

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	3
7 2 Data Collection Methods	4
8 3 Indices and Packages	7
9 3.1 Acoustic Complexity Index:	7
10 3.2 Acoustic Diverstiy Index:	8
11 3.3 Acoustic Evenness Index:	8
12 3.4 Bioacoustic Index:	9
13 3.5 Biophony:	10
14 4 Initial visualization	11
15 4.1 Creating a cleaned data frame	13
16 4.2 Using the HPC	17
17 4.3 Analyzing one full file from an audio recorder	18
18 5 Full data visualization	26
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	45
³⁰	7 Conclusion (ADD MORE AND CONNECT POINTS)	47
³¹	References	48

³² 1 Abstract

³³ From March 30 to April 16th, Erika Barthelmess and students from the St. Lawrence biology
³⁴ department collected data from 20 audio recording devices in the Northern New York area.
³⁵ These devices were strategically located to observe potential audio changes in wildlife at the
³⁶ time of the total solar eclipse on April 8th, 2024. We processed this audio data to produce
³⁷ organized data frames, centered around 5 commonly used audiological indices. From this
³⁸ tabular data, we constructed many visualizations displaying patterns across time, day, and
³⁹ recording device. Due to the non-linearity present in our plots, generalized additive models
⁴⁰ based on date, time of day, and recording device were implemented to assess patterns that
⁴¹ may have occurred during the time of the eclipse. From these visuals and models, we conclude
⁴² the 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. All locations were fully
⁵⁰ 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
⁵² ended at 15:27:05 and the partial eclipse ended at 16:35:38 (all times local, times from
⁵³ <https://www.timeanddate.com/eclipse/in/@5111484?iso=20240408>) for a total eclipse dura-
⁵⁴ tion of 2 hours 24 minutes, with totality lasting 3 minutes and 13 seconds.

⁵⁵ Each AudioMoth was configured to record within four temporal windows on each day of the
⁵⁶ deployment. The first window was from 05:45 – 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

Times	Reasoning
14:00 - 16:00	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)

Parameters	Setting
Sleep Duration (seconds)	5
Recording Duration (seconds)	55

69 We used ArcGIS Pro (version XXXX, ESRI Incorporated, Redlands, California) to identify
 70 areas of forest-wetland interface or forested wetland occurring on public or University-owned
 71 land. Our intention was to place the recorders at locations where they could capture the
 72 sounds of both forest birds and pond-breeding amphibians (as well as other biotic sounds
 73 including any active insects or bats). Due to our northern location within the path of the
 74 eclipse in North America, the onset of spring was just beginning. Red-winged blackbirds
 75 (*Agelaius phoeniceus*) had returned to the area and were establishing breeding territories and
 76 both wood frogs (*Lithobates sylvaticus*) and spring peepers (*Pseudacris crucifer*) had begun
 77 to chorus at least 5 days prior to deployment of the recorders. To reduce the time required to
 78 deploy units, we located the devices near but out of view of hiking trails and within 20 miles
 79 of the St. Lawrence University campus (44.58931027483651° N, -75.1613716006626 ° W).
 80 In our analysis, we took a lot of inspiration from a paper published in 2020 titled “*Soundscape*
 81 *shifts during the 2017 total solar eclipse: An application of dispersed automated recording*
 82 *units to study ephemeral acoustic events*” by Jacob E. Gerber, Dakota Howard, and John E.
 83 Quinn.(Gerber, Howard, and Quinn (2020))

⁸⁴ **3 Indices and Packages**

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 the recording devices, all data was subsetting
⁸⁸ from the left channel as the right channel values were not available.

⁸⁹ **3.1 Acoustic Complexity Index:**

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

⁹³ A higher ACI value indicates a more diversity 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* pack-
⁹⁸ age. The specific numerical values used are from subsetting `$acl_left_vals` (Villanueva-Rivera
⁹⁹ and Pijanowski (2018)).

¹⁰⁰ **3.2 Acoustic Diverstiy Index:**

¹⁰¹ The Acoustic Diversity Index (ADI) assesses both the variety and evenness in the sound
¹⁰² distribution 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 often summed to produce
¹¹⁴ the 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 To interpret Acoustic Evenness, a high AEI value signifies a more even soundscape, within the
121 audio there isn't any sound domination from one type of sound/species. A low AEI signifies
122 that the sound is not as even, there may only be a few loud or similar sounds which disrupt
123 the consistency of the sound level.

124 In the paper by Gerber et. al in 2017, this was found to be a statistically significant predictor
125 in their model. There they found that the AEI was highest right before the eclipse began and it
126 rose until period of totality. (In the paper the authors believe this is likely due to the increase
127 in human sounds near their recording locations) (Gerber, Howard, and Quinn (2020))

128 This index is found using the acoustic_evenness() function in the soundecology package. The
129 specific numerical value used is from subsetting \$aei_left (Villanueva-Rivera and Pijanowski
130 (2018)).

131 **3.4 Bioacoustic Index:**

132 The Bioacoustic Index (BEI) provides a measurement of the diversity and abundance of biolog-
133 ical noise. This index is calculated by segmenting the data into bins between 2-8kHz, and then
134 assessing the variation in relation to the lowest frequency from that bin. The total Bioacoustic
135 Index is the average across all of these bins.

¹³⁶ Interpretation of this index can be understood as higher values signal an increase in species
¹³⁷ diversity, suggesting more species are making noise during the clip.

¹³⁸ Within the Gerber et. al paper (2017), this was found to be a non-significant predictor but, it
¹³⁹ seemed to be greatest at the period of totality (Gerber, Howard, and Quinn (2020)).

¹⁴⁰ This is obtained using the bioacoustic_index() function in the soundecology package. The
¹⁴¹ specific numerical value used is from subsetting \$left_area (Villanueva-Rivera and Pijanowski
¹⁴² (2018)). The default minimum hertz value of 2kHz was set for this index.

¹⁴³ **3.5 Biophony:**

¹⁴⁴ Biophony is an index which calculates the average frequency of biotic sound (between 2-8kHz).
¹⁴⁵ This index is much more straightforward, high biophony means that there is more sound and
¹⁴⁶ low values mean less.

¹⁴⁷ Within the Gerber et. al paper in 2017, the biophony was found to increase at the begin-
¹⁴⁸ ning and end of the eclipse, akin to the dawn/dusk choruses. In this paper, biophony was a
¹⁴⁹ significant predictor in their model (Gerber, Howard, and Quinn (2020)).

¹⁵⁰ This index is obtained as a byproduct from the Normalized Difference Soundscape Index
¹⁵¹ function and the numerical values used is from subsetting \$biophony_left (Villanueva-Rivera
¹⁵² and Pijanowski (2018)).

¹⁵³ **4 Initial visualization**

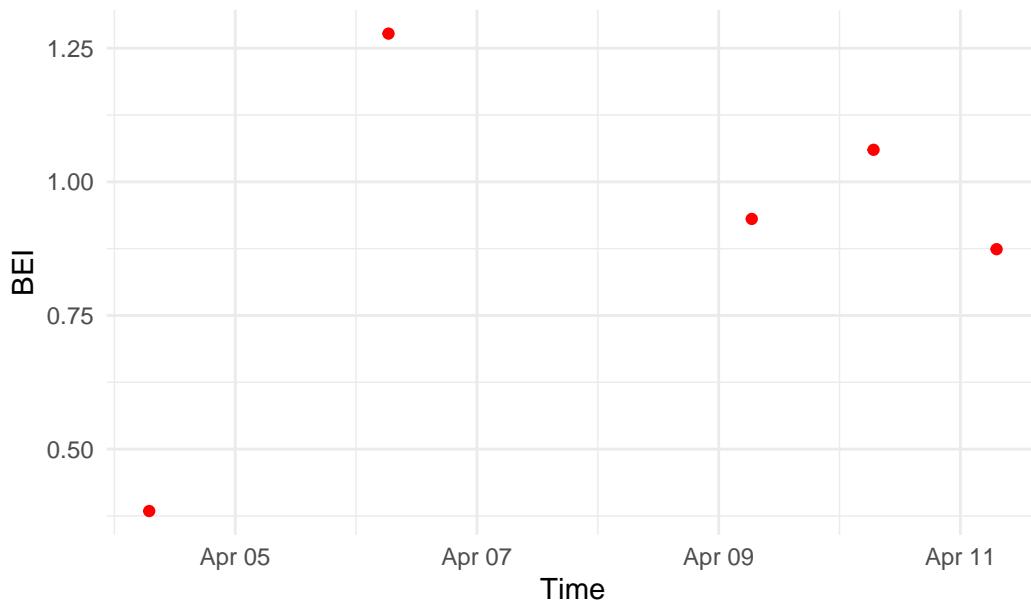
¹⁵⁴ Due to the large amount of recordings at our disposal, we began creating graphics using a
¹⁵⁵ subset of five audio files (.WAV) from a single audio recorder to explore the possibilities of the
¹⁵⁶ graphics and to search for the best ways to display the indices.

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

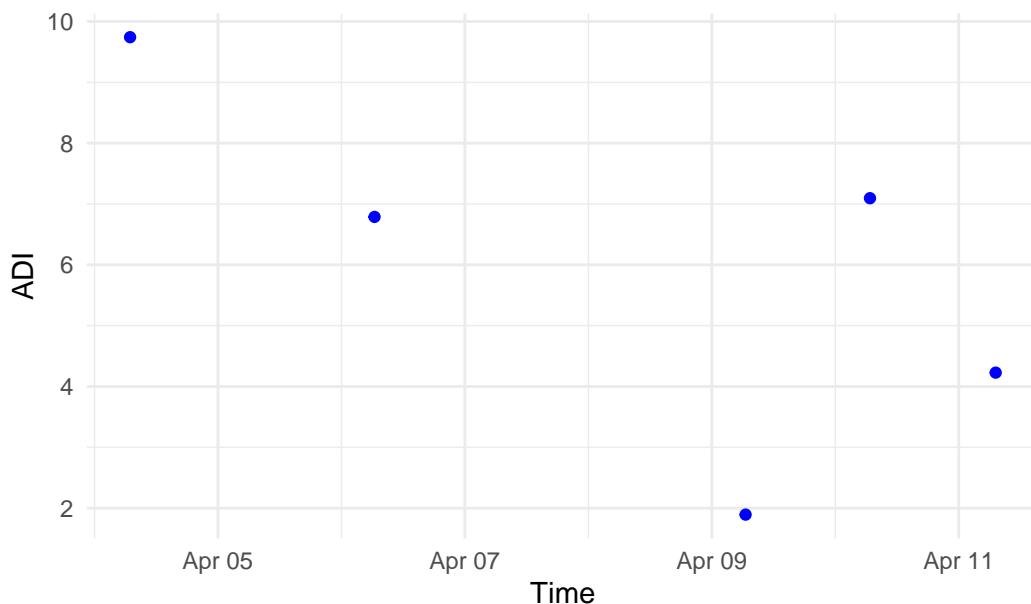
ACI_all	ADI_all
9.01389387925447,6.42427026282213	0.9896,0.980612
8.8657794310719,6.4421208288537	0.844655,0.716814
9.06604640985277,6.49691812601518	0.473127,0.217534
8.85770232363042,6.52467721488273	0.857982,0.745923
8.8272143441091,6.39791145784707	0.691836,0.472565

First Graphic of Bioacoustic Index



157

First Graphic of Acoustic Diversity



158

¹⁵⁹ To save space we have included only the Bioacoustic index and Acoustic Diversity example

¹⁶⁰ plots, but this was the initial visualization done for each of our indices. Although these initial

₁₆₁ visualizations did not help us assess patterns, the graphics helped us gain an understanding
₁₆₂ 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

₁₆₃ **4.1 Creating a cleaned data frame**

₁₆₄ To expand the collection of data, we generalized our code and created a function, “eclipse_df()”
₁₆₅ which would use a folder containing all .WAV files for one audio recorder as input and would
₁₆₆ return a cleaned data frame. This data frame would then be saved in our environment and
₁₆₇ would contain all the data that we would need from this point forward.

