⁰ A high AEI value signifies a more even soundscape: within the audio, there is not any sound
¹²¹ domination from one type of sound/species. A low AEI signifies that the sound is not as even:
¹²² there may only be a few loud or similar sounds which disrupt the consistency of the sound
¹²³ level.

¹²⁴ The paper by Gerber et. al found that the AEI was highest right before the eclipse began
¹²⁵ and it rose until the period of totality. The authors believe this is likely due to the increase in
¹²⁶ human sounds near their recording locations (Gerber, Howard, and Quinn (2020)).

¹²⁷ This index is found using the `acoustic_evenness()` function in the `soundecology` package.
¹²⁸ The specific numerical value used is from subsetting `$aei_left` (Villanueva-Rivera and Pi-
¹²⁹ janowski (2018)).

¹³⁰ **3.4 Bioacoustic Index**

¹³¹ The bioacoustic index (BEI) provides a measurement of the diversity and abundance of biolog-
¹³² ical noise. This index is calculated by segmenting the data into bins between 2-8kHz, and then

¹³³ assessing the variation in relation to the lowest frequency from that bin. The total bioacoustic

¹³⁴ Index is the average across all of these bins.

¹³⁵ Higher values of this index signal an increase in species diversity, suggesting more species are

¹³⁶ making noise during the clip.

¹³⁷ In our reference paper, the authors found that BEI was greatest at the period of totality

¹³⁸ (Gerber, Howard, and Quinn (2020)).

¹³⁹ This is obtained using the `bioacoustic_index()` function in the soundecology package. The

¹⁴⁰ specific numerical value used is from subsetting `$left_area` (Villanueva-Rivera and Pijanowski

¹⁴¹ (2018)). The default minimum hertz value for this index was set at 2kHz.

¹⁴² **3.5 Biophony**

¹⁴³ Biophony is an index which calculates the average frequency of biotic sound (between 2-8kHz).

¹⁴⁴ The interpretation of this index is much more straightforward: high biophony means that there

¹⁴⁵ is more sound and low values mean less sound.

¹⁴⁶ Gerber et al. found that biophony increased at the beginning and end of the eclipse, similar

¹⁴⁷ to the dawn and dusk choruses (Gerber, Howard, and Quinn (2020)).

¹⁴⁸ This index is obtained as a byproduct from the Normalized Difference Soundscape Index

¹⁴⁹ function and the numerical values used is from subsetting `$biophony_left` (Villanueva-Rivera

¹⁵⁰ and Pijanowski (2018)).

¹⁵¹ **4 Initial visualization**

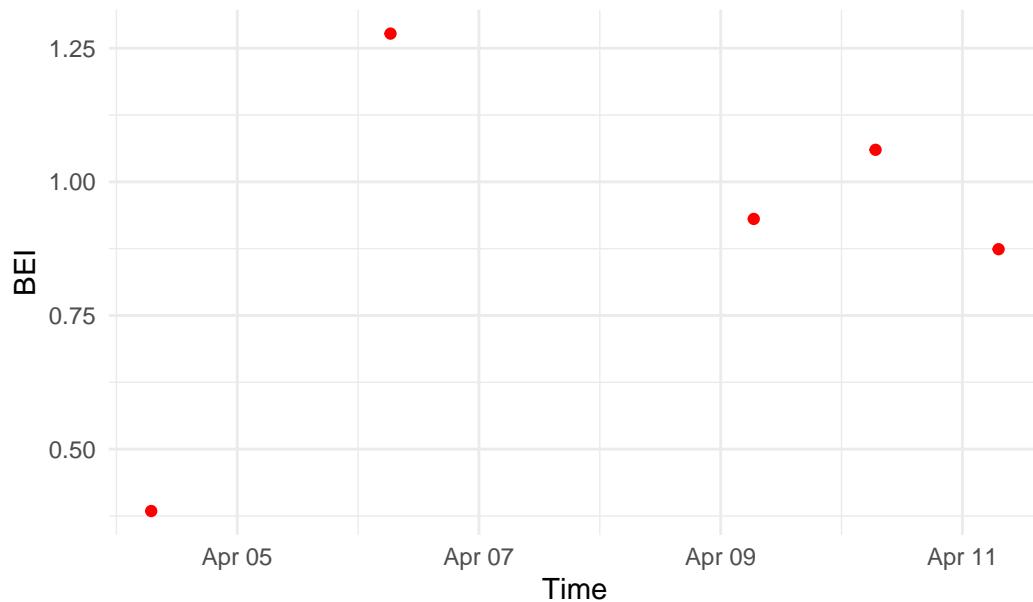
¹⁵² Due to the large amount of recordings at our disposal, we began creating graphics using a
¹⁵³ subset of five audio files (.WAV) from a single audio recorder to explore the indices. Table 3
¹⁵⁴ shows indices for 5 minutes from a single audiomoth on different days.

Table 3: First 5 Index values calculated (continued below)

| time | aei | bei | biophony |
|---------------------|----------|--------|----------|
| 2024-04-04 06:55:00 | 0.004024 | 0.3841 | 2.063 |
| 2024-04-06 06:26:00 | 0.06941 | 1.277 | 2.169 |
| 2024-04-09 06:34:00 | 0.4253 | 0.9306 | 1.071 |
| 2024-04-10 06:45:00 | 0.06465 | 1.06 | 1.646 |
| 2024-04-11 07:11:00 | 0.1946 | 0.874 | 1.375 |

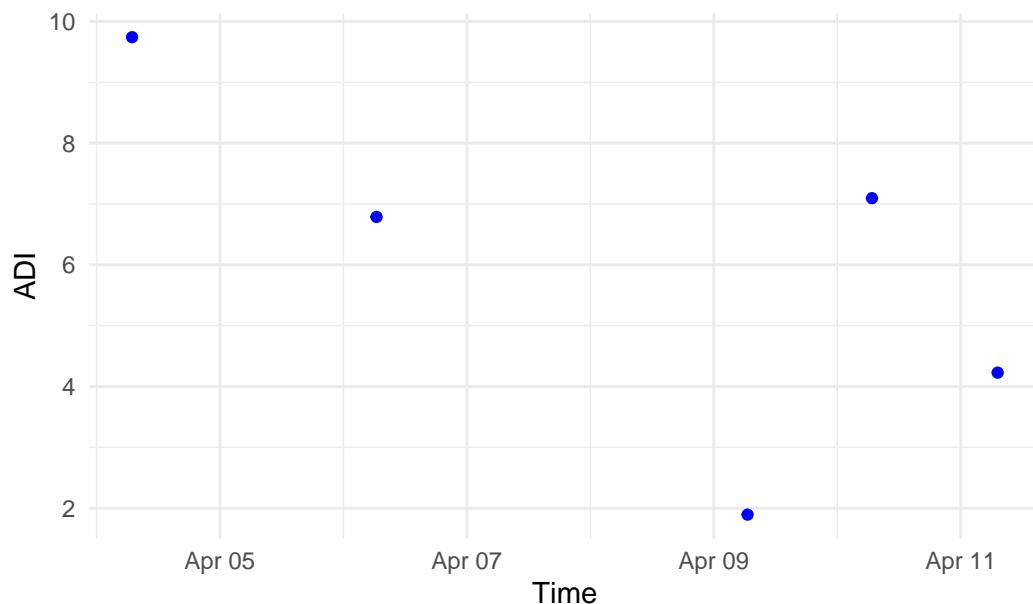
| ACI_all | ADI_all |
|-----------------------------------|-------------------|
| 9.01389387925447,6.42427026282213 | 0.9896,0.980612 |
| 8.8657794310719,6.4421208288537 | 0.844655,0.716814 |
| 9.06604640985277,6.49691812601518 | 0.473127,0.217534 |
| 8.85770232363042,6.52467721488273 | 0.857982,0.745923 |
| 8.8272143441091,6.39791145784707 | 0.691836,0.472565 |

First Graphic of Bioacoustic Index



155

First Graphic of Acoustic Diversity



156

157 The plots above show the Bioacoustic index and Acoustic Diversity values. Although these initial visualizations did not help us assess patterns, the graphics helped us gain an understanding

¹⁵⁹ of the ranges that we could expect these values to fall in.

Table 5: Assessment of Index Ranges

| Index | Range |
|---------------------|-------------|
| Acoustic Complexity | 1600 - 1700 |
| Acoustic Diversity | 0 - 10 |
| Acoustic Evenness | 0 - 0.5 |
| Bioacoustic Index | 0 - 1.5 |
| Biophony | 1 - 3 |

¹⁶⁰ 4.1 Creating a cleaned data frame

¹⁶¹ To expand the collection of data, we generalized our code and created a function, `eclipse_df()`

¹⁶² which would use a folder containing all .WAV files for one audio recorder as input and would

¹⁶³ return a cleaned data frame. This data frame would then be saved in our environment and

¹⁶⁴ would contain all of the data that we would need to create our visuals and models.

¹⁶⁵ Initially, indices were calculated from all .WAV files and stored in their own data frames. For

¹⁶⁶ the bioacoustic index, acoustic evenness index, and biophony, all values are subsetted from the

¹⁶⁷ original output of the `soundecology` function. These three indices are then bound into one

¹⁶⁸ tibble.

```

BEI<-as.data.frame(BEI_ALL)|>
  select(starts_with("left_area"))|>
  pivot_longer(everything(), names_to= "bei_name", values_to = "BEI")

```

169

170 For the acoustic complexity and the acoustic diversity, to obtain the full list of values for each
 171 folder, a for loop is incorporated into the function. This obtains all the left channel values and
 172 combines them into a list.

```

ACI_all<-vector("list", length(c(1:n)))
ADI_all<-vector("list", length(c(1:n)))

for (i in 1:n){
  ACI_all[i]<-(as.data.frame(ACI[[i]]$aci_fl_left_vals))
  ADI_all[i]<-(as.data.frame(ADI[[i]]$left_band_values))
}

```

173

174 After they have been calculated the data frames are bound together into one. The code listed
 175 below is for the final data frame, which creates all our indices, date/time information, and the
 176 audio recorder that the data corresponds to.

```

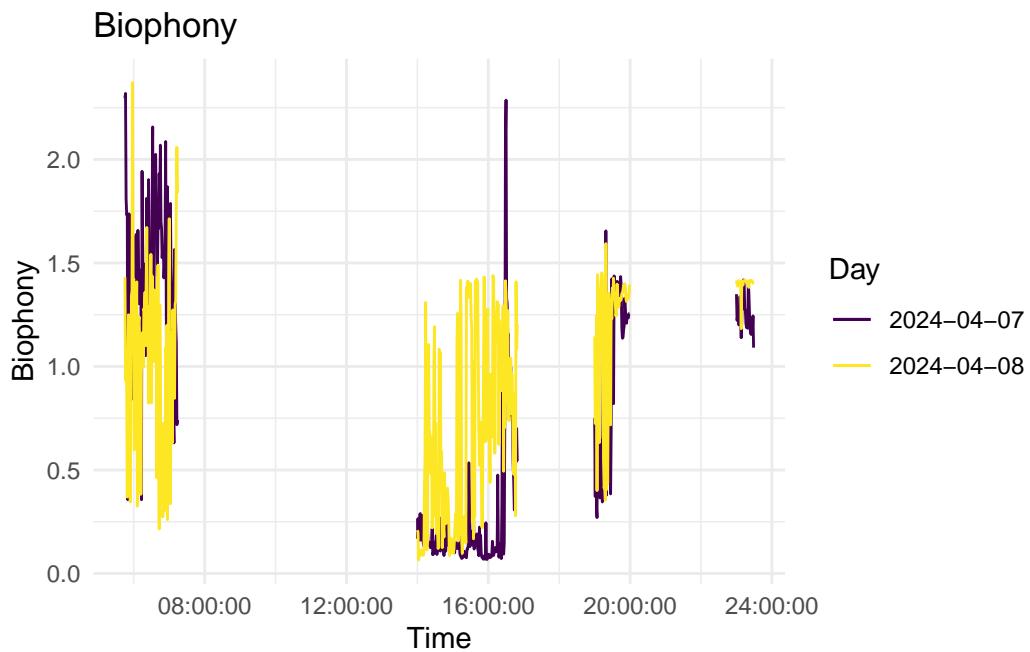
full|> mutate(biophony = as.numeric(biophony),
              aei = as.numeric(aei),
              bei = as.numeric(bei))|>
  separate(paths_date, into = c("date", "time_hms"), sep = "_")|>
  separate(time_hms, into = c("time", "wav"), sep = "\\.")|>
  separate(time, into = c("hours", "other"), sep = 2)|>
  separate(other, into = c("min", "sec"), sep = 2)|>
  mutate(date = parse_number(date))|>
  unite("time", c("date", "hours", "min", "sec"), sep = ":")|>
  mutate(time= ymd_hms(time))|>
  select(-wav)|>
  mutate(folder_name = deparse(str_remove(folder, here()))))|>
  select(folder_name, everything())

