UNIVERSITY OF CALIFORNIA

SANTA CRUZ

Estimating Rates of Natural Mortality for Fish Using Bayesian Linear Regression Methods: A Comprehensive Approach

A thesis submitted in partial satisfaction
of the requirements for the degree of

MASTER OF SCIENCE

in

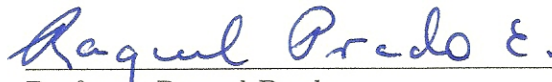STATISTICS AND APPLIED MATHEMATICS

by

Robert Curzon

Applied Mathematics and Statistics

December 2011

The Thesis of Robert Curzon
is approved:

_____
Professor Marc Mangel, Chair

_____
Professor Raquel Prado

ABSTRACT

Rates of natural mortality (M) in marine species are important in stock assessment and fishery management, yet they can be difficult to estimate with any degree of certainty. To obtain a more comprehensive way of estimating M, I compare three size-dependent mortality rate models that are log-linear between M and dry weight (W_{dry}) of the form $\log(M) = \beta_0 + \beta_1 \log(W_{dry})$. I show that Bayesian linear regression methods give more informative estimates over traditional methods of least squares linear regression when looking at datasets of 40 observations or less, or when considering uncertainty in estimation. Using three previously described models from the literature, I construct informative priors that can be used both as the priors for a Bayesian linear regression, and as a stand-alone model for non-Bayesians that is very different from any previous model used before. For example, in previous models the slope term (β_1) is generally close to -0.25, whereas in my constructed prior model the slope term is close to -0.34. I construct a hierarchical Bayesian regression model using the informative priors and investigate how this differs from a model using vague priors. I compare the posterior estimates, analyze the sensitivity of the models to changes in the prior selection and in the number of data points, and then look at the predictive properties of each model. I show that Bayesian linear regression methods give more conservative estimates than the least squares method of regression, and that, for a smaller number of data points, using informative priors is superior to using vague priors.

INTRODUCTION

The rate of natural mortality (M) in aquatic species refers to the rate of death caused by all forms of natural phenomena while excluding fishing mortality (F). There are many differing views on how to estimate M; many of these were reviewed by Vetter (1988). The methods reviewed by Vetter include estimates based on life history, catch analysis, and predation models. Since that time there have been other models introduced and past models have been improved upon.

Several studies have shown that natural mortality rates in fish stocks are linearly related to body size on the log-scale (Peterson and Wroblewski, 1984; McGurk, 1986; Lorenzen, 1996). These studies fit differing linear relationships between natural mortality rate and body size (either by length or weight relationships); however there is no clear way to determine which of these models is best to use in general. In order to obtain a more robust model for estimating size-dependent natural mortality rates, I construct a model that combines information from three previously known models and use it as the basis for the priors in a Bayesian linear regression model. In particular, I take three prominent size-dependent natural mortality rate models from the literature (Gislason et al 2010, McCoy and Gillooly 2008, Peterson & Wroblewski 1994) to infer priors on the intercept and slope terms $(\beta_0 \ and \ \beta_1)$ of the overall regression model $y = \beta_0 + x\beta_1$, where $y = \log (M)$, and $x = \log(W_{dry})$. The prior for each term is chosen to be normal with the mean and variance terms based on the three models I have selected. The mean terms of the prior on $\boldsymbol{\beta}$ are weighted averages of the means of the model parameter estimates, and the prior standard deviation terms are weighted

averages of the standard error of measurements (SEMs) for the model parameter estimates. I

then compute the posterior distributions for the slope and intercept terms using data compiled

by Lorenzen (1996). I then compare the Bayesian regression model using the informative priors

with a similar model that uses vague priors, meaning that I still have normal prior on β, but that

I set prior means to be zero and the prior standard deviations to be arbitrarily large at 1000.

This method leads to an 'all inclusive' approach for estimating M, with credibility regions

that are conservative, taking into account all of the uncertainty within each model being

included. This approach also allows for future updating of the model (given a new dataset) by

using the latest calculated posterior distribution as the new prior distribution.

In section 1, I give a brief overview of Bayesian statistics and describe how Bayesian

methods are used both as a way of reconciling incompatible measurements (Press, 1997), and

as a tool for analyzing data in conservation biology (Wade, 1999). In section 2, I describe the

von Bertalanffy Growth Function (VBGF), its properties, and how it is used in mortality rate

models. Section 3 I show how M is important in stock assessment. In section 4, I lay out various

mortality rate estimation methods and models. In section 5, I consider three size-dependent

natural mortality rate models, converting them as necessary so they are on the same allometric

scale and showing how they give very different linear models for estimating M. In section 6, I

give the setup for two hierarchical Bayesian regression models, one which employs informative

priors constructed from the three models given in section 5, and the other which employs

vague priors. Section 6 also contains an overview of the Lorenzen (1996) data that is used in the

model and gives some analysis and criteria used in the cleaning of the data. In section 7, I

compare the informative prior model and vague prior model, giving some sensitivity and posterior predictive analysis. I utilize a hypothetical dataset to illustrate that for a smaller number of observations the informative prior model gives very different estimates from the vague prior and least squares regression models.