₁₆₈ Initially indexes are formed for all the WAV files and stored in their own data frames. For
₁₆₉ the Bioacoustic index, Acoustic evenness index, and biophony all values are subsetted from the
₁₇₀ original output of the soundecology function. These three indices are then bound into one
₁₇₁ tibble.

```

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

```

172

173 For the Acoustic Complexity and the Acoustic diversity to obtain the full list of values for each
 174 folder a for loop is incorporated into the function. This obtains all the left channel values and
 175 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))
}

```

176

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

```

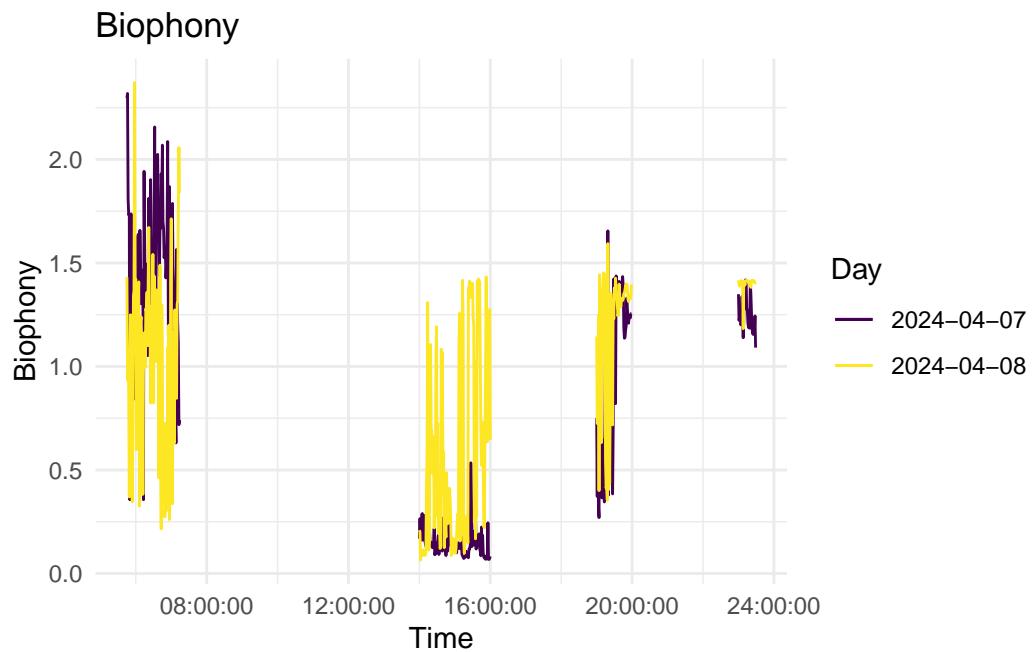
full|> mutate(biophony = as.numeric(biophony),
              aei = as.numeric(aei),
              bei = as.numeric(bei))|>
  separate(paths_date, into = c("date", "time_hms"), sep = "_")|>
  separate(time_hms, into = c("time", "wav"), sep = "\\.")|>
  separate(time, into = c("hours", "other"), sep = 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())

```

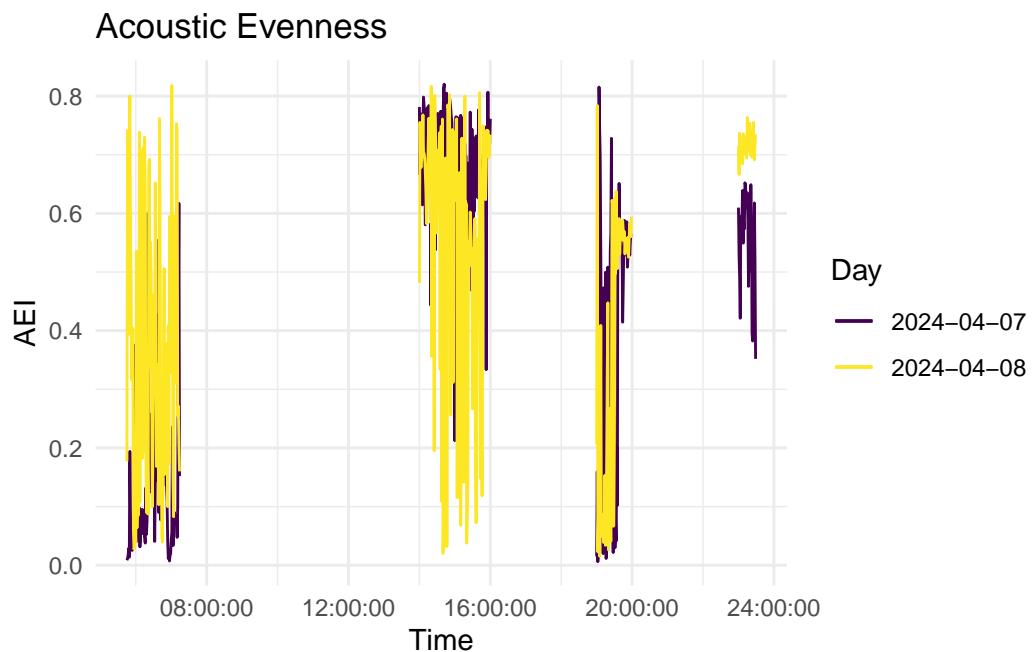
180

181 Initially we tested this function on two folders which are a part of the larger collection of data

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

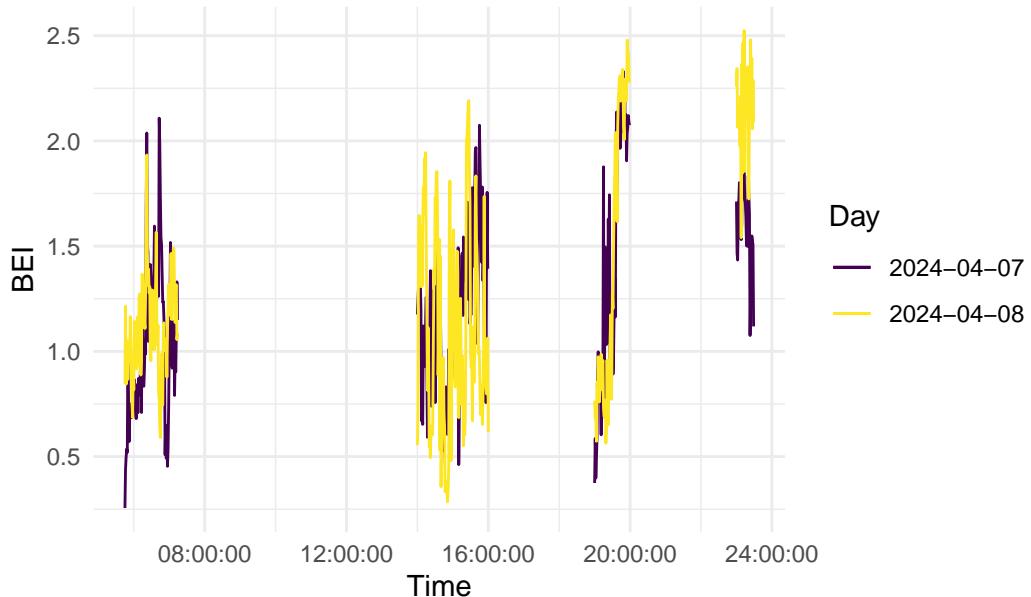


184



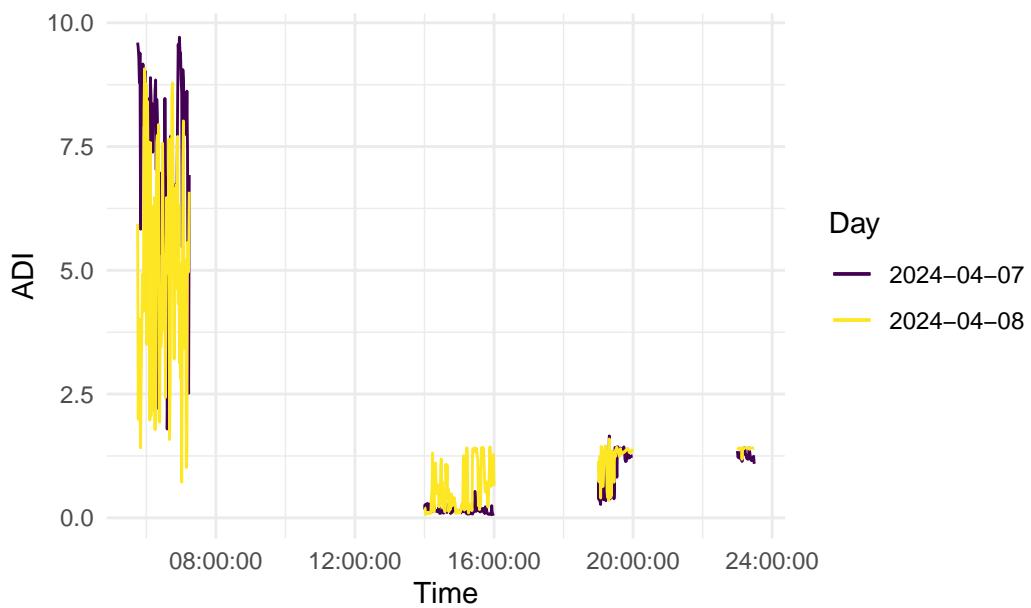
185

Bioacoustic Index

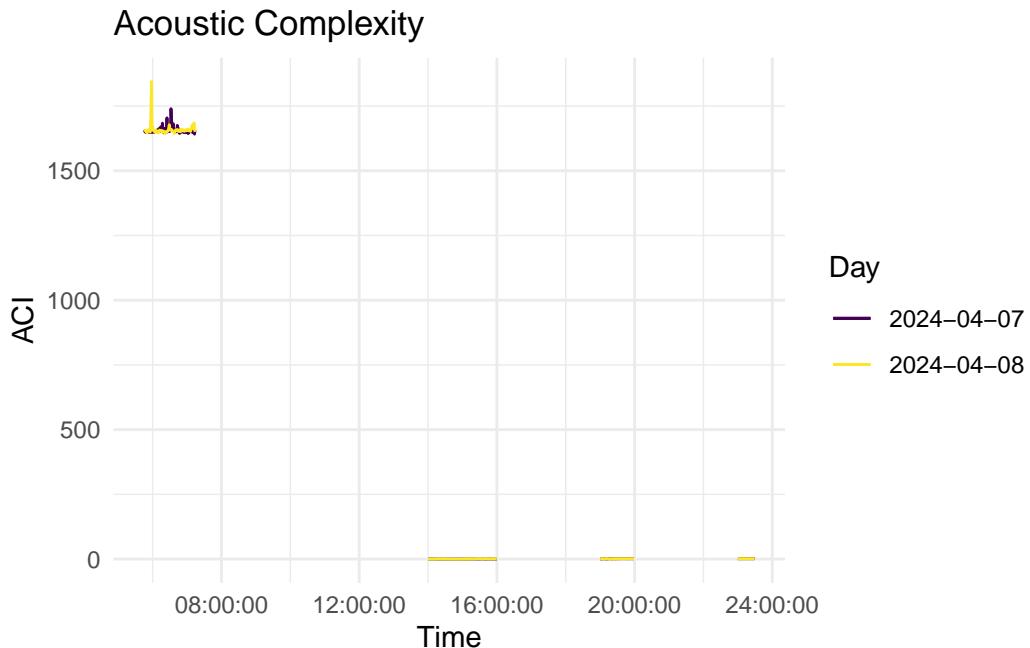


186

Acoustic Diversity Index



187



188

189 Visualizing these two data sets help us understand the patterns that occur across a full day
 190 of recording. For example, a pattern we can see is that ADI, ACI, and biophony are higher
 191 during the dawn segment and then decrease to similar low levels across the rest of the day.
 192 Another thing these types of graphs can shed light on are the potential differences between a
 193 normal day and the day of the eclipse. However, as this is just a comparison from two folders,
 194 we can't yet make any conclusions about differences being due to the eclipse.

195 **4.2 Using the HPC**

196 To assist in efficiently expanding to the full set of our audio files, we decided to utilize the
 197 High Powered Computer (HPC) to run our function on the large files for each audio recorder
 198 and to extract the cleaned data. This would assist in speeding up the computation time, and

199 decreasing the amount of space utilized by WAV files on a desktop. To avoid any chance of
200 overwriting through this process, the function created a final output of an RDS file, so that
201 it could be renamed as the audio recorder name upon loading into the environment. We ran,
202 visualized, and analyzed all of the data, belonging to 20 recorders. Each was processed through
203 our function ‘eclipse_df’, per folder this task took roughly 10.5 hours.

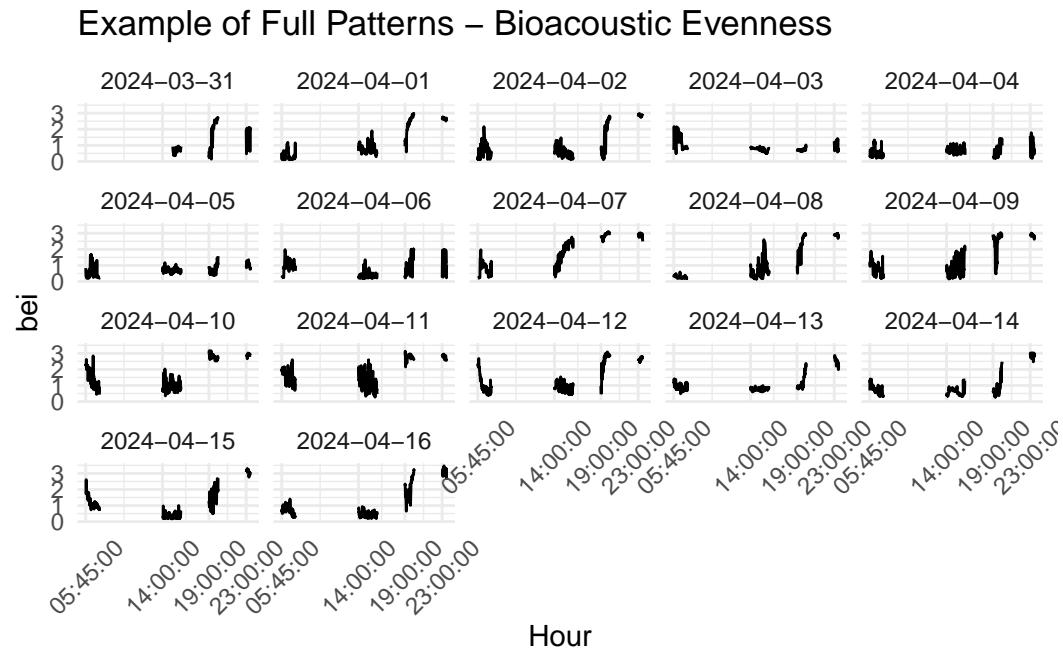
204 Through using the HPC, it was discovered that some of the folders contained empty .WAV