```

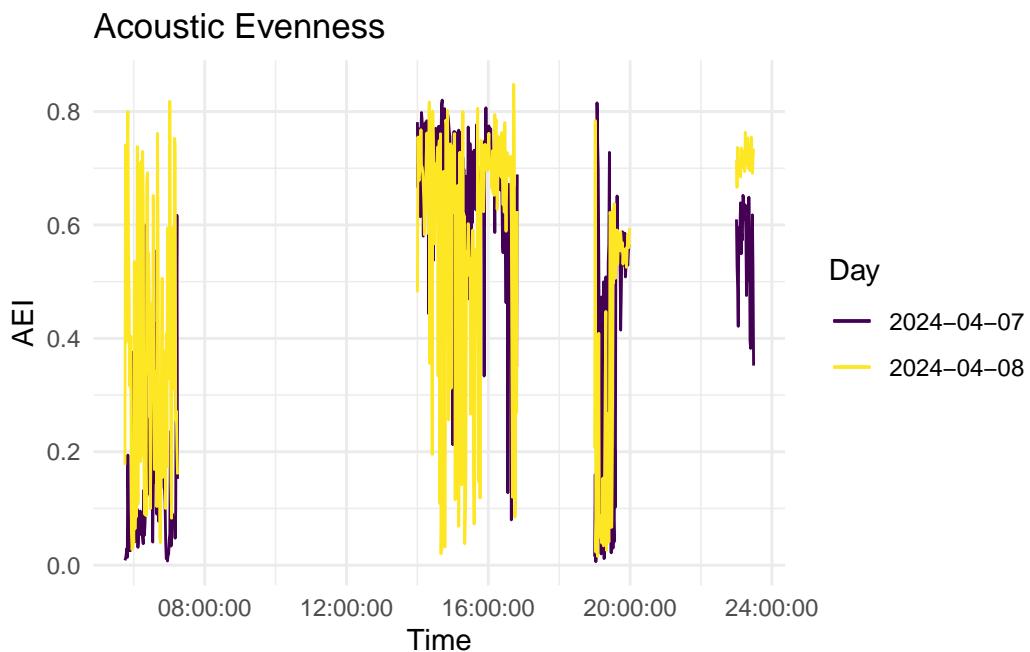
177

178 Initially, we tested this function on two folders which were a part of the larger collection of

¹⁷⁹ data from audio recorder 4. The two folders we chose correspond to April 7th and 8th. Graphs
¹⁸⁰ from the indexes across the four different recording times are shown below.

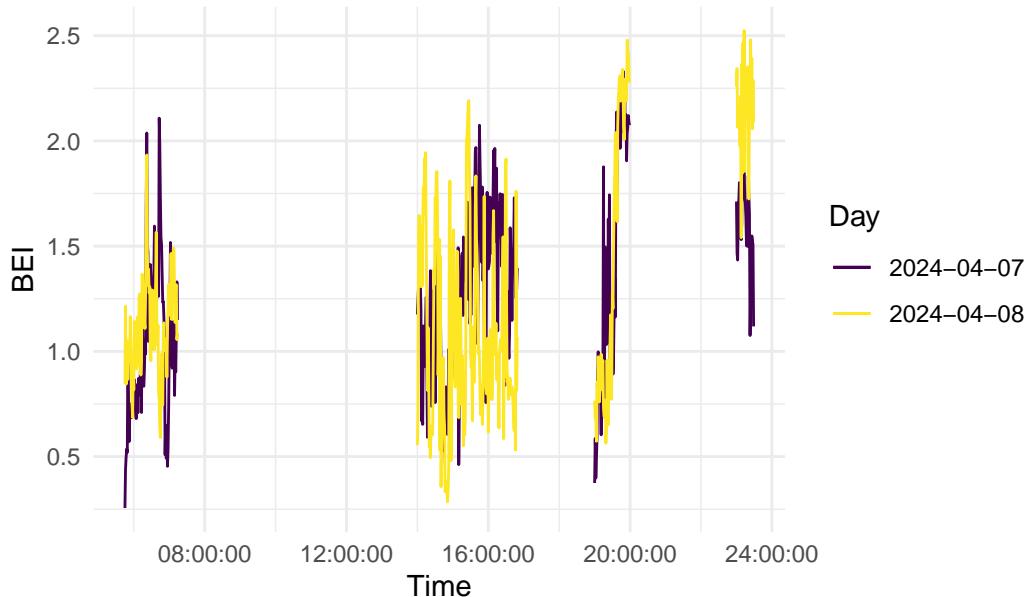


¹⁸¹



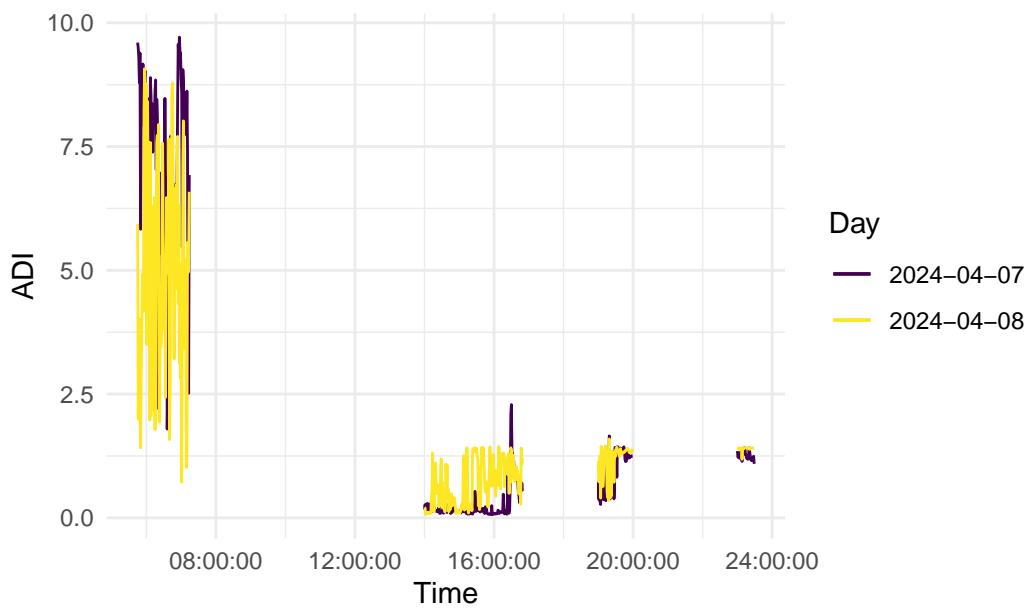
¹⁸²

Bioacoustic Index

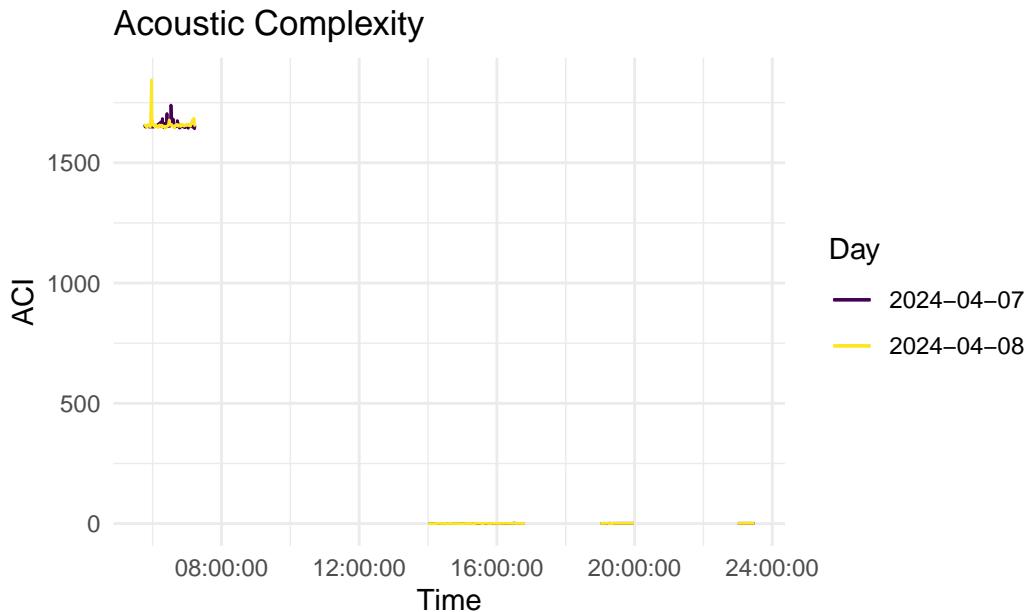


183

Acoustic Diversity Index



184



185

186 Visualizing these two data sets helped us understand the patterns that occur across a full day
 187 of recording. For example, a pattern in this visual above is that ADI, ACI, and biophony are
 188 higher during the dawn segment and then decrease to similar low levels across the rest of the
 189 day. Another thing these types of graphs shed light on are the potential differences between
 190 a normal day and the day of the eclipse. However, as this is just a comparison from two
 191 audiomoths, we could not make any conclusions about the changes during the eclipse from
 192 just this graphic.

193 **4.2 Using the HPC**

194 To assist in efficiently expanding to the full set of our audio files, we decided to utilize the
 195 High Powered Computer (HPC) to run our function on the large files for each audio recorder

196 and to extract the cleaned data. This would assist in speeding up the computation time, and
197 decreasing the amount of space utilized by WAV files on a desktop. We ran, visualized, and
198 analyzed all data from the 20 audio recorders. Running each audiomoth's data on the HPC
199 took roughly 10.5 hours.