1. BAYESIAN METHODS

1.1 A Brief Overview of Bayesian Statistics

Bayesian statistics is based on Bayes theorem (1763) and is different from traditionally used statistical methods (Jeffreys 1939). Bayesian methods allow one to calculate the probability of a parameter having a certain value that is based upon the data that is observed. The traditional (frequentist) statistical paradigm of statistics is based on calculating the probability of the observed data given a particular value of the parameter. For example, given that the null hypothesis is true, the p-value is the probability of observing data in the long-run with repeated sampling, which is more extreme than the data that has been observed. For Bayesians, the data can give direct information on what the parameter value is and allows for probability statements to be made about the value of that parameter. Both branches of statistics use the likelihood function, but they use them in different ways. Bayesians calculate the posterior probability distribution of a parameter by integrating the product of the likelihood and the prior distribution. The prior distribution is a predetermined probability distribution of the parameter based upon past knowledge, previous data, or intuition.
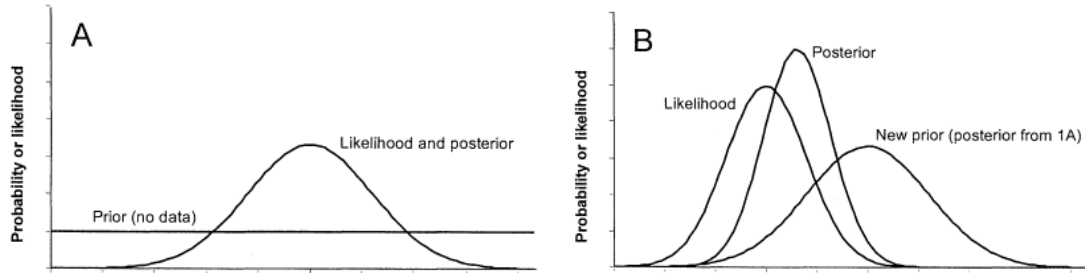
Figure 1. A. Uniform prior with normal likelihood and resulting posterior distribution.
B. Informative normal prior with normal likelihood and resulting posterior distribution. Wade (1999)

In Figure 1.A. I see that using a uniform prior yields a posterior distribution that is essentially equivalent to the likelihood, which is the basis for the frequentist estimate. In Figure 1.B. the prior is chosen to be informative and has a substantial effect on the posterior distribution. After the posterior is calculated all statistical inference is made from it (Wade 1999).

1.2 Bayesian Methods and Incompatible Measurements

One example which shows how Bayesian methods are used is given by Press (1997). He notes that the various independent measurements of Hubble's constant are in fact incompatible, and he suggests using Bayesian methods to "construct a well-posed statistical framework that allows apparently incompatible measurements to be combined in a useful way" (Press, 1997; pg 51).

Press gives the 'traditional' method for combining experimental measurements, which is the average, divided by the variance of each particular observation. Using the values of the individual observations $H_1, \dots, H_N$ with standard deviations $\sigma_1, \dots, \sigma_N$ the tradition method for obtaining a best estimate (H_0) is given in Eq. (1).

$$H_0 = \frac{\sum_{i=1}^{N} H_i/\sigma_i^2}{\sum_{i=1}^{N} 1/\sigma_i^2} \tag{1}$$

Press argues that this is not a desirable method and makes a case for using Bayesian methods instead.

He proposes the following assumptions and notations: that the probability of measurement i being correct is given by p_i and that the probability of a measurement being incorrect is $(1 - p_i)$, that $\boldsymbol{v}$ is a vector of length N where the ith component is one or zero based on whether or not the ith measurement is correct or incorrect, that $P(H_i)$ is the prior probability, D is the data, and that P_G and P_B are the probability distributions of "good" and "bad" measurements .

Applying Bayes rule, Press derives the probability for the desired value H_0 given the incompatible data D:

$$P(H_0|D) \propto P(H_0) \sum_p P(p) \sum_v \left[\prod_{v_i=1} P_{G_i} p\right] \left[\prod_{v_i=0} P_{B_i}(1-p)\right]$$

$$\propto P(H_0) \sum_p P(p) \prod_i \left[p P_{G_i} + (1-p) P_{B_i}\right] \tag{2}$$

Assuming the probability p_i for each experiment is a vector $\boldsymbol{p}$, Eq. (2) becomes:

$$P(H_0|D) \propto P(H_0) \sum_p P(p) \prod_i \{\int_0^1 dp_i P(p_i) \left[p_i P_{G_i} + (1-p_i) P_{B_i}\right] \tag{3}$$

Using a uniform prior for p_i gives:

$$P(H_0|D) \propto P(H_0) \prod_i \frac{1}{2} \left(P_{G_i} + P_{B_i}\right) \tag{4}$$

The coefficient of $1/2$ inside the product is the average of the $p_i s$, and is in the place of the integral over p in Eq. (3).

The results achieved by Press shows that the Hubble constant has a distribution with a 95% confidence interval of $66 < H_0 < 82 (km\ s^{-1} Mpc^{-1})$ and the probability that all the measurements were wrong was 0.3%.

Press illustrates how Bayesian methods can be useful in cases where there are incompatible or conflicting measurements. Taking a Bayesian approach as my framework I will combine the mortality rate models that are seemingly incompatible to obtain a model that is comprehensive and robust.

