Capstone 1 - Milestone Report
Burns, Echelle
Dec 2019

1. Introduction

Acoustic telemetry is a common tool used in research projects that focus on the movement patterns of species that are resident to seawater environments. Individuals are tagged with small, acoustic transmitters (hereby transmitters) that emit an ultrasonic signal that can be heard by acoustic receivers (hereby receivers) that are deployed nearby (within 0 m to 1 km away, depending on the environmental conditions). Similar to the FasTrak lane on the highway, receivers act as a toll booth, listening for transmitters that come into its detection radius, and transmitters act as the FasTrak beeper in your car, sending out a signal to be picked up by the toll booth (receiver). When a transmitter is within the detection range of a receiver, the receiver then records the date, time, and transmitter ID code of the tagged animal in the area. This method of acoustic telemetry is termed “passive” acoustic telemetry, because it does not require researchers to actively follow individuals in real-time. Instead, the receivers remain stationary in a single area for as little as a few days and to as much as an entire year, collecting data streams as they come. When receivers are recovered and downloaded, the resulting dataset tells researchers when a particular individual was within a specific area.

Research studies that focus on large, predatory species like sharks typically attract more media attention and public enthusiasm than other, less charismatic species. White Sharks (*Carcharodon carcharias*), in particular, appear to be one of the more popular species, as they are commonly addressed by news broadcasters and the entertainment industry (e.g., Jaws). In southern California, juvenile White Sharks show seasonal and sometimes annual fidelity to particular coastal beach habitats, meaning that individuals return to the same beaches year after year. These beaches are often coined ‘hotspots’ for juvenile White Sharks, and individuals may stay in these hotspot areas for up to a few months before moving to another region. Researchers hypothesize that these movements are primarily temperature driven, as juvenile White Sharks are not yet capable of fully controlling their internal body temperatures; the sharks move towards warmer waters when it becomes too cold. However, no studies have explored whether this hypothesis is correct, or how other environmental parameters may affect White Shark movement. By determining which parameters are most influential in White Shark movement, we may be able to predict when and where future hotspots will arise. Because juvenile White Sharks frequently use coastal beaches that are shared with beach-goers, predicting White Shark movements may help to reduce the probability of shark-human interactions by giving lifeguards advanced notice of when White Sharks are likely to be present. Therefore, the purposes of this project are 1) to determine which combinations of environmental conditions are most likely correlated with juvenile White Sharks presence in southern California, and 2) to predict the ranges and categories of conditions that are most likely to result in the presence of a White Shark.

In addition, although this project will be using White Shark specific data, the methods used in this project will be applicable to researchers who study similar species or who are trying to answer similar research questions. These results may also be useful to lifeguards and other

public safety personnel who benefit from advanced knowledge of potential threats within their jurisdictional region.

2. *Materials and Methods*

2.1 Data Sources

Passive acoustic telemetry data were collected by the California State University, Long Beach Shark Lab between 2012 and 2019 from regions as far north as Morro Bay, CA and as far south as San Diego, CA. However, receiver coverage during these times varied greatly on the spatial scale. The dataset provided for this project includes data from several juvenile White Sharks that were tagged opportunistically during the study period. Variables within this dataset include: DateTime, Latitude, Longitude, and Transmitter Code. A detection filter for each individual of two detections per day at a single receiver was run by the Shark Lab prior to providing the dataset. This was done in order to reduce the probability of false positives during a receiver's deployment. A log of acoustic receiver deployments was also provided. This file includes data on the receiver serial number, the latitude and longitude of deployment location, the time the receiver was put in the water, and the time the receiver was retrieved from the water. The geospatial range of receiver data was used later to gather environmental datasets.

A transmitter deployment log was also provided by the Shark Lab, which includes the following data: Transmitter Code, Species, Size at Tagging, Tagging Location (City), Tagging Date, Sex (M, F, or Unknown). For the purpose of this study, acoustic receiver data will only include the transmitter codes that are present in the transmitter deployment log (White Sharks only).

Daily environmental datasets for sea surface temperature[1], chlorophyll-A[2], sea surface salinity[3], and overall seafloor depth gradient[4] were gathered from NOAA CoastWatch's ERDDAP and were downloaded to a local server. Each of these environmental datasets have been hypothesized to influence the behavior of other marine species. Due to the large spatial extent of these datasets, data were saved in one-year increments. Moon phases were also calculated using the *pylunar* package.