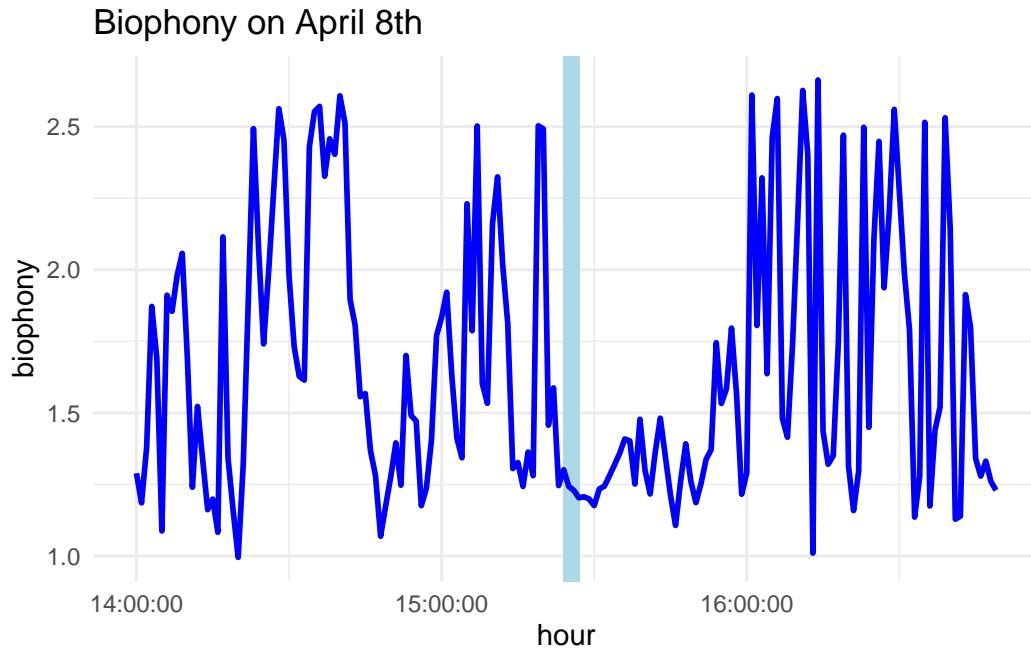
200 Through using the HPC, it was discovered that some of the folders contained empty .WAV
201 files which would kill the HPC job before the RDS was finished and saved. To combat this,
202 any folder which contained empty .WAV files, small .WAV files (<1 MB), and alternate file
203 types were deleted from the folder before running on the HPC. As this was only about 5-8
204 files, this did not cause concern about missing data or unequal RDS file sizes.

205 **4.3 Analyzing one full file from an audio recorder**

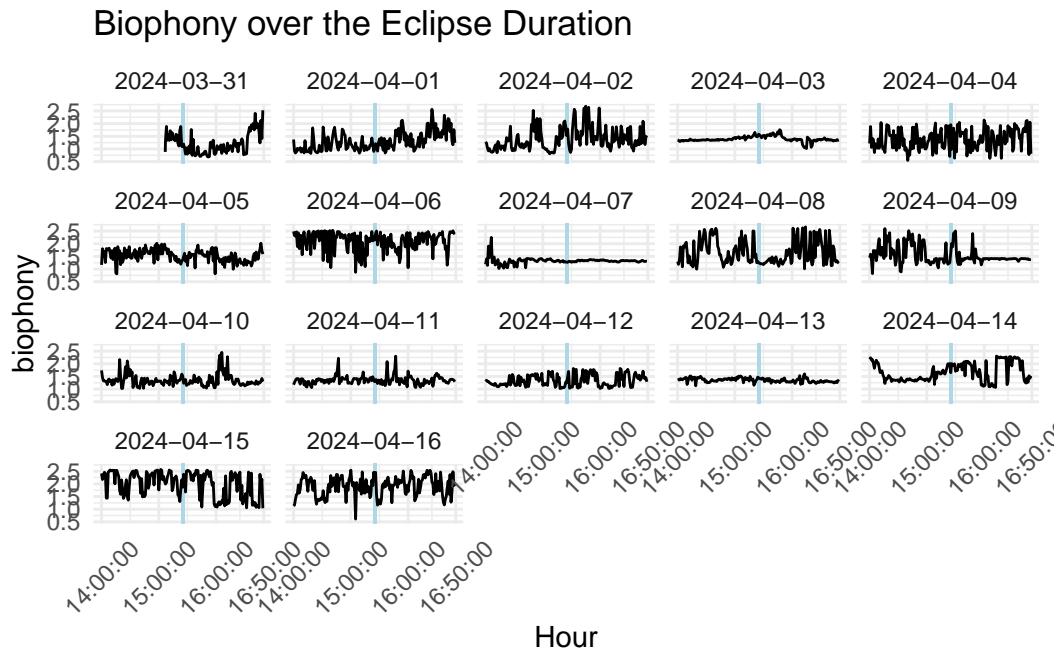
206 We began by running one folder corresponding to the first recorder (A001_SD001) on the
207 HPC to assess the efficiency of our function, understand how the HPC would work, and begin
208 our visualization of the full recording time. To analyze these full files we decided to focus on
209 looking for patterns across the days that the audio recorders were running. We wanted to see
210 if the eclipse data looked different from the other days when sampling occurred, and if the
211 time during the eclipse was similar to the dawn.

212 We started our analysis on the full time series, but this created a lot of crowding and was
213 difficult to read. We decided to focus on the time interval 14:00:00 to 16:50:00, which corre-
214 sponded to the eclipse. We then created two visuals: one displays only the pattern from the

215 day of the eclipse, and then a similar visual shows all of the recording days. In these visuals
216 we added a shaded region to show the time of totality (15:27:05 to 15:23:52).



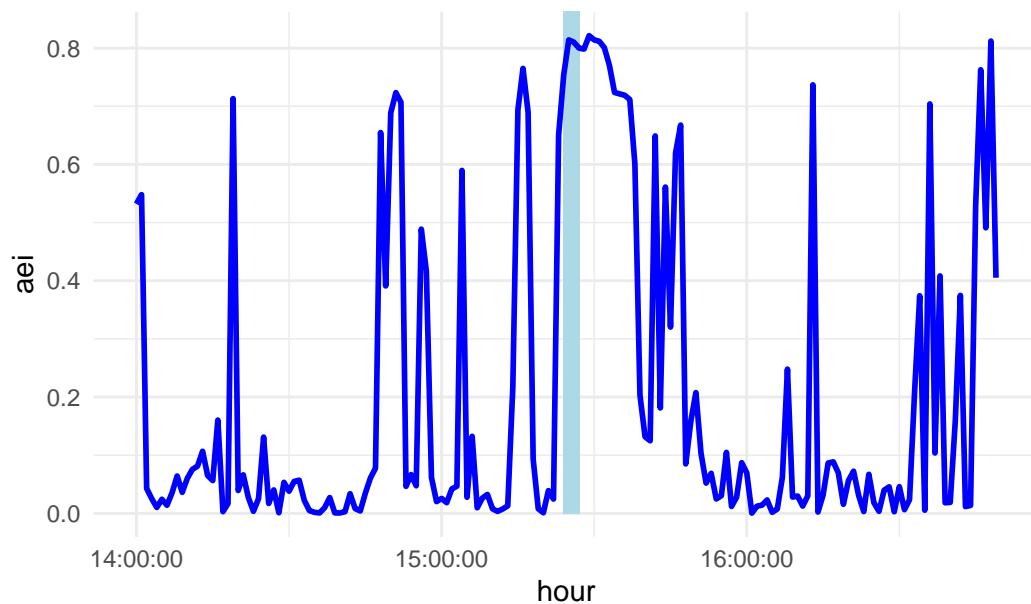
217



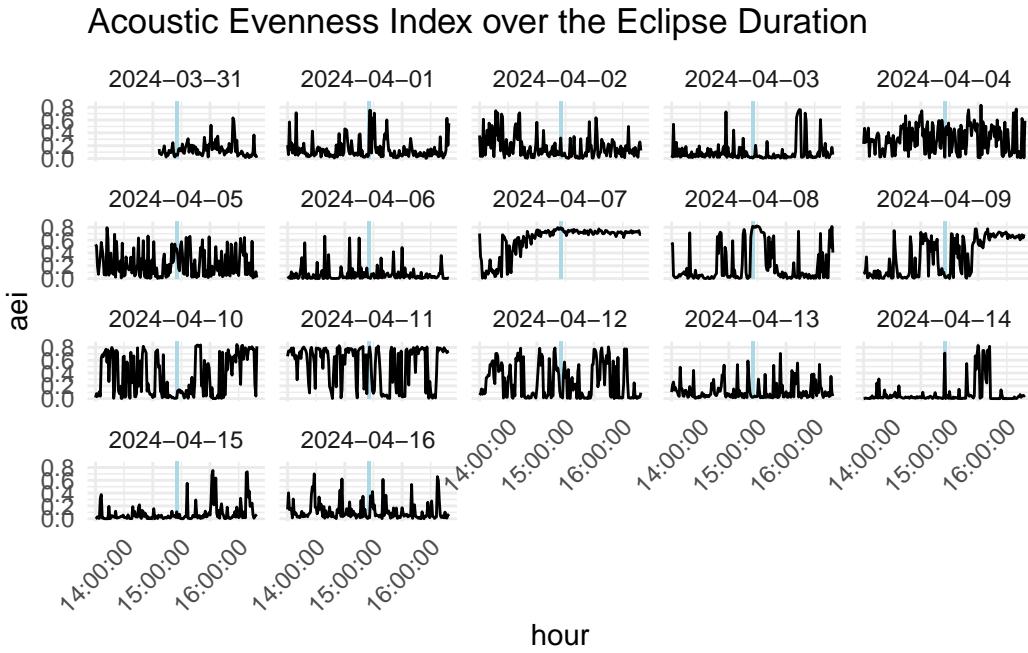
218

219 From the graph depicting biophony on the day of the eclipse, we can see there is a sharp
220 decrease before totality begins, and during/after totality the biophony from this recorder
221 seems to stay pretty low. When we look at the total recording days, there are days where the
222 biophony is at similar levels. From this one audiomoth, it is hard to tell if biophony will be
223 an index which shows potential patterns across the eclipse. We can also see that, even on the
224 non-eclipse days, there is a lot of variability with how much biophony is present.

Acoustic evenness on April 8th



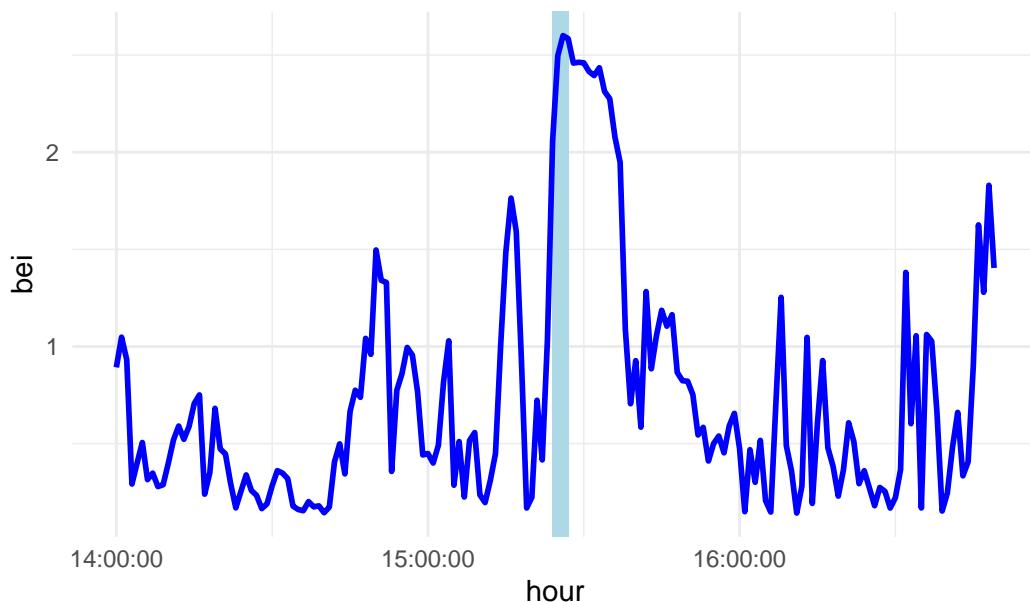
225



226

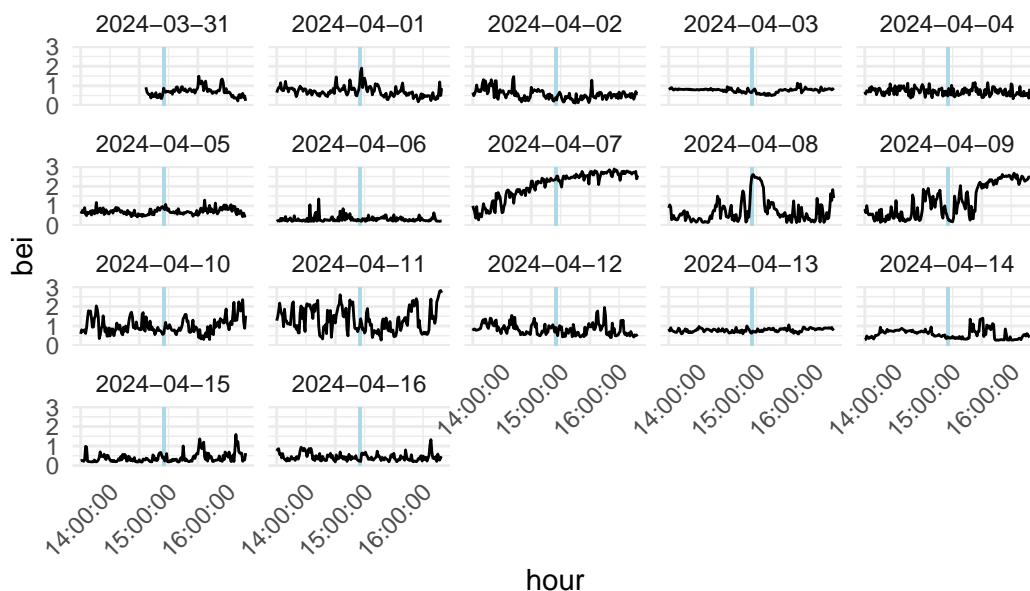
227 On the day of the eclipse, we can see an increase of AEI right at the beginning of totality. This
 228 increase is one of the starker increases in the full time interval. When we then look at all of
 229 the days, we can see that this type of increase is pretty infrequent at this time. The pattern
 230 on the 8th during totality seems to stick out. Again, there is a lot of between-day variability
 231 in acoustic evenness.