1.2 Bayesian Methods in Conservation Biology

Bayesian methods and statistical inference are a novel way to approach problems in conservation biology. Wade (1999) contrasts Bayesian methods with traditional hypothesis testing techniques by looking at trend analysis of two hypothetical populations. Wade points out that standard statistical hypothesis testing procedures may not detect certain aspects of the data and that in some cases can be misleading due to the underestimating of uncertainty. The fact that Bayesian methods incorporate uncertainty more thoroughly enables decisions to be made with more care.

Wade gives an example of Bayesian linear regression to highlight the differences between Bayesian and frequentist methods. Wade analyzes the trends of 10 years of abundance data for two hypothetical populations. He shows that using the frequentist test, the slope for population 1 (Figure 2.A.) is significantly different from zero at the 0.05 level, with a p-

value of 0.048, but that the test for population 2 (Figure 2.B.) showed that there in the slope

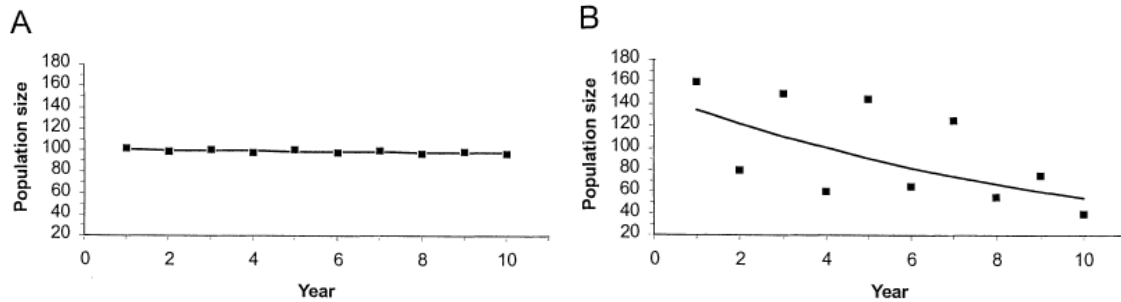was not significantly different from zero with a p-value of 0.053.

Figure 2. Log abundance data for two hypothetical populations (Wade 1999).

Wade points out that, in contradiction to the frequentist findings, population 1 appears

to be stable and that population 2 appears to be declining and is therefore a greater

conservational problem. Population 2 is estimated to be declining at a rate of 10% per year with

large uncertainty, but population 1 is fairly precisely estimated to be declining at a rate of

0.33% per year. The frequentist method did not catch which population is actually at greater

risk of significant decline due to the uncertainty in the data from population 2.

In using Bayesian linear regression methods, Wade shows that the posterior

distributions for the slope terms of the two populations differ dramatically (Figure 3).
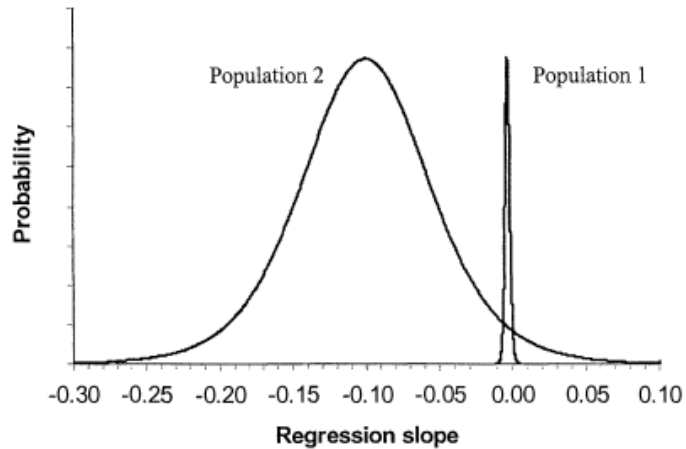
Figure 3. Regression slope posterior distributions for populations 1 and 2 (Wade 1999).

The intuition from the posterior distributions is the same as before, that population 2 is at the greatest risk. He also points out that if a 5% decline rate were the threshold for concern then it could be asked: "what is the probability that the population is declining faster than 5%?" For population 1 the probability is zero and for population 2 the probability is 0.86. This shows that frequentist methods often fail to give useful and/or realistic results, and that bayesian methods are often preferable for many biological problems.

2. THE VON BERTALANFFY GROWTH FUNCTION

The von Bertalanffy Growth Function (VBGF) and its parameters are an essential part of many natural mortality rate models as well as many other aspects of stock assessment. Therefore, to gain a basic understanding of the mortality rate models presented here, one must first gain an understanding of the VBGF and the meaning of its parameters. With this in mind I will first show a brief derivation and explanation of the VBGF and its parameters.