2.2 Data Cleaning:

The receiver deployment log was used to parse the receiver data to create a new dataset that included animal detections only from receivers during times in which the individual receiver was confirmed to have been in the water. This reduces the bias of false detections, in which a receiver indicates an animal is present when it is not, and also reduces the inconsistencies that may be associated with acoustic receivers that were 'lost' and recovered in a new location. In addition, when acoustic receivers are initialized for the field, researchers have the option to provide a series of latitude/longitude coordinates for the receiver's location. However, these coordinates are often not the true location of the acoustic receiver, once deployed. Therefore, the acoustic detection data coordinates were replaced by the coordinates

---

[1] Multi-scale Ultra-high Resolution (MUR) SST Analysis fv04.1, Global, 0.01°, 2002-present, Daily

[2] Chlorophyll, NOAA VIIRS, Science Quality, Global, Level 3, 2012-present, Daily (for data between 2012 and 2014) and Chlorophyll a, North Pacific, NOAA VIIRS, 750m resolution, 2015-present (1 Day Composite) (for data post 2014)

[3] Sea Surface Salinity, Near Real Time, Miras SMOS Daily Composite (smosSSS3ScanDailyAggLoM), CoastWatch v6.62, 0.25°, 2010-present

[4] Estimated Seafloor Depth Gradients: ETOPO (US West Coast)

given for each receiver deployment in the receiver deployment log. These data cleaning techniques limited the amount of receiver data that were available for analyses to data between years 2014 to 2019.

Using ArcGIS, a 'fishnet' (grid) shapefile that partitioned the oceans within the southern California jurisdiction from ~ 10 m onshore to ~ 300 km offshore was created. Each individual grid cell (0.01 x 0.01 decimal degrees) was paired with an identification code. The shapefile was used to standardize and merge the environmental datasets with the shark detection data. The grid ID over which individual points were located was saved to each dataset. This allowed for the comparison of data across grid cells, as opposed to across individual latitude/longitude combinations. Receiver density was calculated across grid cells as the number of unique receiver stations deployed in each grid cell each day.

Several environmental datasets did not extend far enough into shallow waters to overlap with grid cells that contained receivers or shark detections. Therefore, missing data were filled using values from the nearest grid cell for that parameter during that day. In order to accomplish such filling, the data had to be ordered by date, latitude, and longitude. However, each individual dataset had its own unique set of latitude and longitude values. Therefore, the mean latitude and longitude for each zone were calculated as the average latitude and longitude across all geospatially-referenced datasets. Additionally, in order to calculate moon phase, latitude and longitude coordinates had to first be converted from decimal degrees to tuples of (degrees, minutes, seconds) and datetime had to first be converted to tuples of (year, month, day, hour, minute, second). The appropriate conversions were conducted before *pylunar* was used. Environmental values were calculated for all grid cells for each day between 2014 and 2019, regardless of if a shark was detected in that grid.

After initial data cleaning took place, the new dataset included the fields: Zone (grid cell), Date, Transmitter ID, weight, length, sex, tagging location, tagging cohort, receiver density, latitude, longitude, seafloor depth gradient, sea surface temperature, sea surface salinity, chlorophyll-A, and moon phase. NAs remained in the dataset for Transmitter ID, weight, length, sex, tagging location, tagging cohort to indicate a date and zone in which no sharks were present.

However, due to the high abundance of zeros in the dataset, a jackknifing technique was performed to produce a dataset with a relatively equal number of data points between times when sharks were present and when sharks were absent. Zone IDs in which sharks were detected each year were gathered and a list of nearby zones (± 10 zone IDs away) was generated. A random sample of data with no sharks detected (non-shark dataset) was then taken. The random sample only collected data from zones that were within the zone list that was generated. The size of random samples varied across years, but were equal to the number of rows in the dataset in which sharks were present. Sometimes, the non-shark dataset with zones within the zone list was not large enough to perform the random sampling. In this case, all data within the non-shark dataset with relevant zones were gathered, and the remaining rows of data were randomly grabbed from the non-shark dataset with non-relevant zones. Making sure the data were gathered across similar zones was important to reduce biases that may arise from comparing shark data that primarily occurs in nearshore waters with non-shark data that may extend several hundred miles offshore.

After jackknifing took place, the individual shark data was transformed into a dataset that featured a count of total individuals per grid cell per day. In order to determine whether the number of sharks in a particular area increased or decreased across seasons or years, month and year columns were generated. The final dataset used for analyses, therefore, included the following columns: Date, Month, Year, Zone ID, Number of Sharks, Receiver Density, Sea Surface Temperature, Sea Surface Salinity, Chlorophyll-A, Seafloor Depth Gradient, and Lunar Phase.

2.3 Visual Assessments:

The data were visually assessed for outliers and descriptive statistics were conducted. For continuous and numeric datasets, histograms were plotted and the mean and all four quartiles were calculated. For categorical data (lunar phase), a barplot was created and the counts in each category were calculated. For data that were associated with time (number of sharks detected, receiver density, Sea Surface Temperature, Sea Surface Salinity, Chlorophyll-A), time series plots were also created in order to see annual and seasonal trends.