205 files which would kill the HPC job before the RDS was finished and saved. To combat this,
206 any folder which contained empty .WAV files, small .WAV files (<1 MB), and alternate file
207 types were deleted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

209 **4.3 Analyzing one full file from an audio recorder**

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

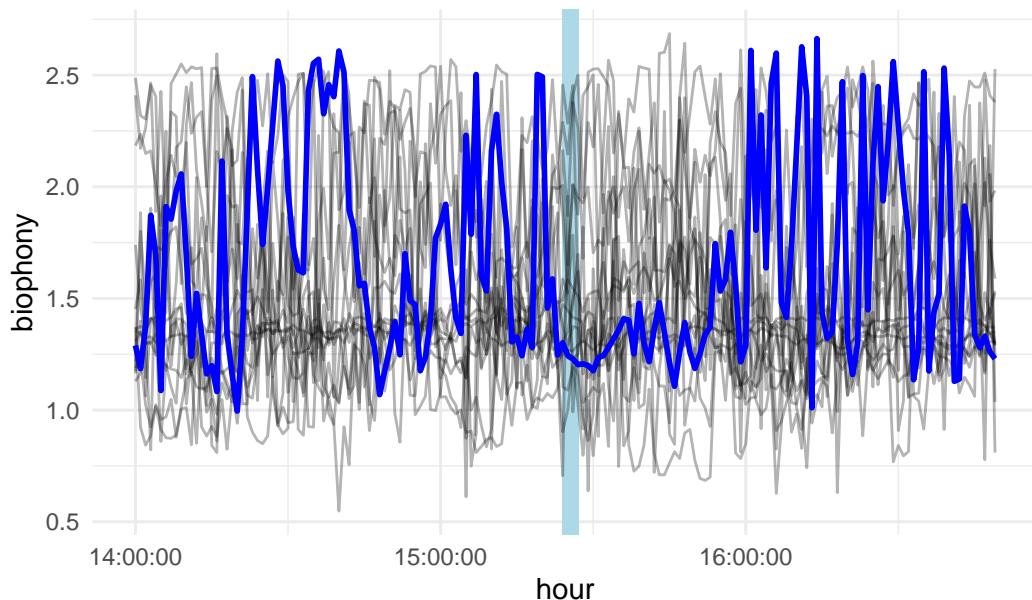
216 We started our analysis both on the full time series and the time segment that corresponded
217 to the eclipse. The creation of faceted graphics allowed us to look for patterns across days.

218 We learned that day to day the indices fluctuated, and that looking at the complete set of
219 recording intervals per day made it difficult to hone in on particular patterns between the 8th
220 and other days.



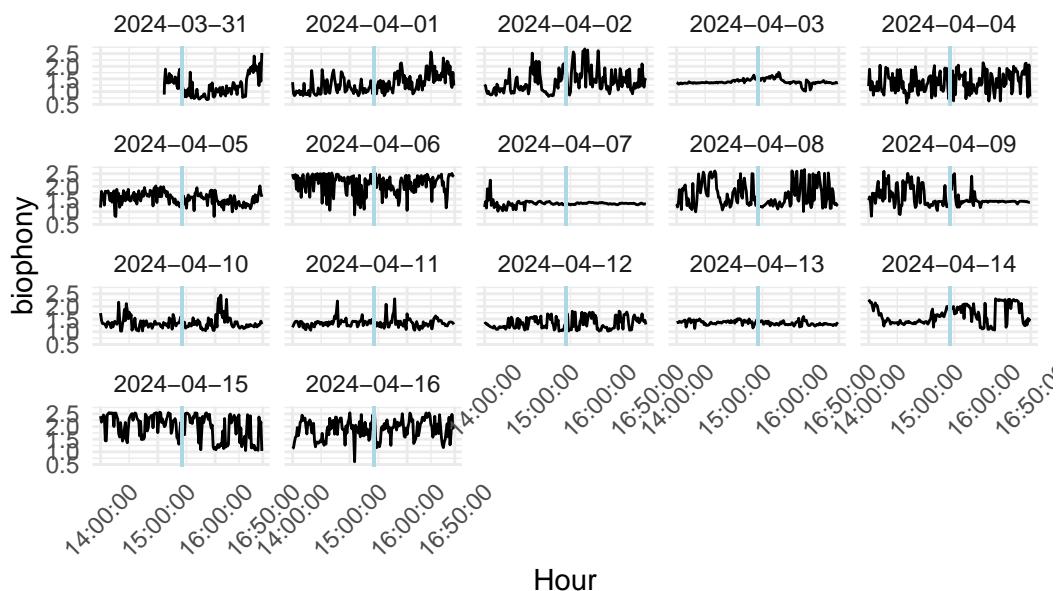
221
222 Focusing on variation across the eclipse portion, we subsetted that specific interval from
223 14:00:00 to 16:00:00, changed the color of the line corresponding to April 8th and added a
224 rectangle which highlights the time of totality (15:27:05 to 15:23:52). From this graph we can
225 see a strong decrease around the time that totality occurs on April 8th. It is unclear whether
226 this pattern occurs only on the 8th, but we can see some shading which shows other days have
227 higher biophony levels during this time.

Biophony over Eclipse Duration



228

Biophony over the Eclipse Duration

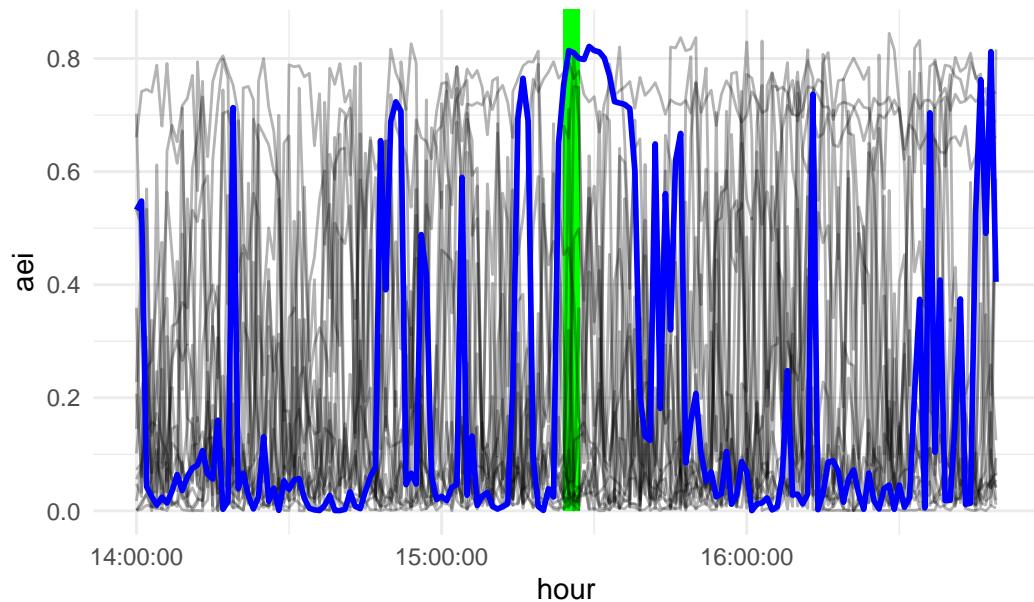


229

230 Creating this graph across all of the days, we can see that this stark decrease in sound is not
231 common and does not seem to occur on any day other than April 8th. This is interesting, as

²³² it seems to suggest that there isn't an eclipse-driven reaction in biophony.

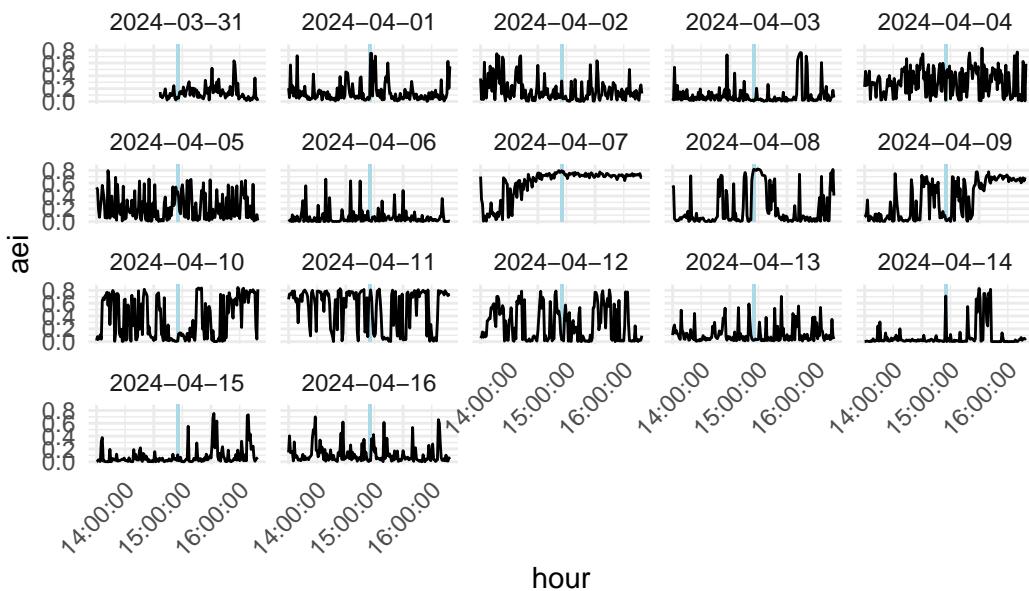
Acoustic evenness over the Eclipse Duration



²³³

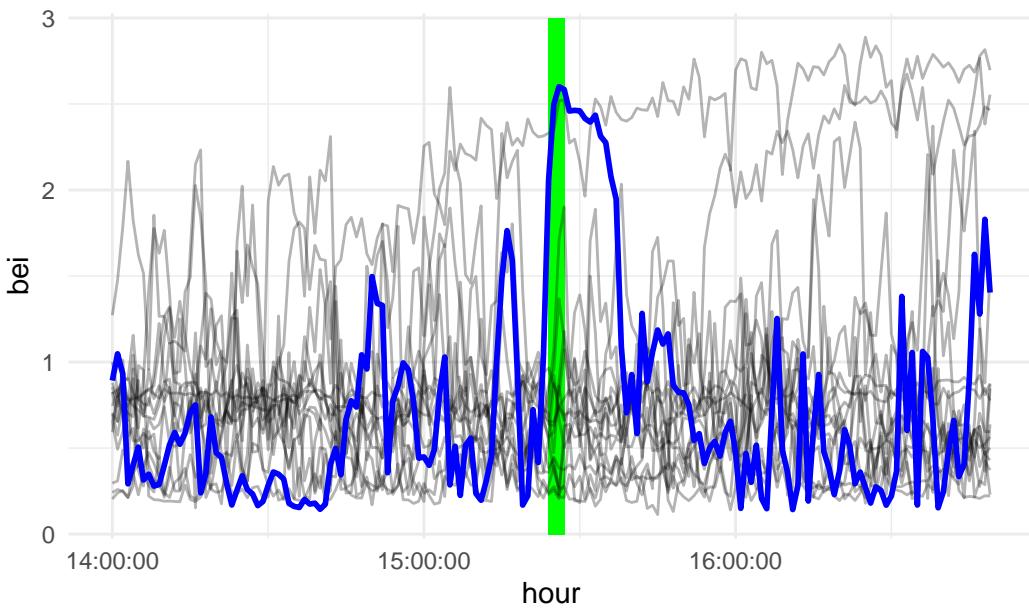
²³⁴ In acoustic evenness we can see an increase of AEI during the time of totality. We can also say
²³⁵ that on this day that peak is the highest that the AEI gets. Below when we look at each day
²³⁶ separately, it seems that this pattern we see on the day of the eclipse is infrequent. During
²³⁷ that time corresponding to totality, we don't really see a lot of similar increases and holds like
²³⁸ we do on the day of the eclipse.

Acoustic Evenness Index over the Eclipse Duration



239

Bioacoustic Index over the Duration of the Eclipse

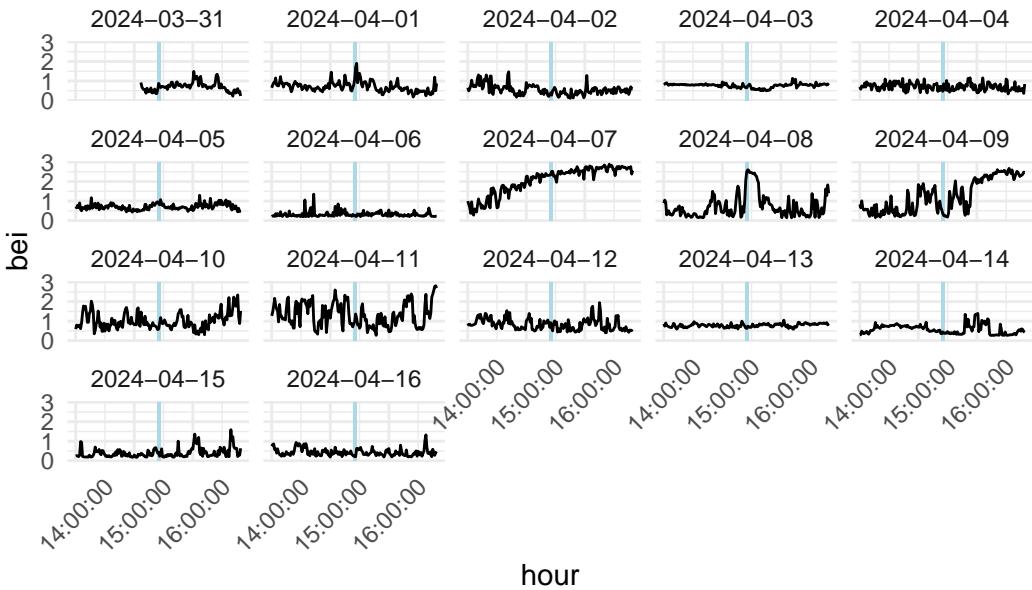


240

- ²⁴¹ In this graph we see another striking pattern, the BEI increasing significantly at the beginning of totality compared to the rest of this period. We can also see there are not a lot of days
- ²⁴²

²⁴³ which have this high BEI at this time, or during this period of day.

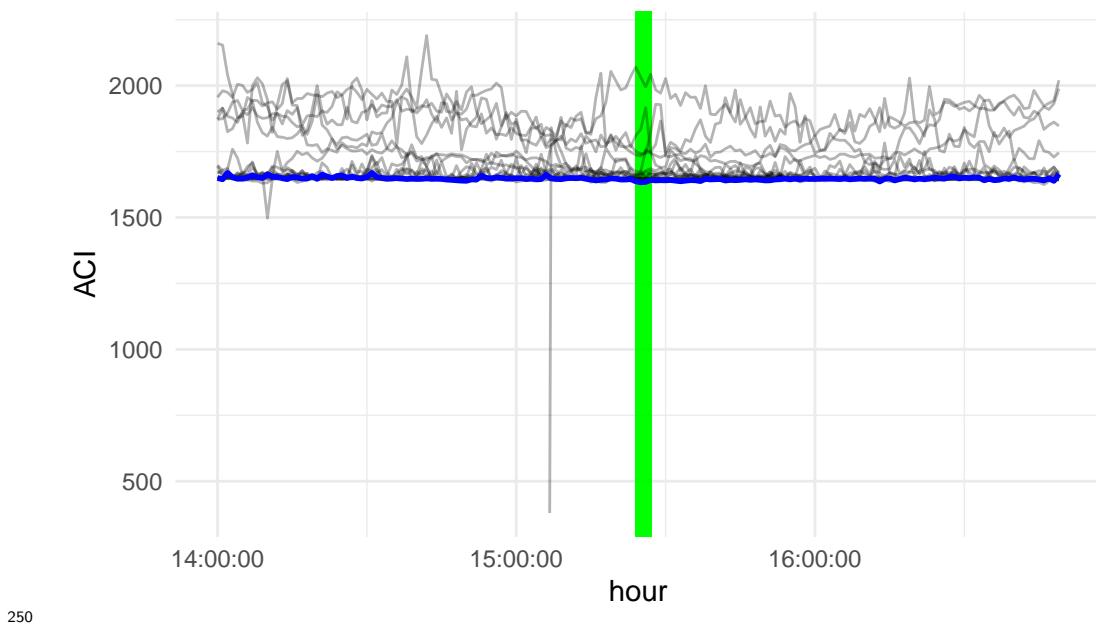
Bioacoustic Index over the Eclipse Duration



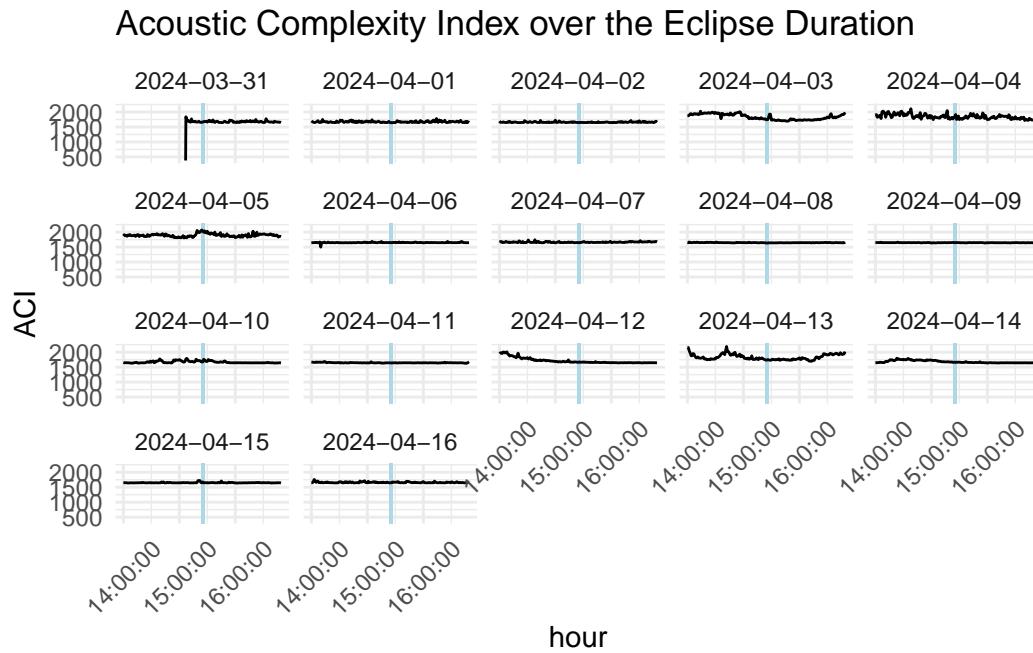
²⁴⁴

²⁴⁵ Faceting the graph we can get a better look at the patterns occurring at the times corresponding
