University of South Florida

DIGITAL COMMONS
@ UNIVERSITY OF SOUTH FLORIDA

Digital Commons @ University of South Florida

USF Tampa Graduate Theses and Dissertations

USF Graduate Theses and Dissertations

November 2018

Quantifying the Probability of Lethal Injury to Florida Manatees Given Characteristics of Collision Events.

B. Lynn Combs
University of South Florida, blynncombs@gmail.com

Follow this and additional works at: https://digitalcommons.usf.edu/etd

Part of the Other Oceanography and Atmospheric Sciences and Meteorology Commons

Scholar Commons Citation

Combs, B. Lynn, "Quantifying the Probability of Lethal Injury to Florida Manatees Given Characteristics of Collision Events." (2018). *USF Tampa Graduate Theses and Dissertations.*
https://digitalcommons.usf.edu/etd/7488

This Thesis is brought to you for free and open access by the USF Graduate Theses and Dissertations at Digital Commons @ University of South Florida. It has been accepted for inclusion in USF Tampa Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact digitalcommons@usf.edu.

Quantifying the Probability of Lethal Injury to Florida Manatees

Given Characteristics of Collision Events.

by

B. Lynn Combs

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Oceanography
with a concentration in Marine Resource Assessment
College of Marine Science
University of South Florida

Co-Major Professor: Cameron Ainsworth, Ph.D.
Co-Major Professor: Julien Martin, Ph.D.
Ernst Peebles, Ph.D.

Date of Approval:
September 13, 2018

Keywords: citizen science, Florida manatee, marine mammals, Markov Chain Monte Carlo, protection zones, wildlife collisions

Copyright © 2018, B. Lynn Combs

DEDICATION

I dedicate this work to public school teachers who make science accessible.

ACKNOWLEDGMENTS

I extend the deepest gratitude to all the biologists and managers I collaborated with at Florida Fish and Wildlife Research Institute on this project, including: Tim Gowan, Jeff Hostetler for the diligent and determined statistical brainstorming sessions; Chip Deutsch, Holly Edwards, Leslie Ward-Geiger for their insights and contributions; Scott Calleson and Brandon Bassett for sorting through piles of collision reports and for sharing the data; and to Jay Gorzelany et al., for sharing his radar-gun dataset. I am grateful to the Florida Fish and Wildlife Research Institute Marine Mammal and Turtle research groups for their tireless dedication to conservation – this institution challenged me and fostered my growth as a researcher. I thank my committee, especially my two co-advisors, Drs. Julien Martin and Cameron Ainsworth, for their mentorship. I am honored to have been chosen to work on this project and consider this the most important opportunity of my life. Many thanks to my committee member Dr. Ernst Peebles, whose wisdom and expertise on all things Tampa Bay were incredibly appreciated. I have my family to thank for feverishly believing in me, my mother Cheryl and father Ray for their love and sacrifice, my siblings Courtney, Harry, and Hannah for constantly propelling me forward with encouragement. Thanks to my partner, Timmy Hintze, for his love, support, and hours of discussion that fueled creativity and stamina for this work. Finally, I thank all of the lab members, classmates, colleagues and friends that I have had the pleasure to meet through this endeavor and who's names I couldn't list. The dedication, intelligence, passion, and encouragement that comes so natural to many of these people are the kinds of things that shape the foundation of a person. This study was funded by the Florida Fish and Wildlife Conservation Commission and the Florida Sea Grant Program. This study was supported by the National Sea Grant College Program of the USA. Department of Commerce's National Oceanic and Atmospheric Administration (NOAA), Grant No. NA 14OAR4170108, and by the Save the Manatee Trust Fund. Work was conducted under USFWS Federal research permit #MA773494. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Wherever wildlife share space with boaters, collisions are a potential source of mortality. Establishing protection and speed zones are the primary actions taken to mitigate collision risk. However, creation of protection zones may be a point of contention with stakeholders as new zones can have significant socioeconomic impacts. The Florida Manatee is a prime example of a species whose abundance and viability are constrained by this balance between the needs of humans and wildlife on a shared landscape. The goal of this work is to help further understand the risk to manatees by quantifying the probability of lethal collisions. I hypothesized that higher boat speeds increase the probability of lethal injury to manatee during a collision and aimed to characterize the relationship between vessel speed and the probability of lethal injury to manatee. I used a logistic regression model implemented with a Bayesian approach and fitted to citizen reported collision data as a feasible method for examining the influence of vessel speed in contributing to lethal injury to a manatee when struck. I also present a method for dealing with uncertainty in data used to report collisions. To conduct this analysis, I used citizen reported collision data. These data are typically collected opportunistically, suffer from low sample sizes, and have uncertainty in reported vessel speeds. Uncertainty associated with speed values in reported collision events was assessed by performing a multiple imputation to replace qualitative vessel speed – in other words, "missing data" – with quantitative values. This procedure involves fitting log-normal distributions to radar data that contained precise vessel speeds along with a physical description like 'planing', 'plowing', or 'idle'. For each imputation of the data, a quantitative value was selected randomly from that distribution and used in place of its initial corresponding speed category. I evaluated issues related to quasi-

separation and model fit using simulated data sets to explore the importance of sample size and evaluated the effect of key sources of error. The prediction that greater strike speed increases the probability of lethal injury was supported by the data that I analyze.

CHAPTER 1:

INTRODUCTION

Wildlife-vehicle collisions affect the populations of a large range of taxa. Collisions with cars and terrestrial species have been well documented (Forman and Alexander, 1998) and are a large source of mortality for many mammals (Allen and McCullough, 1976; Hell et al., 2005), birds (Hell et al., 2005), reptiles (Langen et al., 2009), amphibians (Puky, 2006; Langen et al., 2009) and insects (Rao and Girish, 2007). Many large marine animals suffer from strikes from both commercial and recreational watercraft. This includes sirenians, (Aipanjiguly et al., 2003; Calleson & Frohlich, 2007; Maitland et al., 2006) including Dugongidae and Trichechidae (manatees), Cetacea, including North Atlantic right whales (*Eubalaena glacialis*) (Kraus, 1990; Ward-Geiger et al., 2005; Fonnesbeck et al., 2008; Vanderlaan et al., 2008), fin whales (*Balanenoptera physalus*), humpback whales (*Megaptera novaeangliae*), sperm whales (*Physeter macrocephalus*), grey whales (*Eschnichtus robustus*) (Laist et al., 2001), and some Delphinidae (Wells & Scott, 1997; Stone & Yoshinaga, 2000); as well as green turtles (*Chelonia mydas*) (Hazel et al., 2007). In the case of the Florida manatee (*Trichechus manatus latirostris*), collisions with vessels are a primary source of mortality (Runge et al., 2007).

The creation of protection zones that regulate the speed and operation of vessels can reduce wildlife-vehicle collisions and decrease the resulting impact on wildlife populations (Allen and McCullough, 1976; Calleson & Frohlich, 2007; Hazel et al., 2007; Martin et al. 2016). Laist and Shaw (2006), and Calleson and Frohlich (2007) are among the authors that have argued that slower vessel speeds make up an important component of risk mitigation for vessel strikes of manatees. In fact, the creation of protection zones to regulate speed and operation of vessels is considered a primary management action to protect marine mammals (USFWS, 2001; Calleson and Frohlich, 2007; FWC, 2007; Fonnesbeck et al., 2008; Bauduin et al., 2013; Martin et al. 2016).