2.4 Statistical Analyses:

The data were then assessed to determine how each predictor variable was related to the response variable. Instances in which no sharks were detected were removed from analyses in order to reduce biases associated with zero-inflated data. Receiver Density, Year, Lunar Phase, and Zone ID were converted to categorical values and ordinary least squares regression ANOVAs were conducted. The remaining datasets (sea surface temperature, sea surface salinity, chlorophyll-A, depth gradient) contained continuous data. However, initial visual assessments showed that the variation of the number of sharks detected at each predictor variable had high levels of variability. In order to maintain the integrity of the question at hand (what environmental conditions attract sharks to an area), each predictor variable was rounded to the nearest 0.1 (sea surface temperature, sea surface salinity, chlorophyll-A) or 0.001 (depth gradient), respectively. For each rounded value, the maximum number of sharks was kept. Correlation tests were run and linear regression lines were generated for each environmental parameter to determine whether there was a linear relationship between each environmental parameter and the maximum number of sharks per day. For relationships that did not appear linear, higher order polynomial functions were fit to the data.

3. Results and Discussion

3.1 Shark Density and Year

Between 2014 and 2019, the majority of shark detections occurred during 2015 and 2017; there were more than double the number of shark detections in these two years compared to all other years (Figure 1). Results from an ordinary least squares ANOVA indicate that there was a significant relationship between year and shark density. Years 2015, 2017, and 2018 had significantly higher shark densities when compared to 2014 ($p < 0.001$). However, 2016 exhibited statistically similar shark densities when compared to 2014 ($p = 0.153$) and 2019 had significantly lower shark densities compared to 2014 ($p = 0.028$). Perhaps years with higher densities corresponded to years with more receivers in the water or more ideal environmental conditions. Because 2019 is not yet over, seeing significantly smaller shark densities is not surprising.

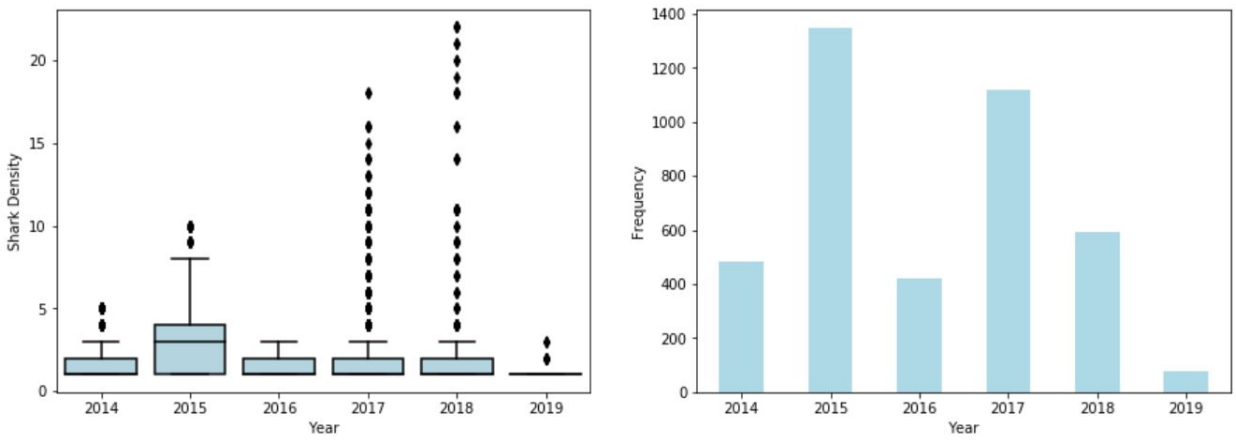

Figure 1. Left: Boxplot of shark density across years; Right: Frequency of shark detection data points across years.

3.2 Shark Density and Receiver Density

On most days, there were no receiver deployed in a particular zone and receiver density ranged across zones from 0 receivers to 3 receivers (Figure 2). Results from an ordinary least squares ANOVA indicate a significant relationship between shark density and receiver density. Having two receivers in the water in the same zone resulted in significantly higher shark density values, compared to results from when one or three receivers were in the water ($p < 0.001$). It makes sense that having two receivers in the water at the same time would increase the ability to detect a shark in a particular zone; two receivers cover a larger proportion of the entire zone region. However, under this logic, one would expect that three receivers would result in even higher shark densities, which is not supported by the data ($p = 0.625$). This can be explained by looking at the distribution of receiver density data. There are few instances in which three receivers are present in a single zone, which could explain why we do not see higher shark density values when three receivers are present.

