**Abstract:**

**Introduction:**

**Materials & Methods:**

The Manatee Rescue and Mortality Response team is very thorough. Every manatee that is reported injured or dead is recovered. Trained biologist throughout Florida assess the necropsy of deceased manatees, record the size, determine the sex, record the date retrieved and document location information surrounding the manatees regardless of the state they are in when retrieved. The resulting data of this collection, sex, size, location information, and dates are used in this analysis.

There were three location classes reflecting the location of death within the data: the waterway, county, and city that the manatees were found in. This analysis focusses on determining how the environment surrounding deceased manatees influence deaths by watercraft collision however, due to restrictions in the data a uniform classification for location was created. The limitation in using the waterway class fell in the frequency of deaths in the waterways. There were 838 unique waterways that manatees were found in throughout Florida. Of those 838 waterways 506 out of 838 of them had less than or equal to two manatees that were retrieved in their waters and 655 out of 838 of the waterways had less than or equal to five manatees that were retrieved in their waterways[[1]](#footnote-1). Given that the original dataset contained 9 classifications for cause of death and 20 percent of the data results in collisions it was determined that there were not enough data points per waterway to learn much from these waterways. Given that the lack of records occurred in more than 75 percent of the waterways using waterways as a metric was omitted. Similar results were found in city where 326 out of 559, over 58 percent, cities had a less than or equal to five manatees retrieved within the city limits. City as a result was also omitted. In order to gather location information, the location classes were reduced to a single classification, “Region”, which indicates whether the manatee was found in Northwest, Northeast, Southwest, or Southeast Florida. The regions are generated based on the location of the county in Florida that a given deceased manatee was retrieved in.

The dataset includes some information regarding the anatomy of the manatees. The size, given in centimeters, refers to the total body length of the manatee from its head to its tail’s end. The data contained some missing data entries for the size of the manatees. Records with missing size information were omitted reducing the size of the dataset. Size was included in the analysis to examine potential relationships between size and deaths by watercraft collision. The manatees sex is also provided. The sex is broken down into three categories: male, female and undetermined. Manatees of an undetermined sex were removed from the analysis as it was determined harder to highlight differences in death by watercraft collision to sex if the sex is not known. The combination of these omissions reduced the dataset to 9,507 total data points from 10,057.

The date that every manatee was retrieved was recorded. It was desired to extrapolate some meaning from this. The dates were taken and mapped into seasons: winter, spring, summer, and fall and included in the analysis.

**Generalized Linear Modeling**

The data contains three categorical explanatory variables: sex, region, and season as well as a single continuous variable size(cm). The response that was studied is binomial capturing whether or not a manatee was killed from watercraft collision (T/F). Generalized linear modeling was used to analyze possible effects of independent variables, that is, anatomical (sex and size), and environmental (season and region), and their interactions among one another with death by watercraft collision. The maximal model including all of the anatomical and environmental factors and their interactions was reduced using two approaches: Step AIC and Dredging.

Approach 1: Step AIC

Using the Step functionality in R the Step AIC function was used to reduce the maximal model. Step continually adds and removes elements of the model until the Akaike Information Criteria(AIC) is reduced. AIC is a measurement for the fittingness of a model to the data. The model was reduced to include size, region, sex and season with 6 interactions: Size and sex; region and sex; size and season; region and season; sex and season; size, sex and season.

Model 1: Maximal Model

Model 2: Region + Season + Sex + Size + Sex:Size + Region:Season + Region:Sex + Season:Sex + Size:Season + Sizecm:Sex:Season

*Note: “:” Denotes interaction.*



**Analysis of Deviance Table** for Model 1 and Model 2.

There is sufficient evidence that there is not a significant difference between Model 1, the AIC Step reduced model, and Model 2, the maximal model. Therefore, model 1, the simpler model was chosen for the analysis.

Approach 2: Dredging

The Step AIC method takes a greedy approach to minimizing the AIC by continually adding and removing elements from the model until the AIC is reduced. This creates a reduced model but not necessarily the best possible model. There is an R library package called MuMIn that contains a function called Dredge which takes the maximal model and then breaks it apart into all possible combinations of models with interactions and independence. It then ranks the models based on the AIC of each of the models reporting them in order from lowest to highest AIC. This second approach served as validation for the model. The result of the dredge technique produced a model that was two steps simpler than the model reached in the Step AIC function containing interactions that were all present in both Dredge and Step AIC. The final minimal model from the Dredge function was size, region, sex and season with 4 interactions: region and sex, region and season, season and sex, and sex with size. Effectively removing the interaction of size and season and sex size and season.