However, the potential impacts on waterway users caused by such management actions can present a point of contention as they may have socioeconomic impacts (Aipanjiguly *et al.*, 2003). Speed-limited zones, restrictions on boat access, restrictions on dock construction, and locations of boat ramps that increase boater access to manatee habitat are among some things concerning vessel operation that have been challenged in court (Marsh et al., 2011). Despite this controversy, the total area in which vessel speeds are regulated for manatee protection in Florida is a small fraction of the total thousands of kilometers of available waterways (FWC, 2007). In addition, the general boating public favors speed zones for manatee protection (Aipanjiguly et al., 2003). To meet the needs of humans and wildlife on a shared landscape, it is important to identify optimal management policies that balance multiple objectives: e.g., protecting marine mammals while minimizing impacts on waterway users (Martin et al. 2016; Udell 2016; Udell et al. in review). For manatees, it has been proposed that limits on vessel speed in high traffic areas reduces the risk of lethal collisions by allowing the manatee and the vessel a greater amount of time for reaction. It has also been proposed that reducing speed decreases the severity of injuries if a collision does occur (Calleson and Frohlich, 2007). The following work addresses the latter.

Although there is a general understanding that the risk of injury to a manatee increases with increasing vessel speeds, estimates of the probability of death at speed have not been made. Here I estimated the probability of lethal injury to manatees as a function of vessel speed at the time of collision. I propose this to be a first step towards understanding the effectiveness of these protection zones. In order for credible management decisions regarding speed zones to be made, it is necessary to understand the precise characteristics of a collision that lead to lethal injury in manatees. Martin et al. (2016) described a modeling framework known as a Bayesian Belief Network (BBN) which links regulation of vessel speed to marine mammal mortality. This network incorporates uncertainty from encounter, collision, and mortality rates (Fig. 1). Given information about the probability of lethal injury given collision speed, it may be possible to optimize the design of speed zones in order to minimize collision risk while considering the burden these impose on vessel operators and waterway users (Martin et al. 2016; Udell 2016). This analysis should hopefully fill an important gap and ultimately help improve the management of vessel regulations for manatees and other species affected by vessel collisions. While quantifying lethal injury speed is essential for management decisions, it is equally necessary to effectively describe the uncertainty surrounding the collision event (i.e. vessel speed), which is intrinsic given the current reporting system. I predict that vessel speed increases the probability of lethal injury to manatees during collisions with vessels. I examined the effect of vessel speed on contributing to either a lethal or non-lethal injury to Florida manatees using a logistic regression model implemented with a Bayesian approach. Vanderlaan and Taggart (2007) and Conn and Silber (2013) also used a logistic regression approach to estimate the probability of death given strike speed of the North Atlantic right whale, but they did not account for uncertainty about strike speed. When combined with other data, results of this analysis can be used to derive the number of potential deadly collisions between manatees and vessels, and the potential impact on the manatee populations (Martin *et al.* 2016).

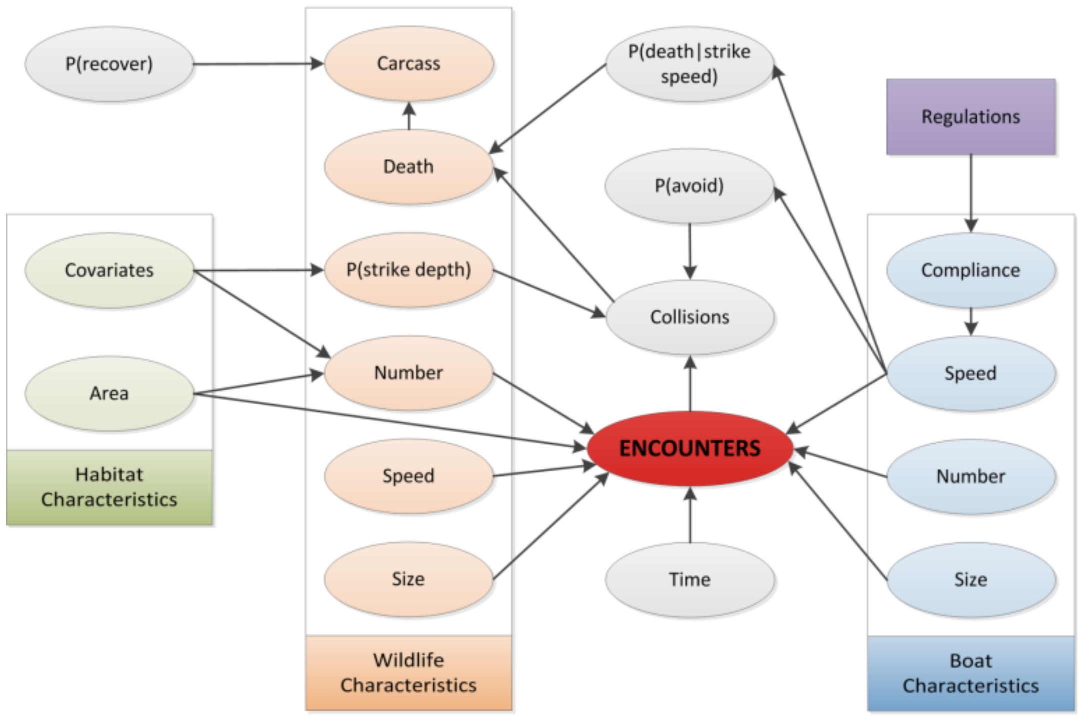

Figure 1: Bayesian Belief Network (BBN) which links regulation of vessel speed to marine mammal mortality from Martin et al. 2016.

To conduct this analysis, I used citizen reported vessel collision information. Human interactions with wildlife are a regular occurrence, and it is possible to exploit the tendency for the public to report collisions with species of concern, such as the threatened Florida manatee. For instance, McClintock et al. (2015) describe a novel methodology using public reports of collisions between Florida panthers and vehicles along with routine telemetry monitoring data to produce the first defensible population estimates for the Florida panther. Carcass recovery models have also been used to estimate survival probability (Brownie *et al.*, 1985) and mortality rate (Bellan *et al.*, 2013). These types of data are underutilized in ecology but contain valuable information that can be extracted. Citizen science, with its "many eyes", is an invaluable resource for answering ecological and conservation questions (recent examples in Dickinson *et al.*, 2012). Citizens can log rare events and obtain data that would not be recorded otherwise. Along with the many benefits, citizen science presents analysis-related challenges (e.g. sampling bias, observer variability, and detection probability) that are not easily addressed with statistical hypothesis testing or model selection approaches (Weir et al. 2005). Manatee collision events are typically collected as opportunistic reports, suffering from low sample sizes and uncertainty in reported traveling speeds. For these types of data, the challenge lies in estimating the uncertainty associated with the reported characteristics of the event, particularly vessel speed. In order to address uncertainty and low sample size, I used simulated data sets to explore the importance of sample size and evaluated the effect of key sources of error. Uncertainty for collision events with speeds reported as qualitative descriptors was assessed by performing a multiple imputation (MI) approach. This was conducted by fitting log-normal distributions to ancillary radar speed gun data with a corresponding category ('planing', 'plowing', or 'idle' speed), iteratively drawing speed values from these distributions for each collision event and re-fitting the regression model. I will describe MI more in depth in the following paragraphs. I also evaluated issues related to quasi-separation, an issue affecting binomial data of small sample sizes, and model fit.

CHAPTER 2:

METHODS

About Citizen Reported Collision Data

I used a collection of observational records detailing vessel-manatee collisions occurring throughout Florida that were reported by either the operator of the vessel or by a witness. These records cover a time span from 1978 to 2014. I excluded records where the vessel was greater than 64 feet. This allowed for exclusion of large barges, work boats, and other heavy vessels where lethality can be attributed to sheer vessel mass acting on the manatee at slow speeds. Similarly, tugboats were excluded due to large propeller size.