2.1. Derivation of the von Bertalanffy Growth Function

Here I give a general overview of the basic background and derivation of the von Bertalanffy Growth Function (VBGF) (Mangel, 2006; Essington et al, 2010). The original version of the VBGF was introduced by von Bertalanffy in 1938 and was given in the differential equation form

$$\frac{dW(a)}{da} = Anabolic\ factors - Catabolic\ factors \tag{5}$$

where $W(a)$ is the weight at age a. Assuming that the anabolic factors are proportional to the surface area of the organism I have that the anabolic factor $= AL(a)^2$, where A is the scaling factor with units of $kg/year \cdot cm^2$. It is also assumed that the catabolic factors are dependent on the metabolism of the organism and is therefore proportional to the volume of the organism. I assume that the volume is approximately spherical so the catabolic factors $= CL(a)^3$, where C is the scaling factor with units of $kg/year \cdot cm^3$. Putting the anabolic and catabolic equalities into Eq. (5) I obtain

$$\frac{dW(a)}{da} = AL(a)^2 - CL(a)^3 \tag{6}$$

Using the allometric relationship $W(a) = \rho L(a)^3$ I can solve for $dW(a)/da$

$$\frac{dW(a)}{da} = 3\rho L(a)^2 \frac{dL(a)}{da} \tag{7}$$

and setting Eq. (6) equal to Eq. (7) I obtain

$$3\rho L(a)^2 \frac{dL(a)}{da} = AL(a)^2 - CL(a)^3 \tag{8}$$

Dividing through by $3\rho L^2$ in Eq. (8) gives

$$\frac{dL(a)}{da} = \frac{A}{3\rho} - \frac{C}{3\rho}L(a) \tag{9}$$

Defining $k = C/3\rho$ and $L_\infty = A/C$ (where L_∞ is the asymptotic size) results in

$$\frac{dL(a)}{da} = \frac{C}{3\rho}\left(\frac{A}{C} - L(a)\right) = k(L_\infty - L(a)) \tag{10}$$

Setting the initial condition $L(0) = L_0$ and solving the differential equation in Eq. (10), I obtain the following result

$$L(a) = L_0 e^{-ka} + L_\infty(1 - e^{-ka}) \tag{11}$$

An equivalent formulation of the solution of Eq. (11) is

$$L(a) = L_\infty\left(1 - e^{-k(a-a_0)}\right) \tag{12}$$

2.2 VBGF Variables and Their Properties

Now consider some of the properties and relationships between the VBGF parameters. The variables that I will be using are summarized in Table 1.

Table 1. Summary of Variables

Parameter	Definition	Units
$L(a)$	Fish length at age a	cm
L_∞	Mean length at maximum age	cm
$W(a)$	Fish mass at age a	g
$W_{dry}(a)$	Dry mass at age a	g
W_∞	Mean mass at maximum age	g
k	Growth coefficient	$year^{-1}$
$M(a)$	Natural mortality rate at age a	$year^{-1}$

The relationships between the allometric variables $L(a)$, $W(a)$, and $W_{dry}(a)$ are

$$W(a) = \rho \cdot L(a)^3; \tag{13}$$

$$W_{dry}(a) = c \cdot W(a) \tag{14}$$

In Eq. (13) ρ is the proportionality constant between length and weight, and in Eq. (14), c is the proportionality constant between "wet" weight ($W(a)$) and dry weight. In choosing the proportionality constants c and ρ, one can find many applicable values depending on the species or subset of species being investigated. Here I use values that are applicable across fish species, but one could also use specific values for a given species depending on the analysis at hand. I use the mean length-weight proportionality constant $\rho = 0.022$, given by Letourneur (1998) using data from 28 marine fish species. McGurk (1986) and Peterson & Wroblewski (1984) both assume that the dry weight of a fish is 20% of the original wet weight, so I take $c = 0.2$. Given this I will used the following length-to-weight, and weight-to-dry weight conversion formulas

$$W(a) = 0.022 \times L(a)^3 \tag{15}$$

$$W_{dry}(a) = 0.2 \times W_{wet}(a) \tag{16}$$

From Eqs. (15) and (16) I attain

$$L(a) = \sqrt[3]{\frac{W(a)}{0.0044}} \qquad (17)$$

To see how the VBGF parameters interact with one another I consider a few scenarios which give some insight into how the various parameters relate to one another. Letting $k = (0.1, 0.15, 0.3.0.6)$, $L_\infty = 100cm$, and running age from 0 to 40 years, I see how k affects length at age (Figure 4).
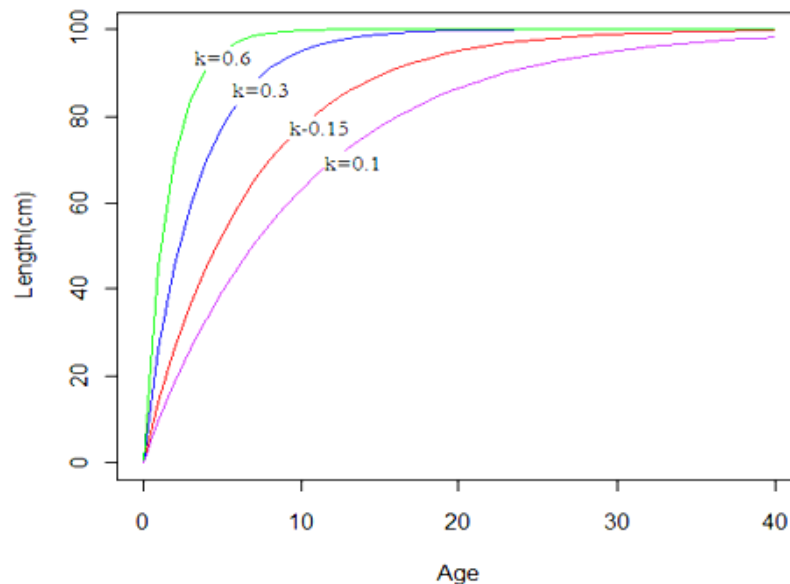

Figure 4. VBGF Length-at-age plot with varying **k** values

This shows that lower values of k give slower rates of growth and higher values of k give faster rates of growth, as expected.