Bioacoustic Index over the Duration of the Eclipse



232

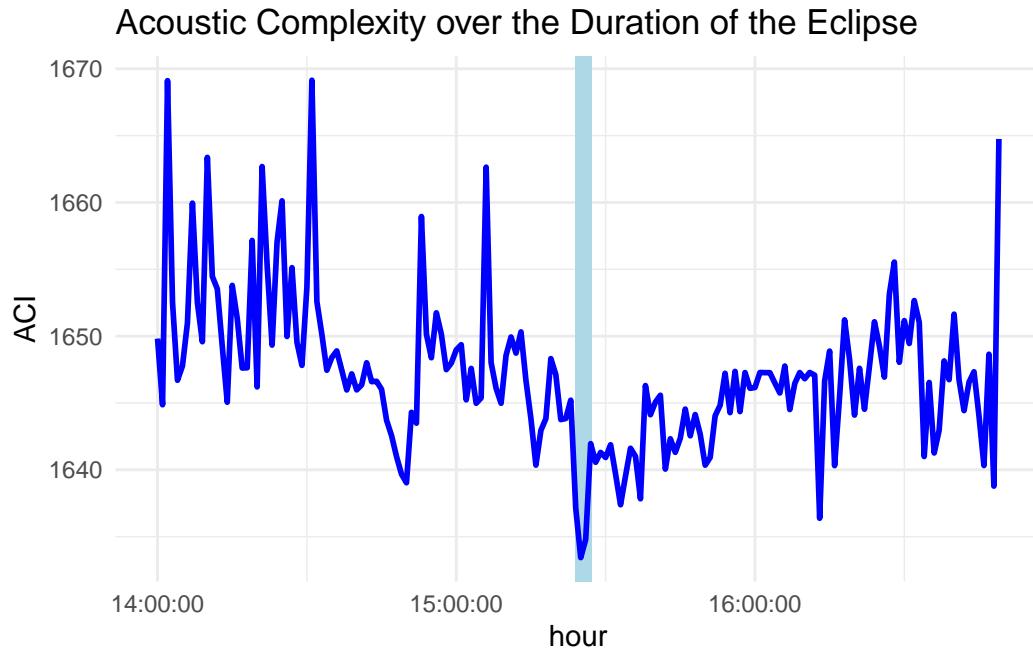
Bioacoustic Index over the Eclipse Duration



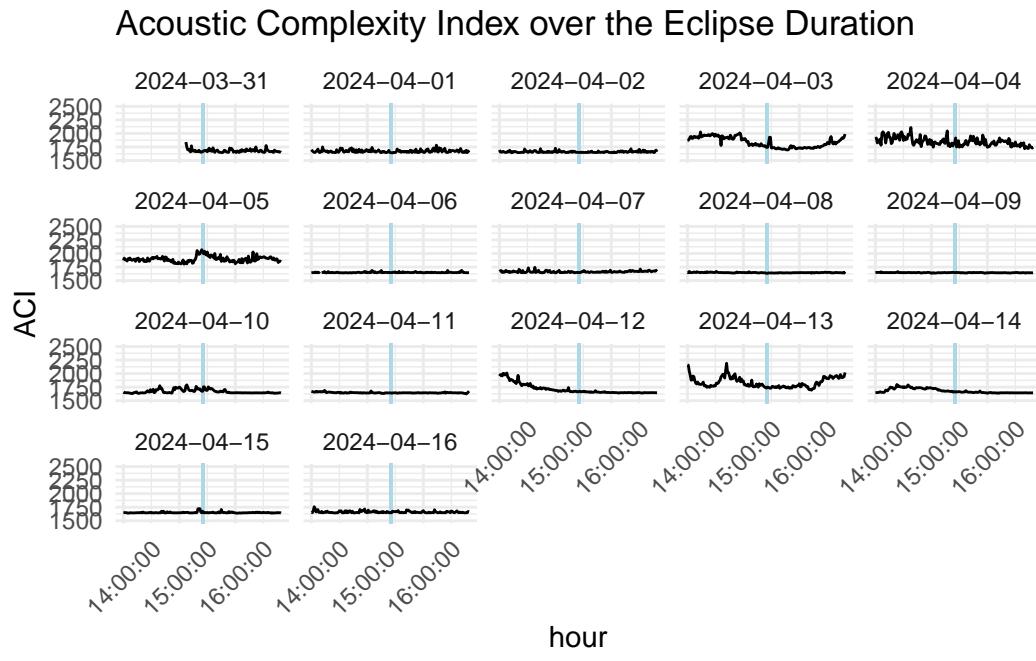
233

- ²³⁴ In this graph we see another striking pattern, the bioacoustic index increasing significantly at the beginning of totality compared to the rest of this period. This peak is the highest across
- ²³⁵

236 this eclipse interval, and the rest of the clip has pretty low BEI in comparison. Across all days,
237 we notice during this time the BEI is often much lower, and the stark increase pattern at this
238 time doesn't occur at any other day (besides potentially on April 1st).



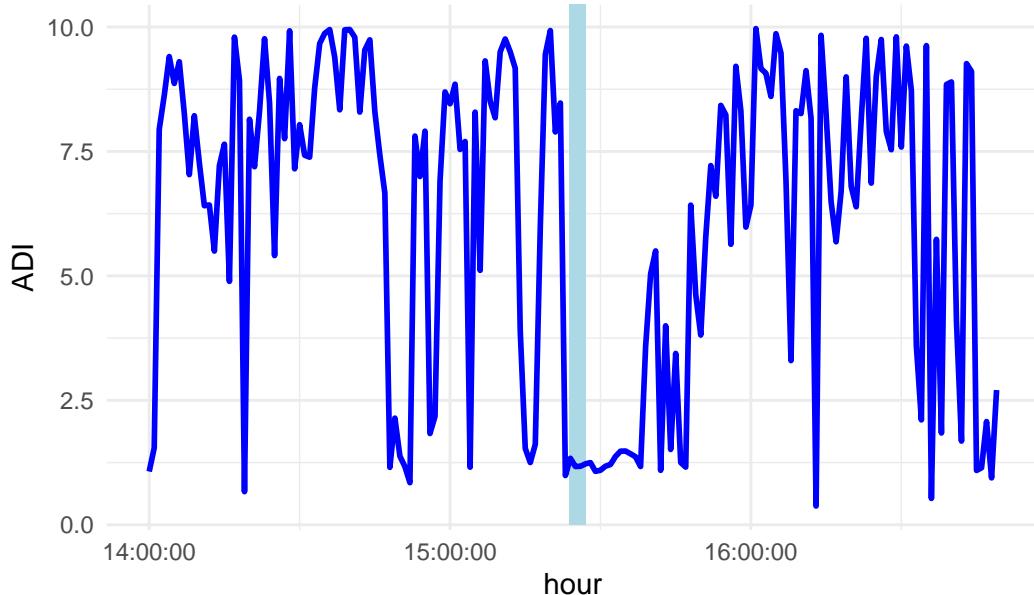
239



240

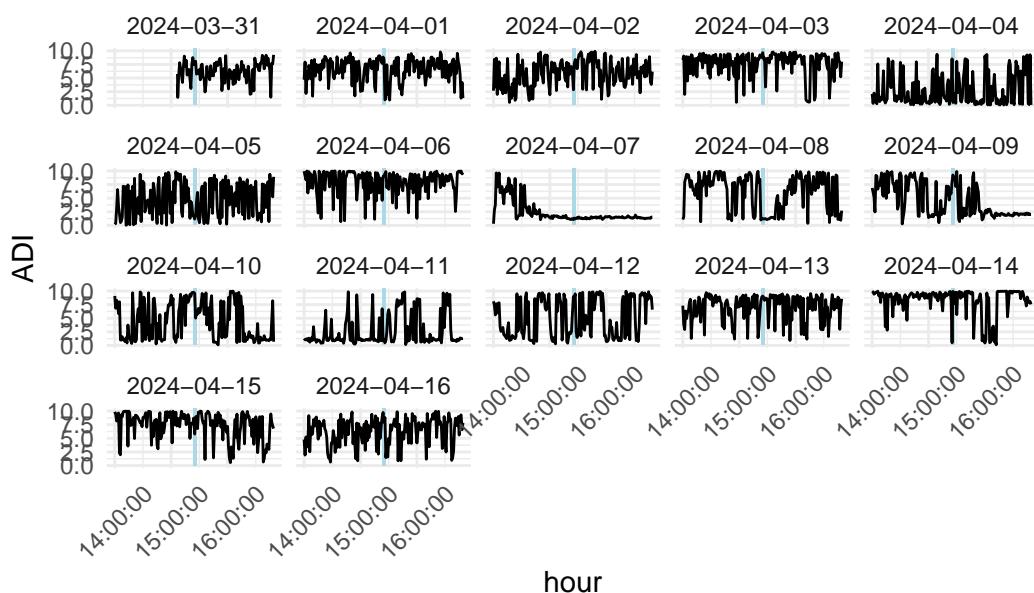
241 For the ACI, there seems to be the potential for an interesting pattern: on on the 8th, there
 242 is a sharp decrease during the time of totality. However, when we compare this to the rest
 243 of the days, we can see that the ACI for this day is quite linear, and other days have more
 244 variation.

Acoustic Diversity Index over the Duration of the Eclipse



²⁴⁵

Acoustic Diversity over the Eclipse Duration



²⁴⁶

²⁴⁷ Lastly, looking at the ADI in this audiomoth, we can see that there is potentially a decrease at
²⁴⁸ the point of totality. Like biophony, after totality we see the acoustic diversity stay pretty low.

249 When we try and look at the rest of the days we can see a large amount of variation in this
250 index. This variation makes it difficult to really compare the patterns, but this low stagnant
251 period around totality doesn't seem to occur in the other days, other than potentially April
252 7th.

253 This set of visuals allowed us to look into whether we could expect to observe any potential
254 patterns across the time of the eclipse for our five indices from a single audiomoth. From these
255 visuals, we felt confident that as we got into more visuals and modeling we would see some
256 type of audiological changes during the time of the eclipse.