Vessel-manatee collision reports were originally classified as: 'known vessel lethal collision events', 'suspected vessel lethal collision events', and 'uncertain collision events' (see Table 1 for detailed definitions of each category). In the first category, lethal collision events with a known vessel (n = 22), collisions were reported by the vessel operator, a carcass was matched to the event, or the collision was reported by a witness with a high level of confidence. These records describe a lethal collision and the associated vessel with the highest level of certainty. While in most cases, a lethal collision report describes certain death at the time of the collision, a few of these cases describe an instance where the manatee was severely injured and died later. For this analysis, I excluded vessels that were only suspected, but not known for certain, to be involved in lethal collision (n=8) in order to reduce the level of uncertainty in the details about the involved vessel (e.g. speed, size, characteristics, etc.). Of the initial reports, 141 subset records were considered for the analysis described hereafter.

Table 1: Original collision report data classification. Suspected vessel, lethal collision category was excluded from this analysis.

Collision type	n	Definition
Known vessel, lethal collision	22	collisions were reported by the vessel operator, a carcass was matched to the event, or the collision was reported by a witness with a high level of confidence; records describe events with a high level of certainty that a collision with the associated vessel resulted in a lethal injury to the manatee
Suspected vessel, lethal collision	8	collisions were reported predominantly by law enforcement; records describe events with less certainty that a collision with the associated vessel resulted in a lethal injury to the manatee
Uncertain collisions	119	collision reports contain varying degrees of missing information and were reclassified using expert opinion
Total =	**149**	

Addressing uncertain collision outcomes

To address uncertain collisions, I recruited the help of two experts in manatee biology who are particularly familiar with the data set of recovered carcasses linked to watercraft collisions, S. Calleson (US Fish and Wildlife Service, Jacksonville FL) and B. Bassett (Florida Wildlife Research

Institute, St. Petersburg FL). They worked independently to reclassify the 119 uncertain collision records into 6 categories (called ID) based on the original event reports (Table 2). The main purpose of this was to identify records that were unlikely to have resulted in a lethal injury, letting us define a negative outcome. Since this type of probability analysis requires a binary response variable (lethal/not lethal) to proceed, it was necessary to have a set of "not lethal" or "severe injury unlikely" records, backed by good evidence that the manatee survived. In some cases, severe injury was categorized by open, bleeding wounds or blood in the water which was based on an actual observation of the animal (Table 2 category 6). However, in most cases, severe injury was determined as the outcome based on descriptions in the report (Table 2 category 5). The same was true for the cases where severe injury was 'unlikely' (Table 2 categories 4 and 3, respectively). In either "severe injury likely/unlikely" cases, a manatee observation constitutes a more certain outcome. While the advice on the matter from each expert was considered useful, one of datasets resulting from classification 1 reclassification experienced complete separation (this is discussed more in a following section). Therefore, for the purpose of this analysis, I utilized only one of their reclassification efforts (classification 2, Table 2). For the purposes of the logistic model, a positive outcome (a fatality) was based on 'known vessel, acute lethal collision' (n=22, Table 1). A negative outcome was based on 'Severe injuries unlikely observational' (n=5, Table 2), which relies on an observation of the manatee. These cases are the best possible representations of lethal/non-lethal incident. Note, however, that even in the case of 'Severe injuries unlikely observational', there is uncertainty about the ultimate outcome. Indeed, it is possible that the animal was classified in that category when in fact the animal suffered a deadly injury that was invisible to the observer. Nevertheless, the tradeoff for a smaller sample size is a more reliable dataset. Since the goal of this work is to demonstrate the effectiveness of MI for estimating cases where there was no data for the speed of the vessel, this is an appropriate trade off to make.

ID	Name	n class. 1	n class. 2	Definition
1	Collision unknown	10	22	vessel-manatee collision uncertain (no further assessment)
2	Injuries unknown	20	20	vessel-manatee collision likely but injuries unknown
3	Circumstantial: severe injuries unlikely	31	20	severe injuries unlikely based on vessel speed and/or other described circumstances
4	Observational: severe injuries unlikely	6	5	severe injuries unlikely based on manatee observation
5	Circumstantial: severe injuries likely	48	43	severe injuries likely based on vessel speed and/or other described circumstances
6	Observational: severe injuries likely	4	4	severe injuries likely based on manatee observation

Table 2: Categories resulting from reclassification of uncertain collision events. Uncertain collision events are represented by classification (class) 1 and 2, unlikely severe injury categories 3 and 4, and severe injury categories 5 and 6. Classification 2 sums to 114 and not the total number of "uncertain" records (N=119) because five were excluded by the reviewer.

Statistical model

I employed a logistic regression approach to model the probability of manatees suffering lethal injury during a collision as a function of boat speed. The response predicts probability of a lethal outcome given vessel speed at collision $P(\varphi \mid x)$, where x is vessel speed. The outcome of the collision event (denoted c_i, where $c_i = 1$ if lethal) was modeled as a Bernoulli response variable:

(Eq. 1) $c_i \sim \text{Bernouilli}(\varphi_i)$

where φ_i is the probability of lethal injury given strike speed for event *i*. The probability of injury is assumed independent among collision events. The probability of injury was modeled as a function of vessel speed using a logit link:

(Eq. 2) $\text{Logit}(\varphi_i) = \beta_0 + \beta_1 * x$

where β_0 is the regression intercept, and β_1 is the slope parameter for x (speed). Several studies have used this formulation when addressing mortality associated with vessel collision, specifically for whales (Vanderlaan & Taggart 2007; Conn & Silber 2013), but it has not yet been conducted for manatees. The logistic regression is a special case of a generalized linear model and is appropriate here since the response variable is discrete valued (binomial), and the predictor variable is a vector containing discrete or continuous variables. In this analysis, inferences were drawn using a Bayesian approach and Markov Chain Monte Carlo (MCMC) simulation methods (e.g. Link et al. 2002). This method allowed for the uncertainty around lethality risk φ at vessel speed x to be treated probabilistically as $P(\varphi \mid x)$. The parametric form used for the logistic regression is the same form implied by Bayesian inference, which incorporates prior knowledge about the model or model-parameters. At the root of this type of inference is Bayes' theorem, where in this case:

(Eq. 3) $P(\varphi \mid x) = P(x \mid \varphi) * P(\varphi) / P(x)$ 8 *likelihood * prior*

In other words, the joint posterior probability, or the probability of obtaining the response parameter given the data, is proportional to the product of the likelihood and the prior distribution. Bayes theorem can be employed for this problem because the sample space in this case is partitioned into a set of mutually exclusive collision events, which lends information about the prior distribution (the probability of getting the data given the parameter of interest). These estimates can then be used with Bayes theorem to determine the joint posterior probability, ultimately obtaining a predictive model that estimates the conditional probability of lethal injury at any new instance of *x* (vessel speed).

I ran three parallel chains with initial values picked randomly from their prior distributions (Uniform(-10,10)) for each parameter (β_0 and β_1), each with 10000 iterations discarding the first 5000 (so-called 'burn in') iterations. I assessed convergence of the chains to their stationary distributions using the Brooks-Gelman-Rubin diagnostic (also called R-hat Gelman et al. 2004). This analysis was conducted with JAGS version 3.4.0 (Plummer 2013), in a batch mode using the R package RJAGS version 0.5-6 (Su & Yajima 2015) which imports R2WinBUGS (Sturtz et al. 2005; see Kéry 2008 for similar approach). All analysis was done in program R version 3.1.4. The results of these logistic regressions were used to draw inferences based on the inflection points.

Addressing uncertainty in vessel speeds

Vessel speed was typically reported as continuous quantitative values with values ranging from 0-50 knots. Figure 2 shows the distribution of quantitative vessel speeds that are present in the original data. Two cases reported 25+ and < 5 knots for vessels at the time of strike, and speeds of 25 and 5 knots, respectively, were assumed for the analyses. A portion of the vessel speeds were reported as a categorical "qualitative descriptor" (n=62), and some as a quantitative range (n=26).