²⁴⁶ to totality on April 8th. From this we can see that this stark increase is not common across the
²⁴⁷ other days, with the only other day where this pattern looks similar is on April 1st. Besides
²⁴⁸ this one occurrence, this pattern does not appear in any other day, and even on April 1st it is
²⁴⁹ not as severe as the day of the eclipse.

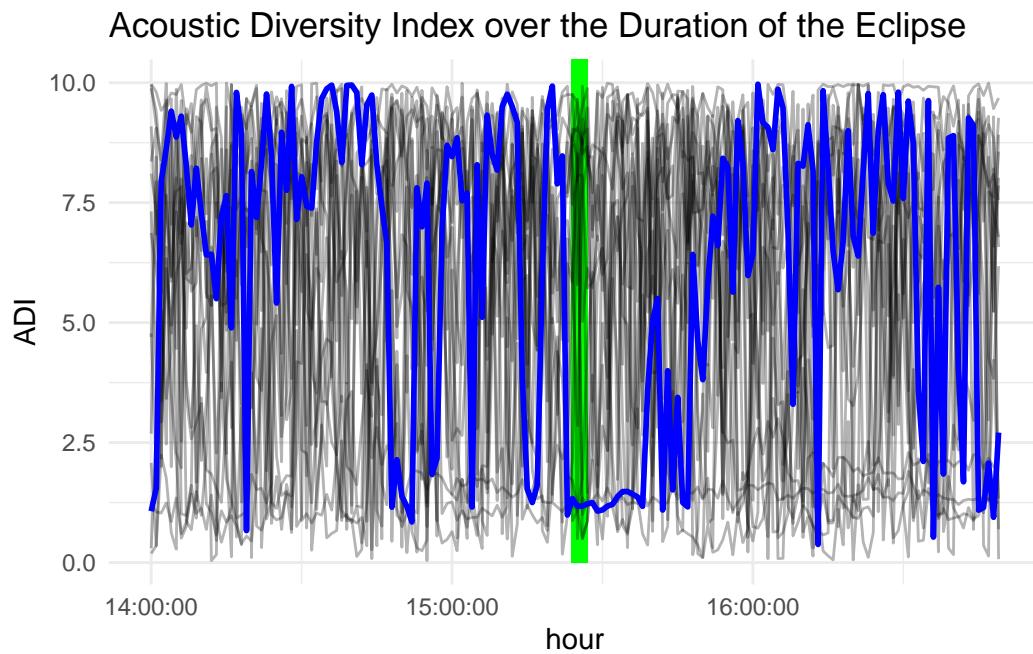
Acoustic Complexity over the Duration of the Eclipse



250
251 For the ACI there seems to be the potentially for an interesting pattern, however when we
252 create a faceted graph we can see that the ACI is often at this 1600 level and is actually quite
253 common across all the days.



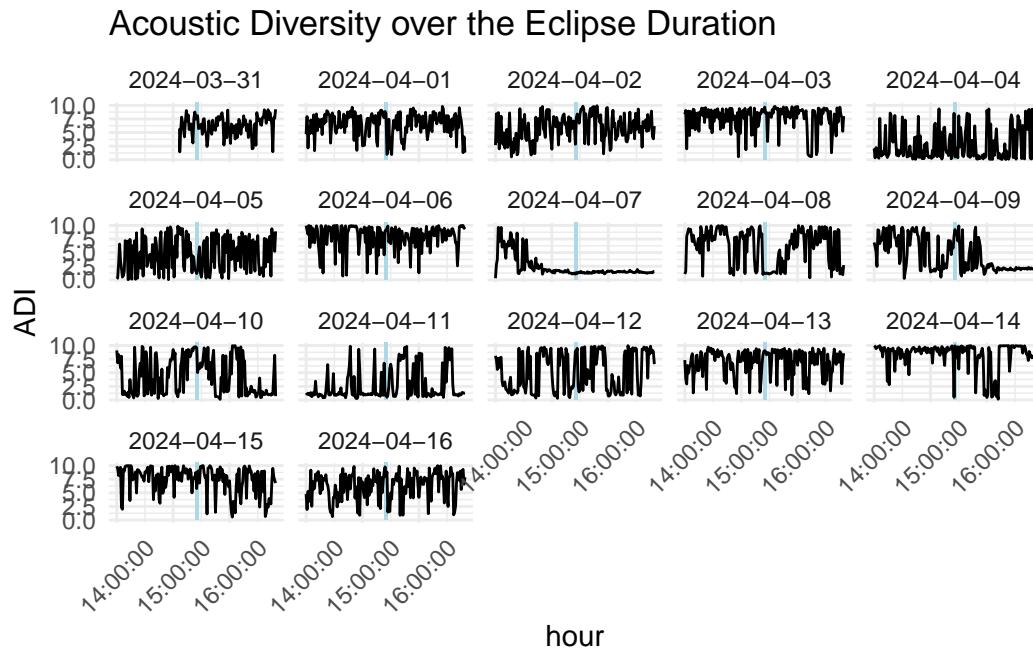
254



255

256 Lastly, looking at the ADI across this folder we can see that there is potentially a decrease
 257 at the point of totality. Due to the variation in this index, it is hard to discern whether this

²⁵⁸ pattern is infrequent from the type of graphic below, and looking into other representations
²⁵⁹ will give us more clarity.



²⁶⁰

²⁶¹ This set of visuals allowed us to look into whether we could expect to observe any potential
²⁶² patterns across the time of the eclipse for our five indices. We saw that there was definitely
²⁶³ the potential for some of them to display patterns later on, as we got into modeling.

²⁶⁴ 5 Full data visualization

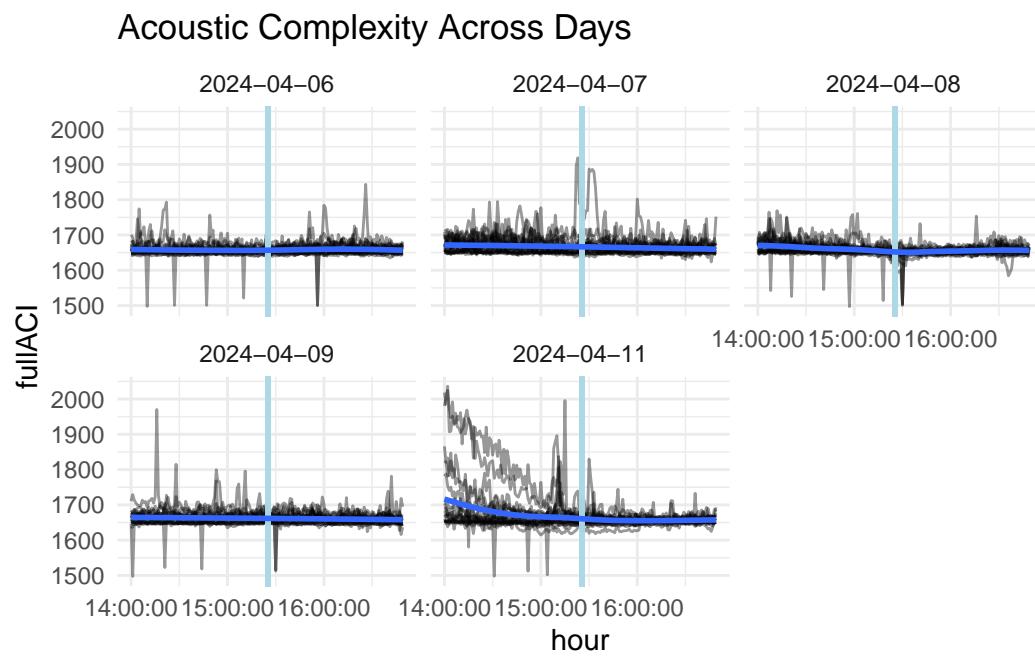
²⁶⁵ Due to concerns about changes in weather, migration around this time of year, readability we
²⁶⁶ utilized a subset of the recording days. To assess the weather conditions across the days we
²⁶⁷ used the website:

268 <https://www.wunderground.com/history/daily/us/ny/ogdensburg>

269 From looking at this website across our recording days, we decided to use the 2 days buffering
270 the eclipse, however due to wind gusts around midday on the 10th we opted for the 11th. This
271 created a subset of days which had similar weather patterns to the day of the eclipse, and
272 were close to April 8th, this decreased concern about avian or amphibian migration that could
273 alter the soundscape. This created a subset including the days: April 6th, 7th, 8th, 9th, and
274 11th.

275 We approached the full visualization of our data by looking across our five day subset to search
276 for patterns that differed between the 8th and the rest of our chosen days. In our visualizations,
277 we used a geom_smooth() layer to display a smoothed line so we could see overall patterns
278 more clearly.

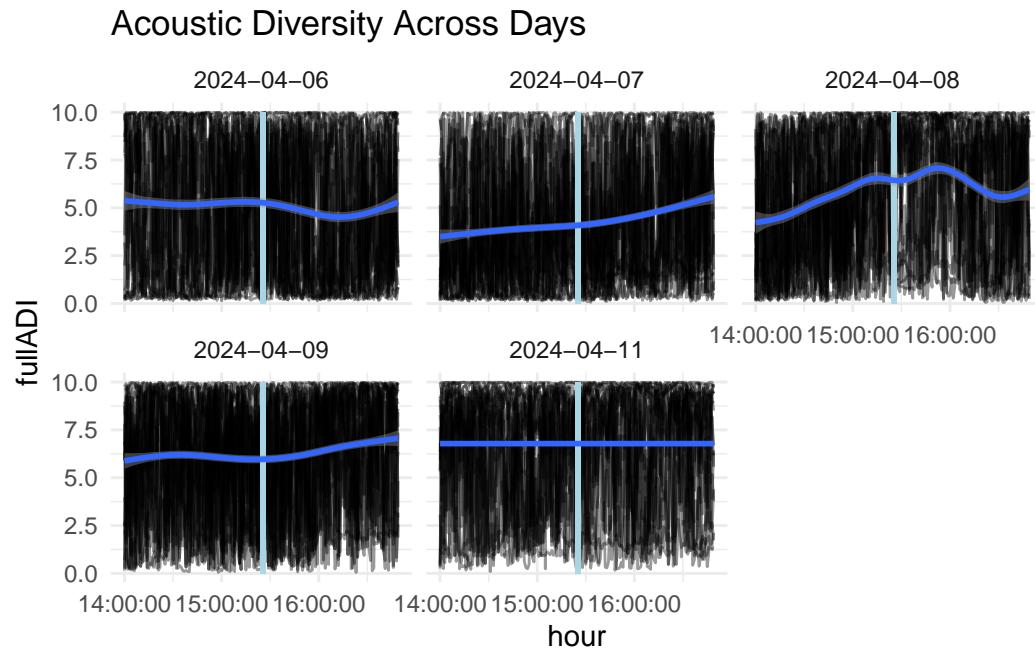
²⁷⁹ **5.1 Acoustic Complexity**



²⁸⁰

- ²⁸¹ In this graph it seems there isn't a pattern in this index. Along with the other days, April 8th
²⁸² has a pretty linear smoothing curve, and this signals that the eclipse on April 8th likely did
²⁸³ not have any influence on Acoustic Complexity.

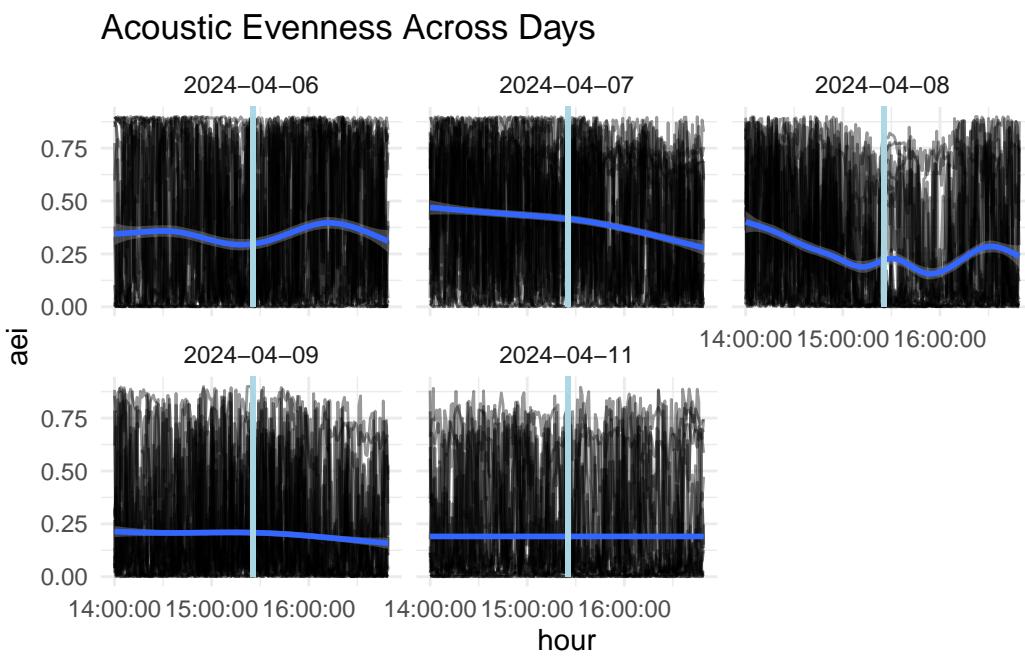
284 **5.2 Acoustic Diversity**



285

286 When we look at this visual, we can see basically all of the days from our subset are flat,