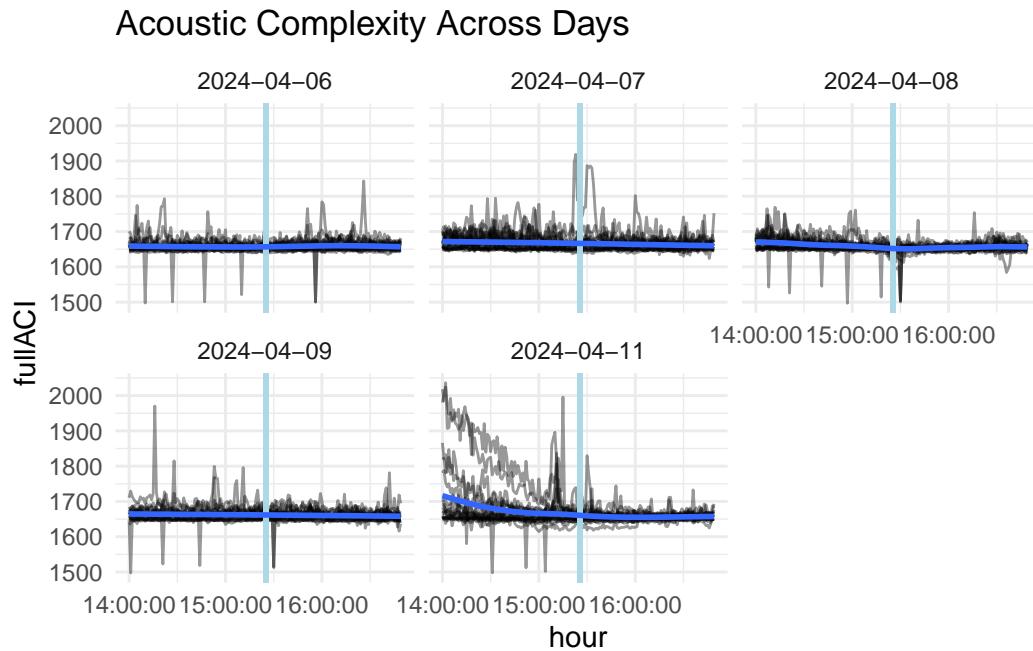
257 **5 Full data visualization**

258 Due to concerns about changes in weather, migration around this time of year, and readability,
259 we utilized a subset of the recording days. To assess the weather conditions across the days, we
260 used the website: <https://www.wunderground.com/history/daily/us/ny/ogdensburg>. From
261 looking here across our recording days, we decided to use the 2 days buffering the eclipse.
262 However, due to wind gusts around midday on the 10th, we opted for the 11th. This created
263 a subset of days which had similar weather patterns to the day of the eclipse: partly cloudy
264 with no rain and little wind. These days were close to April 8th, so we had less concern about
265 avian or amphibian migration that could alter the soundscape. Our final subset included the
266 following days: April 6th, 7th, 8th, 9th, and 11th.

267 We approached the full visualization of our data by looking across our five day subset to search

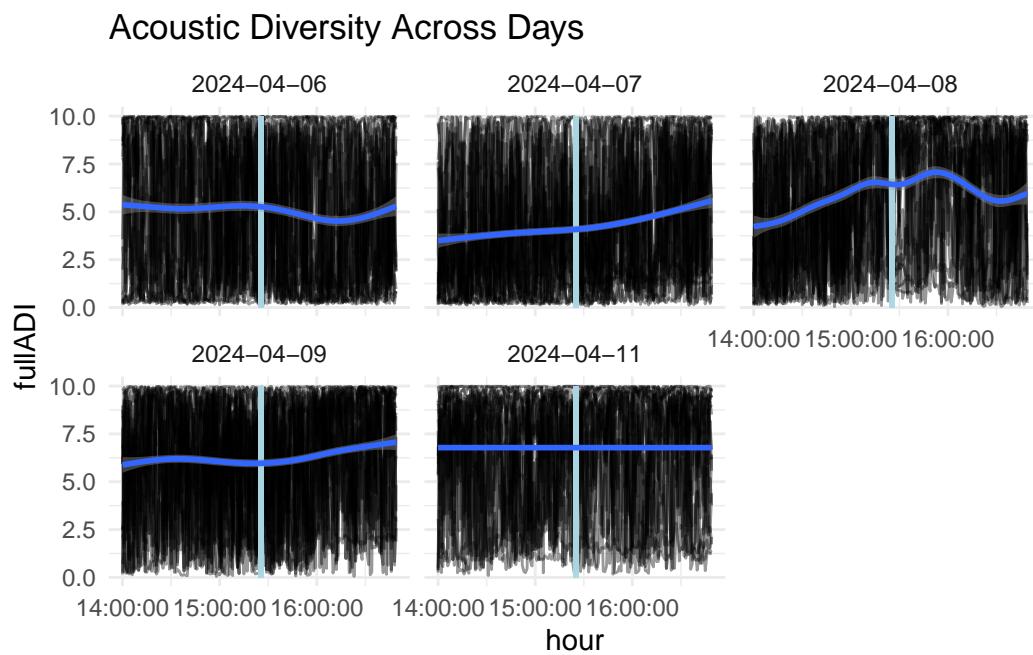
268 for patterns during the eclipse that differed between the 8th and the rest of our chosen days.
269 In our visualizations, we used a `geom_smooth()` layer to display a smoothed line so we could
270 see overall patterns more clearly.

271 **5.1 Acoustic Complexity**



272
273 In this graph, we note that all the days including the eclipse have a pretty linear smoothing
274 curve. This signals that the eclipse on April 8th likely did not have any influence on Acoustic
275 Complexity.

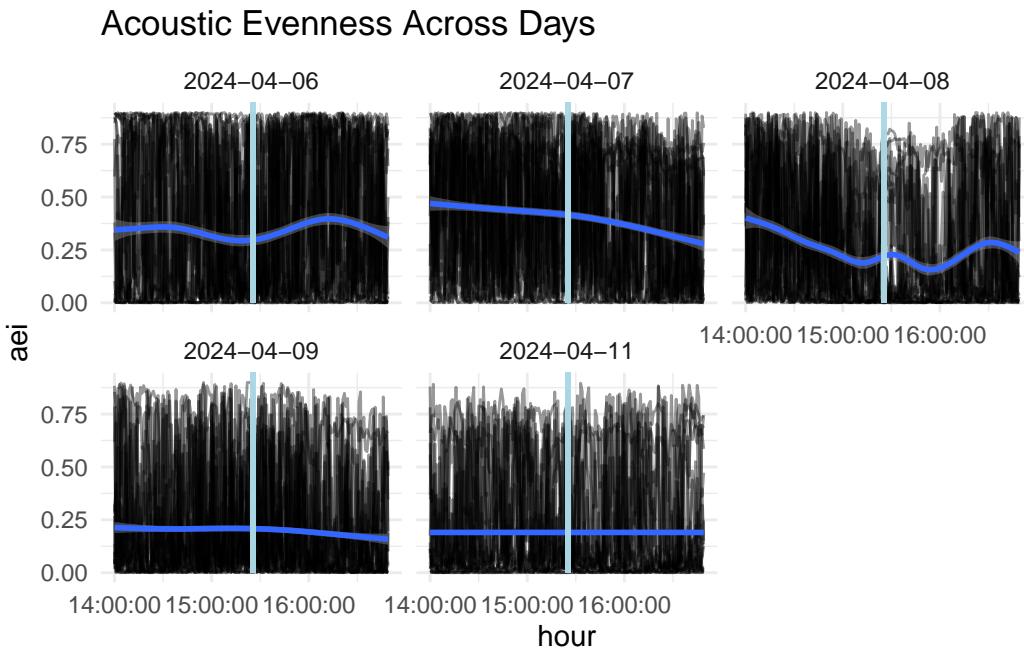
²⁷⁶ **5.2 Acoustic Diversity**



²⁷⁷

²⁷⁸ When we look at this visual, we can see basically all of the days from our subset are flat.
²⁷⁹ However, on April 8th, there seems to be two peaks close to the time of totality. There is
²⁸⁰ also some non-linearity around the time of totality from the 6th and the 9th. This makes it
²⁸¹ difficult to make a claim about whether or not this index is showing a potential pattern at this
²⁸² time.

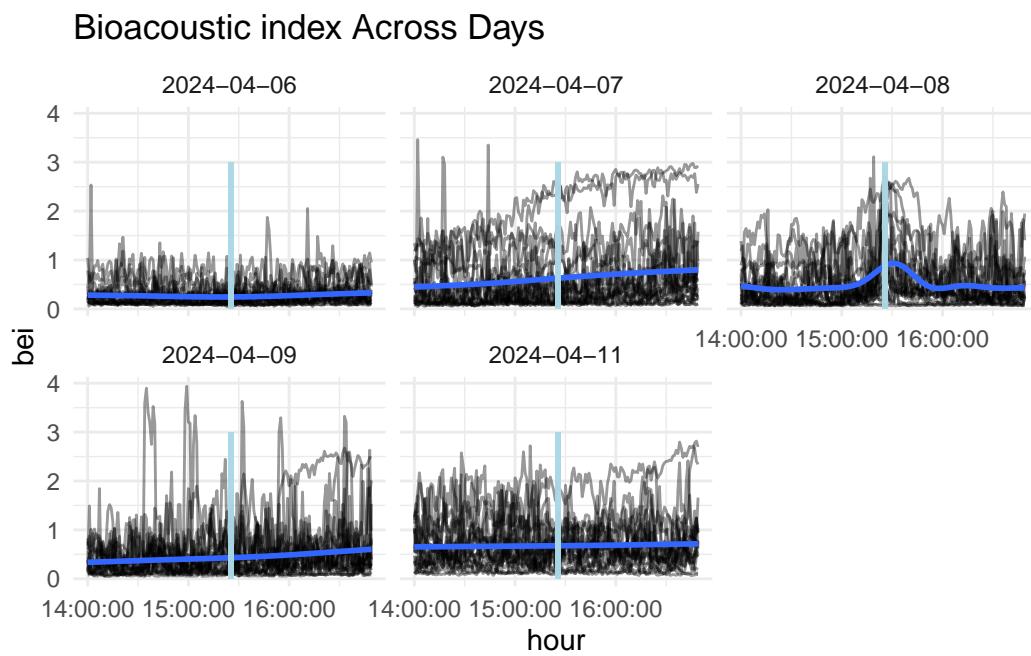
²⁸³ **5.3 Acoustic Evenness**



²⁸⁴

²⁸⁵ Again, we can see that most days have flat smoothing curves, with fluctuations only on the
²⁸⁶ 8th and 6th. We can see that on the 8th, there is a much steeper rise around that time of
²⁸⁷ totality, with minimums on either side of its beginning and end. However, since we see some
²⁸⁸ curvature around the time of totality on the 6th, it is hard for us to conclude from this visual
²⁸⁹ whether the eclipse did influence Acoustic evenness or not.