MI is a procedure for treating missing data by replacing each missing datum with a set of m >1 plausible values (Rubin, 1987). MI involves three steps: data imputation, routine analysis, and pooling results for parameter estimation. The imputation step is a "filling-in" process that replaces missing data. This step assumes that data are missing at random. This step is repeated m times, each set resulting in unique imputed missing values. The second step involves each "complete" dataset being analyzed. The third step aggregates parameter estimates from each analysis (Rubin, 1987). The most common approach to dealing with missing data is to delete cases containing missing observations. However, this approach reduces statistical power and increases estimation bias (Nakagawa & Freckleton 2011). Recent studies show how data can be biased if "missing fraction" is removed (Hadfield 2008). Data imputation can be split into types: single imputation and MI. Single imputation is worse than data deletion in terms of parameter estimation and especially in estimation of uncertainty (i.e. standard errors) because single imputation ignores any uncertainty of imputed values (Nakagawa & Freckleton 2008). Here we analyze the effects of MI and single imputation (no-MI).

I used a MI procedure to account for uncertainty in vessel speed described as follows: For records provided as "qualitative descriptors", I treated speed values (e.g. planing, plowing, and slow/idle) as "missing values" and, for each iteration of the MI, replaced them with values randomly selected from a corresponding lognormal distribution fit to ancillary vessel speed data (Gorzelany 2013). These data were obtained from Gorzelany et al.'s boater compliance study in which a radar gun was used to measure the speeds of vessels throughout Florida (N= 7418). Each record also included a physical description of the vessel's position in the water that corresponded to the "qualitative descriptors" in the collision data. This allowed me to obtain a range of possible speeds for each description. For records reported as a range, I replaced them with values randomly selected from a uniform distribution constructed from the range for each speed category.

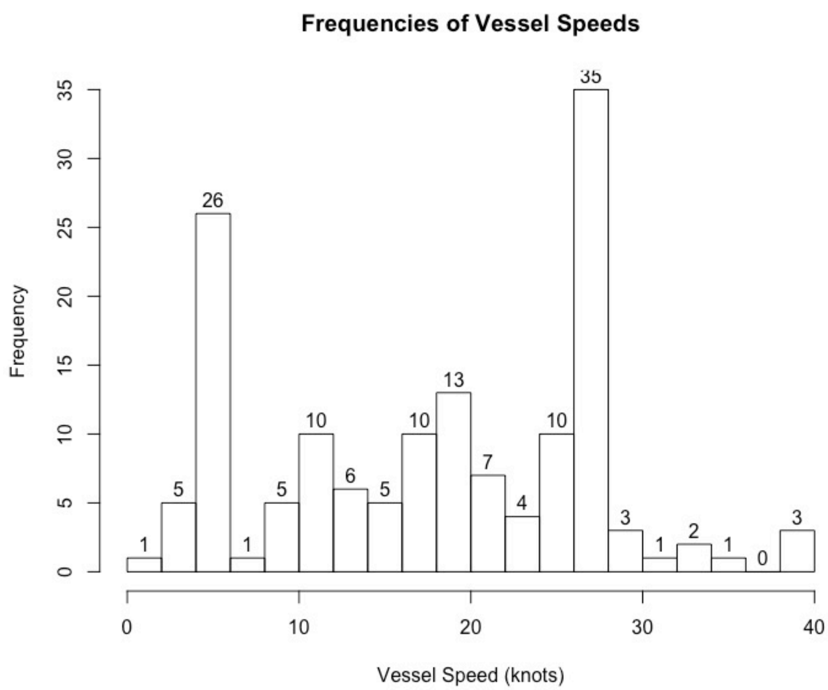

Figure. 2: Frequencies of observed quantitative vessel speeds present in collisions report data

Category	Mean (knots)	Standard deviation	Radar gun	Possible definitions
1	5.62	1.15	Slow (SL), Idle (ID)	no wake, very slow, out of gear, less than idle
2	10.96	3.80	Plowing (PW), Cruising (CR)	not on plane, cruising
3	26.52	6.63	Planning (PL)	on plane, high speed, fast, coming up on plane

Table 3: Vessel speed categories used in this analysis. Includes definitions from the original collision report data that were classified into these categories.

Table 3 shows the three vessel speed categories used in this analysis and the possible definitions from the original collision report that were classified into each of these categories. The ancillary radar gun data were separated into three categories for comparison with collision categories: category 1 – slow and idle; category 2 – plowing and cruising; category 3 – planing. In this way, the radar gun data served as a link between the qualitative categories and quantitative speed estimates. Each corresponding category from the collision data was then assigned a lognormal distribution based on the fitted empirical speed data. The MI treatment was performed 200 times, each imputation containing 5000 MCMC iterations described previously. To assess the effects of ignoring the uncertainty in qualitative speed values, I also fit the logistic regression model using the mean of corresponding qualitative speed categories (rather than multiple random draws) or the mean of the range of speed values given. This is also called single imputation and is referred to as no-MI hereafter.

To assess the effectiveness of the MI routine, the validity of the parameter estimates, and the effects of ignoring uncertainty in the qualitative speed values, simulations were performed. 500 datasets were generated based characteristics of the original data – including, proportion of qualitative speeds, number of records expressed as ranges, and the number of lethal and non-lethal records. 200 MI runs were then applied to each simulated dataset, and parameter estimates, bias, mean squared error (MSE), and coverage were tracked for each iteration.

About separation in data

Due to the small size of the observational data, complete and quasi-complete separation existed in a proportion of the simulated datasets. Complete separation occurs when a linear combination of the predictors yields a perfect prediction of the response variable, and quasi-complete separation occurs when the predictors yield a perfect prediction of the response for most values of the predictors but not all (Heinze, 2006). In other words, in complete separation the outcome variable separates a predictor variable completely. When this occurs, the maximum likelihood estimate for the separated predictor variable does not exist. This is common in small datasets. Data with separation are weak data that tend to not provide enough information about the parameters of statistical models. Informing the prior allows us to regularize the coefficients and pull them just slightly towards zero — reducing the standard deviation of possible outcomes (Rainey, 2016). In other words, an informative prior is a reasonable and mathematically sound way to regulate separated data. The prior distribution should be reflective of prior information for a range of situations — this doesn't mean it is always perfect, sometimes it provides too little prior information and sometimes too much (Rainey, 2016). Selecting a reasonable prior is the responsibility of the researcher, here I used expert advice to determine this range. In large datasets, the data tend to overwhelm the contribution of the prior so that the specific choice of the prior has little effect on the posterior distribution. In the case of separation, the prior distribution is an important choice that affects the inference. Bayesian inference, more specifically incorporating prior information, is helpful in situations where separation creates implausible MLEs (Rainey, 2016). In Bayesian inference, a posterior probability derived from a prior probability and the likelihood function, which is derived from a statistical model for the observed data. Here, Bayesian inference is used not only to obtain a posterior probability for predictive purposes, but to mitigate the quasi and complete separation encountered when simulating data from the small original dataset. Bayesian inference treated issues with separation in simulated datasets, however in an additional attempt to quell the effects of separation on parameter estimates, one of the reclassified original datasets (classification 1) was not considered for this analysis because it exhibited complete separation (Fig. 3). It is one thing for simulated datasets to have separation, it is another to simulate datasets from an original set that is completely separated, where parameter estimates are ultimately doomed from the start.

I conducted the MI routine on sets of simulated data in which the instances of complete and quasi-complete separation were retained, and for comparison, sets of simulated data in which instances of complete separation were removed and new simulations generated until 500 were reached.