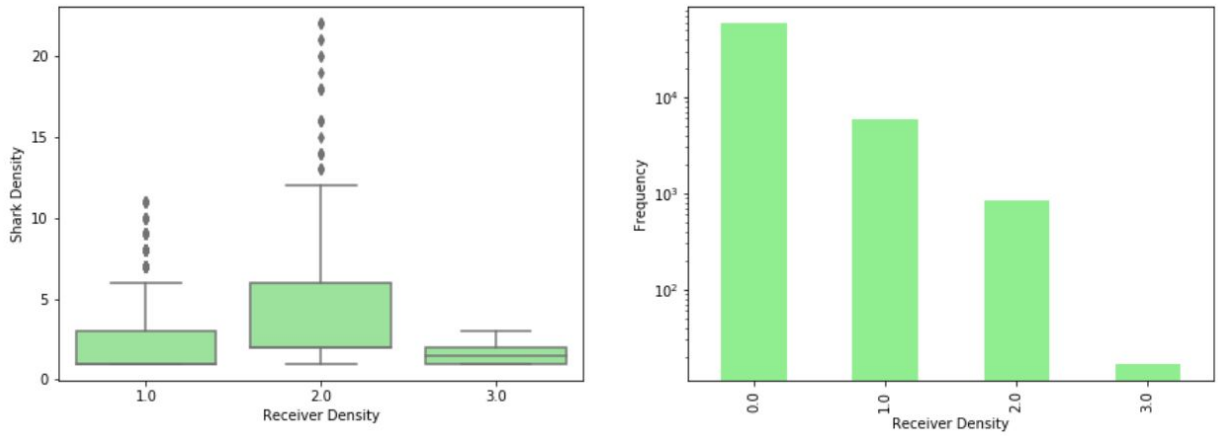

Figure 2. Left: Boxplot of shark density at different levels of receiver density; Right: Frequency of shark detection data points across different receiver densities.

3.3 Shark Density and Zone ID

A total of 50 zones were present in the final dataset that was used for analyses. Some of these zones appeared to have larger proportions of shark detection data, but several detections in a zone do not necessarily represent more sharks in that zone (Figure 3). Results from an ordinary least squares ANOVA indicated that at least 8 zones had significantly higher shark densities than the other 42 zones that were present in the model. These zones may represent nearshore 'hotspot' areas in which juvenile white sharks spend their time.

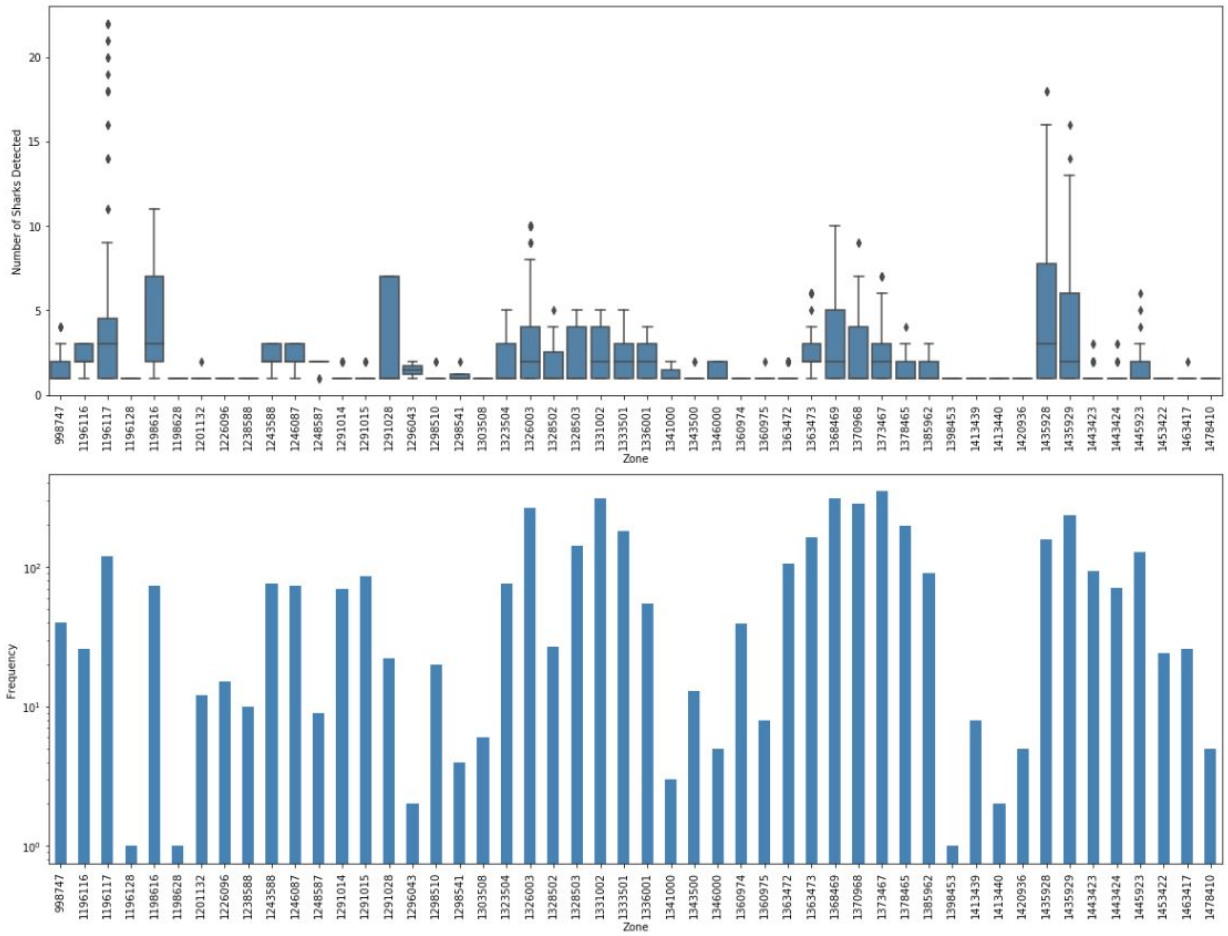

Figure 3. Top: Boxplot of shark density across zone ids (grid cells); Bottom: Frequency of shark detection data points across zones.