²⁹⁰ **5.4 Bioacoustic Index**

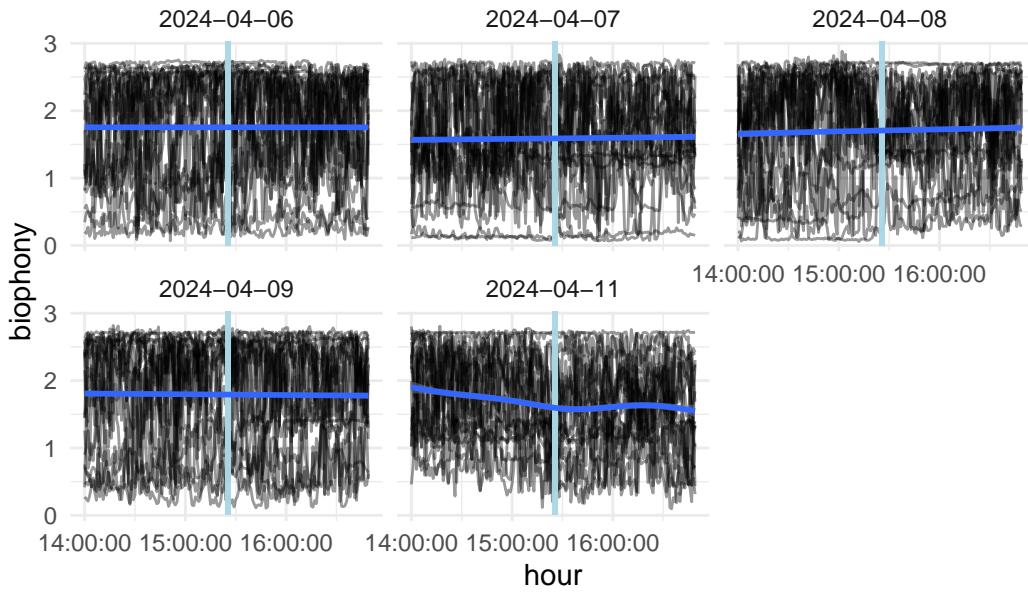


²⁹¹

²⁹² In this index, we can clearly see a potential pattern, where the bioacoustic index increases
²⁹³ around the time of totality. This pattern is very striking, as all the other days are quite linear
²⁹⁴ and this increase on the 8th is really the only curvature across our subset. This increase is
²⁹⁵ also very localized, the peaks rise and fall are quite close to its maximum. This signals that
²⁹⁶ the bioacoustic index may be affected by the eclipse and totality.

²⁹⁷ **5.5 Biophony**

Biophony Across Days



²⁹⁸

²⁹⁹ For biophony, when we look at the smoothing curves across the days we can see that basically
³⁰⁰ all of the days besides the 11th are quite linear. Since April 8th is so similar to the other days,
³⁰¹ this suggests that the eclipse may not be influencing this index.

³⁰² **6 Modeling**

³⁰³ To model this data, we decided to use Generalized additive models so we could incorporate
³⁰⁴ non-linear patterns that we were seeing while still maintaining additivity between the variables.
³⁰⁵ A GAM model resembles linear regression, but replaces the linear components with smooth
³⁰⁶ non-linear functions (James et al. (2013)):

$$y_i = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + f_3(x_{i3}) + \dots + f_p(x_{ip}) + \epsilon_i$$

307 where y_i is the response, β_0 is an intercept, $f_j(x_{ij})$ is a smoothing function for predictor j

308 and ϵ_i is assumed to be $N(0, \sigma^2)$ and independent of all other ϵ_i

309 Utilizing a smoothing spline can help us capture non-linear patterns which are present in the
 310 model. A smoothing spline specifies a function $g(x)$ that minimizes the formula below (James
 311 et al. (2013)). The function $g(x)$ is known to be a differentiable cubic polynomial function at
 312 knots for each training observation x_i (James et al. (2013)).

$$\sum_{i=1}^n (y_i - g(x_i))^2 + \lambda \int g'' t^2 dt$$

313 This formula is similar to the idea of minimizing the residual sum of squares which can be
 314 seen in the first portion of the formula, but it also specifies a shrinking parameter λ , which
 315 will penalize the variability of our function $g(x)$ and will control the “wiggleness” of the curve
 316 (James et al. (2013)). In other words, the second portion of the above formula, which is tuned
 317 by λ , influences the effective degrees of freedom which will appear in our model summaries.
 318 This can then be used as an assessment of the flexibility in the curves. From the `s()` function
 319 of the mgcv package that is used to create a smoothing spline, this λ is optimized using
 320 generalized cross-validation (Wood (2011)).

321 Our final generalized additive models were made using a smoothing spline for hour sepa-

322 rated by day, an additional term for audiomoth which would help us account for the different
 323 recorders and their respective locations, and lastly a term for day. To create these models
 324 we used the `gam()` function from the `mgcv` package, and to create our visualization, we used
 325 the `data_grid()` function (Wood (2011)) as well as the `augment()` function from the `broom`
 326 package. This allowed us to create predicted values based on our model and to then visualize
 327 them alongside our GAMs. To produce these visuals, we also needed to specify a specific
 328 recorder which lied around the “center” of our observed index values. The fitted curve for a
 329 different recorder would follow the same pattern, but would be shifted up or down. A table of
 330 the chosen audio recorders and the indices that they were used for is below:

| Index | Audio Recorders |
|----------|-----------------|
| ACI | A017_SD024 |
| ADI | A016_SD022 |
| AEI | A16_SD022 |
| BEI | A014_SD021 |
| Biophony | A001_SD001 |

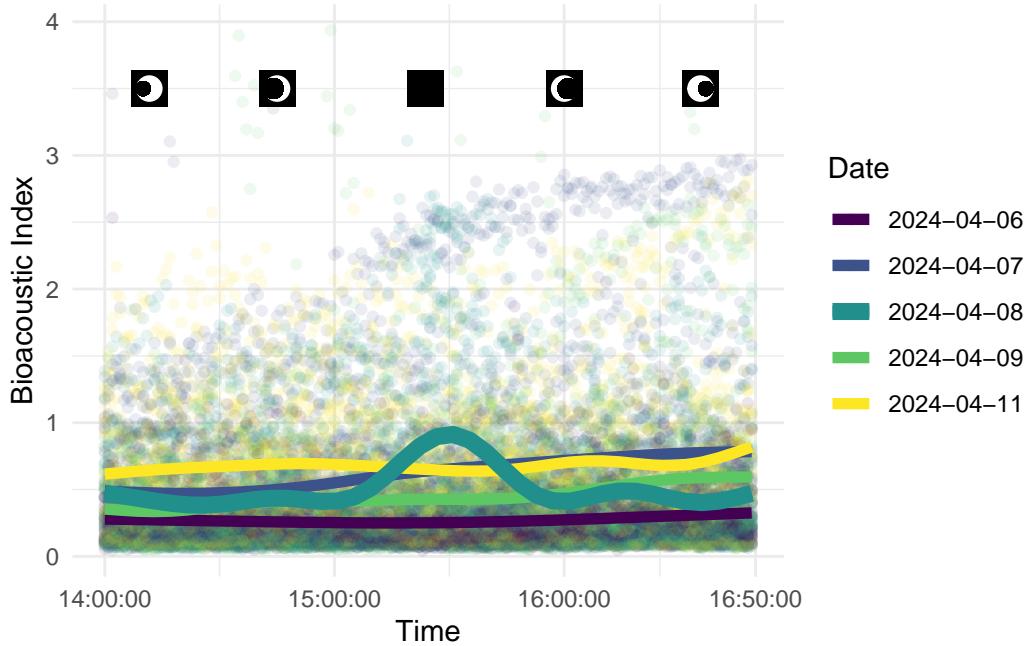
331 6.1 Bioacoustic Index

```

 332 # A tibble: 6 x 5
 333   term          edf ref.df statistic p.value
 334   <chr>        <dbl> <dbl>     <dbl>    <dbl>
 335 1 s(hour_numeric):day_factor2024-04-06  2.25   2.80      4.14  0.00873
 336 2 s(hour_numeric):day_factor2024-04-07  4.55   5.59     51.8   0
 337 3 s(hour_numeric):day_factor2024-04-08  8.86   8.99     57.5   0
 338 4 s(hour_numeric):day_factor2024-04-09  6.52   7.66     17.9   0
 339 5 s(hour_numeric):day_factor2024-04-11  7.46   8.42      3.59  0.000415
  
```

340 6 s(folder_name) 19.0 19 527. 0

341 To analyze our model outputs, we focused on the estimated degrees of freedom, as this provides
342 us with an estimate of the flexibility in the lines. We were able to compare this numerical
343 flexibility on the 8th with the other days we have included in our subset, as well as compare
344 the trend lines we see in our model visualizations. From our model for Bioacoustic index, we
345 can see that April 8th possesses the most complex line, with an edf of 8.86. The other days
346 vary in their complexity, but we can see that they all have non-linear effects and April 7th is
347 the second most complex to the day of the eclipse (edf: 7.46).



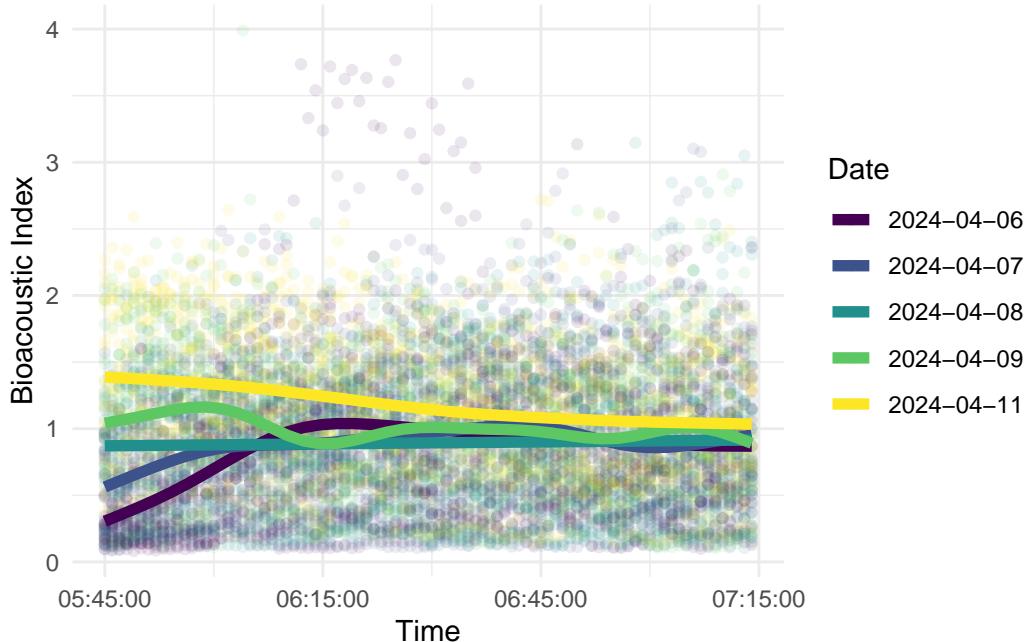
348

349 Note the icons that are in our model visuals. These align specifically with the changes during
350 the eclipse. The first and last icon correspond to the beginning (14:11:38) and end (16:35:38)
351 of the partial eclipse, and the center corresponds to peak totality (15:25:29).

352 In this visualization, we can really see the flexibility on April 8th compared to the other days.
353 Although in the model summary, the other days had some high edf values, their trend lines
354 appear more linear in our graphic. On April 8th, we can see an abrupt peak that aligns with
355 that peak at totality. This peak is pretty localized as well, with the rest of the trend line
356 showing minimal curvature. These aspects suggest that this index is displaying some type of
357 reaction wildlife is having to the time of totality.