VBGF variables are used in many mortality rate models since it has been shown that rates of mortality depend on the particular life stages of fish. Since life stages and behaviors

within them are integrally tied to body size, growth rate and maximum age and length/weight
the VBGF parameters are ideal for mortality rate estimation (Pauly 1980).

3. THE IMPORTANCE OF M IN STOCK ASSESSMENT

The estimation of M is an important aspect of fishery management and stock
assessment and bet it is one of the most difficult elements to obtain in a reliable and consistent
way (Hewitt & Heonig, 2005). Estimates of M are important in the Beverton and Holt (1957)
model, playing a critical role in computing Maximum Sustainable Yield (MSY) and Optimal
Sustainable Yield (OSY) which are used by fishery managers to set seasonal and annual fishing
quotas, as well as setting size limits on mesh size and determining the minimum/maximum
sizes of fish that can be kept. These estimates are therefore critical in predicting population
trends in fish stock populations in the planning of conservation efforts and the measuring of
their effectiveness.

3.1 The Role of M in Optimal Age Maturity

It is known that as growth rate (k) increases, the optimal age of maturity (a_m)
decreases, and as M decreases, a_m decreases as well. Mangel (2006) gives that

$$a_m = \frac{1}{k}\log\left(\frac{M + bk}{M}\right) \tag{18}$$

where b is the allometric parameter which I will assume is equal to 3. Plotting a_m verses k for
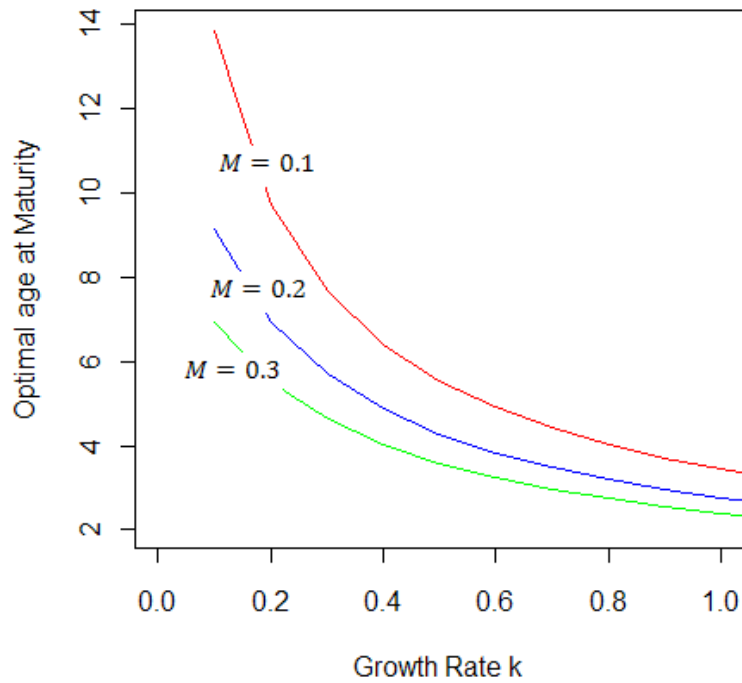various values of M I obtain Figure 5.

Figure 5. Optimal age at maturity as a function of k for $M = 0.1, 0.2, 0.3$

Here I see that M is very influential in effecting optimal age at maturity especially for lower values of k which are more common in nature.

3.2 The Role of M in the Calculation of Yield Per Recruit (YPR)

In a YPR model, taking $N(a)$ as the number of individuals at age a, and $M(a)$ and $F(a)$ as the natural and fishing mortality rates for individuals at age a, then the number of fish at age a is

$$N(a)e^{-(M+F(a))} \tag{19}$$

where $M + F(a)$ is the total mortality. Assuming a 'knife edge' fishing mortality I set $F(a) = 0$ for fish less than age a_r (age of recruitment to the fishery), and for fish of age a_r or greater I assume $F(a) = F$, where F is a constant. So I have that

$$N(a+1) = N(a)e^{-M(a)}; \quad for \ a = 0,1,2, \dots, a_r - 1 \tag{20}$$

$$N(a+1) = N(a)e^{-(M(a)+F)}; \quad for \ a = a\_r \ , \dots, 9 \tag{21}$$

From Eq. (19) I can conclude that the number of fish of age a that die is given by

$$N(a)\left(1 - e^{-(M(a)+F)}\right)$$

If I assume that the fishing catch rates are proportional to the fraction of total mortality