3.4 Shark Density and Sea Surface Temperature

Sea surface temperature ranged between 10.99ºC and 25ºC, with a median (± standard deviation) of 17.88ºC ± 2.66ºC. When using all data values, it was difficult to determine a clear relationship between sea surface temperature and shark density; the variation in shark density at different temperature values is quite large (Figure 4). Such variation can be described partially by looking at the relationship between sea surface temperature and shark density across years; in 2014, 2015, and 2016, shark densities are higher at ~ 16ºC, but in 2017 shark densities are higher around 18-22ºC, and in 2018 shark densities are higher at 12-14ºC (Figure 5). However, even when data are split up by year, the variation at each temperature value remains high. To get a clearer idea of the relationship between sea surface temperature and

shark density, the sea surface temperatures were rounded to the nearest 0.1 degree and the maximum shark density that was observed in that temperature range was kept. This approach allowed for the determination of specific temperatures that may be used more frequently by several individuals. Using this distribution, the data do not appear to have a linear relationship, which was supported by the results of a correlation test (correlation coefficient = -0.1938, p = 0.030, Figure 6). It did appear that there were two peaks in the dataset, where shark densities were higher around 12-14ºC and 18-22ºC. Therefore, I used bootstrapping (sampling with replacement) techniques to fit a fourth degree polynomial to the dataset. I ran the bootstrap 100 times to get an idea of how missing data values could influence the resulting curve (Figure 6). I also ran the polynomial fit on the entire dataset. The resulting formula from the model that used all data values was:

$$y = -0.013x^4 + 0.93x^3 - 24.5x^2 + 280x - 1169$$

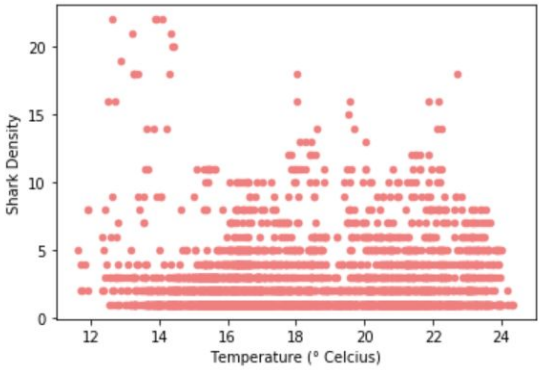

Figure 4. The relationship between shark density and different temperatures for all data points.

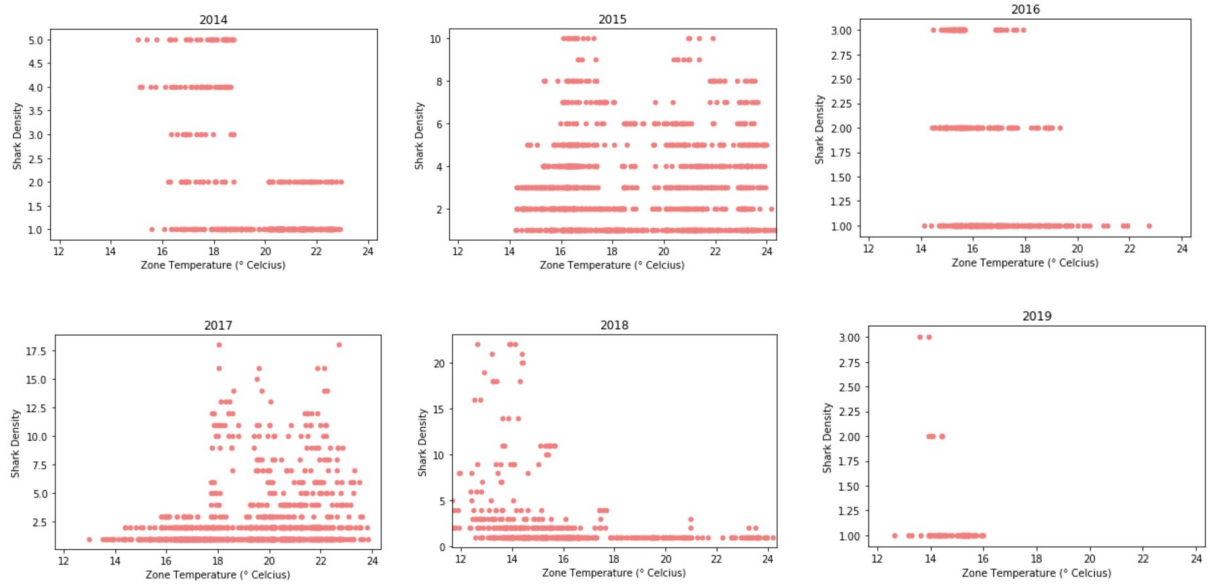

Figure 5. The relationship between shark density and different temperatures for all data points across each year.