358 To understand if the only oddity across the 8th was at the time of the eclipse, we created a
359 similar model visualization that corresponds to the dawn.

```
360 # A tibble: 6 x 5
361   term                  edf ref.df statistic p.value
362   <chr>                <dbl> <dbl>    <dbl>    <dbl>
363 1 s(hour_numeric):day_factor2024-04-06  6.40    7.55    66.4     0
364 2 s(hour_numeric):day_factor2024-04-07  6.89    7.98    18.4     0
365 3 s(hour_numeric):day_factor2024-04-08  1.00    1.00     2.24   0.135
366 4 s(hour_numeric):day_factor2024-04-09  8.31    8.87     8.98     0
367 5 s(hour_numeric):day_factor2024-04-11  3.35    4.16    48.2     0
368 6 s(folder_name)           18.9     19      516.     0
```



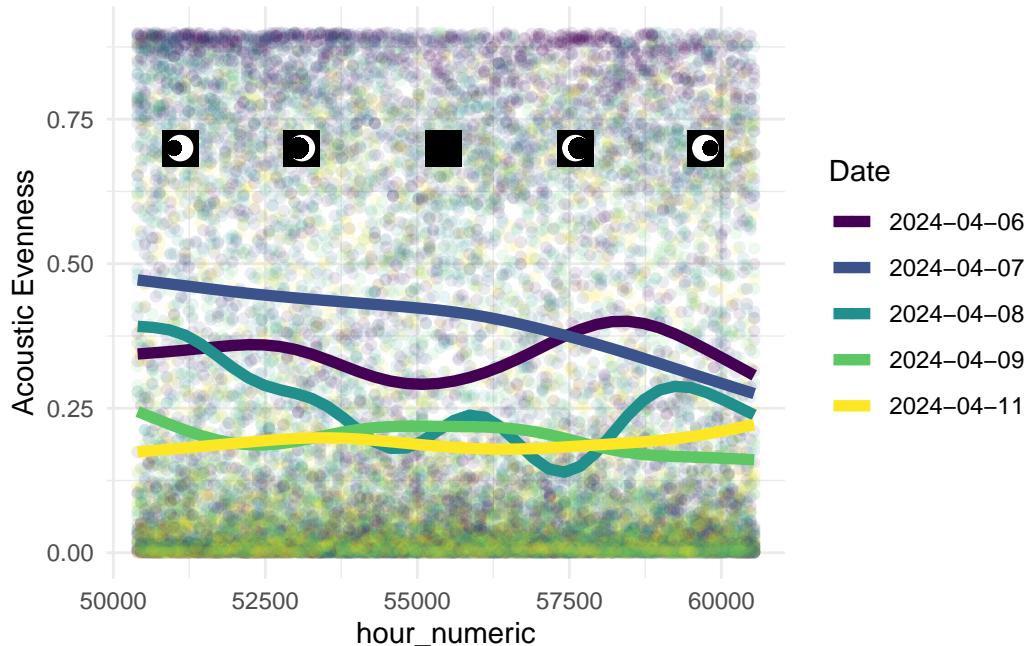
369

370 Looking at the model summary and the visual, we can see that there isn't anything distin-
 371 guishable about April 8th. Based on its low edf value and the trend line, we can see that this
 372 day was pretty linear, and it didn't stick out compared to the other days. This helps provide
 373 evidence that the eclipse is causing a change in this index, since the day itself prior to the
 374 eclipse does not show any overall differences that could cause the irregular peak we saw.

375 6.2 Acoustic Evenness

```
376 # A tibble: 6 x 5
377   term                  edf ref.df statistic p.value
378   <chr>                 <dbl> <dbl>    <dbl>    <dbl>
379 1 s(hour_numeric):day_factor2024-04-06  5.46   6.60     8.59  0
380 2 s(hour_numeric):day_factor2024-04-07  2.94   3.66    41.0   0
381 3 s(hour_numeric):day_factor2024-04-08  8.68   8.97    24.9   0
382 4 s(hour_numeric):day_factor2024-04-09  5.37   6.50     3.68  0.000814
383 5 s(hour_numeric):day_factor2024-04-11  3.79   4.69     1.21  0.299
384 6 s(folder_name)           19.0    19       278.   0
```

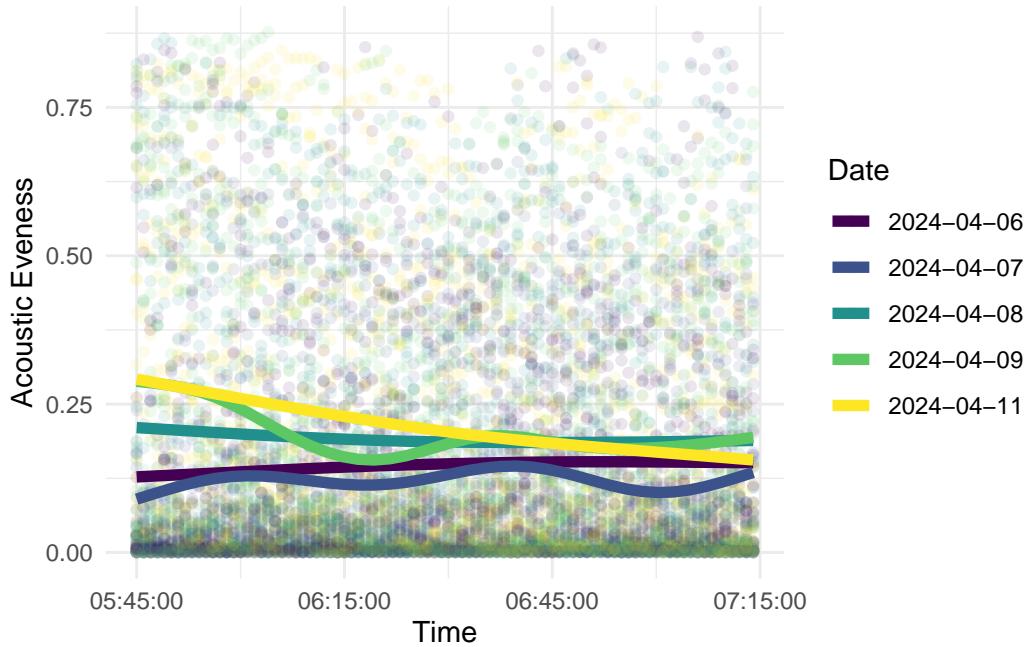
385 Similar to the bioacoustic index model, here we can see that April 8th is again the most flexible
 386 (edf: 8.68). In this model, the 8th is actually much more flexible than the other days, with
 387 the second highest being the 6th (edf: 5.46). We also see a close third on April 9th, with an
 388 edf of 5.37. As we recall from our full visualization for AEI, this makes sense as we ended up
 389 seeing a lot of curvature in the 6th, 8th, and 9th.



390
 391 Looking at this visual, we can see that that the lines from the 6th, 8th, and 9th are those
 392 with the most curvature. Looking at the 6th and 9th, we see some fluctuations during the
 393 eclipse, but these are not as steep as the patterns from the 8th. On the 8th, we can see there
 394 are two local minimums which fall before and after totality, and we see a local maximum
 395 appearing likely as totality was just ending. Aside from totality this line is pretty flexible,
 396 with lots of curvature across this full time interval. This provides some reason to suggest that

397 the acoustic evenness index is affected by the eclipse, but due to the non-linearity in other
 398 days, the evidence isn't as strong as we saw in the bioacoustic index.

```
399 # A tibble: 6 x 5
400   term                  edf ref.df statistic p.value
401   <chr>                <dbl> <dbl>    <dbl>    <dbl>
402  1 s(hour_numeric):day_factor2024-04-06  1.62    2.01     1.93  0.143
403  2 s(hour_numeric):day_factor2024-04-07  5.59    6.73     1.89  0.0694
404  3 s(hour_numeric):day_factor2024-04-08  1.74    2.17     1.73  0.187
405  4 s(hour_numeric):day_factor2024-04-09  5.89    7.05    11.0   0
406  5 s(hour_numeric):day_factor2024-04-11  1.97    2.46    34.9   0
407  6 s(folder_name)           18.9    19      138.    0
```



408
 409 After creating an acoustic evenness model for the dawn, we can see again that April 8th is
 410 pretty linear (edf: 1.74). In our visual, we see the 8th doesn't stick out compared to our
 411 other days, again suggesting that there is not anything off about this index or the day itself.
 412 Although this helps show that the 8th was a "normal" day, it doesn't provide a lot more
 413 evidence to our claim that the eclipse affected this index.

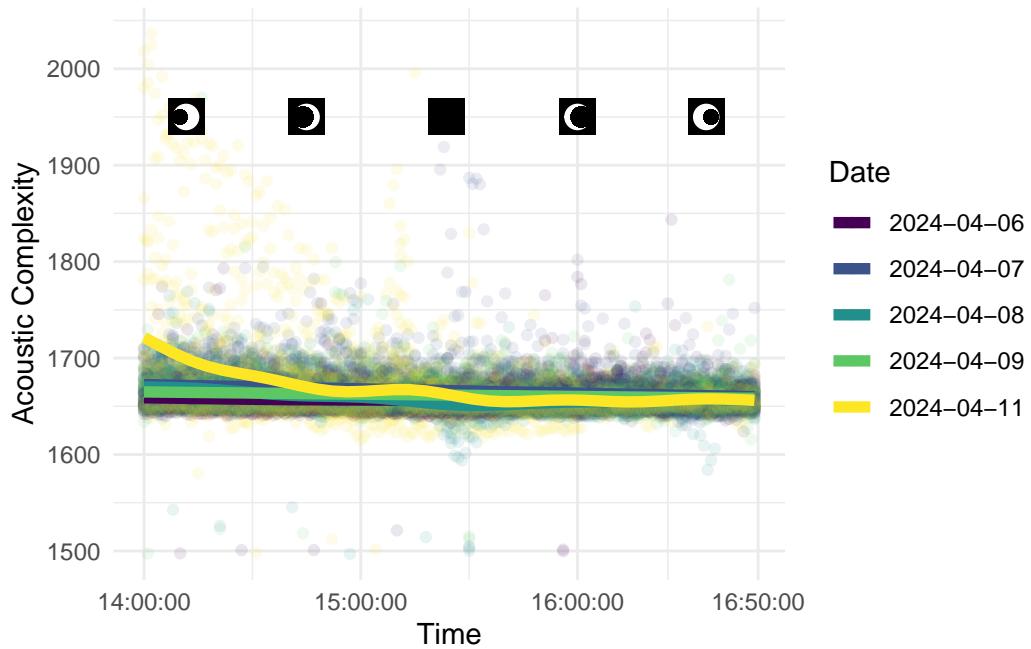
414 **6.3 Acoustic Complexity**

```

415 # A tibble: 6 x 5
416   term                  edf ref.df statistic p.value
417   <chr>                 <dbl> <dbl>    <dbl>    <dbl>
418 1 s(hour_numeric):day_factor2024-04-06  3.82    4.72     2.29  0.0399
419 2 s(hour_numeric):day_factor2024-04-07  1.00    1.00    102.   0
420 3 s(hour_numeric):day_factor2024-04-08  7.50    8.44     28.4  0
421 4 s(hour_numeric):day_factor2024-04-09  1.00    1.00     28.6  0
422 5 s(hour_numeric):day_factor2024-04-11  8.75    8.98    226.   0
423 6 s(folder_name)          18.9     19      316.   0

```

424 Like the other days, we can see that the edf for April 8th is pretty high, compared to the other
 425 days (edf: 7.497). This is not the most complex day, with April 11th having an edf value of
 426 8.748. This possibly could suggest that although the data from April 8th is showing a lot of
 427 non-linearity, it may be due to factors other than the eclipse.



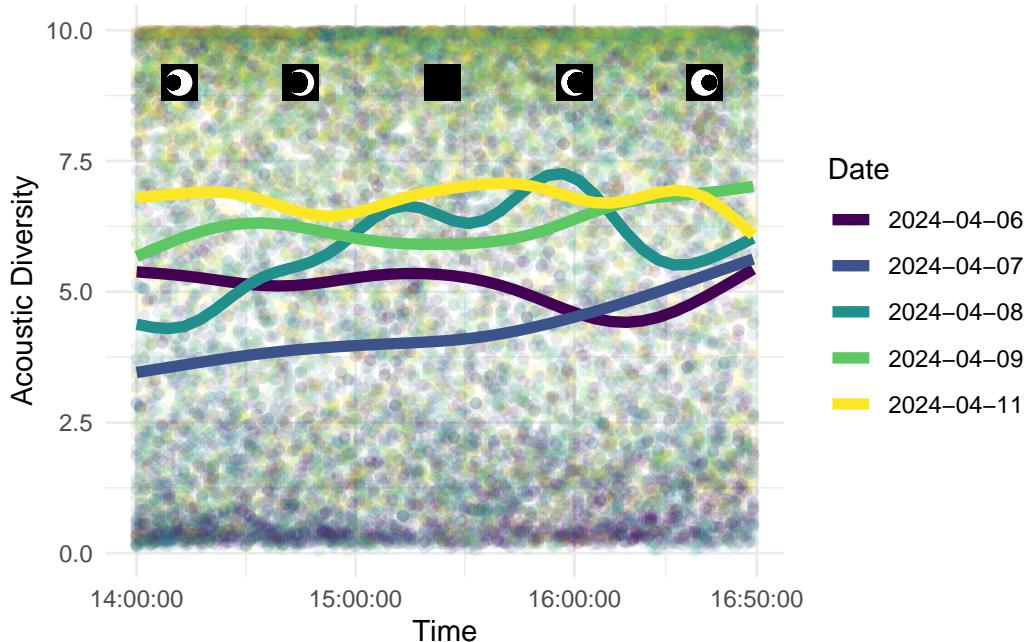
428
 429 Although the estimated degrees of freedom values suggest a lot of non-linearity, the line for

430 April 8th in this visual is pretty flat. We can see that there isn't a lot of differences between
431 the lines, besides April 11th. We conclude that this metric is likely not having a reaction to
432 the eclipse.