$M(a) + F$, I can conclude that the fraction of fish caught is proportional to $F/(M(a) + F)$. So

the yield of fish of age a is

$$Y(a_r, F) = \frac{F}{M(a) + F}\left(1 - e^{-(M(a)+F)}\right)N(a)W(a) \tag{22}$$

Following a particular age cohort through time, and fixing $N(0) = N_0$, I have that the YPR for a

single cohort is

$$Y(a_r, F) = \frac{1}{N_0}\sum_{a=a_r}\frac{F}{M(a) + F}\left(1 - e^{-(M(a)+F)}\right)N(a)W(a) \tag{23}$$

To see how YPR is effected by variation in $M(a)$, I use a simulation study, plotting

various YPRs given various values of M. I assume that $a_r = 3$ and set $N_0 = 500{,}000$, $W(a) =$

$0.022L(a)^3$, $a_0 = 0$, $a_{max} = 10$, $L_0 = 1$, $L_\infty = 50$, and $k = 0.25$. I simulated various lengths

for given ages $a = 0, \dots, 10$ and held $M(a) = 0.1, 0.2, 0.3$. The resulting simulations for YPR vs
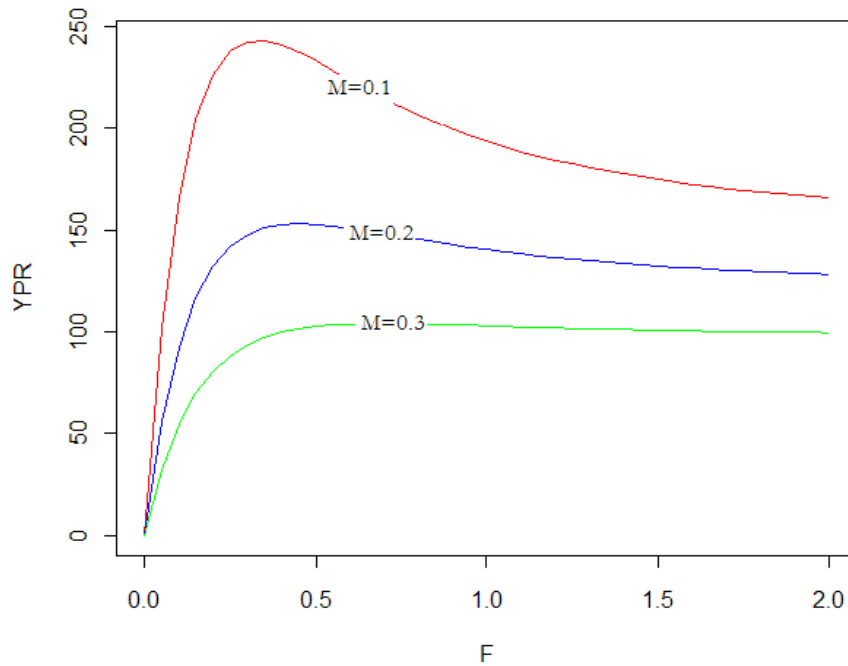
F are seen in Figure 6.

Figure 6. YPR vs fishing effort (F) for $M = (0.1, 0.2, 0.3)$

This shows how natural mortality has a significant impact on yield. It is also apparent that for increasing mortality rate the yield is decreased as there are fewer fish available to catch. It is also apparent that the point at which maximum yield is obtained, denoted by F^*, is shifting to the right. For the three values of M, $F^* = (0.35, 0.4, 0.7)$.

Assuming that M is constant over a lifespan is not considered to be prudent since M varies with age and/or size. In general, natural mortality in fish is higher for younger fish and decreases as the fish grows larger reaching a fairly steady state up until the fish approaches maximum age whereby the mortality rate increases exponentially (Jennings et al, 2001). Given that M and the length of a fish are tightly bound together, I consider

$$M(a) = m_0 L(a)^{-m_1} \tag{24}$$

To see how m_0 and m_1 each effect YPR, first I fix m_1 at 0.25 in Eq. (24) and vary m_0 between 0.1 and 0.7 and plot YPR vs F (Figure 7).
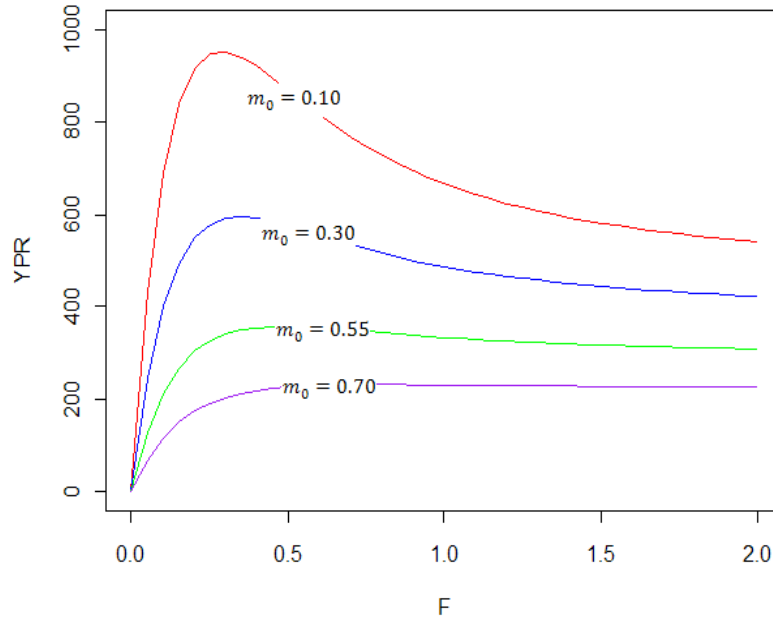

Figure 7. With $m_1 = 0.25$ and $m_0 = (0.10, \ 0.30, 0.55, 0.70)$, I Use Eq. (24) in the YPR formula (Eq.(23)) to assess how YPR depends on m_0.