287 however on April 8th there seems to be two peaks close to the time of totality. However, we
288 also see some non-linearity around the time of totality from the 6th and the 9th. This makes
289 it difficult to make a claim about whether this index is showing a potential pattern at this
290 time.

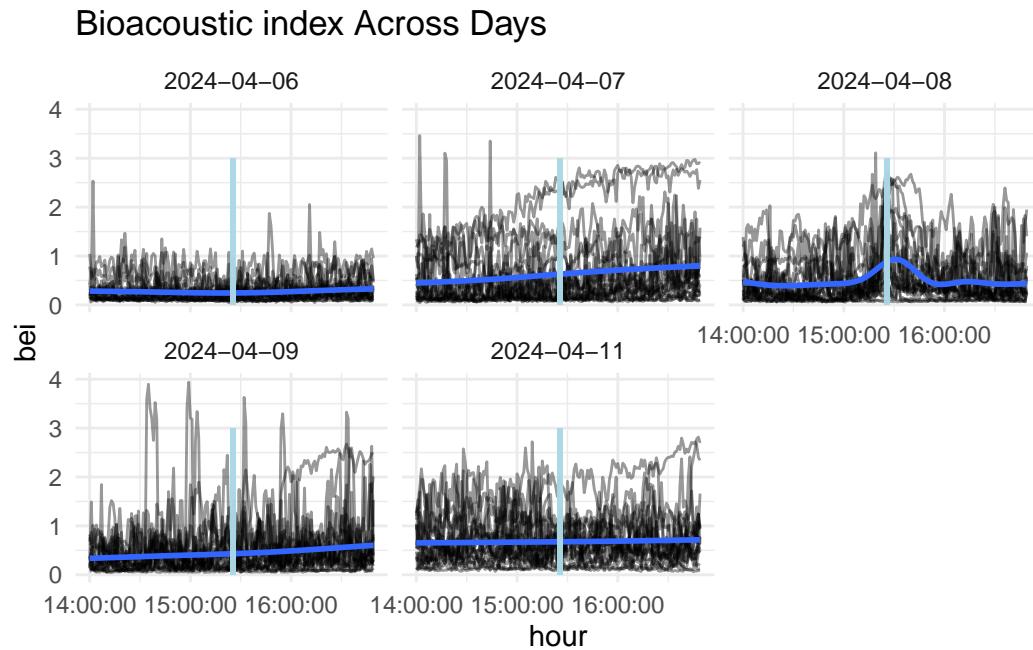
²⁹¹ **5.3 Acoustic Evenness**



²⁹²

²⁹³ Here, we can see that most days have flat smoothing curves, with fluctuations only on the 8th
²⁹⁴ and 6th. We can see that on the 8th, there is a much steeper rise around that time of totality,
²⁹⁵ with minimums on either side of its beginning and end. However, since we see some curvature
²⁹⁶ around the time of totality on the 6th, it is hard for us to conclude from this visual whether
²⁹⁷ the eclipse did influence Acoustic evenness or not.

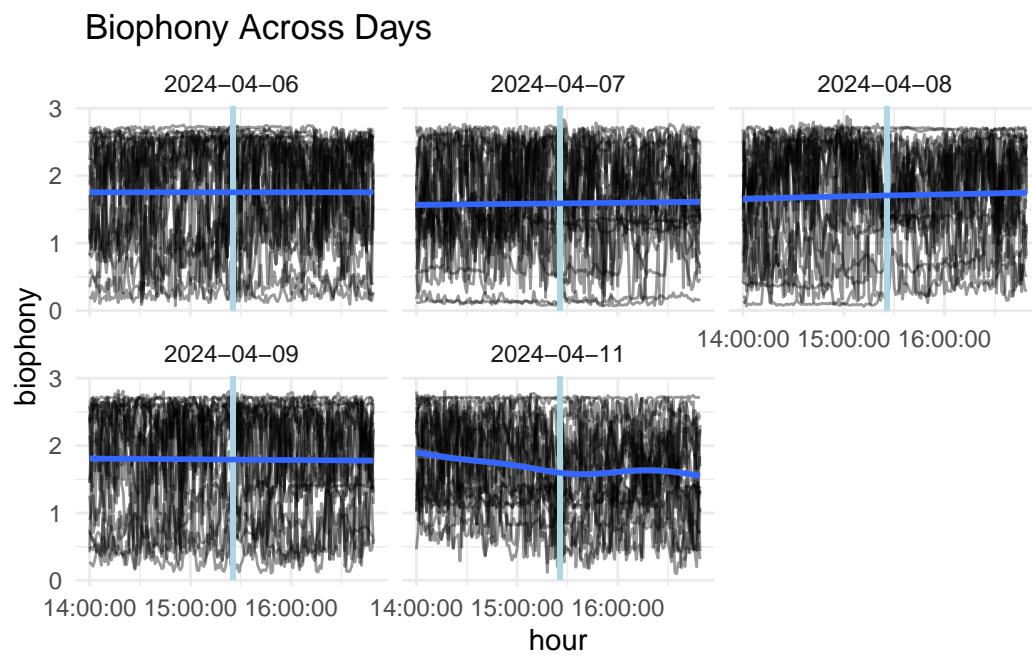
298 **5.4 Bioacoustic Index**



299

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

305 **5.5 Biophony**



306

307 For biophony, when we look at the smoothing curves across the days we can see that the
308 basically all of the days besides the 11th seem quite linear. Since April 8th is so flat, this
309 suggests that the eclipse may not be influencing this index.

310 **6 Modeling**

- 311 • Generalized Additive Modeling

312 To model this data we decided to use Generalized additive models so we could incorporate
313 non-linear patterns that we were seeing while still maintaining additivity between the variables.

314 A GAM model resembles linear regression, but replaces the linear components with smooth
315 non-linear functions (James et al. (2013)):

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

316 * Smoothing Splines

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

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

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

329 • Final Models

330 Our final generalized additive models were made using a smoothing spline for hour separated
331 by day, an additional smoothing spline for folder_name which would help us account for the
332 different recorders and their respective locations, and lastly a term for day. To create these
333 models we used the gam() function from the mgcv package, and to create our visualization
334 we used the data_grid() function (Wood (2011)) as well as the augment() function from the
335 broom package. This allowed us to create predicted values based on our model and to then
336 visualize them alongside our GAMs. To produce these visuals we also needed to specify a
337 specific recorder which lied around the “center” of our observed index values. A table of the
338 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

339 **6.1 Bioacoustic Index**

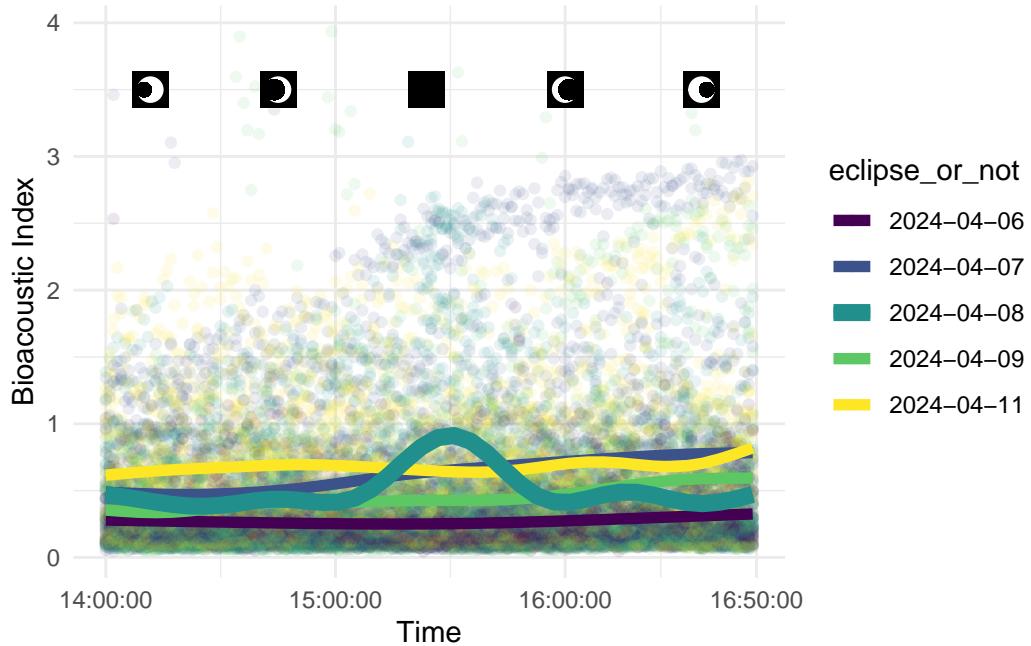
```
340 # A tibble: 6 x 5
341   term                  edf ref.df statistic p.value
342   <chr>                 <dbl> <dbl>     <dbl>    <dbl>
343 1 s(hour_numeric):day_factor2024-04-06  2.25    2.80      4.14  0.00873
```