Selection:

Given that Dredge and Step AIC resulted in very similar models with Dredge being two parameters simpler the model produced by Dredge was used to derive the relationship between the response, death by watercraft collision, and the explanatory variables.

Model 1: Maximal Model

Model 3: Region + Season + Sex + Size + Region:Season + Region:Sex + Season:Sex + Sex:Size

*Note: “:” Denotes interaction.*



**Analysis of Deviance Table** for Model 1 and Model 3.

There is sufficient evidence that there is not a significant difference between Model 3, the dredge reduced model, and Model 2, the Maximal model . Therefore, model 3, the simpler model was chosen for the analysis.

Model 1(Maximal):

Model 2(Step AIC): Region + Season + Sex + Size + Sex:Size + Region:Season + Region:Sex + Season:Sex + Size:Season + Sizecm:Sex:Season

Model 3(Dredge): Region + Season + Sex + Size + Region:Season + Region:Sex + Season:Sex + Sex:Size

Model 4(Intermediary Step 1): Region + Season + Sex + Size + Sex:Size + Region:Season + Region:Sex + Season:Sex + Sizecm:Sex:Season

Model 5(Intermediary Step 2): Region + Season + Sex + Size + Sex:Size + Region:Season + Region:Sex + Season:Sex + Size:Season

*Note: “:” Denotes interaction.*



**Analysis of Deviance Table** for Model 1 (Maximal), Model 2 (Step AIC), Model 3(Dredge), Model 4 (Model one step simpler than Step AIC Model), Model 5 (Alternate Model one Step simpler than Step AIC model).

There is sufficient evidence that there is no significant difference between the simplest model, model 3(Dredge) and more complex models. The Analysis of Deviance table shown above shows a breakdown from the Maximal model to Model 2 (Step AIC). Models 4 and 5 are formed from one of the two terms present in Step AIC, Size:Sex:Season and Size:Season (“:” Denotes interaction).

**Non-Metric Multidimensional Scaling**

In order to visualize the differences in the manatees that died from collision versus those that did not a Non-Metric Multidimensional Scaling (NDMS) of the data was performed. Two populations were identified. The population of manatees that died by watercraft collision and those that did not. Given that there were three categorical explanatory variables that were being examined they were converted to numerical factors from 1 to n, where n is the number of unique classifications per categorical explanatory variable. The Vegan package in R was used to generate an ordination of the categorical and continuous explanatory variables. The results of the ordination were plotted and color coordinated by Collision(T/F). It was expected that if the manatees that died from collision are different than the manatees that do not there would be some clustering in the NDMS that distinguish the Collision population from the Non-Collision population.

Given the large dataset, 9507 points, a NDMS could not be generated. In order to generate an approximation of the collision and non-collision populations 500 data points of manatees that died from collision and 500 data points of manatees that did not were randomly sampled and used in the multidimensional scaling. A few iterations of this random sampling based ordination were carried through to get the best spread of visualizing these populations to help determine if they are distinct.



Figure: Non-Metric Multidimensional Scaling. Red: Collision Population. Blue: Non Collision Population.

To view more iterations of the NDMS generated through random sampling view Appendix C.

**Results:**

The study used the data collected by the Florida Fish and Wildlife Commission from 1974-2015. The dataset contained 10,057 entries which were reduced to 9,507 after pruning the data. The data showed that a large percentage of manatees were experiencing death by collision and the Generalized Linear Model and NDMS suggest that there is not much of a relationship between

The results of the GLM modeling tell us that none of the explanatory variables are significant on their own. The selected model had interactions with region, season, sex and size with at least one of the other explanatory variables which meant that the independent variables could not be used to determine the overall relationship between them and the response variable. A manatee found in the Southeast Region has a log likelihood of 1.16 +/- 0.32 (1 S.E) with respect to the intercept(Region Northeast) which shows a significant comparison. There was also a significant relationship shown for the Winter Season of -1.24 +/- 0.31 that cannot be interpreted on its own. A table detailing the full estimated log likelihoods of the selected model are detailed in appendix D.

**Discussion:**

**Acknowledgments:**

**Literature Cited**

**Appendix A (Full Dredge Table):**

**Appendix B ():**

1. Refer to Appendix B for Frequency Distributions of Deceased Manatees by Waterway. [↑](#footnote-ref-1)