As m_0 increases YPR decreases. This makes sense as one considers that as more fish are dying with increased natural mortality, there are fewer fish remaining to be able to catch. I also see that F^* is shifting to the right as m_0 increases. I get that $F^* = (0.3, 0.35, 0.45, 0.80)$

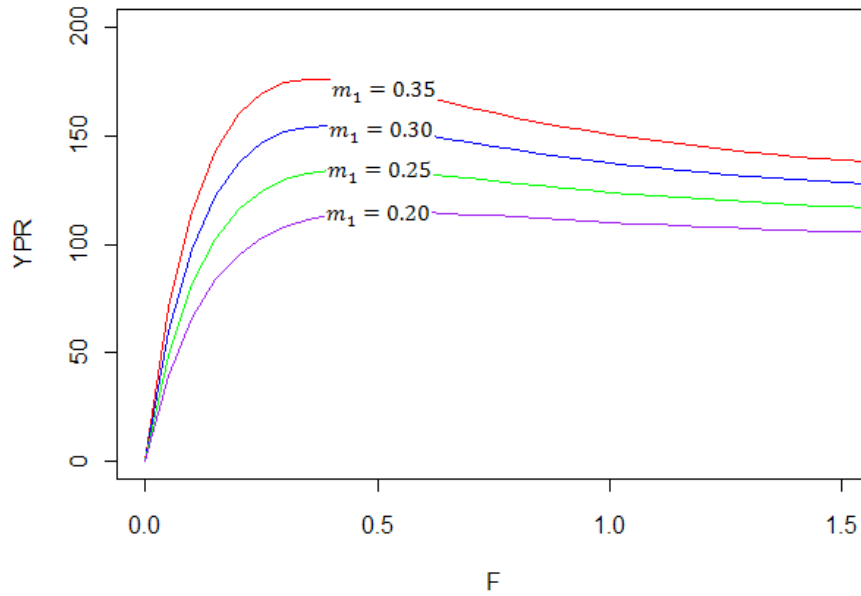Fixing m_0 at 0.5 in Eq. (24) and varying m_1 I obtain Figure 8.

Figure 8. With $m_0 = 0.5$ and $m_1 = (0.35,\ 0.30, 0.25, 0.20)$, I Use Eq. (24) in the YPR formula (Eq. (23)) to assess how YPR depends on m_1.

As m_1 increases, YPR increases as well due to the fact that m_1 is a negative exponent variable in Eq. (24). In this case, F^* shifts to the right as m_1 decreases. I get that

$$F^* = (0.35, 0.40, 0.45, 0.55).$$

There are significant differences in the degree to which YPR is affected by m_0 and m_1 as seen in Figures 7 and 8. YPR appears to be more sensitive to changes in m_0 and less responsive to changes in m_1. Due to these differences it makes sense to estimate mortality based upon Eq. (24) and to try to find separate estimates for m_0 and m_1 so that the variation in each is considered carefully. Another reason to use Eq. (24) as the basis for estimating mortality rate is the fact that mortality rates are not constant for different sized fish of the same species.

4. A COLLECTION OF MODELS FOR ESTIMATING M

There are many ways of estimating M in marine species; some use anatomical traits such as gonad size or vertebral growth rings, others use life history traits such as age, age of maturity or maximum age, and some use size-dependent models that usually based upon VBGF variables or length and weight data.

4.1 Models Based on Anatomical Traits

Gunderson and Dygert (1988) give a method that relates reproductive investment with mortality rate as measured by gonadal mass in relation to total body mass. This relation is termedthe gonado-somatic index (GSI). They give the following estimate for M:

$$M = 0.03 + 1.68 \cdot GSI \tag{25}$$

No units were given for this estimate but since it is a mortality rate I can assume it have units of $(year^{-1})$.

There are other models based on anatomical traits such as vertebral growth rings and various other anatomical features. These methods require sampling methods that take more time and have higher costs than other methods based on life history traits or body size.

4.2 Models Based on Life History Traits

Hewitt and Hoenig (2005) give a method for estimating M based on maximum age (a_{max}). They use the estimate as given in Eq. (26).

$$M \approx \frac{4.22}{a_{max}} (year^{-1}) \tag{26}$$

There are differences in how to estimate or define maximum age. In general, maximum age is defined as the age at which 1% or 5 % of the population is as old or older.

Using the Beverton and Holt invariants, Jensen (1996) gives the estimate for natural mortality rate as

$$M = \frac{1.65}{a_m} (year^{-1}) \tag{27}$$

where a_m is age at maturity in years.