```

344 2 s(hour_numeric):day_factor2024-04-07 4.55 5.59      51.8  0
345 3 s(hour_numeric):day_factor2024-04-08 8.86 8.99      57.5  0
346 4 s(hour_numeric):day_factor2024-04-09 6.52 7.66      17.9  0
347 5 s(hour_numeric):day_factor2024-04-11 7.46 8.42      3.59 0.000415
348 6 s(folder_name)                      19.0   19      527.   0

```

349 To analyze our model outputs, we will focus on the estimated degrees of freedom, as
 350 this will provide us with an estimate of the flexibility in the lines. We can also compare this
 351 flexibility on the 8th with the other days we have included in our subset, and can compare
 352 this the trend lines we see in our model visualizations. From our model for Bioacoustic index,
 353 can see that April 8th possesses the most complex line, with an edf of 8.86. The other days
 354 vary in their complexity, but we can see that they all have non-linear effects and April 7th is
 355 the second most complex to the day of the eclipse (edf: 7.46).



356

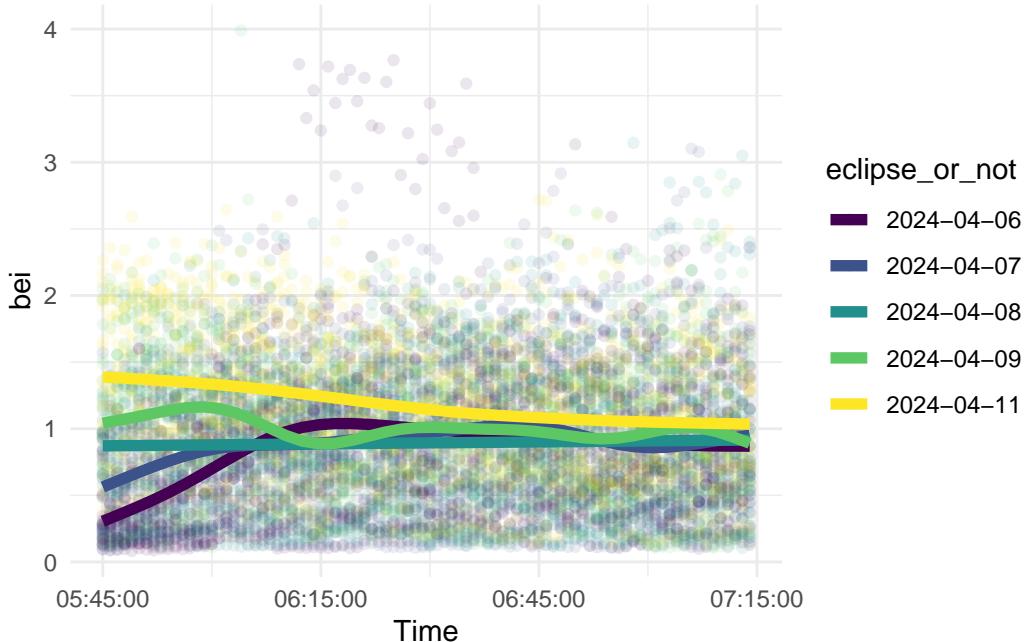
357 Note the icons that are in our model visuals, these align specifically with the changes during

358 the eclipse. The first and last icon correspond to the beginning (TIME) and end (TIME) of
359 the partial eclipse, and the center corresponds to peak totality (TIME).

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

366 To expand our understanding of if this day was potentially an anomaly or had different patterns
367 throughout the day, we created a model visualization that corresponds to the dawn.

```
368 # A tibble: 6 x 5
369   term                  edf ref.df statistic p.value
370   <chr>                <dbl> <dbl>    <dbl>    <dbl>
371 1 s(hour_numeric):day_factor2024-04-06  6.40    7.55    66.4     0
372 2 s(hour_numeric):day_factor2024-04-07  6.89    7.98    18.4     0
373 3 s(hour_numeric):day_factor2024-04-08  1.00    1.00     2.24    0.135
374 4 s(hour_numeric):day_factor2024-04-09  8.31    8.87     8.98     0
375 5 s(hour_numeric):day_factor2024-04-11  3.35    4.16    48.2     0
376 6 s(folder_name)           18.9     19      516.     0
```



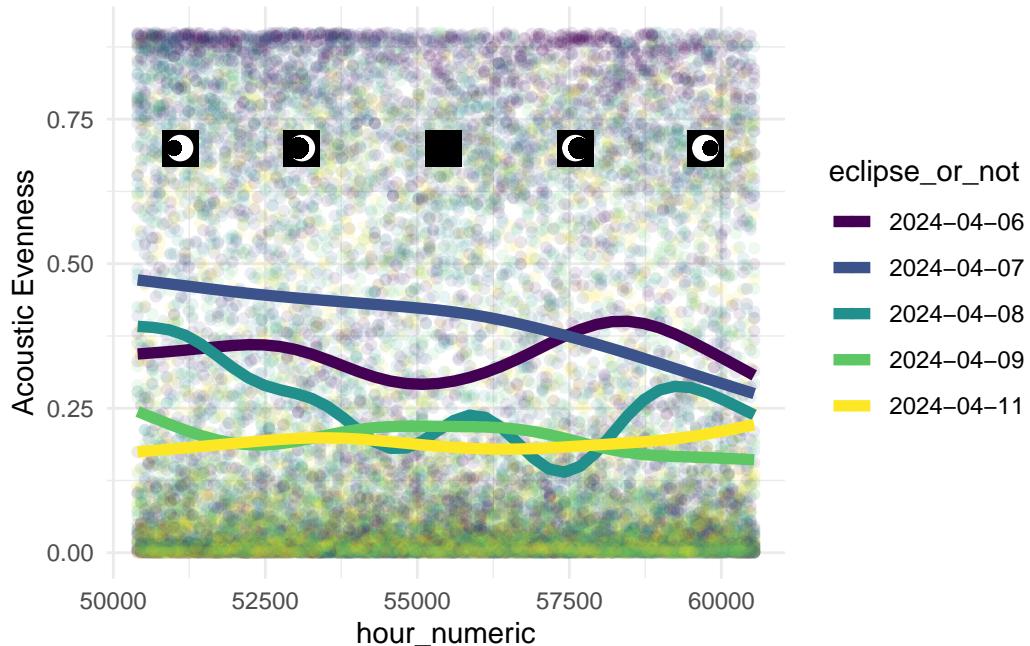
377

378 Looking at the model summary and the visual, we can see that there isn't anything off about
 379 April 8th. Based on its low edf value and the trend line, we can see that this day was pretty
 380 linear, and it didn't stick out compared to the other days. This helps provide evidence that
 381 the eclipse is causing a change in this index, since the day itself doesn't show any overall
 382 differences that could cause the peak we saw.

383 6.2 Acoustic Evenness