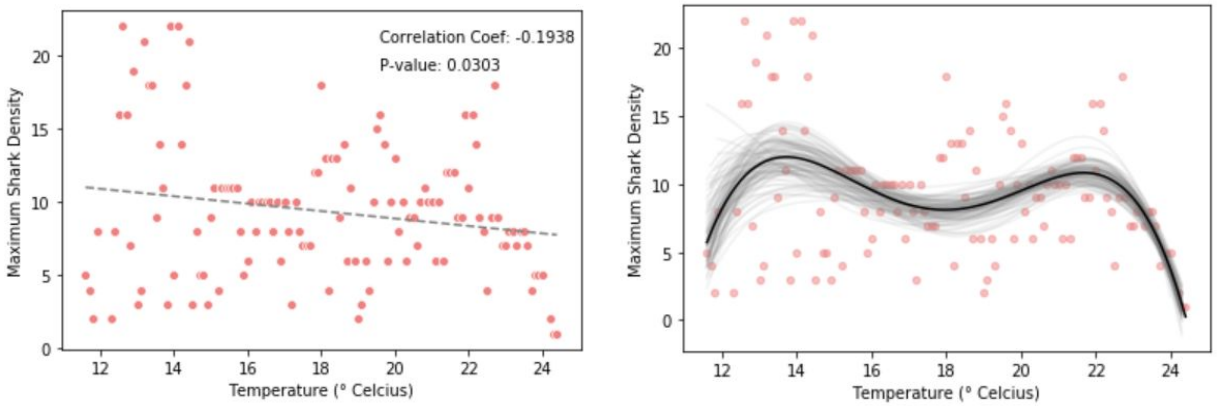

Figure 6. Left: The maximum value for shark density at each 0.1 degree of temperature, with a fitted linear regression line (gray dashed line). Right: The maximum value for shark density at each 0.1 degree of temperature, fitted with 4th degree polynomial functions (100 iterations using bootstrapped samples - gray lines; all data - black line).

3.5 Maximum Shark Density and Sea Surface Salinity

Sea surface salinity ranged between 30.00 PSU and 39.44 PSU, with a median (± standard deviation) of 33.27 ± 1.34 PSU (Figure 7). There was no linear relationship between salinity and maximum shark density (correlation coefficient = -0.2736, p = 0.012, Figure 8) and there appeared to be a peak in shark density at 33-35 PSU. These data were fit to a third degree polynomial and models were run for 100 bootstrapped samples and for the entire dataset (Figure 8). The resulting formula from the model that used all data values was:

$$y = 0.14x^3 - 15x^2 + 531x - 6245$$

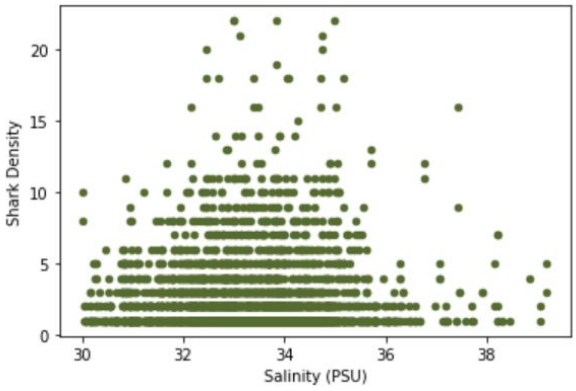

Figure 7. The relationship between shark density and different salinities for all data points.

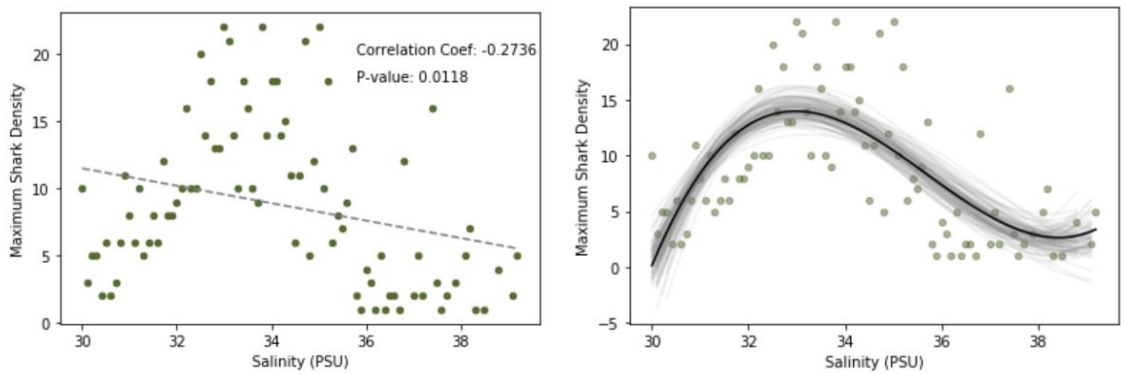

Figure 8. Left: The maximum value for shark density at each 0.1 value of salinity, with a fitted linear regression line (gray dashed line). Right: The maximum value for shark density at each 0.1 value of salinity, fitted with 3rd degree polynomial functions (100 iterations using bootstrapped samples - gray lines; all data - black line).

