Title: Automated acoustic detection of river herring *(Alosa pseudoharengus* and *A. aestivalis*) spawning activity

Time period: 2021

Location: Choptank River, Maryland

Purpose: To use passive acoustic monitoring (PAM) and automatic detection of spawning splashes to examine the timing and environmental drivers of spawning in river herring (Alewife *Alosa pseudoharengus* and Blueback Herring *A. aestivalis*).

Access: These data are not sensitive or classified.

Owners:Matthew Ogburn (Smithsonian Environmental Research Center)

Data compiled by: Matthew Ogburn (Smithsonian Environmental Research Center)

Amber Staples (Smithsonian Environmental Research Center)

Henry Legett (Smithsonian Environmental Research Center)

Keira Heggie (Smithsonian Environmental Research Center)

Jessica Deichmann (Smithsonian’s National Zoo and Conservation Biology Institute)

Cross reference: These data are used in an analysis in a published manuscript:

Staples S, Legett HD, Deichmann JL, Heggie K, Ogburn MB. Automated acoustic detection of river herring (*Alosa pseudoharengus* and *Alosa aestivalis*) spawning activity.

File names: Data: spawning\_splash.csv

Status: Ongoing

Methodology: Acoustic recordings of spawning splashes were collected from March to May 2021 in the Choptank River, Maryland, using an AudioMoth recorder. Recordings were analyzed using a random forest model on the RFCx-ARBIMON platform to determine hourly presence-absence of splashes. Recording were also analyzed for the presence of predators (Great Blue Herons *Ardea herodias*, Bald Eagles *Haliaeetus leucocephalus*, and Ospreys *Pandion haliaetus*) using the RFCx Arbimon Pattern Matching algorithm.

Completeness: These data are ongoing

***spawning\_splash\_data.csv***

Description:This dataset contains hourly presence-absence of spawning splashes produced by river herring, sonar counts of river herring, associated environmental data, and the presence-absence of avian predator calls in the Choptank River.

File type: Comma-separated values file

Processing steps: River herring spawning activity was monitored by documenting splashing throughout the spawning season from March 8 to May 25, 2021 using an AudioMoth acoustic recording device (version 1.1.0). The AudioMoth was attached to a tree on the bank of the river approximately 2.5 m from ground level. Acoustic recordings were collected 24-hours a day for 1 min every 10 mins (six samples per hour), yielding a total of 11,215 recordings. These recordings were analyzed using a random forest model (RFM) on the RFCx-ARBIMON platform (https://arbimon.rfcx.org) to automatically determine the presence or absence of splashes. The RFM was trained using a subsample of 400 haphazardly chosen recordings in which the presence or absence of splashing had been manually determined. This subsample was evenly split, such that 200 recordings had splashes and the other 200 did not. In each set of 200 recordings, 140 were used in model fitting and 60 were used to verify the model by to assessing precision (true positives / [true positives + false positives]) and accuracy ([true positives + true negatives] / total). Following the RFM analysis, the per-hour presence of spawning splashes was established. An hour was scored as having splashes if they were detected in at least one of the six recordings for that hour. Daily splash activity was calculated as the number of hours in which splashing was detected each day (0-24h).

Hourly measures of water temperature (ºC) were obtained using a HOBO Pendant water temperature logger (Onset Computer Corp, UA-002-08, accuracy: ± 0.53°C from 0° to 50°C). Values of river discharge (m3/s) were obtained from USGS gauging station 01491000 and water levels (m) relative to the Mean Lower Low Water (MLLW) were obtained from NOAA tide predictions (ID: 8572669). Values for solar elevation (º, angle from the horizon) were calculated in R (version 4.0.3) using the “astrocalc4r” function in the fishmethods package (https://cran.r-project.org/web/packages/fishmethods/fishmethods.pdf). Daily lunar cycle values were obtained using the “lunar.illumination” function in the lunar R package (https://cran.r-project.org/web/packages/lunar/lunar.pdf). The lunar values are calculated from MULTIFAN-CL fisheries stock assessment models from the R4MFCL project. All fish, predator, and environmental data were set to a local standard time of GMT/UTC-5.

The presence or absence of the calls of three avian predators (Great Blue Heron, Bald Eagle, and Osprey) were assessed from the recordings of spawning splashes. The calls from these species were too rare in the recordings to build subsamples to train RFMs. Instead, the presence of calls for each species was detected using the RFCx Arbimon Pattern Matching algorithm to identify regions of interest that were similar (correlation coefficient > 0.2) to example recordings of each species. Recordings with regions of interest above the correlation threshold were manually reviewed to confirm avian call presence.

Counts of adult river herring migrating upstream were conducted using imaging sonar at a site 240 meters upstream from the recorded spawning splashes. Counts were collected for the entire spawning season but only a one-week period (April 15 to April 21, 2021) of processed data during the estimated peak of the spawning run was available at the time of this study to make hourly comparisons between spawning splashes and fish counts. Imaging sonar video recordings were collected for 10-min segments every hour using a Dual-frequency Identification Sonar unit (DIDSON: Sound Metrics Corp, Bellevue, WA). These recordings were processed following established protocols, wherein individual fish moving upstream in the recordings were manually counted by a trained observer using DIDSON V5.25.52 software. The observer was trained using a standard set of files.