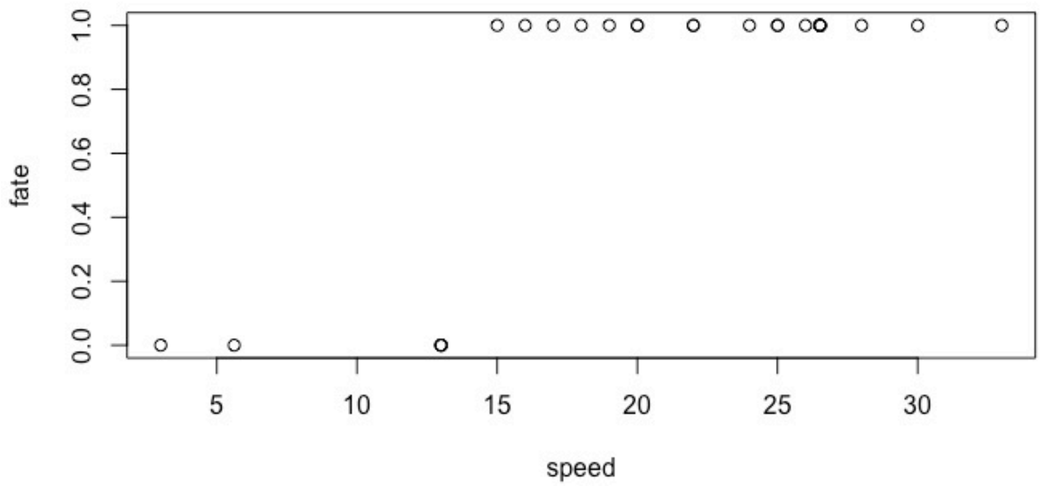

Figure 3: Demonstration of the complete separation present in "classification" treatment of collisions report data. Complete separation means that there is no overlap of the prediction variable for the positive and negative response. This was the motivation behind using only one of the re-classification efforts ("classification 2") for the analysis hereafter.

Data was simulated based on parameter values estimated from the original data using MI, for the purpose of this analysis I refer to these as the 'true values'. This acts as a baseline by which comparisons of simulation results can be made. In order to determine if simulations were effective, I assessed a model outcome where the β_1 values were forced to be zero, i.e. no relationship, by setting the 'true value' for the β_1 to zero at the simulation stage.

CHAPTER 3:

RESULTS

The logistic model (no-MI) of manatee fate given speed shows that there is a positive relationship between the probability of lethal injury and boat speed (β_0 = -5.13 and β_1 = 0.35 95% CDI [-9.55 -0.76; 0.12 0.61]). That is, as speed increases the probability of lethal injury increases. The logistic relationship treated with MI (where uncertainty in vessel speed is accounted for using ancillary radar gun data) had parameter estimates: β_0 = -4.68 and β_1 = 0.34 95% CDI [-9.32 -0.03; 0.07 0.60]. The main difference between these two estimation techniques is best understood by the credible intervals. Figure 4 shows a comparison of the main effect modeled with MI and no-MI. Not surprisingly the uncertainty is larger with MI.

Both the MI and no-MI treatment demonstrate that the greatest rate of change in the probability of lethal injury to manatees occurs between 10 and 20 knots, or somewhere after the first inflection of the first derivative of the logistic. Large credible intervals (CDI) are observed at very low speeds, decreasing as speed increases. MI results show a larger range of probabilities (larger CDI), indicating that uncertainty in vessel speed data may be more accounted for. MI treatment also decreases the inflection point by a small amount, which is indicative of a more conservative estimate. Figure 4 shows the distribution of the β_1 parameter from 500 simulation runs. The true values for the beta parameters are based on the results from the MI performed on the original collision report data. Figures 5.a and 5.b compare parameter results and difference between expected and true values from simulated datasets to simulated datasets exhibiting no separation. With separation eliminated, the expected value is more similar to the true value.

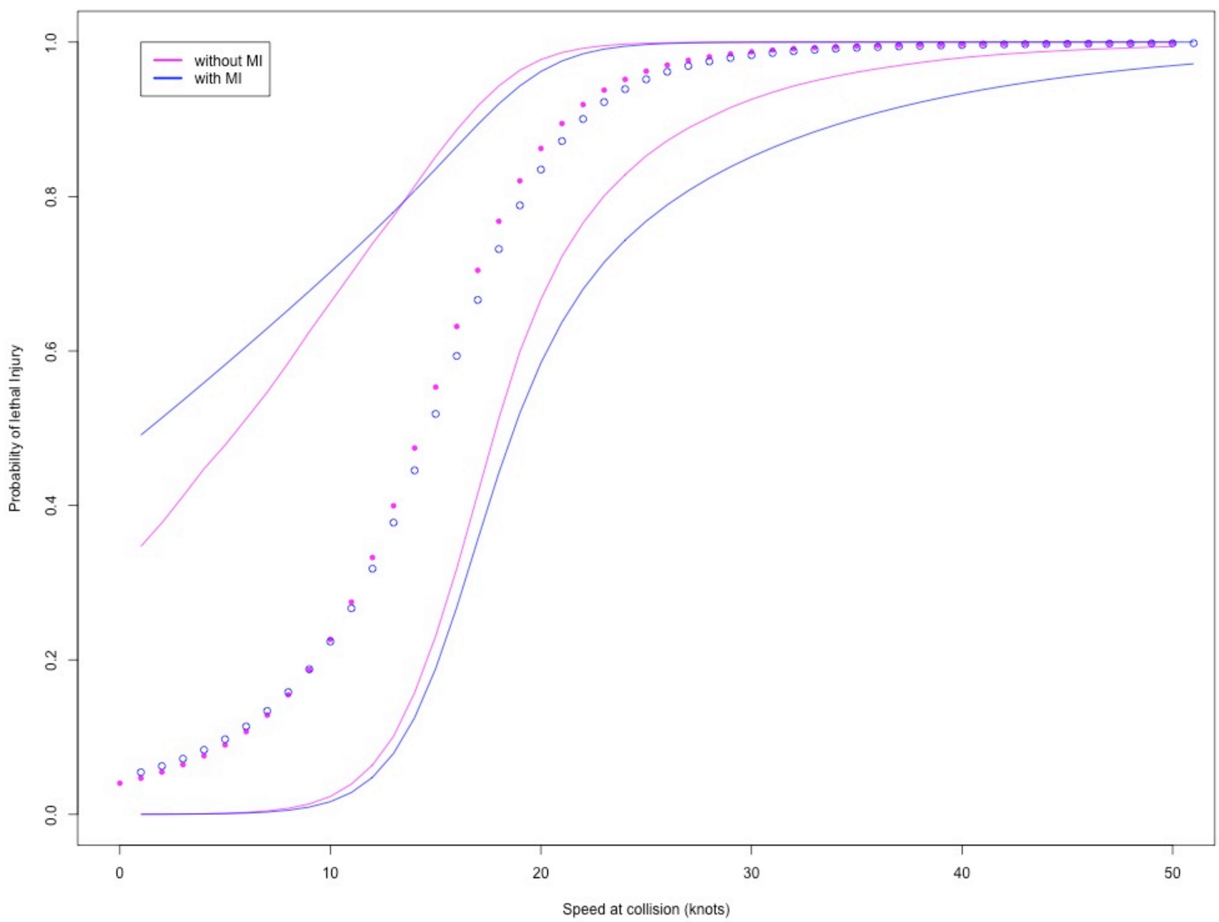

Figure 4: Probability of a lethal injury resulting from a vessel collision to manatee as a function of vessel speed. Model used is a simple logistic regression without multiple-imputation (MI) (solid circles) and 95% credible intervals (CDI) (solid thin lines) and the logistic fitted with multiple-imputation (empty circles) and 95% CDI for each distribution (solid thin lines).