```
384 # A tibble: 6 x 5
385   term                  edf ref.df statistic p.value
386   <chr>                 <dbl> <dbl>    <dbl>    <dbl>
387 1 s(hour_numeric):day_factor2024-04-06  5.46   6.60     8.59  0
388 2 s(hour_numeric):day_factor2024-04-07  2.94   3.66    41.0   0
389 3 s(hour_numeric):day_factor2024-04-08  8.68   8.97    24.9   0
390 4 s(hour_numeric):day_factor2024-04-09  5.37   6.50     3.68  0.000814
391 5 s(hour_numeric):day_factor2024-04-11  3.79   4.69     1.21  0.299
392 6 s(folder_name)           19.0    19       278.   0
```

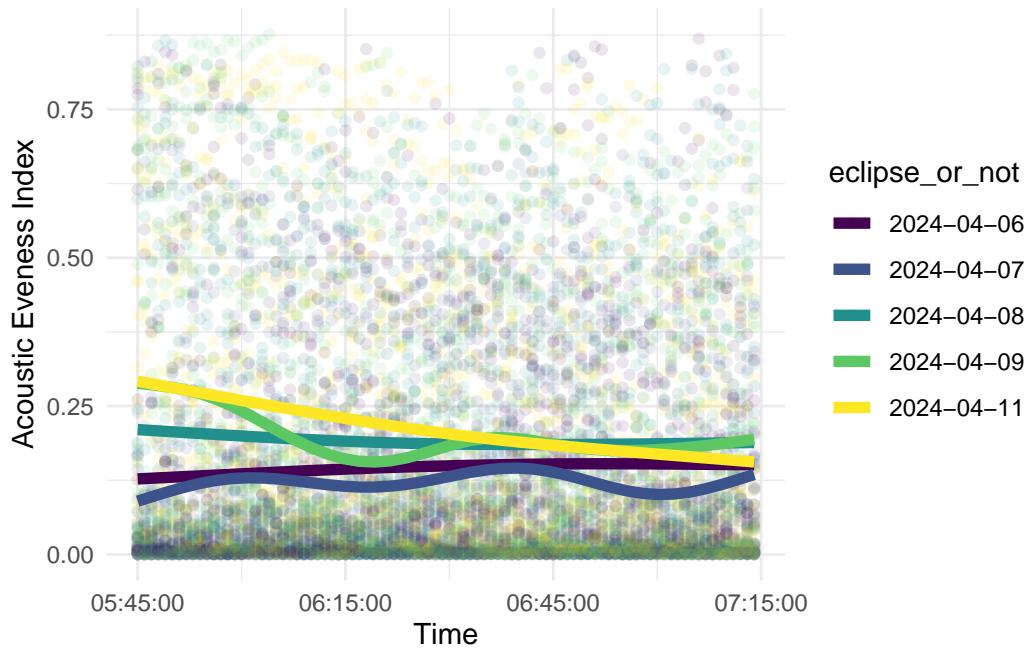
393 Similar to the bioacoustic index model, here we can see that April 8th is again the most
 394 flexible (edf: 8.68). In this model, the 8th is actually extremely more flexible than the other
 395 days, with the second highest being the 6th (edf: 5.46). As we recall from our full visualization
 396 for AEI, this makes sense as we ended up seeing a lot of curvature in both the 8th and the
 397 6th. Interestingly we see a close third with April 9th, with an edf of 5.37.



398
 399 Looking at this visual, we can see that that the lines from the 6th, 8th, and 9th are those with
 400 the most curvature. Looking at the 6th and 9th we see some fluctuations during the eclipse,
 401 but these are not as steep as the patterns from the 8th. On the 8th we can see there are two
 402 local minimums which fall before and after totality, and we see a local maximum appearing
 403 likely as totality was just ending. Aside from totality this line is pretty flexible, with lots of
 404 curvature across this full time interval. This provides some reason to suggest that the acoustic

405 evenness index is affected by the eclipse, but due to the non-linearity in other days and the
 406 amount of curvature we see, the evidence isn't as strong as we saw in the bioacoustic index.

```
407 # A tibble: 6 x 5
408   term                  edf ref.df statistic p.value
409   <chr>                <dbl> <dbl>    <dbl>    <dbl>
410  1 s(hour_numeric):day_factor2024-04-06  1.62    2.01     1.93  0.143
411  2 s(hour_numeric):day_factor2024-04-07  5.59    6.73     1.89  0.0694
412  3 s(hour_numeric):day_factor2024-04-08  1.74    2.17     1.73  0.187
413  4 s(hour_numeric):day_factor2024-04-09  5.89    7.05    11.0   0
414  5 s(hour_numeric):day_factor2024-04-11  1.97    2.46    34.9   0
415  6 s(folder_name)           18.9     19      138.    0
```



416

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

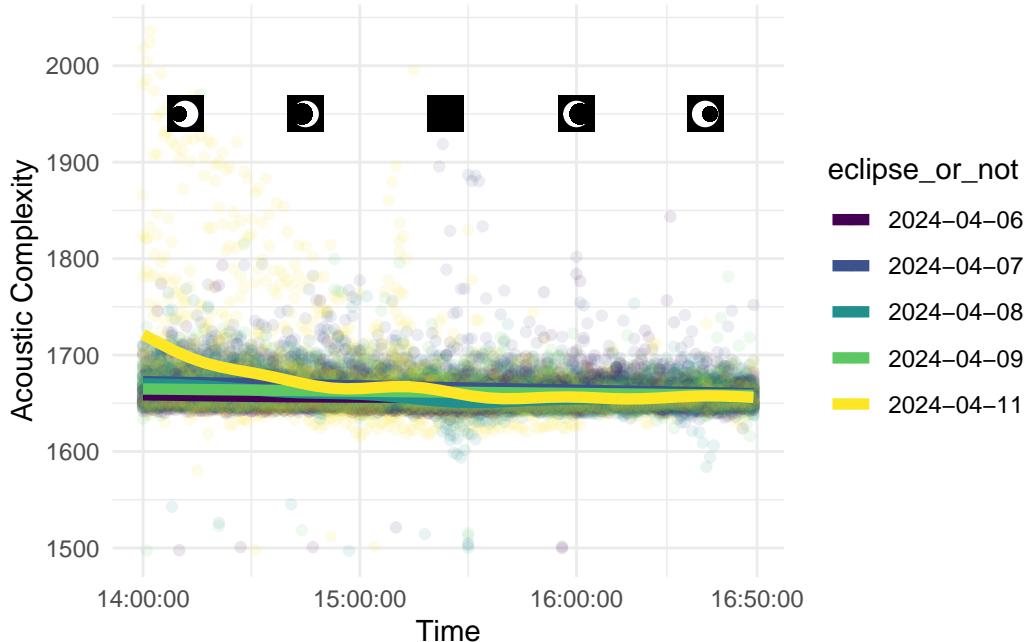
422 **6.3 Acoustic Complexity**

```

423 Family: gaussian
424 Link function: identity
425
426 Formula:
427 fullACI ~ s(hour_numeric, by = day_factor) + day_factor + s(folder_name,
428   bs = "re")
429
430 Parametric coefficients:
431                               Estimate Std. Error t value Pr(>|t|)
432 (Intercept)           1657.8533    2.8706 577.519 < 2e-16 ***
433 day_factor2024-04-07     8.3847    0.5066 16.550 < 2e-16 ***
434 day_factor2024-04-08     0.4213    0.5066  0.832  0.406
435 day_factor2024-04-09    3.7115    0.5066  7.326 2.48e-13 ***
436 day_factor2024-04-11    9.8398    0.5066 19.422 < 2e-16 ***
437 ---
438 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
439
440 Approximate significance of smooth terms:
441                               edf Ref.df      F p-value
442 s(hour_numeric):day_factor2024-04-06 3.815  4.719  2.29  0.0399 *
443 s(hour_numeric):day_factor2024-04-07  1.000  1.000 102.05 <2e-16 ***
444 s(hour_numeric):day_factor2024-04-08  7.497  8.440  28.35 <2e-16 ***
445 s(hour_numeric):day_factor2024-04-09  1.000  1.000  28.60 <2e-16 ***
446 s(hour_numeric):day_factor2024-04-11  8.748  8.981 225.82 <2e-16 ***
447 s(folder_name)                  18.940 19.000 316.04 <2e-16 ***
448 ---
449 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
450
451 R-sq.(adj) =  0.347  Deviance explained = 34.9%
452 GCV = 437.54  Scale est. = 436.36  n = 17000

```

453 Like the other days, we can see that the edf for April 8th is pretty high, compared to the other
 454 days (edf: 7.497). This is not the most complex day, with April 11th having an edf value of
 455 8.748. This possibly could suggest that although the data from April 8th is showing a lot of
 456 non-linearity, it may be due to factors other than the eclipse.



457

458 In this visual, we can see that although the estimated degrees of freedom values suggest a lot
 459 of non-linearity, the line for April 8th seems pretty flat. We can see that there isn't a lot of
 460 differences between the lines, besides April 11th. We conclude then that this metric is likely
 461 not having a reaction to the eclipse.

462 6.4 Acoustic Diversity

```

463 Family: gaussian
464 Link function: identity
465
466 Formula:
467 fullADI ~ s(hour_numeric, by = day_factor) + day_factor + s(folder_name,
468     bs = "re")
469
470 Parametric coefficients:
471                               Estimate Std. Error t value Pr(>|t|)
472 (Intercept)             5.04928   0.39293  12.85 <2e-16 ***
473 day_factor2024-04-07 -0.77722   0.07139 -10.89 <2e-16 ***