3.6 Maximum Shark Density and Chlorophyll-A

Chlorophyll-A values ranged between 2.269e-13 and 359.254 mg/m^3, with a median (± standard deviation) of 0.3171 ± 7.383 mg/m^3. Upon visual inspection of the data, there appeared to be only one instance in which shark density exceeded 0 at high chlorophyll levels (> 50 mg/m^3, Figure 9). Therefore, it appears that as long as chlorophyll levels are low, sharks are present. I also calculated the maximum shark density at each 0.1 value of chlorophyll-A. However, there appears to be no correlation between these two values (correlation coefficient = -0.1458, p = 0.203, Figure 10). After conducting a log transformation on shark density and chlorophyll-A values, however, a strong, negative correlation was present (correlation coefficient = -0.6036, p = 0.0, Figure 10). This suggests that as chlorophyll-a values increase logarithmically, maximum shark density values decrease logarithmically.

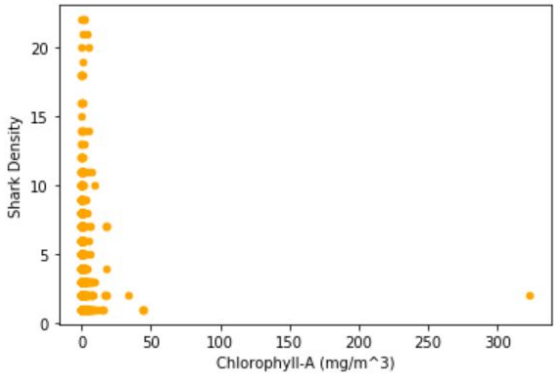

Figure 9. The relationship between shark density and different chlorophyll-A values for all data points.

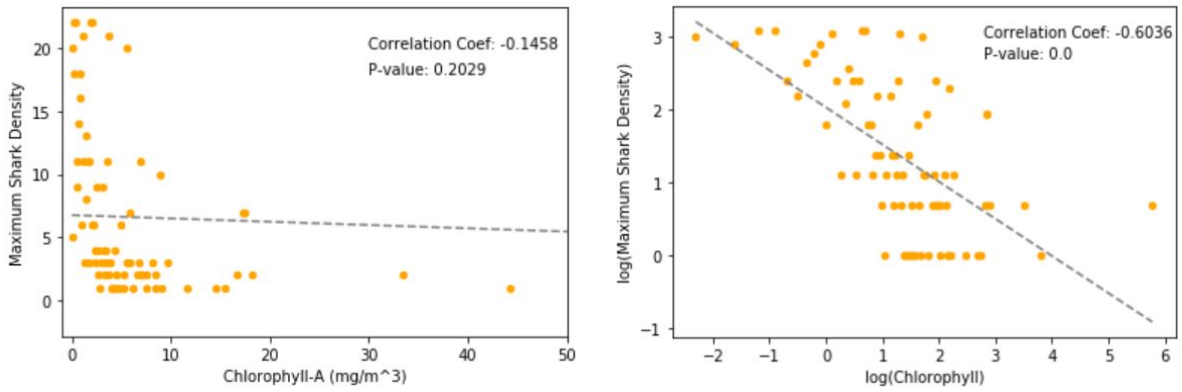

Figure 10. Left: The maximum value for shark density at each 0.1 value of chlorophyll-A, with a fitted linear regression line (gray dashed line). Right: The log-transformed maximum value for shark density at each 0.1 value of chlorophyll-A (also log-transformed), fitted with a linear regression line (gray dashed line).

3.7 Maximum Shark Density and Depth Gradient

Depth gradient values ranged between 0.0009 m and 0.023 m with a median (± standard deviation) of 0.007 ± 0.005 m. Shark density was highly variable across all values of depth gradient (Figure 11), so the maximum shark density per 0.001 was calculated. However, because depth gradient values occurred between a narrow range (0 m to 0.025 m), there was no linear relationship between maximum shark density and depth gradient (correlation coefficient = 0.359, p = 0.157, Figure 12). These results did not change when maximum shark density values were log-transformed (correlation coefficient = 0.282, p = 0.273, Figure 12).

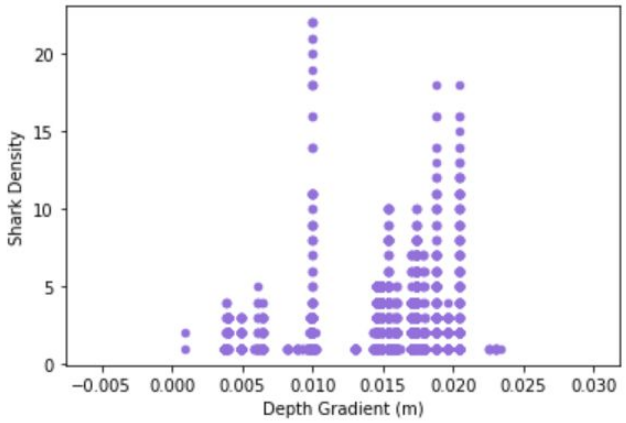

Figure 11. The relationship between shark density and different depth gradients for all data points.