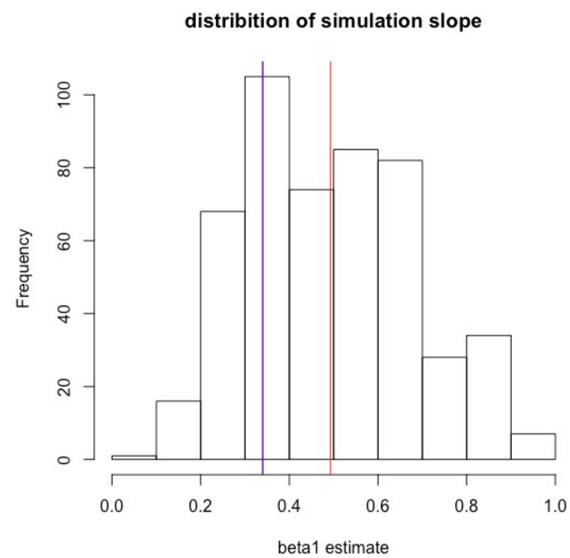

Figure 5.a: Histogram of β_1 MI-treated relationship for 500 simulated datasets. The blue line represents the "true value" of the parameter based on the MI estimates, and the red line is the mean of the simulated results for the parameter.

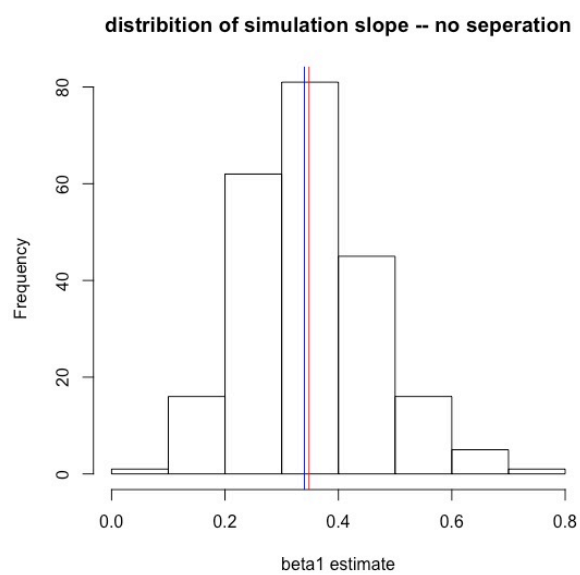

Figure 5.b: Histogram of β_1 MI-treated relationship for 500 simulated datasets with datasets exhibiting separation removed. The blue line represents the "true value" of the parameter based on the MI estimates, and the red line is the mean of the simulated results for the parameter.

Scenario		True	Mean	Bias	MSE	Coverage (%)
With MI	beta0	-4.6816	-5.408	-0.7267	3.8967	96.6
	beta1	0.3404	0.4930	0.1525	0.0601	96.6
Without MI	beta0	-4.6816	-5.9991	-1.3174	4.3801	98.4
	beta1	0.3404	0.5343	0.1938	0.0706	96.8
With MI - no relationship	beta0	1.4816	1.7609	0.2793	1.8328	98.4
	beta1	0.0000	0.0017	0.0017	0.0050	98.8
With MI - no separation	beta0	-4.6816	-4.0476	0.6340	2.8380	93.3
	beta1	0.3404	0.3472	0.0068	0.0135	96.5

Table 4: Simulation results for scenario of interest (classification 2). Comparison of results with and without MI (MI) replacing qualitative speed values, scenario where no-relationship was forced (i.e. β_1 (slope) = 0), and one scenario where I excluded any simulated dataset that exhibited separation, re-simulating until I reached 500 non-separated datasets.

Table 4 shows the comparisons of all model estimates. Effectiveness of the simulation is shown by a high coverage (>90%) and low MSE of the no treatment β_1 for simulated datasets. Figure 6 shows a comparison of the β_1 parameter for the three MI models. Once more, it is noted that removing separated data sets brings the estimated slope closer to the true MI estimate of 0.34, also reducing uncertainty. It is also clear that the credible intervals and estimates are more conservative in the MI results, with the mean value slightly closer to the true value than without MI.

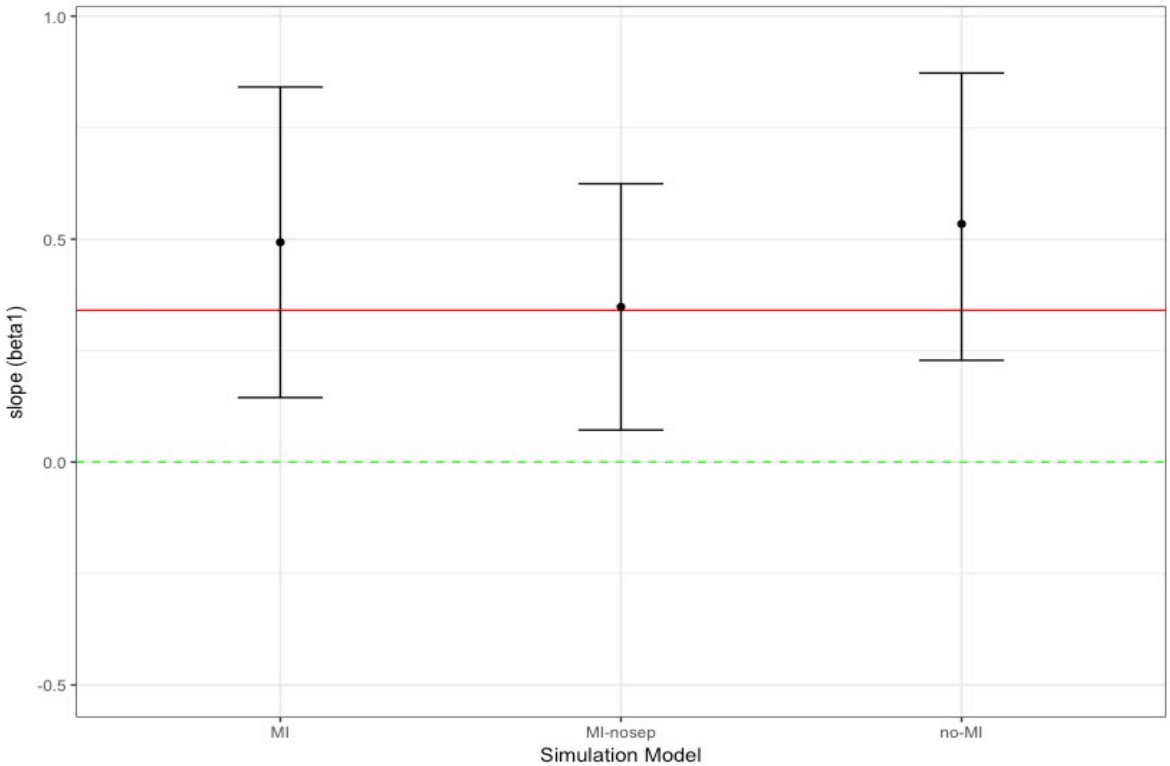

Figure 6: Comparisons of average β_1 values three simulation models. With multiple-imputation (MI), with multiple-imputation and no separated datasets (MI-nosep), and without multiple-imputation (no-MI). The red line indicates the 'true value' or β_1 estimated from original collision report data using multiple-imputation to replace unknown vessel speed values.

CHAPTER 4:

DISCUSSION

This work aimed at evaluating the hypothesis that vessel speed has an effect on the probability of lethal injury to manatees. Until now, managers have assumed that the probability of lethal injury increases with vessel speed, but this assumption is based on anecdotal evidence and not on any rigorous statistical analysis (Calleson and Frochlich, 2007). Here, I used citizen reported information to evaluate this hypothesis, and quantify the relationship. It is important to note that data obtained from citizen-based surveys must be analyzed carefully because of many important sources of uncertainty. For instance, in the case of the vessel-manatee collision data set, uncertainty about vessel speed is a potentially important source of error. My analysis provides information about the relative importance of accounting for speed uncertainty, and accounting for uncertainty in covariate data in general. Using the multiple-imputation approach allows for estimation of risk that are as close to reality as possible given small, flawed samples. Not accounting for this uncertainty can affect the slope of the relationship and associated credible intervals (Nakagawa & Freckleton, 2011). Future work should consider other covariates besides speed, including vessel size, type, and propulsion system as these may have compounding effects on the risk of lethal injury.

Several potential sources of errors could affect my results. First, although the estimated shape of the relationships for the probability of death given strike speed was qualitatively similar, there were numerous discrepancies in the classification of event types among the two observers. This issue illustrates the large uncertainty with the fate of manatees associated with collision events.

For example, an animal may be hit and suffer deadly internal injuries but may be classified as not injured, although it may die later at a different location due to the injuries. Many collision events are obviously not reported, but a bias in the reporting can be particularly problematic. For instance, if collision events are preferentially reported depending on the fate of the animal (e.g. if the cases of minor (or no) injuries are less likely to be reported), this could result in bias of the parameter estimates. Unfortunately, I was not able to assess the extent of this potential source of error. One particular concern of a directional bias is the case of expert classification. The experts may have considered that an event resulted in a serious injury when the collision occurred at a higher speed. Although, we tried to minimize this type of error, we were not able to remove it and we do not have a good way to quantify the extent of the potential bias. The low sample size was another problem for this study. For instance, given the low sample size it would have been difficult to determine the effect of vessel size on the relationship, yet, mechanistically I would expect vessel size to have an impact on this relationship. Given all these caveats, my results should be interpreted with caution.

The probability of death given strike speed is a key parameter to understanding the effectiveness of speed zones for manatee, and this analysis will provide the first quantification of this parameter. This analysis should hopefully fill an important gap and ultimately help improve the management of vessel regulations for manatees and other species affected by vessel collisions. For example, Vanderlaan and Taggart (2007) and Conn and Silber (2013) did not account for speed uncertainty in their analyses, but the quantitative approach presented here can provide a robust means for this. This example also shows a method for making use of citizen data, which is an underused resource, but one that demands a rigorous and transparent method for quantification. Error in variable models are under used in ecology and hopefully my analysis will encourage other scientists to use it for their research.

Data collected by citizens provide many opportunities to integrate data collected at large scales and cheaply. Also, citizen scientists act as "many eyes", and these data are valuable for recording rare events. Given that these data are often collected opportunistically it is important to clearly identify the limitations of the data and to try to properly account for the most important sources of uncertainty. I recommend that to obtain the best dataset on collisions as possible, the collection methodology should be improved and standardized. Citizen science mobile phone applications are becoming more common, and ecologists could leverage this tremendous potential to address many questions in ecology and conservation. Reporters should be given the opportunity to add photographs for quality assurance and control, as well as comment. However only a few variables should be made priority: speed, vessel type, size, and the outcome of the collision. Emphasis should be made on anonymity to help ease concerns that reporting may lead to convictions for speeding, however public outreach on how and what to report along with information on how these data are used should be the priority. It is necessary at this point to find a way to sufficiently explain the reasoning behind parameterizing the relationship between manatees and vessels, not only for wildlife, but for human wellbeing. The benefit to boaters is that this relationship is better understood, and speed zones may be optimized based on actual data. A clear and mathematical explanation for where and why speed zones exist may ease any tensions about the topic. Reporting can be emphasized as a way for boaters to be involved in this optimization process, acting as stewards for the water and wildlife that they appreciate on a daily basis. In order to be beneficial in these ways, the reporting system should be simple and accessible to waterway users.

Even with improved reporting, missing data will be inevitable with citizen reported data. More likely, the witness or operator reporting will describe a vessel as "planing" or "between 20 and

30 knots", for example. This type of data, while vague, still contains valuable information that is lost if the information is ignored or replaced with a mean or other representative value. Multiple imputation (MI) is considered the fastest-growing method in handling missing data and is becoming the standard method in medical and social sciences (Nakagawa & Freckleton 2008). As of this date, it is an underused method in ecology. The critical importance of MI, and the difference between it and other ways of handling missing data, is the capability of MI to provide information, i.e. statistical parameters, regarding the impact of missing data on parameter estimation. Such considerations are crucial for ecological data. As citizen science is used more frequently to answer challenging ecological questions, and in addition as more ecologists begin using model selection and Bayesian approaches, both of which require complete sets of information, we will need to continue to focus on effective ways of dealing with missing data.

This work provides an example of how citizen science data can be used to address the problem of collisions between wildlife and boaters. The methodology used in this paper will help reach the goal of understanding the variables that impact severity of a collision and help develop an adequate treatment of the intrinsic uncertainty that comes with citizen reported data. Moreover, the insights gained here help inform what is needed from collision reports. The goal should be to emphasize the reporting of accurate speeds and other characteristics of the vessel involved from observers of a collision event. Since vessel strikes are the leading cause of mortality, we need to encourage reporting of collisions. We also need to make reporting easy. Reporting by concerned and invested citizens represents our best dataset for understanding where and at what speed collisions occur. Citizen reported data is a potentially valuable resource for resource management, but tool development must take into consideration how to mitigate uncertainty and missing values.

REFERENCES

Aipanjiguly, S., Jacobson, S.K., Flamm, R.O., 2003. Conserving manatees: knowledge, attitudes, and intentions of boaters in Tampa Bay, Florida. Conservation Biology 17: 1098–1105.

Allen, R.E., McCullough, D.R., 1976. Deer–car accidents in Southern Michigan. Journal of Wildlife Management 40: 317–325.

Bauduin, S., Martin, J., Edwards, H. H., Gimenez, O., Koslovsky, S. M., Fagan, D. E., 2013. An index of risk of co-occurrence between marine mammals and watercraft: Example of the Florida manatee. Biological Conservation 159: 127-136

Bellan, S.E., Gimenez, O., Choquet, R. & Getz, W.M., 2013. A hierarchical distance sampling approach to estimating mortality rates from opportunistic carcass surveillance data. Methods in Ecology and Evolution 4: 361–369.

Brownie, C., Anderson, D.R., Burnham, K.P. & Robson, D.S., 1985. Statistical Inference From Band Recovery Data: A Handbook, 2nd edn. U.S. Fish and Wildlife Service, Washington, District of Columbia, USA.

Calleson, C. S., & Frohlich, R. K., 2007. Slower boat speeds reduce risks to manatees. Endangered Species Research 3: 295-304.

Conn, P., & Silber, G., 2013. Vessel speed restrictions reduce risk of collision-related mortality for North Atlantic right whales. Ecosphere 4(4): 43.

Dickinson, J.L., Shirk, J., Bonter, D., Bonney, R., Crain, R.L., Martin, J., Phillips, T., Purcell, K., 2012. The current state of citizen science as a tool for ecological research and public engagement. Ecol Environ 2012; 10(6): 291–297.

Fonnesbeck, C.J., Garrison, L.P., Ward-Geiger, L.I., Baumstark, R.D., 2008. Bayesian hierarchical model for evaluating the risk of vessel strikes on North Atlantic right whales in the SE United States. Endangered Species Research 6: 87–94.

Forman, R.T.T., & Alexander, L.E., 1998. Roads and their major ecological effects. Annual Review of Ecology and Systematics 29: 207–231.

FWC, 2007. Florida Manatee Management Plan. Florida Wildlife Commission. St Petersburg, FL. https://www.nrc.gov/docs/ML1432/ML14328A280.pdf

Gorzelany, J., 2013. A Characterization of Recreational Boating Activity and Boater Compliance with Posted Speed Zones in Palm Beach County. Final Report to Paul Davis, Sea to Shore Alliance. Palm Beach County Department of Environmental Resources Management 2300 North Jog Road West Palm Beach, FL 33411

Hadfield, J.D., 2008. Estimating evolutionary parameters when viability selection is operating. Proceedings of the Royal Society of London B: Biological Sciences: 275, 723–734.

Hazel, J., Lawler, I.R., Marsh, H.D. and Robson, S., 2007. Vessel speed increases collision risk for the green turtle Chelonia mydas. Endangered Species Research 3: 105–113.

Heinze, G., 2006. A comparative investigation of methods for logistic regression with separated or nearly separated data. Statistics in Medicine. 25(24): 4216-4226.

Hell, P., Plavy, R., Slamecka, J., Gasparik, J., 2005. Losses of mammals (Mammalia) and birds (Aves) on roads in the Slovak part of the Danube Basin. Eur. J. Wildlife Research 51: 35–40.

Ke´ry, M., 2008. Estimating abundance from bird counts: binomial mixture models to uncover complex covariate relationships. The Auk 125: 336–345.

Kraus, S.D., 1990. Rates and potential causes of mortality in North Atlantic Right whales (*Eubalaena glacialis*). Marine Mammal Science 6: 278–291.

Laist, D.W., Knowlton, A.R., Mead, J.G., Collet, A.S., Podesta, M., 2001. Collisions between ships and whales. Marine Mammal Science 17: 35–75.

Laist, D.W. & Shaw, C., 2006. Preliminary Evidence That Boat Speed Restrictions Reduce Deaths of Florida Manatees. Marine Mammal Science 22: 472–479.

Langen, T.A., Ogden, K.M., Schwarting, L.L., 2009. Predicting hot spots of herpetofauna road mortality along highway networks. Journal of Wildlife Management 73: 104–114.

Link, W.A., Cam, E., Nichols, J.D. & Cooch, E.G., 2002 Of Bugs and birds: Markov Chain Monte Carlo for hierarchical modeling in wildlife research. Journal of Wildlife Management 66: 277–291.

Maitland, R.N., Lawler, I.R., Sheppard, J.K., 2006. Assessing the risk of boat strike on dugongs at Burrum Heads, Queensland, Australia. Pacific Conservation Biology 12: 321–326.

Martin, J., Sabatier, Q., Gowan, T. A., Giraud, C., Gurarie, E., Scott Calleson, C., Ortega-Ortiz, J. G., Deutsch, C. J., Rycyk, A., Koslovsky, S. M., 2016. A quantitative framework for investigating risk of deadly collisions between marine wildlife and boats. Ecology and Evolution 7: 42-50.

Marsh, H., O'Shea, T.J., Reynolds, J.E., 2011. Ecology and Conservation of the Sirenia: Manatees and Dugongs. Conservation Biology 18: 543

McClintock, B.T., Onorato, D.P., Martin, J., 2015. Endangered Florida panther population size determined from public reports of motor vehicle collision mortalities. Journal of Applied Ecology 52: 893–901.

Nakagawa, S., Freckleton, P., 2008. Missing inaction: the dangers of ignoring missing data. Trends in Ecology and Evolution 23(11): 592-596.

Nakagawa, S., Freckleton, P., 2011. Model averaging, missing data and multiple imputation: a case study for behavioral ecology, Behavioral Ecology and Sociobiology 65(1): 592-596.

Ng, A., Jordan, M., 2002. On discriminative vs. generative classifiers: a comparison of logistic regression and naive Bayes. NIPS'01 Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic: 841-848.

Plummer, M., 2013. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). ISSN1609-395X.

Puky, M., 2006. Amphibian road kills: a global perspective. In: Irwin, C.L., Garrett, P., McDermott, K.P. (Eds.), Proceedings of the 2005 International Conference on Ecology and Transportation. Center for Transportation and the Environment, North Carolina State University, Raleigh, 325–338.

Rao, R.S.P. & Girish, M.K.S., 2007. Road kills: assessing insect casualties using flagship taxon. Current Science 92(6): 830–837.

Rubin, D.B. & Schenker, N., 1987. Interval estimation from mulitply-imputed data: a case study using agriculture industry codes. Journal of Official Statistics 3, 4: 75-387.

Runge, M.C., Sanders-reed, C.A., Fonnesbeck, C.J., 2007. A Core Stochastic Population Projection Model for Florida Manatees (Trichechus manatus latirostris). U.S. Geological Survey Open-File Report 2007.

Stone, G.S., Yoshinaga, A., 2000. Hector's Dolphin *Cephalorhynchus hectori* calf mortalities may indicate new risks from boat traffic and habituation. Pacific Conservation Biology 6: 162–170.

Sturtz, S., Ligges, U., Gelman, A., 2005. R2WinBUGS: a package for running WinBUGS. R. Journal of Statistical Software 12: 1–16.

Su, Y.S., Yajima, M., 2015. Package 'R2jags'. R package version 0.03-08, URL http://CRAN. R-project. org/package= R2jags.

USFWS, 2001. Florida manatee recovery plan, (*Trichechusmanatus latirostris*), third revision.US Fish and Wildlife Service, Atlanta, GA

Udell, B. 2016. Integrating lethal collision risk from watercraft into optimal design of proection zones for marine mammals, Thesis, University of Florida, Gainesville, FL.

Udell, B., Martin, J., Fletcher, R.J., Bonneau, M., Edwards, H., Gowan, T., Hardy, S. K., Gurarie, E., Calleson, C. S., Deutsch, C.J.. Integrating encounter rate theory with decision analysis to evaluate collision risk and determine optimal protection zones for wildlife. In review in Journal of Applied Ecology.

Vanderlaan, A.S.M. & Taggart, C.T. 2007. Vessel collisions with whales: the probability of lethal injury based on vessel speed. Marine Mammal Science, 23, 144–156.Wegner, S. J., Freeman, M. C., 2008. Estimating Species Occurrence, Abundance, and Detection Probability Using Zero-Inflated Distributions. Ecology 89, 10: 2953-2959.

Vanderlaan, A.S.M., Taggart, C.T., Serdynska, A.R., Kenney, R.D., Brown, M.W., 2008. Reducing the risk of lethal encounters: vessels and right whales in the Bay of Fundy and on the Scotian Shelf. Endangered Species Res. 4: 283–297.

Ward-Geiger, L.I., Silber, G.K., Baumstark, R.D., Pulfer, T.L., 2005. Characterization of ship traffic in right whale critical habitat. Coastal Management 33: 263– 278.

Weir L.A., Royle J.A., Nanjappa P., and Jung R.E. 2005. Modeling anuran detection and site occupancy on North American Amphibian Monitoring Program (NAAMP) routes in Maryland. J Herpetol 39: 627–39.

Wells, R.S., Scott, M.D., 1997. Seasonal incidence of boat strikes on bottlenose dolphins near Sarasota, Florida. Mar. Mammal Sci. 13: 475–480.

ABOUT THE AUTHOR

Brittany (Bea) Combs grew up one of four children in a military family with a vague dream of studying the ocean. She is incredibly passionate about environmental sustainability, science communication, and education. Bea chose her college – The University of Mary Washington in Fredericksburg, VA – based on cross country and aesthetics. However, quickly developed a deep love for science there and found the mentorship she needed to help her succeed in obtaining her Bachelor's of Science in Biology in addition to a minor in Environmental Science and Studio Art. Immediately after graduation she moved to Florida for an internship at Florida Fish and Wildlife Research Institute which morphed into 4 years of research work in both the Marine Mammal and Turtle Departments. At FWRI she specialized in spatial analysis of populations using GIS and statistical methods. It was also while at FWRI that she had the privilege of joining the project that became her Master's research. She worked as a research contractor throughout grad school, including a year stint at the United States Geological Survey.

In addition to her career, Bea is passionate about volunteering in her Saint Petersburg community, particularly as it pertains to communicating science and science-informed policy. Her other interests include Astronomy, adventuring, crewing sail boats to faraway places, gardening, bird watching, and hanging out with her partner in crime, 3 dogs, Chinese water dragon, scorpion, and tarantula. She is currently teaching biology at St. Petersburg High school.