433 **6.4 Acoustic Diversity**

```
434 # A tibble: 6 x 5
435   term                  edf ref.df statistic    p.value
436   <chr>                <dbl> <dbl>     <dbl>      <dbl>
437 1 s(hour_numeric):day_factor2024-04-06  5.40   6.54     6.35  0.00000217
438 2 s(hour_numeric):day_factor2024-04-07  3.20   3.98     32.8   0
439 3 s(hour_numeric):day_factor2024-04-08  8.45   8.92     30.9   0
440 4 s(hour_numeric):day_factor2024-04-09  5.29   6.41     8.36   0
441 5 s(hour_numeric):day_factor2024-04-11  7.37   8.35     2.24  0.0294
442 6 s(folder_name)          18.9    19       299.    0
```

443 Looking at the acoustic diversity, the estimated degrees of freedom are the highest yet again
444 on the day of the eclipse (8.454). For this metric, the 2nd highest is from April 11th at 7.369.
445 From our model summary, we also note that all of the other days are also showing some type
446 of non-linearity through their edf values.



447

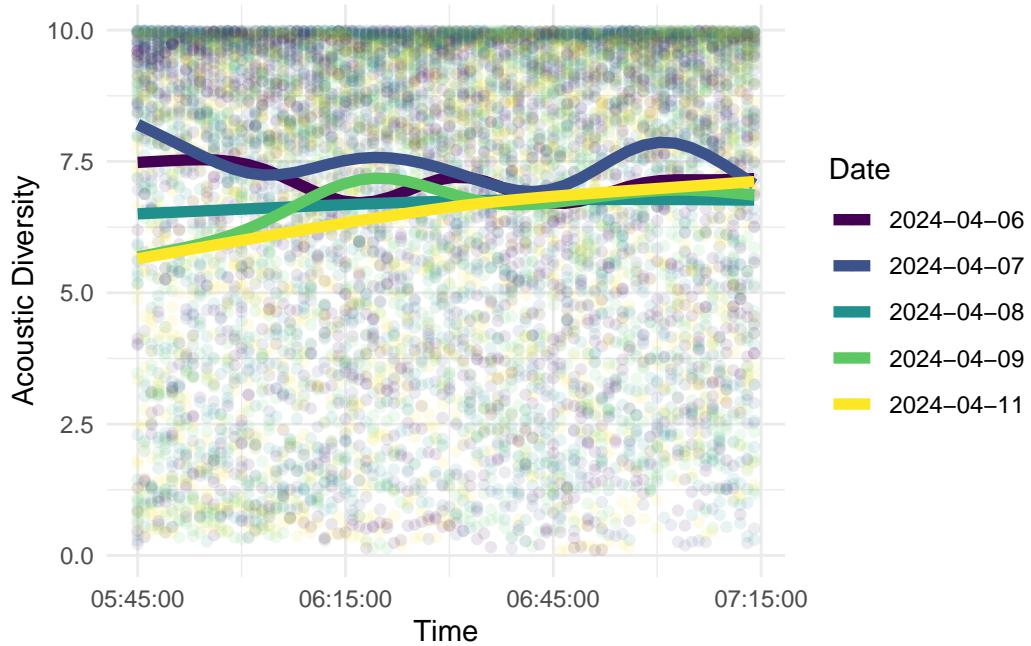
448 As we noted, we can see that all of the days are showing some type of flexibility. When we
 449 look specifically at the 11th, we can see it is quite wavy, and we see some dips around this
 450 period of the eclipse. When we look at the trend from the 8th, we note that there are two
 451 maximums surrounding the time of totality. In between these peaks is a minimum around the
 452 time that totality was ending. Compared to the trends we see from the 11th, these patterns
 453 align more with the changes occurring because of the eclipse. This suggests to us that the
 454 acoustic diversity is being affected by the eclipse, yet due to the curvature in the other lines,
 455 we do not have the strongest evidence.

```
456 # A tibble: 6 x 5
457   term                  edf ref.df statistic p.value
458   <chr>                 <dbl> <dbl>    <dbl>    <dbl>
459 1 s(hour_numeric):day_factor2024-04-06  6.80    7.91     2.94  0.00329
460 2 s(hour_numeric):day_factor2024-04-07  6.48    7.62     4.04  0.000129
461 3 s(hour_numeric):day_factor2024-04-08  1.51    1.86     1.07  0.256
462 4 s(hour_numeric):day_factor2024-04-09  5.47    6.61     8.07  0
```

```

463 5 s(hour_numeric):day_factor2024-04-11 1.93 2.41 23.7 0
464 6 s(folder_name) 18.9 19 175. 0

```



```
465
```

466 To try and get some more evidence, we looked at a model from the dawn. Like the other
 467 metrics we calculated dawn models for, April 8th displays a pretty linear trend line and a low
 468 edf value (1.514). Since the 8th does not vary too much from the other days, April 8th seems
 469 to be a “regular” day for acoustic diversity measurement.

470 6.5 Biophony

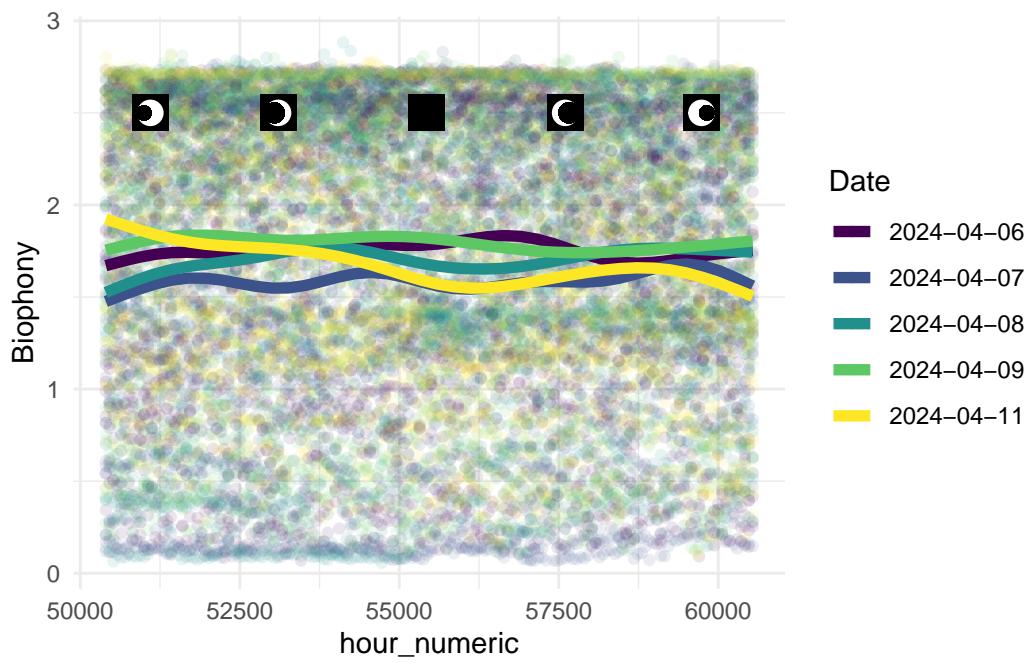
```

471 # A tibble: 6 x 5
472   term                  edf ref.df statistic p.value
473   <chr>                 <dbl> <dbl>     <dbl>     <dbl>
474 1 s(hour_numeric):day_factor2024-04-06 7.38  8.36    3.60 0.000291
475 2 s(hour_numeric):day_factor2024-04-07 8.11  8.79    3.32 0.00159
476 3 s(hour_numeric):day_factor2024-04-08 6.56  7.69    6.28 0
477 4 s(hour_numeric):day_factor2024-04-09 5.88  7.03    2.43 0.0167
478 5 s(hour_numeric):day_factor2024-04-11 6.29  7.45    21.9 0

```

```
479 6 s(folder_name)          19.0   20      1377.   0
```

480 Lastly, in our biophony model, all of the days have higher non-linearity with edf values above
481 five. Interestingly, in this summary, the 8th is not one of the most complex lines (edf: 6.558).
482 This began to suggest that we were not going to see a pattern that would align with the
483 eclipse on this day, and if we did, we could potentially also see similar non-linearity in the
484 other days.



485

486 This visualization shows very clearly that all of the days during this time frame are very
487 similar, and we don't see any particular patterns from April 8th or any other day. Like we
488 expected, all are displaying a lot of curvature, and we are not seeing any patterns that really
489 align with totality, or the beginning and end of the partial eclipse. This provides no evidence
490 that biophony was impacted during the time of eclipse.

⁴⁹¹ **7 Conclusion**

⁴⁹² From our analysis, we are at a place to believe that the bioacoustic index, acoustic evenness
⁴⁹³ index, and acoustic diversity index may display patterns across the time of the eclipse. Looking
⁴⁹⁴ at model summaries and visualizations, we say that the bioacoustic index has the most evidence
⁴⁹⁵ of a potential pattern with a pretty pronounced maximum around peak totality. The acoustic
⁴⁹⁶ diversity and acoustic evenness indices show a bit less evidence of patterns, but the curvature
⁴⁹⁷ around totality provide enough evidence for us to claim there is the potential these are affected
⁴⁹⁸ as well. We believe that these potential patterns display that wildlife is impacted by some
⁴⁹⁹ type of eclipse-driven change, and based on the patterns, we can suggest that they are affected
⁵⁰⁰ by specifically a totality-driven change. Whether this is due to something we cannot perceive
⁵⁰¹ or it is the changes in light, we cannot be sure at this time.

⁵⁰² We do want to note that we made the assumption that the one minute audio clips were not
⁵⁰³ correlated with each other. This assumption is likely to be incorrect, and we believe that some
⁵⁰⁴ of these clips would have autocorrelation between them. Looking to future work with this
⁵⁰⁵ data, it would be interesting to account for this within the models and to see if and how the
⁵⁰⁶ patterns would change.

⁵⁰⁷ **References**

⁵⁰⁸ Gerber, Jacob E, Dakota Howard, and John E Quinn. 2020. “Soundscape Shifts During
⁵⁰⁹ the 2017 Total Solar Eclipse: An Application of Dispersed Automated Recording Units to

- 510 Study Ephemeral Acoustic Events.” *Biodiversity* 21 (1): 41–47.
- 511 James, Gareth, Daniela Witten, Trevor Hastie, Robert Tibshirani, et al. 2013. *An Introduction*
512 *to Statistical Learning*. Vol. 112. 1. Springer.
- 513 Villanueva-Rivera, Luis J., and Bryan C. Pijanowski. 2018. *Soundecology: Soundscape Ecology*.
- 514 <https://CRAN.R-project.org/package=soundecology>.
- 515 Wood, S. N. 2011. “Fast Stable Restricted Maximum Likelihood and Marginal Likelihood
516 Estimation of Semiparametric Generalized Linear Models.” *Journal of the Royal Statistical*
517 *Society (B)* 73 (1): 3–36. <https://doi.org/10.1111/j.1467-9868.2010.00749.x>.