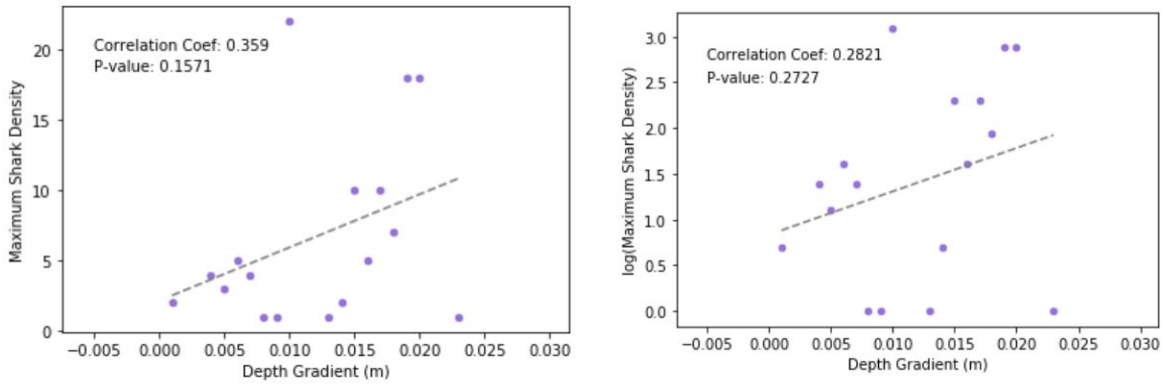

Figure 12. Left: The maximum value for shark density at each 0.001 value of depth gradient, with a fitted linear regression line (gray dashed line). Right: The log-transformed maximum value for shark density at each 0.001 value of depth gradient, fitted with a linear regression line (gray dashed line).

3.8 Shark Density and Lunar Phase

Most shark detection data were present during the waxing or waning crescent and the waxing or waning gibbous. However, several rows of data (~200-400) were present during the new moon, first quarter, full moon, and last quarter (Figure 13). Results from an ordinary least squares ANOVA indicated that there was not a significant difference in shark density across different lunar phases ($p > 0.05$).

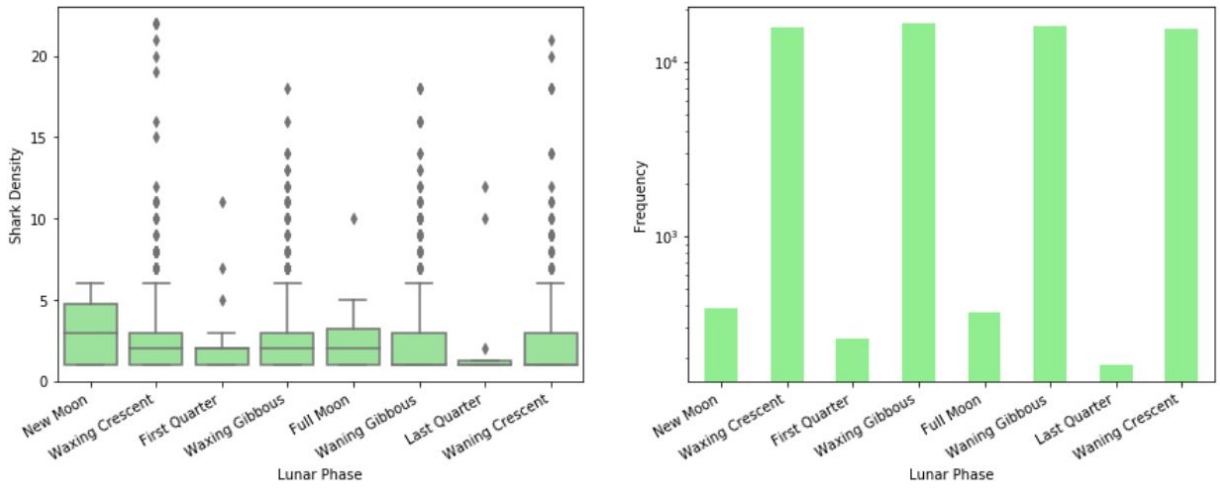

Figure 13. Left: Boxplot of shark density at different lunar phases; Right: Frequency of shark detection data points across different lunar phases.

4. Supplementary Materials

Github repository can be found here: https://github.com/echelleburns/SpringBoard-Capstone1
Presentation Slides can be found here:
https://drive.google.com/file/d/15ecWr9sxPtVIffaKJYAJ--GqMpEduznI/view?usp=sharing