Source inputs: AudioMoth acoustic recording device (version 1.1.0)

Entity and attributes:

Column heading: date

Label: Date

Description: Date of count

Data type: Date

Measurement unit: None

Comments: Format: Month/Day/Year

Column heading: month

Label: month

Description: Month in which counts were collected

Data type: Numeric

Measurement unit: None

Comments:

Column heading: day

Label: Day

Description: Day of the month on which counts were collected

Data type: Numeric

Measurement unit: None

Comments:

Column heading: year

Label: Year

Description: Year in which counts were collected

Data type: Character

Measurement unit: None

Comments: 2021

Column heading: hour

Label: Hour

Description: Hour of the day in which counts were collected

Data type: Numeric

Measurement unit: None

Comments:

Column heading: lat

Label: Latitude

Description: Latitude of sampling site

Data type: Numeric

Measurement unit: Decimal degrees (DD)

Comments:

Column heading: lon

Label: Longitude

Description: Longitude of sampling site

Data type: Numeric

Measurement unit: Decimal degrees (DD)

Comments:

Column heading: timezone

Label: Timezone

Description: Timezone in which hourly data was aligned

Data type: Numeric

Measurement unit: None

Comments: GMT/UTC-5

Column heading: splash\_sum

Label: Sum of splashes per hour

Description: Number of recordings per hour in which splashing was present. There were 6 recording per hour, thus possible counts range from 0 to 6.

Data type: Numeric

Measurement unit: Count

Comments: 0 to 6

Column heading: splash\_pres

Label: Presence of splashes

Description: Presence or absence of splashing per hour

Data type: Numeric

Measurement unit: Binary

Comments: 0 = absent, 1 = present

Column heading: temp

Label: Temperature

Description: Temperature

Data type: Numeric

Measurement unit: degrees Celsius (°C)

Column heading: flow

Label: Flow

Description: Stream discharge measured by USGS gauging station.

Data type: Numeric

Measurement unit: Meter cubed per second (m3/s)

Comments: Choptank River station no. 01491000. Values were converted from ft3/s.

Column heading: tide\_cambridge

Label: Tide Cambridge MD

Description: Predicted water level (Mean Lower Low Water) measured by NOAA tides and currents station in Cambridge, MD.

Data type: Numeric

Measurement unit: Meters (m) of Mean Lower Low Water (MLLW)

Comments: Station ID 8571892

Column heading: tide\_hillsboro

Label: Tide Hillsboro MD

Description: Predicted water level (Mean Lower Low Water) measured by NOAA tides and currents station in Hillsboro, MD.

Data type: Numeric

Measurement unit: Meters (m) of Mean Lower Low Water (MLLW)

Comments: Station ID 8572669

Column heading: solar\_zenith

Label: Solar Zenith

Description: Solar zenith in degrees from vertical, GMT-5. Angles are corrected for atmospheric diffraction.

Data type: Numeric

Measurement unit: Degrees (°)

Comments:

Column heading: solar\_elevation

Label: Solar Elevation

Description: Solar elevation in degrees above horizon, GMT-5. Angles are corrected for atmospheric diffraction.

Data type: Numeric

Measurement unit: Degrees (°)

Comments:

Column heading: heron\_sum

Label: Sum of Great Blue Heron calls per hour

Description: Number of recordings per hour in which Great Blue Heron calls were present. There were 6 recording per hour, thus possible counts range from 0 to 6.

Data type: Numeric

Measurement unit: Count

Comments: 0 to 6

Column heading: heron\_pres

Label: Presence of Great Blue Heron calls

Description: Presence or absence of Great Blue Heron calls per hour

Data type: Numeric

Measurement unit: Binary

Comments: 0 = absent, 1 = present

Column heading: eagle\_sum

Label: Sum of Bald Eagle calls per hour

Description: Number of recordings per hour in which Bald Eagle calls were present. There were 6 recording per hour, thus possible counts range from 0 to 6.

Data type: Numeric

Measurement unit: Count

Comments: 0 to 6

Column heading: eagle\_pres

Label: Presence of Bald Eagle calls

Description: Presence or absence of Bald Eagle calls per hour

Data type: Numeric

Measurement unit: Binary

Comments: 0 = absent, 1 = present

Column heading: didson\_upstream

Label: Upstream DIDSON Count

Description: Count of the number of fish moving upstream in the 10 min sonar recording. All fish within the size range of river herring, 200–350 mm, were counted.

Data type: Numeric

Measurement unit: Count

Comments:

Column heading: didson\_downstream

Label: Downstream DIDSON Count

Description: Count of the number of fish moving downstream in the 10 min sonar recording. All fish within the size range of river herring, 200–350 mm, were counted.

Data type: Numeric

Measurement unit: Count

Comments: