

The University of Maine

DigitalCommons@UMaine

Honors College

5-2013

Bayesian Analysis of Data on Nest Success for Marsh Birds

Sean Hardy

University of Maine - Main

Follow this and additional works at: <https://digitalcommons.library.umaine.edu/honors>



Part of the [Applied Statistics Commons](#), and the [Poultry or Avian Science Commons](#)

Recommended Citation

Hardy, Sean, "Bayesian Analysis of Data on Nest Success for Marsh Birds" (2013). *Honors College*. 128.
<https://digitalcommons.library.umaine.edu/honors/128>

This Honors Thesis is brought to you for free and open access by DigitalCommons@UMaine. It has been accepted for inclusion in Honors College by an authorized administrator of DigitalCommons@UMaine. For more information, please contact um.library.technical.services@maine.edu.

BAYESIAN ANALYSIS OF DATA ON NEST SUCCESS FOR MARSH BIRDS

By

Sean Taylor Hardy

A Thesis Submitted in Partial Fulfillment of the
Requirements for a Degree with Honors
(Mathematics)

The Honors College

University of Maine

May 2013

Advisory Committee:

William Halteman, Professor and Co-Chair of Mathematics and Statistics,
Advisor

Brian Olsen, Assistant Professor of Biology and Ecology

Kathleen Ellis, Adjunct Assistant Professor in Honors (English)

David Hiebeler, Associate Professor in Mathematics and Statistics

Andrew Knightly, Associate Professor in Mathematics and Statistics

Abstract

Bayesian methods are an increasingly popular form of statistical analysis which uses informative prior distributions to help calculate posterior distributions of models that represent different hypotheses. Frequentist methods are contrasting methods that are used more commonly and more well known, but have come under recent criticism. I examined data gathered by Ellen Robertson, who used information theoretic methods for a Masters' Thesis in Ecology and Environmental Science at the University of Maine to analyze the daily survival probabilities of marsh birds with a Bayesian perspective in order to get a sense of the Bayesian analysis. Results were as expected; when using uninformative prior distributions, the Bayesian analysis had almost the same results as Robertson's. With the use of Robertson's calculated parameter estimates as informative prior distributions, the Bayesian analysis still ended with similar results. The conclusions in all three versions of statistical analysis were the same. Hence, Bayesian methods can construct models representing hypotheses effectively.

TABLE OF CONTENTS

LIST OF TABLES.....iv

LIST OF FIGURES.....v

Chapter

1. BIOLOGY OF RAILS.....1

 1.1. Introduction.....1

 1.2. Background on the Virginia Rail and the Sora.....2

 1.3. Robertson’s Model Parameters.....3

2. BAYESIAN METHODS.....8

 2.1. Introduction.....8

 2.2. Frequentist Methods Summary.....9

 2.3. Bayesian Methods Summary.....12

 2.4. Differences in Methods.....20

3. COMPARISONS OF MODELS.....23

 3.1. Introduction.....23

 3.2. Methods.....24

 3.3. Results.....25

 3.4. Discussion/Conclusions.....33

BIBLIOGRAPHY.....36

BIOGRAPHY OF THE AUTHOR.....38

LIST OF TABLES

Table 1.	Interpretation of AIC Values.....	18
Table 2.	Rail Nest Survival Model-Averaged Parameter Estimates.....	29
Table 3.	Rail Nest Survival Models Parameters.....	31
Table 4.	Rail Nest Survival Model Comparisons.....	32

LIST OF FIGURES

Figure 1.	Example of a Probability Distribution.....	11
Figure 2.	Example of an Auto-Correlation Function Plot.....	17
Figure 3.	Model-Averaged Mean Plus or Minus One Standard Error for Intercept.....	26
Figure 4.	Model-Averaged Mean Plus or Minus One Standard Error for Age.....	27
Figure 5.	Model-Averaged Mean Plus or Minus One Standard Error for Water Depth Change.....	27
Figure 6.	Model-Averaged Mean Plus or Minus One Standard Error for Nest Height Change.....	28
Figure 7.	Model-Averaged Mean Plus or Minus One Standard Error for Impoundment.....	28
Figure 8.	Model-Averaged Mean Plus or Minus One Standard Error for the Interaction Between Water Depth Change and Nest Height Change.....	29

CHAPTER ONE:
BIOLOGY OF RAILS

1.1 Introduction

This paper is based on research done by Ellen Robertson for a Masters' Thesis in Ecology and Environmental Science at the University of Maine on the virginia rail (*Rallus limicola*) and the sora (*Porzana carolina*), which are both birds in the family *Rallidae*, or rails. To fully understand the implications of this paper, it is necessary to learn about the rails themselves and the biology of the variables and methods involved in the statistical models analyzed by Robertson and reanalyzed by myself. Many of the species in this family prefer wetland habitats, and both the virginia rail and the sora can be found in these sorts of locations throughout most of North America (Alderfer 2006). It is important to study the habitat and survival probabilities of these birds as they are facing habitat loss and population decline over recent years (Conway et al. 1994). In the US, twenty-two states have lost half or more of the wetland habitat they once had (Fletcher 2003). The potential deterioration of populations of rails also can also have other lasting effects on the ecosystems (Fletcher 2003). Although these birds are in the "Least Concern" category of conservation status ("All About Birds" 2011), we need to understand the impact that this habitat loss may have on their avian populations. By studying what primarily affects these rails' daily nest survival probabilities and how habitat loss will affect these variables we can potentially find ways to prevent these

species from becoming higher on the category of conservation status. We must be careful when studying these probabilities as there is a bias against failed nests that is not properly accounted for, which leads to an apparent nest success probability that is likely higher than the true value (Mayfield 1975). To combat this, we calculate daily survival probabilities and use those to determine an overall nest success probability (Mayfield 1975). The variables studied in Robertson's thesis were nest age, the water depth change at the location of the nest and the nest height change between visits to the nest, whether the marsh where the nest was found had water levels managed through human intervention (a technique known as impoundment), and the interaction between water depth change and nest height change. She studied 72 virginia rail and 22 sora nests in Maine from 2010-2011.

1.2 Background on the Virginia Rail and the Sora

The virginia rail and sora are rails that have many similarities. While they are very common marsh birds in North America, they have a secretive nature that makes observation difficult (Kaufmann 1989). They both are found in marshes and wetlands throughout North America, and nest in hidden locations in very thick vegetation near the edge of water in the marsh (Alferder 2006). The nests are made from nearby emergent vegetation, or aquatic plants that also reach above water level into the air (Lowther 1977). Their nests are constructed so similarly that the primary way to tell the difference is by identifying the color of the eggs (Kaufmann 1989). The sora is typically around 19 to 25

centimeters in length while the virginia rail is around 22 to 27 centimeters (Alferder 2006). The sora and virginia rail on average weigh between 49 to 112 grams and 65 to 95 grams respectively (“All About Birds” 2011). Due to their secretive nature, they are more often heard than seen (Melvin and Gibbs 2012, Conway 1995). They both feed on seeds, and insects found in the marshes, although the virginia rail also will eat fish, frogs and small snakes (“All About Birds” 2011). To breed, they both regularly migrate to a northern range from the Northwest Territories to California in the west and from New England to Pennsylvania in the east in the spring (typically around April and May) (Alferder 2006). They migrate in the early fall (typically around August or September) for the winter as far south as northern South America (Alferder 2006). Soras tend to more rigorously and aggressively defend a territory than the virginia rail (Kaufman 1989). During breeding season, nests are built from nearby vegetation near the water level of the marsh (Alferder 2006). For both species, both the male and female parents incubate eggs, feed and protect the young (Kaufmann 1989). Virginia rails and soras very often thrive in the same marshes throughout their range of distribution (Conway 1995) and even respond to each other’s calls (Kaufmann 1989). The average clutch size, or the total number of eggs incubated, is around 10.5 eggs for soras and 8.5 for virginia rails (Kaufmann 1989). The average incubation period for the virginia rail is around 20 days and for the sora between 16 and 22 days (Massachusetts Division of Fisheries & Wildlife, 2010).

Due to the many similarities, in habitat, diet, and nesting/breeding, and their few differences, it makes sense to combine the data gathered on the two birds and study them

together as has been done in several studies. We also have to consider sample sizes of studies; both the virginia rail and the sora can be difficult to find because of their covert nesting behaviors. Combining data collected on the two birds can help make larger sample sizes and in turn lead to more precise estimates of the parameters of any statistical model that is developed.

1.3 Robertson's Methods and Model Variables

An important method used during the data gathering from Robertson's research, as well as any research done on probabilities on nest success, is the Mayfield method. This is used to counter a bias in data collecting that leads to a perceived higher nest success probability than is the true probability. When collecting data on nests, the researcher is randomly collecting data which includes nests of all ages. What would be ideal is when attempting to estimate a population mean, in this case, the probability of nest success for the entire population, we could observe all of the nests from start to finish (success or failure) and take a simple ratio to determine the sample proportion of nest success (Mayfield 1975). Most often this is impossible, and the data gathered by Robertson was not collected in this manner. Instead, we use the Mayfield method, which is calculating daily survival probabilities and using those to estimate the overall rate of nest success for the population to compensate for a bias towards successful nests (Mayfield 1975). For example, if we observe 10 nests in a marsh, 9 of which we consider successful, we have an apparent nest success rate of 90%. However, if there were 8 nests

that failed and went unobserved, the true nest success rate would have been 50%, which is significantly different than the apparent nest success rate. Using the Mayfield method, we look at the sum of days we observed each nest — that is, if we observed one nest for 3 days, one nest for 4 days and one nest for 5 days, we would have 12 observation days. Of the 10 nests observed in this example, let's say we had around 30 total observation days. Since we had only one nest fail in these 30 observation days, we see that there is a $1/30$, roughly .033, probability of any individual nest failing daily. To calculate the daily success probability, we subtract the daily failure probability from one, which comes out to roughly .967. Now that we have our daily survival probability for this example, since we know that the virginia rail and the sora both have incubation periods of around 20 days, we can raise that probability to the 20th power to estimate the probability of overall nest success. So for this example, $.967^{20}$ is approximately 0.51, which is a much better estimate of the true value of the population described above.

It is also crucial to understand why we research the particular parameters we do, otherwise studies may make statistical sense but not make any sense biologically. Likewise, once we determine what is significant and what is not, we must understand what the significance means.

Of the models researched by Robertson, one of the variables was the age of the nest. We then ask, what affect, if any, does the age of a nest have on it's daily survival probability? There could be two possible answers to this question. We may see that the older a nest is, the longer it is exposed and the less likely it will be to survive each day.

On the contrary, it would be possible that a nest has a higher probability of daily survival the older it gets because it was hidden and constructed very well. These may not be the only positive or negative relationships between age of a nest and its daily survival probability, but we need to consider these before performing any kind of statistical analysis.

Another variable in the models is whether the area the nest was found in was in an impoundment or not. An area that is in an impoundment has a structure, either a dam or a pit, built to control water levels in an area for reasons such as maintaining wetland environments and providing water for livestock. A positive or negative relationship with daily nest survival probabilities would likely mean that human intervention of water levels has an impact.

The final three parameters in the models are water depth change at the location of the nest and nest height change between visits to the nest, and the interactions between them. The water depth was measured at the base of the nest, and the nest height was measured from the base of the nest to the lip of the top. The interactions term in the models is simply the multiplication of the two for each nest. The reason for including the interactions term is that the nest height change depends largely on the depth of the water. Rails have the ability to increase the height of their nests using nearby nesting material, and is typically done to help eggs survive flooding conditions (Pospichal and Marshall 1954). While flooding can cause nest failures, it also makes it less likely that a predator can reach the nests and cause failure through predation (Conway 1995). A strong

correlation with daily nest survival probabilities would suggest that more variability in water levels and nest heights would increase the chances that a rail would survive day to day.

CHAPTER TWO: BAYESIAN METHODS

2.1 Introduction

Scientific papers rely heavily on statistical inferences to validate findings from research. The frequentist method of statistics, the most commonly used approach, is one of many different ways to use data to interpret relationships in science and otherwise. It is known as the “frequentist” method because the philosophy behind it relies on the expected frequency that collected data would yield the same results given that the data were gathered in the same fashion and analyzed in the same way (McCarthy 2011). Although it is the most commonly used, frequentist statistics has more recently come under criticism for allowing for conflicting conclusions to be formed from the same data under different circumstances (McCarthy 2011). Similarly, in order to form a conclusion based on the data, frequentist statistics requires the setting of a controversial “p-value” which is used as a guideline to determine if our hypothesis is accepted or rejected. While no form of analysis is perfect, the abandonment of frequentist methods for the use of Bayesian methods, which take into account perceived prior information, has been on the rise (McCarthy 2011). These methods contrast in both fundamental philosophies and logic (McCarthy 2011). The driving philosophy behind Bayesian methods is based on the “likelihood” of competing hypotheses, where the likelihood function for each individual hypothesis is defined as the probability of the hypothesis being true given the observed

data (McCarthy 2011). Together, the likelihood function and perceived prior information give way to the fundamental philosophy of Bayesian methods, that we are calculating the probability of each competing hypothesis being true given that the data observed is true (McCarthy 2011). Even though the two approaches also tend to realize conclusions that are not largely different from each other when using uninformative priors (McCarthy 2011), there are several reasons which I present later to use Bayesian methods over frequentist (Link and Barker 2010).

2.2 Frequentist Methods Summary

Frequentist methods are based on an approach called null-hypothesis significance testing in performing statistical inference (McCarthy 2011). The central philosophy is that the researcher will assume a null-hypothesis, which is a base statement about a mean or a relationship between two means, is true, and calculate the probability of obtaining a collected data set given that the null-hypothesis is true (McCarthy 2011). First, one is required to define a null-hypothesis and an alternative hypothesis, which is the opposite of the null, in the case the null is rejected. Very often these null-hypotheses are designed to have a low probability of being true in the first place, making their rejection relatively uninformative (McCarthy 2011). These null-hypotheses, deemed trivial null-hypotheses, are used due to the difficulty in forming a non-zero null-hypotheses (McCarthy 2011). For example, a trivial null-hypothesis might be that the difference between the average number of math majors and psychology majors at the University of Maine each year is zero. The two variables in the hypothesis that we would concern ourselves with are the

average number of math majors and the average number of psychology majors at the University of Maine. We may believe it to be extremely likely that the two variables are in fact different, but crafting and testing a null-hypothesis with a more defined estimate is difficult (McCarthy 2011). Next is the gathering of your data, which has essentially the same concerns for all methods of statistics; randomization of sampling and adhering to any assumptions that must be made in your modeling process are critical (McCarthy 2011).

After obtaining the data set so we can test the null-hypothesis, the frequentist method calls for the calculation of a p-value, which is the probability of observing the recorded data set or more extreme data with the assumption that the null-hypothesis is true (McCarthy 2011). This calculation is done by standardizing the data set and using standard deviations which takes into account random variability in the data (Quinn and Keough 2002). Standard deviations for sample data are calculated as:

$$\sqrt{\sum_i \frac{(x_i - \bar{x})^2}{n - 1}} \quad (1)$$

where each x_i is an individual data point from your data set, \bar{x} is the mean of the sample set, and n is the size of the sample set (Quinn and Keough 2002). The sample standard deviation helps set up a probability distribution around the sample mean, which describes the probability of every possible value for a variable we are estimating (Quinn and Keough 2002). Let's say we are studying the number of heads we would observe in 8 coin flips and our null-hypothesis is that we would have a proportion of .25, or 2 heads,

with a standard deviation of 0.15, and our alternative hypothesis is that we would find a higher proportion of heads on average. Suppose we perform the 8 coin flips and find a proportion of .625 heads. The probability distribution for our hypothesis above is shown in Figure (1).

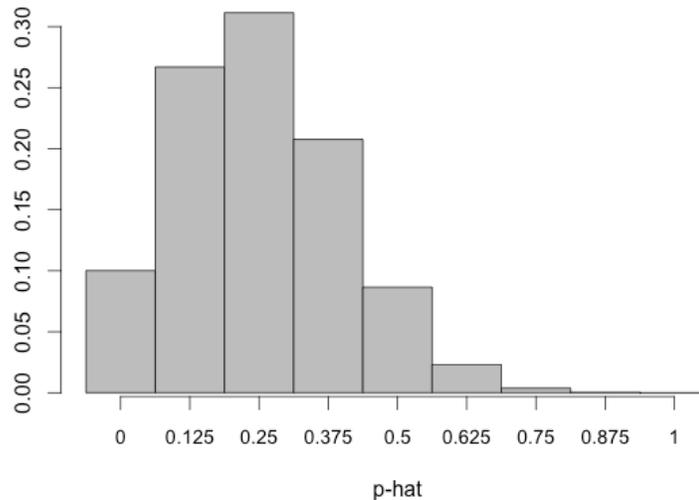


Figure 1. Example of a probability distribution. Assuming our null hypothesis, this distribution represents the probabilities (y values) of seeing certain proportions of heads (x values) on 8 coin flips, assuming that the true mean for the proportion of heads is 0.25.

Using this sample probability distribution, we can calculate the probability that we would obtain this sample data given that the true mean of the proportion is .25. Once the p-value is calculated, you compare it to an arbitrary value, conventionally set to 0.05, to determine whether or not you reject the null-hypothesis (Quinn and Keough 2002). As you can see from Figure (1), the p-value for .625 is extremely low, and is calculated to be approximately .004. Frequentist thinking leads us to say that the probability of obtaining this data set just by random chance, given that the true mean is actually .25, is this

p-value. Since this probability is so small and this data set seems extremely unlikely by random chance, a frequentist thinker would say that we reject the null-hypothesis and the true mean is larger than .25.

However, the p-value is not always so straightforward. Critics claim that the strict choice of 0.05 or any other value to compare with a p-value has no formal basis and therefore exposes itself to higher rates of error (McCarthy 2011). In this example comparing a p-value to 0.05 only means that we are only dissatisfied with the null-hypothesis when we gather data that we would only expect to observe with less than 5% frequency if our null-hypothesis were true (McCarthy 2011). For example, had our p-value in the above example come out to be 0.06, we still have a very low probability of obtaining this data sample by chance, but by conventional standards this is not seen as statistically significant. If our p-value is larger than the chosen comparative value, we can only fail to reject the null hypothesis – it is not proven. If the null hypothesis is rejected due to a small enough p-value, we are forced to accept the alternative hypothesis (Quinn and Keough 2002). This is another point of criticism, as alternative hypotheses that are not carefully considered may be even less representative of the data than the null hypothesis that was just rejected since they are accepted without necessarily confirming it matches the data (McCarthy 2011).

2.3 Bayesian Methods Summary

Bayesian methods differ vastly in almost all of these major areas. Fundamentally,

instead of being concerned with the probability of obtaining a data set given a specific hypothesis, Bayesian thinkers will construct their hypotheses and calculate the probabilities of each hypothesis given that the data collected are a good representation of the real data, known as the population. Involved in this process is the use of prior probability distributions, which is the researcher's prior beliefs about the data (Dennis 1996). Instead of constructing null and alternative hypotheses, in Bayesian analysis you can form multiple hypotheses and compare them (McCarthy 2011). These hypotheses can be represented as hierarchical models in complex cases where we are studying the effects of certain variables on another (such as later in this paper). These models assign parameter values to the data, and the idea is to find the model that most accurately predicts one variable in terms of these parameters. The model hypotheses are most often chosen based on reality and what the researcher believes are the most realistic hypotheses (McCarthy 2011). While there are forms of frequentist methods which set up multiple models to represent hypotheses as well, the underlying thought process that Bayesians are looking for the probability of each hypothesis being true given the data is the major difference between the methods.

Once the set of hypotheses to be tested has been chosen, Bayesian thinkers also take into account prior beliefs about the parameters or means, which is impossible in frequentist analysis (McCarthy 2011). This is another major advantage of Bayesian analysis, as frequently scientists compare their results to that of similar studies done in the past (McCarthy 2011). To incorporate prior belief into the statistical models, a

Bayesian thinker will assign probability distributions to each parameter. Just like the probability distribution described in the frequentist section, these distributions represent what the researcher believes the probabilities are for the true value of the parameter. This can be done with uninformative prior distributions, where the range of any possible value with equal probability is assigned, or with informative prior distributions taken from what is believed to be the true value of the parameters (Link and Barker 2010). In the case of uninformative priors, the models usually have similar results as frequentist methods (Link and Barker 2010). When all the prior distributions have been chosen, the data are gathered similar to any other method, and we use the principle of Bayes' Theorem to calculate a posterior distribution, or the new probability distributions for each parameter (McCarthy 2011).

Bayes' Theorem, derived from the relationship of two related probabilities, is then applied to the prior distributions and the data. Bayes' Theorem states that for discrete hypotheses:

$$Pr(B_i | A) = \frac{Pr(A | B_i) Pr(B_i)}{\sum_i Pr(A | B_j) Pr(B_j)} \quad (2)$$

where $Pr(B_i|A)$ is the probability of an event B_i occurring given event A , $Pr(B_i)$ is the probability of event B_i , $Pr(A|B_i)$ is the probability of event A occurring given event B_i , and all B_j and B_i are mutually exclusive events of a set of possible events B (Link and Barker 2010). For continuous hypotheses:

$$Pr(B_i | A) = \frac{Pr(A | B_i) Pr(B_i)}{\int_0^{\infty} Pr(A | x) Pr(x) dx} \quad (3)$$

where the integral in the denominator represents all possible events in the set B (McCarthy 2011). For the continuous case, these are probability densities, meaning that instead of direct probabilities, they are probability distributions. The coin flipping example from the frequentist section is an example of a continuous hypothesis, because there are an infinite amount of numbers between 0 and 1 that could be the true proportion. An example of a discrete hypothesis might be detecting if a species is present or absent in a habitat; there is a finite number of outcomes that the species is either present or absent. In both cases the denominator on the right hand side of the equation is the probability distribution of A, which is calculated differently depending on the type of hypothesis you are testing (McCarthy 2011). This probability distribution of A is essentially only a scaling factor known as the normalizing constant; that is, the posterior distribution we are looking for is determined by the relationship in the numerator (McCarthy 2011). This theorem is mathematically and logically sound, and it is only the application of it that involves any controversy in Bayesian methods (Dennis 1996). Here, the philosophy is that $Pr(B_i)$ is the probability distribution that the researcher has assigned as their informative or uninformative prior, which can be subjective and leads to the controversy (Cox 2006).

Due to the difficulty of calculating the denominator directly in most cases, hypothesis testing in Bayesian analysis is usually done with Markov Chain Monte Carlo

(MCMC) sampling methods (McCarthy 2011). This is a sophisticated sampling algorithm which will produce a sequence of values where each is derived from the value before it (McCarthy 2011). The goal of the algorithm is to generate values that look ‘random,’ and after a substantial number of samples have been taken the values generally do (McCarthy 2011). These samples can be calculated very easily with statistical software. If you wish to read about the details of how the algorithm works, I recommend the “MCMC algorithms” appendix of McCarthy’s book, “Bayesian Methods for Ecology,” or chapter 4, “Calculating Posterior Distributions,” of Link and Barker’s book, “Bayesian Inference with Ecological Applications.”

The most prominent issue with this method of chaining samples is correlation. Our drawn samples clearly always depend on the previous sample, which can cause problems in the posterior distribution. It is not useful to have these samples that are heavily dependent and linked between samples, because they can be low precision and be inaccurate in estimating the target distribution (Link and Barker 2010). After enough samples as we approach what will be the posterior distribution, we want the correlation to become close to zero, meaning that each subsequent random number appears randomly generated (Link and Barker 2010). To check this, we view an auto-correlation function plot. This plot calculates a standardized correlational value which is expected to approach zero as the chain approaches the posterior distribution. Typically when running these MCMC tests, three chains will be run simultaneously, to compare the correlational values. It is important that all three chains end up looking relatively the same. An

example of a good auto-correlation function plot is shown in Figure 2.

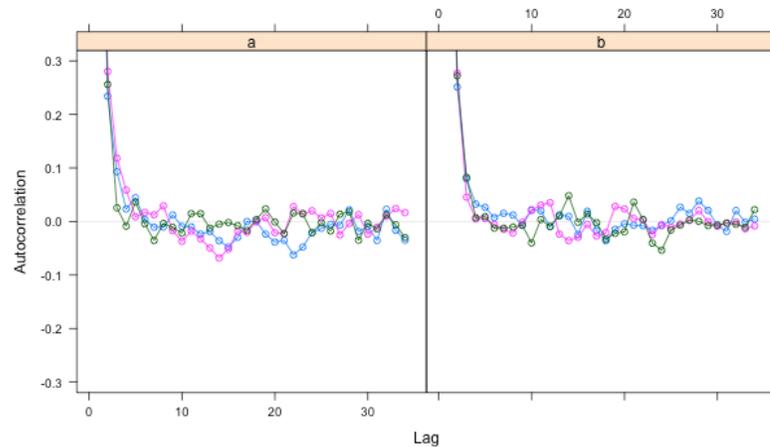


Figure 2. An example of an auto-correlation function plot: You can see all three chains approach zero very rapidly, and continue to vary around 0 randomly. These are good auto-correlation function plots, as there seems to be little correlation between posterior distributions and the first sample distribution.

After ensuring that chains are independent of initial sample distributions, and having run MCMC tests for each of the model hypothesis, it is time to compare the models. This is done simply by using DIC, or deviance information criterion. DIC calculations are the Bayesian alternative to Akaike's information criterion (or AIC), proposed by Spiegelhalter et al. (McCarthy 2011). AIC and DIC values are very similar and when posterior distributions are mostly symmetrical are essentially interchangeable (McCarthy 2011). DIC is calculated by first determining the deviance at the mean of the posterior distributions and can be calculated using statistical software. This calculation provides a gauge at judging what models are the best fit by which ones have the smallest deviance (Link and Barker 2010). However, it is also important to have simplistic models, and the best models are considered the ones with the fewest number of

parameters that provide the most information (McCarthy 2011). This is where DIC comes in, which adds a penalty to models for complexity, where complexity means more parameters. As with deviance, we are looking at models with the lowest DIC values as the best fits (McCarthy 2011). Typically, without this penalty, models with the most parameters would be expected to have the lowest deviance, but it is important to question whether the extra parameters are justified (McCarthy 2011). By adding in the penalty for complexity, a model with lots of parameters must fit particularly well to have a lower DIC value than a competing less complex model.

It is not only of concern which is the best model, but how the models differ amongst themselves. To compare, we look at the Δ DIC, which is the difference between the DIC of a particular model and the minimum DIC of all the candidate models. Table 1 is a table which compares Δ AIC values and how they should be viewed compared to the model with the lowest AIC value; since in most cases DIC values are interchangeable with AIC values, the same comparison rules can apply (McCarthy 2011).

Table 1. Interpretation of the level of support for apparently inferior models relative to the model with the lowest AIC, based on differences in their AIC values (Burnham and Anderson, 2002).

Δ AIC	Degree of support
0-2	Substantial
4-7	Considerably less
>10	Essentially none

Once DIC values are calculated, weights can be calculated and assigned to each model to determine parameter estimates. These weights are normalized values that sum to one and can be interpreted as the probability that particular model is the best of the models that were tested (Johnson and Omland 2004). This is done by calculating the exponential likelihood function and then dividing by the sum of the likelihoods of all the models (Johnson and Omland 2004):

$$w_i = \frac{e^{\left(-\frac{1}{2}\right) (\Delta DIC_i)}}{\sum_i e^{\left(-\frac{1}{2}\right) (\Delta DIC_j)}} \quad (4)$$

Not only are these weights useful when discussing the comparative relevancy amongst models, but they are used to determine the model averaged parameters. Researchers are not only interested in whether or not there are effects in science, but also desire to explain these effects if they are found to be significant. For each parameter inside of each model, we can use the posterior distributions that have already been calculated to do so. Although these distributions vary from model to model, cross averaging them with weights is a direct approach to looking at the overall effects of a single parameter (Link and Barker 2010). Using the weights calculated from the DIC values connects the relevancy of the model and the effect of the parameter to give us a greater understanding. The means for the model average parameters are calculated by (Johnson and Omland 2004):

$$\theta_{map} = \sum_i w_i \theta_i \quad (5)$$

where θ_{map} is the model averaged parameter, and the summation over i is over all the models that contain the parameter θ . Similarly, the standard errors for these model averaged parameters are (Johnson and Omland 2004):

$$SE(\theta_{map}) = \sqrt{\sum_i w_i ((var(\theta_i | g_i) + (\theta_i - \theta_{map})^2))} \quad (6)$$

where $var(\theta_i | g_i)$ is the reported standard error of the distribution of θ_i from model i , squared. By using equations (5) and (6), distributions for the model average parameters are calculated.

To summarize the process of Bayesian thinking in an example, let's take a look at the coin flipping example from the frequentist section. Suppose we did some research and discovered that the accepted proportion of getting a heads when flipping a coin is 0.5 with a standard deviation also 0.5. We now want to test what the probability is that the true proportion of heads is 0.25 given the data we collected. Using the MCMC method described above, we get a posterior distribution with a mean of 0.5018 and a standard deviation of 0.1752. Using these numbers, we calculate an estimated probability of 0.077. This is our new probability of the hypothesis that the true proportion of heads is 0.25. We can compare this to other hypotheses; for example, if we wanted to know the probability of a hypothesis that the true proportion of heads is 0.5, or 4 heads in 8 coin flips on average, we examine the posterior distribution again and have an estimated probability of 0.4959. A Bayesian thinker will compare the probabilities of each hypothesis and make

conclusions based on those results.

2.4 Differences in Methods

The differences between Bayesian and frequentist methods are very clear, and there are advantages and disadvantages to using either. One of the most important advantages to using Bayesian analysis is the ability to combine previous data with each study, making it a much more cohesive process to further research a subject (Dennis 1996). Even without any preemptive knowledge, a Bayesian analysis with uninformative priors will have results similar to what a frequentist's model would (McCarthy 2011). Another major advantage to using Bayesian methods is that the philosophy is more consistent with a natural train of thought. When constructing tests and models in order to understand data, rarely are researchers concerned with “the probability of finding this data *given* that the hypothesis is true” (McCarthy 2011). Realistically, the goal of running the tests is to learn something, which means the true purpose of all the research is to unveil information on the hypothesis. This is exactly what the Bayesian method does, in contrast to frequentist; finding the probability that the hypothesis is true given the observed data (McCarthy 2011). Since this is what researchers truly seek, it follows logically that these methods are more well designed for their purpose (McCarthy 2011). Coming up with multiple hypotheses, represented as models, is also very convenient using this methodology (Dennis 1996). Instead of being concerned with accepted or rejected hypotheses based on an arbitrary probability level, models are compared to each

other as well as averaged to learn more about which one has the best fit, as well as finding good parameter estimates across all models.

If it were that simple, Bayesian analysis would likely be much more widely accepted as the consensus way of analyzing data. There are also controversial drawbacks to these methods, which come almost entirely in the form of the informative priors (Dennis 1996). Just as setting arbitrary probability levels and rejecting null-hypotheses in frequentist statistics can be controversial, informative priors are set up by the researcher's own beliefs, which are not always concrete. Those who oppose Bayesian statistics claim that a disagreement in priors can quite often cause issues in the interpretation of the results (Cox 2006).

No method of statistical inference is perfect and each has its own drawbacks. Bayesian analysis, especially when proper attention is given to the informative priors, can be both useful and convenient in examining relationships in the world. Its use is growing and it can be strongly considered as a viable way of hypothesis testing.

CHAPTER THREE:
COMPARISONS OF MODELS

3.1 Introduction

The goal of this paper is to compare and contrast the use of Bayesian and frequentist methods in the statistical analysis in ecological research. Ellen Robertson wrote a thesis in partial fulfillment of a Degree of Master of Science in Ecology and Environmental Science in May 2012, which discussed the daily nest survival probabilities of two birds in the family *Rallidae*; the Virginia Rail (*Rallus limicola*) and the Sora (*Porzana carolina*), which she groups together to call “rails.” In her research, Robertson used information theoretic methods, which is a form of model hypothesis testing that uses frequentist philosophies, to determine ecological models to explain the probabilities. I examined her data, research and conclusions and used them to perform a Bayesian analysis.

Robertson studied ten wetlands in Maine, five of which were impounded. There, she and other researchers searched for nests of both birds and recorded the date, water depth at the nest, and the nest height (which she defines as the height from the base of the nest to the lip of the nest). The observer would return to the nest once every 3-5 days until the nest was determined successful or failed. Success was defined as at least one hatched egg, where a failure lost all eggs either through predation or nest damage such as flooding. When the age of the nest was not known exactly, it was approximated by using

the number of eggs in the nest and the hatching date for successful nests, and was assumed to be found in the middle of incubation for failed nests.

In Robertson's 16 candidate models, factors that were considered were age of the nest, nest height change, water depth change, whether or not the investigated wetland was impounded or not, and the interaction of nest height change and water depth change.

3.2 Methods

To work on Robertson's models of rail nest daily survival probabilities and reanalyze them using Bayesian methods, I needed to obtain her observational data as well as statistical software which had the capability of running MCMC chains and running Bayesian analysis, and used the statistical software "R" (R Core Development Team 2011). For the Bayesian methods, I wrote models defined in .txt files, which are then loaded into the R workspace to be used in the analysis. These models are defined with prior distributions, and a mathematical definition of the model we are testing. I put all the data together and with all models except for Model 12 ran a burn-in of 1000 samples to be discarded to avoid problems with correlation, following up with a chain of 25,000 samples to be recorded. I then ran a reverse burn-in to observe a final 2,500 samples. The main reason Model 12 required a much larger burn-in, 25,000, was the correlation between MCMC generated samples was too high, and after running the larger burn-in the auto-correlation function produced values much closer to 0. These samples are used to determine a posterior distribution for each parameter which R fed back to me.

Using Robertson's observational data, I first recreated all of her models using uninformative priors. My uninformative priors were a normal distribution with mean 0 and a standard deviation of 1000, which results in an effective distribution from -3000 to 3000 that covers all plausible values for the parameters. I examined each of Robertson's 16 candidate models with a Bayesian perspective. Once the chaining process was completed, I produced auto-correlation function plots, density plots, and trace plots for each parameter, and recorded the means and standard errors associated with each parameter for each model. After this was done for each candidate model, I calculated the DIC values and weights as to compare the fit of the models to Robertson's. Using the weights, I also calculated parameter estimates for each of the variables as well as the standard errors.

Next I searched for informative priors for logistic models on daily nest survival. This process ended up being more difficult as the research done in Robertson's thesis was relatively original. So for prior distributions, I used her results for parameter estimations as informative prior distributions, only multiplying the standard errors by 3. Other than the prior distributions, the calculations and sampling methods were reproduced exactly as they were for the models with uninformative priors.

3.3 Results

Overall, the Bayesian analysis proved to be relatively similar to the frequentist analysis done by Robertson. For the model averaged estimates, none of the parameters,

either from frequentist, Bayesian with informative priors, or Bayesian with uninformative priors were significantly different from each other, although there were a few differences. The standard error for the interactions between water depth change and nest height change, as well as water depth change itself, were larger in both Bayesian analysis tests than the frequentist, suggesting that we are less confident in our estimations of the effect of the variables. Also, Robertson found a significant effect of age while both of my Bayesian analyses included 0 within one standard error of the mean. This suggests that there is a weaker effect of age than was found in the frequentist analysis, notably even more so with the informative prior analysis.

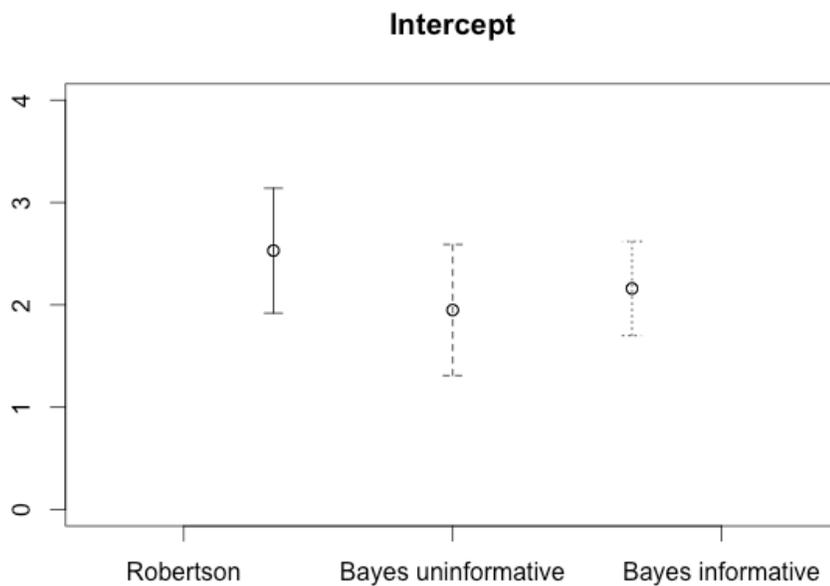


Figure 3. Model-averaged mean plus or minus one standard error for the parameter of intercept in Robertson's models compared to the Bayesian analysis with uninformative and informative priors.

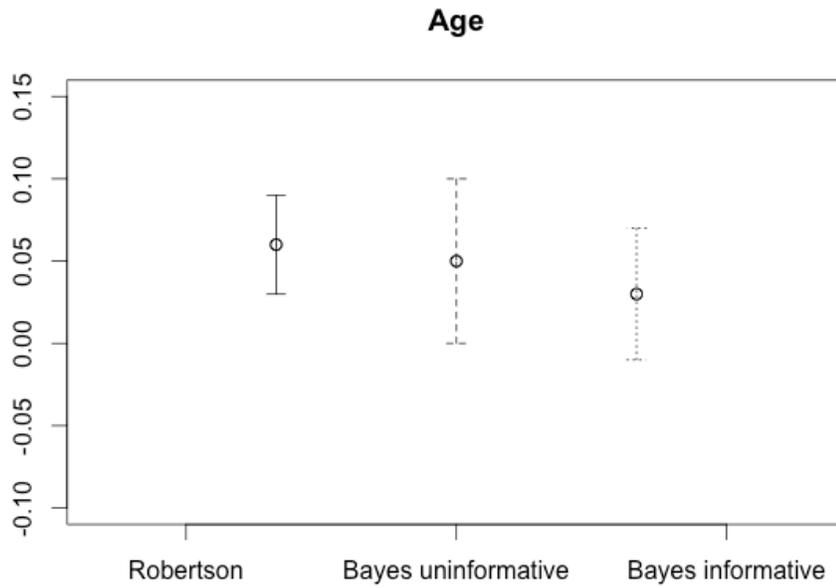


Figure 4. Model-averaged mean plus or minus one standard error for the parameter of age in Robertson's models compared to the Bayesian analysis with uninformative and informative priors.

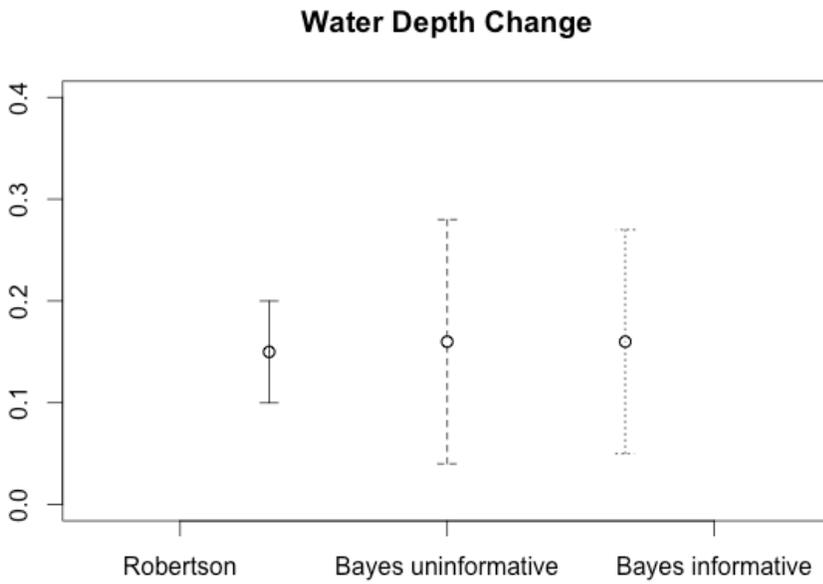


Figure 5. Model-averaged mean plus or minus one standard error for the parameter of water depth change in Robertson's models compared to the Bayesian analysis with uninformative and informative priors.

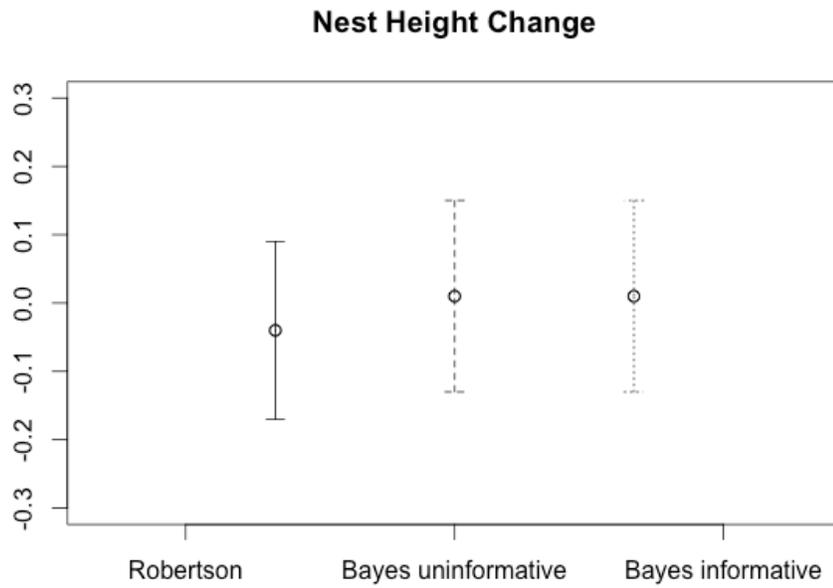


Figure 6. Model-averaged mean plus or minus one standard error for the parameter of nest height change in Robertson's models compared to the Bayesian analysis with uninformative and informative priors.

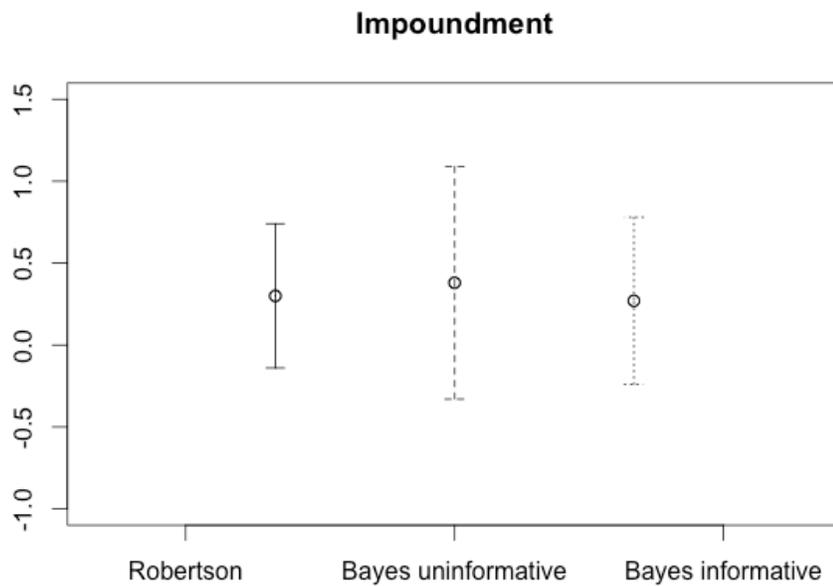


Figure 7. Model-averaged mean plus or minus one standard error for the parameter of impoundment in Robertson's models compared to the Bayesian analysis with uninformative and informative priors.

Water Depth Change * Nest Height Change

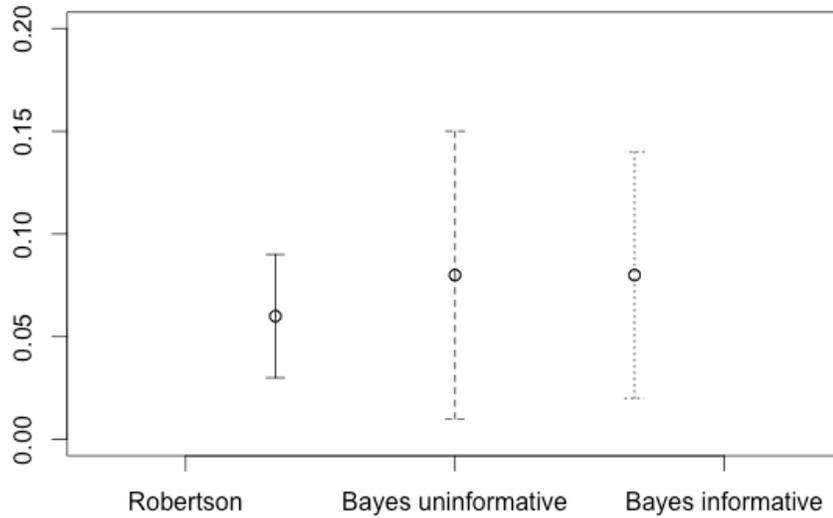


Figure 8. Model-averaged mean plus or minus one standard error for the parameter of the interaction between water depth change and nest height change in Robertson's models compared to the Bayesian analysis with uninformative and informative priors.

Table 2. Model-averaged parameter estimates for Robertson's models and both the Bayesian models with uninformative and informative priors.

Parameter	Robertson's Estimate \pm Standard Error	Bayesian Uninformative Prior Estimate \pm Standard Error	Bayesian Informative Prior Estimate \pm Standard Error
Intercept	2.53 \pm 0.61	1.95 \pm 0.64	2.16 \pm 0.46
Age	0.06 \pm 0.03	0.05 \pm 0.05	0.03 \pm 0.04
Water Depth Change	0.15 \pm 0.05	0.16 \pm 0.12	0.16 \pm 0.11
Nest Height Change	-0.04 \pm 0.13	0.01 \pm 0.14	0.01 \pm 0.14
Impoundment	0.30 \pm 0.44	0.38 \pm 0.71	0.27 \pm 0.51
Water Depth Change * Nest Height Change	0.06 \pm 0.03	0.08 \pm 0.07	0.08 \pm 0.06

Similarly, there was very little variation between the three methods of analysis when it came to the comparisons of the models. In all three methods, Robertson's top model, which combined water depth change, nest height change, age, and the interaction between water depth change and nest height change, was calculated to be the best model with the highest weight. Likewise, Robertson's worst model, which combined impoundment and nest height change, was calculated to be the worst model with the lowest weight also for the Bayesian models. The largest difference in the analysis was the overall differences in the AIC/DIC values. In Robertson's analysis, the top model had a weight of 0.4, which suggested that of the models she tested there was a 40% chance that model was the best, while the weights for the Bayesian analysis were .23 for uninformative priors and .22 for informative priors, suggesting only a 23% and 22% chance respectively that model was the best. The DIC values were much closer together for both of the Bayesian analyses than for the AIC values for Robertson's, as the ΔAIC for her worst model was 10.0, but the ΔDIC for the worst Bayesian models were 6.47 for models with uninformative priors and 6.1 for models with informative priors. This leads to the weights also being much closer together overall, as the variance (a measure of how far the data is spread out) of the weights for Robertson's models was 0.01, but for both Bayesian models was 0.005. The rankings were very similar in all three cases, with R^2 correlational values (with ± 1 being heavily correlated and 0 being not correlated at all) being 0.79 and .71 between Robertson's models and the Bayesian models with uninformative priors and informative priors respectively.

Table 3. The models in order of Robertson's rankings and their parameters. These were used an all three analysis.

Model #	Parameters
1	water depth change + nest height change + age + water depth change * nest height change
2	impoundment + water depth change + nest height change + age + water depth change * nest height change
3	water depth change + nest height change + water depth change * nest height change
4	age + water depth change
5	age
6	impoundment + water depth change + nest height change + water depth change * nest height change
7	impoundment + age + water depth change
8	impoundment + age
9	age + nest height change
10	null (constant intercept)
11	water depth change
12	impoundment + age + nest height change
13	impoundment + water depth change
14	impoundment
15	nest height change
16	impoundment + nest height change

Table 4. Comparisons between models for Robertson's analysis and the Bayesian analysis. Robertson used AIC values in place of DIC values, which are virtually the same in this case. The smaller the AIC/DIC value and the higher the weight, w_i , the higher the probability the model is the best among those listed. (ER) means the column stands for Robertson's analysis, (UN) means the column stands for the Bayesian analysis with uninformative priors and (IN) means the column stands for the Bayesian analysis with informative priors.

Model #	AIC (ER)	Δ AIC (ER)	w_i (ER)	DIC (UN)	Δ DIC (UN)	w_i (UN)	DIC (IN)	Δ DIC (IN)	w_i (IN)	New Rank (UN)	New Rank (IN)
1	160	0	0.40	160	0	0.23	160	0	0.22	1	1
2	162	1.6	0.18	162	1.6	0.10	161	0.7	0.15	3	3
3	163	2.9	0.09	161	0.7	0.20	160	0.3	0.19	2	2
4	164	3.5	0.07	164	3.5	0.04	164	3.5	0.04	6	6
5	164	3.6	0.07	162	1.7	0.10	163	3.0	0.05	5	5
6	165	4.7	0.04	162	1.7	0.10	161	1.1	0.13	4	4
7	165	5.0	0.03	165	4.3	0.03	164	4.3	0.03	10	11
8	166	5.2	0.03	164	4.1	0.03	164	4.2	0.03	9	10
9	166	5.5	0.03	166	5.4	0.02	165	5.1	0.02	13	13
10	167	6.4	0.02	164	3.5	0.04	164	3.6	0.04	7	7
11	167	6.4	0.02	164	3.9	0.03	164	4.0	0.03	8	8
12	168	7.1	0.01	167	6.2	0.01	166	5.9	0.01	15	15
13	169	8.1	0.01	165	4.5	0.02	164	4.3	0.03	12	12
14	169	8.1	0.01	165	4.3	0.03	164	4.1	0.03	11	9
15	169	8.3	0.01	166	5.5	0.01	166	5.6	0.01	14	14
16	170	10.0	0	167	6.5	0.01	166	6.1	0	16	16

3.4 Discussion/Conclusions

We expected both of the Bayesian methods, especially the one with uninformative priors, to have results that were very similar to Robertson's, and they did. All the model averaged parameter estimates were within one standard error of each other, implying that the estimations were all similar. Likewise the top 6 overall models were the same for all three methods, and the rankings between them being so heavily correlated shows that the overall results are the essentially the same. Considering that the best model and the worst model according to AIC/DIC calculations were the same in all three also suggests that the methods come to the same conclusions.

One major difference between the methods was the spread of the AIC/DIC values, resulting in a difference in the spread of the weights as well. It was much more likely that Robertson's best model was the best of those she tested than was the case for either of the Bayesian models. The Bayesian analysis then implies that the models are much more similar to each other than what Robertson's analysis found. From models numbered 10 and higher, her analysis showed almost no support. In the Bayesian analysis, there was not much overall support but considerably higher support that in Robertson's tests seem to have been allocated to the top model. This makes it much less likely that we throw out any of the models from consideration. From the Bayesian tests we see that there appear to be three tiers of interchangeable acceptance; models 1 and 3 appear to be the best, models 2,4,5 and 6 appear to be close, and the rest can all be grouped together. While this is a major difference between the tests, the overall results are extremely similar. We still

accept model 1 as the best model with the highest probability of being the correct model and there were no models that surprised, with the highest change in ranking being model number 7 being ranked 10th and 11th in the uninformative and the informative analysis respectfully.

The inclusion of 0 in the within one standard error of the mean for the age parameter is also an important difference in the models as Robertson's frequentist analysis shows it to have a significant effect on daily nest survival probability but the Bayesian analysis, while not ruling it out, suggests that the evidence of an effect from the age variable is not as strong. More importantly, the Bayesian analysis that used informative priors had a mean even closer to zero. Since we expect the Bayesian analysis with uninformative priors to be similar to frequentist analysis, and the use of informative priors is of high importance to Bayesian methods, this is an important difference. However, in the top rated model, the means and standard errors for the uninformative and informative Bayesian analyses were $.047 \pm .033$ and $.034 \pm .026$ respectively, both intervals excluding zero. This might mean that including models with less support in the model averaging, despite having a lower model weights, could have resulted in larger standard errors.

For the parameters for water depth change, and the interaction between water depth change and nest height change, the standard errors were significantly larger. Unlike the parameter of age, in all three analyses the effect of both variables seems to be significant, but the difference in the standard errors suggests that we know less about how

important the impact they have on daily survival probabilities is.

A small disagreement in the sign of the nest height change parameter, Robertson's being negative and the Bayesian analysis showing it to be slightly positive, can be attributed to the standard error. For all three methods, the standard error was an order of magnitude higher than the estimate for the nest height change parameter, meaning that we are unsure if the true value is positive or negative, despite what the estimates are. While a sign change may seem like a significant difference, after observing the standard errors we see that in reality it is not.

The Bayesian models with the lowest DIC values contained the parameters for the interaction between water depth change and nest height change as well as water depth change. In fact, none of the models without both of these parameters ranked higher than the four models that contained them for either Bayesian analysis. However, there are also models with water depth change that do not contain the interaction term that do not rank as highly, which implies that it is the addition of both of these variables that is important in estimating the daily nest survival probabilities.

We can conclude then that parameter estimations, model rankings, and overall conclusions were nearly identical. For possible further research into the topic, it would have been better to find informative priors from a study separate from the one where the observed data came from. More development on the research of predicting daily nest survival probabilities of rails is necessary to delve better into the informative prior perspective.

BIBLIOGRAPHY

- Alderfer, Jonathan, ed. *Complete Birds of North America*. Washington, D.C.: National Geographic Society, 2006. Print.
- All About Birds*. Cornell University, 2011. Web. 25 Mar. 2013. <<http://www.allaboutbirds.org/>>.
- Billard, R. S. *An Ecological Study of the Virginia Rail and the Sora in Some Connecticut Swamps*. Ames: Iowa State College, 1948.
- Burnham, K.P. and D.R. Anderson. *Model Selection and Multi-Model Inference: a Practical Information Theoretic Approach*. New York: Springer-Verlag, 2002.
- Conway, Courtney J. "Virginia Rail (*Rallus Limicola*)." *The Birds of North America Online*. Cornell Lab of Ornithology, 1995. Web. 25 Mar. 2013. <<http://bna.birds.cornell.edu/bna/species/173/articles/introduction>>.
- Conway, Courtney J., William R. Eddleman, and Stanley H. Anderson. "Nesting Success and Survival of Virginia Rails and Soras." *Wilson Bulletin* 106.3 (1994): 466-73.
- Cox, D. R. *Frequentist and Bayesian Statistics: A Critique*. N.p.: n.p., 2006.
- Dennis, Brian. "Discussion: Should Ecologists Become Bayesians?" *Ecological Applications* 6.4 (1996): 1095-103.
- Dunn, Jon L., and Jonathan Alderfer, eds. *Field Guide to the Birds of North America*. Washington, D.C.: National Geographic Society, 2006. Print.
- Fletcher, Robert. "Loss of Wetlands: How Are Bird Communities Affected?" *Action Bioscience*. American Institute of Biological Sciences, Oct. 2003. Web. 25 Mar. 2013. <<http://www.actionbioscience.org/environment/fletcher.html>>.
- Johnson, Jerald B., and Kristian S. Omland. "Model Selection in Ecology and Evolution." *Trends in Ecology and Evolution* 19.2 (2004): 101-08.
- Kaufmann, Gerald W. "Breeding Ecology of the Sora, Porzana Carolina, and the Virginia Rail, *Rallus Limicola*." *Canadian Field-Naturalist* 103.2 (1989): 270-82.

- Link, William A., and Richard J. Barker. *Bayesian Inference with Ecological Applications*. Amsterdam: Elsevier, 2010. Print.
- Lowther, James K. "Nesting Biology of the Sora at Vermillion, Alberta." *Canadian Field-Naturalist* 91.1 (1977): 63-67.
- Mayfield, Harold F. "Suggestions for Calculating Nest Success." *Wilson Bulletin* 87.4 (1975): 456-66.
- McCarthy, Michael A. *Bayesian Methods for Ecology*. New York: Cambridge University Press, 2011. Print.
- Melvin, Scott M., and James P. Gibbs. "Sora (Porzana Carolina)." *The Birds of North America Online*. Cornell Lab of Ornithology, 2012. Web. 25 Mar. 2013. <<http://bna.birds.cornell.edu/bna/species/250/articles/introduction>>.
- Olson, Storrs L. "A Classification of the Rallidae." *Wilson Bulletin* 85.4 (1973): 381-416.
- Pierson, Elizabeth C., Jan Erik Pierson, and Peter D. Vickery. *A Birder's Guide to Maine*. Camden: Down East, 1996. Print.
- Pospichal, L. B., and W. H. Marshall. "A Field Study of Sora Rail and Virginia Rail in Central Minnesota." *Flicker* (1954): 2-32.
- Quinn, Gerry P., and Michael J. Keough. *Experimental Design and Data Analysis for Biologists*. Cambridge: Cambridge University, 2002. Print.
- R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Robertson, Ellen Padgett. *Responses of Rail Productivity to Water Level Variability and Factors Affecting Rail Broadcast Survey Results*. N.p.: n.p., 2012.
- "Sora." *Natural Heritage Endangered Species Program*. Massachusetts Division of Fisheries & Wildlife, 2010. Web. 30 Mar. 2013. <http://www.mass.gov/dfwle/dfw/nhosp/species_info/nhfacts/porzana_carolina.pdf>.
- "Virginia Rail." *Natural Heritage Endangered Species Program*. Massachusetts Division of Fisheries & Wildlife, 2010. Web. 30 Mar. 2013. <http://www.mass.gov/dfwle/dfw/nhosp/species_info/nhfacts/rallus_limicola.pdf>.

Author's Biography

Sean T. Hardy was born in Danbury, Connecticut on November 30, 1990. He was raised in Norway, Maine and graduated from Oxford Hills Comprehensive High School in 2009 in the top 5% of his graduating class. While originally a Computer Science major, Sean found that he preferred Mathematics and changed majors after the Fall 2010 semester. Contemplating applying for medical schools but unsure of straying far from math and statistics, Sean discussed with his academic advisor Dr. William Halteman the possibilities and chose to apply to graduate schools for biostatistics. Upon graduation he will be studying biostatistics at Boston University with the goal of working as a biostatistician helping in medical research.