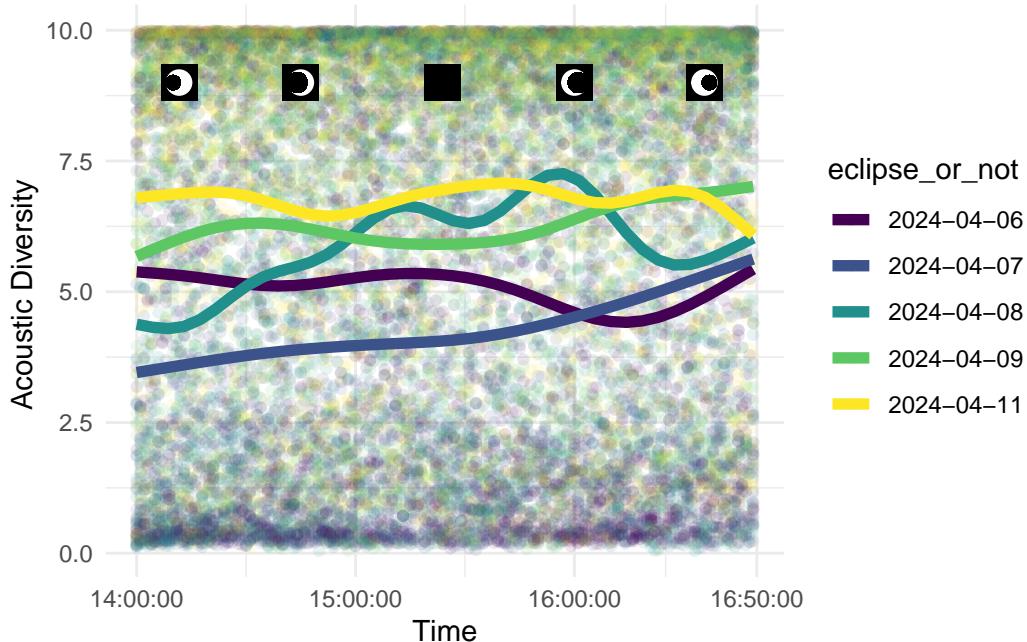
```

```

474 day_factor2024-04-08 0.83484    0.07139   11.69   <2e-16 ***
475 day_factor2024-04-09 1.22697    0.07139   17.19   <2e-16 ***
476 day_factor2024-04-11 1.73250    0.07139   24.27   <2e-16 ***
477 ---
478 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
479
480 Approximate significance of smooth terms:
481                               edf Ref.df      F  p-value
482 s(hour_numeric):day_factor2024-04-06 5.401  6.536  6.352 2.17e-06 ***
483 s(hour_numeric):day_factor2024-04-07 3.202  3.978 32.752  < 2e-16 ***
484 s(hour_numeric):day_factor2024-04-08 8.454  8.916 30.904  < 2e-16 ***
485 s(hour_numeric):day_factor2024-04-09 5.289  6.414  8.357  < 2e-16 ***
486 s(hour_numeric):day_factor2024-04-11 7.369  8.352  2.238   0.0294 *
487 s(folder_name)                      18.936 19.000 299.121 < 2e-16 ***
488 ---
489 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
490
491 R-sq.(adj) =  0.313  Deviance explained = 31.5%
492 GCV = 8.6927  Scale est. = 8.6653    n = 17000

```

493 Looking at the acoustic diversity, the estimated degrees of freedom are the highest yet again
494 (8.454). For this metric the 2nd highest is from April 11th, and it is pretty close at 7.369.
495 From our model summary, we also note that all of the other days are also showing some type
496 of non-linearity through their edf values.



497

498 As we noted, we can see that all of the days are showing some type of flexibility. When we
 499 look specifically at the 11th, we can see its quite wavy, and we see some dips around this
 500 period of the eclipse. When we look at the trend from the 8th, we note that there are two
 501 maximums surrounding the time of totality. In between these peaks is a minimum around the
 502 time that totality was ending. Compared to the trends we see from the 11th, these patterns
 503 align more with the changes occurring because of the eclipse. This suggests to us that the
 504 acoustic diversity is being affected by the eclipse, yet due to the curvature in the other lines,
 505 we don't have the strongest evidence.

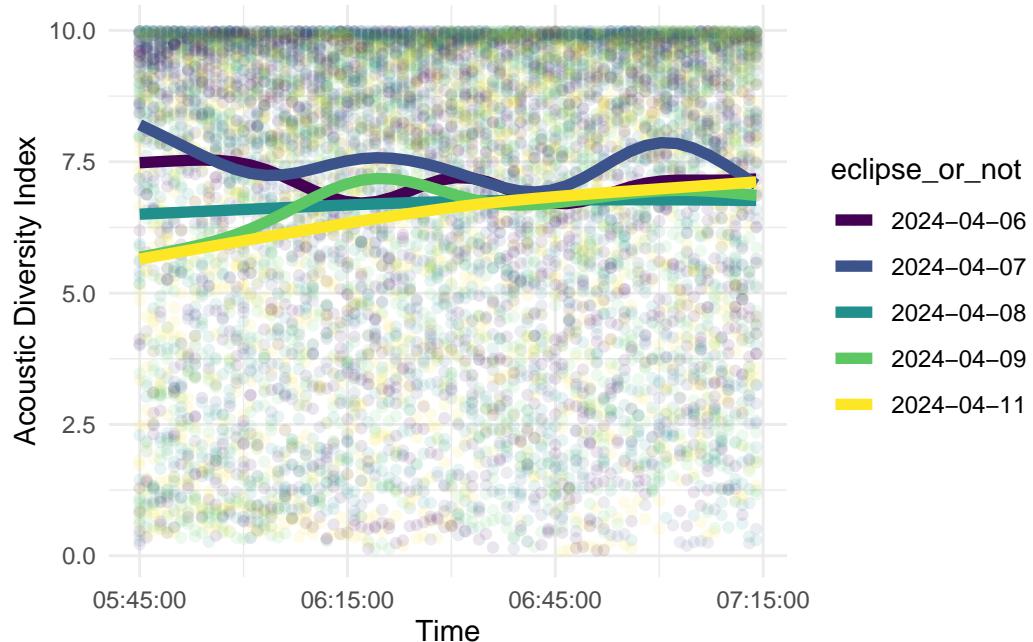
506 Family: gaussian
 507 Link function: identity
 508
 509 Formula:
 510 `fullADI ~ s(hour_numeric, by = day_factor) + day_factor + s(folder_name,`
 511 `bs = "re")`

512

```

513 Parametric coefficients:
514                               Estimate Std. Error t value Pr(>|t|)
515 (Intercept)                 7.09459   0.35118 20.202 < 2e-16 ***
516 day_factor2024-04-07     0.34779   0.08238  4.222 2.45e-05 ***
517 day_factor2024-04-08    -0.39533   0.08238 -4.799 1.62e-06 ***
518 day_factor2024-04-09    -0.41400   0.08238 -5.025 5.12e-07 ***
519 day_factor2024-04-11    -0.57241   0.08238 -6.948 3.96e-12 ***
520 ---
521 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
522
523 Approximate significance of smooth terms:
524                               edf Ref.df      F p-value
525 s(hour_numeric):day_factor2024-04-06 6.800 7.906 2.941 0.003293 **
526 s(hour_numeric):day_factor2024-04-07 6.477 7.618 4.037 0.000129 ***
527 s(hour_numeric):day_factor2024-04-08 1.514 1.863 1.071 0.256363
528 s(hour_numeric):day_factor2024-04-09 5.472 6.612 8.066 < 2e-16 ***
529 s(hour_numeric):day_factor2024-04-11 1.933 2.413 23.665 < 2e-16 ***
530 s(folder_name)                18.893 19.000 174.562 < 2e-16 ***
531 ---
532 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
533
534 R-sq.(adj) =  0.288  Deviance explained = 29.1%
535 GCV = 6.1393  Scale est. = 6.1079  n = 9000

```



536

537 To try and get some more evidence, we looked at a model from the dawn. Like the other
 538 metrics we calculated dawn models for, April 8th displays a pretty linear trend line and a low
 539 edf value (1.514). Since the 8th doesn't vary too much from the other days, this shows April
 540 8th was a regular day for acoustic diversity measurement.

541 6.5 Biophony

```

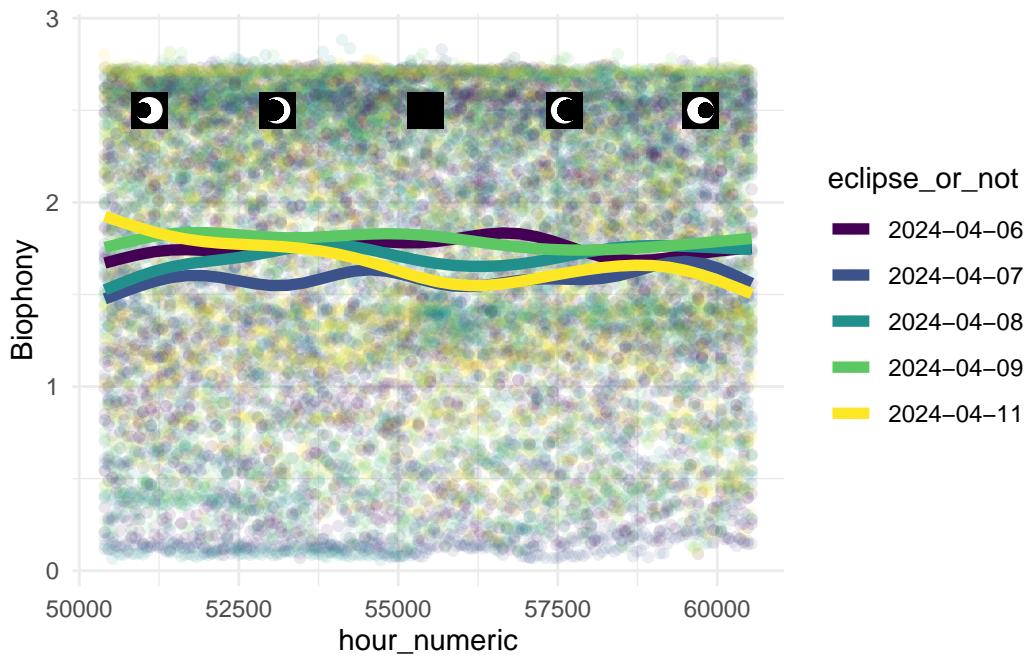
542 Family: gaussian
543 Link function: identity
544
545 Formula:
546 biophony ~ s(hour_numeric, by = day_factor) + day_factor + s(folder_name,
547   bs = "re")
548
549 Parametric coefficients:
550             Estimate Std. Error t value Pr(>|t|)
551 (Intercept) 1.75374  0.13710 12.791 < 2e-16 ***
552 day_factor2024-04-07 -0.16427  0.01156 -14.216 < 2e-16 ***
553 day_factor2024-04-08 -0.04941  0.01156 -4.276 1.92e-05 ***
554 day_factor2024-04-09  0.03912  0.01156  3.385 0.000712 ***
555 day_factor2024-04-11 -0.07651  0.01156 -6.621 3.68e-11 ***
556 ---
557 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
558
559 Approximate significance of smooth terms:
560                               edf Ref.df      F p-value
561 s(hour_numeric):day_factor2024-04-06 7.384  8.362 3.595 0.000291 ***
562 s(hour_numeric):day_factor2024-04-07 8.113  8.788 3.323 0.001594 **
563 s(hour_numeric):day_factor2024-04-08 6.558  7.693 6.281 < 2e-16 ***
564 s(hour_numeric):day_factor2024-04-09 5.876  7.034 2.429 0.016671 *
565 s(hour_numeric):day_factor2024-04-11 6.292  7.445 21.905 < 2e-16 ***
566 s(folder_name)                  18.986 20.000 1376.962 < 2e-16 ***
567 ---
568 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
569
570 R-sq.(adj) =  0.624  Deviance explained = 62.5%
571 GCV = 0.22779  Scale est. = 0.22701 n = 17000
  
```

572 Lastly, in our biophony model all of the days have higher non-linearity with edf values above

573 five. Interestingly, in this summary the 8th is not one of the most complex lines (edf: 6.558).

574 This began to suggest that we were not going to see a pattern that would align with the eclipse

575 on this day, and if we did we could potentially also see similar non-linearity in the other days.



576

577 This visualization shows very clearly that all of the days during this time frame are very similar,

578 and we don't see any particular patterns from April 8th or any other day. Like we expected,

579 all are displaying a lot of curvature, and we aren't seeing any patterns that really align with

580 totality, or the beginning and end of the partial eclipse. This provides evidence that biophony

581 was not impacted during the time of eclipse.

582 **7 Conclusion (ADD MORE AND CONNECT POINTS)**

583 From our analysis, we are at a place to believe that the bioacoustic index, acoustic evenness
584 index, and acoustic diversity index may display patterns across the time of the eclipse. Looking
585 at model summaries and visualizations, we say that the bioacoustic has the most evidence of
586 a potential pattern with a pretty pronounced maximum around peak totality. The acoustic
587 diversity and acoustic evenness indices show a bit less evidence of patterns, but the curvature
588 around totality provide enough evidence for us to claim there is the potential these are affected
589 as well.

590 (EXPLAIN WHAT THE PATTERNS WOULD MEAN)

591 We believe that these potential patterns display that wildlife is impacted by some type of
592 eclipse-driven change, and based on the patterns, we can suggest that they are affected by
593 specifically a totality-driven change. Whether this is due to something we cannot perceive or
594 it is the changes in light, we cannot be sure at this time.

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

600 **References**

- 601 Gerber, Jacob E, Dakota Howard, and John E Quinn. 2020. “Soundscape Shifts During
602 the 2017 Total Solar Eclipse: An Application of Dispersed Automated Recording Units to
603 Study Ephemeral Acoustic Events.” *Biodiversity* 21 (1): 41–47.
- 604 James, Gareth, Daniela Witten, Trevor Hastie, Robert Tibshirani, et al. 2013. *An Introduction
605 to Statistical Learning*. Vol. 112. 1. Springer.
- 606 Villanueva-Rivera, Luis J., and Bryan C. Pijanowski. 2018. *Soundecology: Soundscape Ecology*.
607 <https://CRAN.R-project.org/package=soundecology>.
- 608 Wood, S. N. 2011. “Fast Stable Restricted Maximum Likelihood and Marginal Likelihood
609 Estimation of Semiparametric Generalized Linear Models.” *Journal of the Royal Statistical
610 Society (B)* 73 (1): 3–36. <https://doi.org/10.1111/j.1467-9868.2010.00749.x>.