4.3 Models Based on Body Size

Mortality rate models based on anatomical and life history traits give constant estimates for M. Since mortality rates have been shown to vary with size (Pearcy, 1962; Cushing 1974) due to size-dependent predation and other factors (Ursin 1967, Ware 1975), constant estimates for M are not biologically realistic in most cases. There are many size-dependent mortality rate models that typically depend on length or weight.

Peterson & Wroblewski (1984) give an estimate for M that depends on the dry weight of the fish as

$$M = 1.9W_{dry}^{-0.25} \tag{28}$$

McCoy & Gillooly (2008) give the body mass dependent, and temperature corrected mortality rate for fish as

$$M = 1.17 W_{dry}^{-0.27} \tag{29}$$

McGurk (1986) shows that fish larvae and eggs have 5 to 10 times higher mortality rates than juvenile and adult fish. He finds that larvae and eggs have a mortality rate of

$$M = 2.2 \times 10^{-4} W^{-0.85} \tag{30}$$

For juvenile and adult fish, McGurk (1986) uses Peterson & Wroblewski's (1984) model given in Eq. (28). He attempts to show how the discrepancies between Eq. (30) and Eq. (28) can by accounted for when 'spatial patchiness' is considered.

Gislason et al (2010) set a rigorous criterion on which studies they include into the model in order to ensure statistically meaningful model fit to the relevant data. They explore the same relationship between natural mortality and body size. They recommend using the model

$$\log{(M)} = 0.55 - 1.61 \log\big(L(a)\big) + 1.44 \log(L_\infty) + \log(k) \tag{31}$$

This is different in that this model depends on length at age and VBGF variables.

5. COMPARING THREE SIZE-DEPENDENT NATURAL MORTALITY RATE MODELS

I focus on three size-dependent models; setting $y = \log\big(M(a)\big)$, and $x = \log\big(W_{dry}(a)\big)$, the models are

$$y = a - b\,log\big(L(a)\big) + c\,log(L_\infty) + log(k) \text{ ; Gislason et al (2010)} \tag{32}$$

$$y = 1.17 - 0.27x; \text{ McCoy \& Gillooly (2008)} \tag{33}$$

$$y = 1.92 - 0.25x; \text{ Peterson \& Wroblewski (1984)} \tag{34}$$

The estimates for a, b, and c in Eq. (32), along with their corresponding standard error of measurements (SEMs), are given in Table 2.

Table 2. Gislason et al (2010) Model Parameter Estimates

Parameter	Estimate	SEM
a	0.55	0.27
b	-1.61	0.14
c	1.44	0.12

I incorporate the standard errors of the estimates for Eq. (32) into the Bayesian analysis so that all of the uncertainty is taken into account.

Since Model 1 is not in the dry weight allometric measure, I convert Model 1 using general relationships between length and weight as given in Section 1.2. Inputting Eq. (17) into Eq. (32) I obtain

$$y = 0.55 - 1.61 \, log \left(\sqrt[3]{\frac{W_{dry}(a)}{0.0044}} \right) + 1.44 \, log(L_\infty) + log(k) \tag{35}$$

Simplifying Eq. (35) I obtain

$$y = 0.55 - \frac{1.61}{3} log\big(W_{dry}(a) - log(0.0044)\big) + 1.44 \, log(L_\infty) + log(k) \tag{36}$$

So finally I have

$$y = 1.44 \, log(L_\infty) + log(k) - 2.38 - 0.54 \cdot log\big(W_{dry}(a)\big) \tag{37}$$

Thus, the following models are my final versions:

$$\textbf{Model 1: } y = [1.44\, log(L_\infty) + log(k) - 2.38] - 0.52x \tag{38}$$

$$\textbf{Model 2: } y = 1.17 - 0.27x \tag{39}$$

$$\textbf{Model 3: } y = 1.92 - 0.25\text{x} \tag{40}$$

Note that these are of the form:

$$y = \beta_0 + \beta_1 x \tag{41}$$

Now that all models have dry mass as the allometric measure, the models can now be directly compared with one another. Given the large differences in the slope and intercept terms between these models, and the fact that they cannot all be correct, I will use Bayesian methods, particularly Bayesian linear regression, to reconcile these seemingly incompatible differences.

6. A BAYESIAN LINEAR REGRESSION MODEL

To reconcile the three linear regression models having three different slopes and intercept estimates, I use a Bayesian linear regression to obtain combined estimates for the slope and intercept terms. One method would be to use a Bayesian regression mixture model. This would involve modeling the data as

$$y = \sum_{i=1}^{3} \pi_i\, N(\boldsymbol{X\beta_i}, \boldsymbol{\Sigma}) \tag{42}$$

where π_i is the weight given to the ith model for $i = 1,2,3$. Using this as my data generating mechanism, I would then place Dirichlet priors on the π_is, normal priors on the $\boldsymbol{\beta_i}$s, and Inverse Wishart prior on $\boldsymbol{\Sigma}$. This would mean that my posterior estimates for β_0 and β_1 would each be a

mixture of normals (Fraley, 2007; Zhang, 2004). In turn this would mean that my estimate for

the rate of natural mortality that depends on those estimates would be multimodal instead of
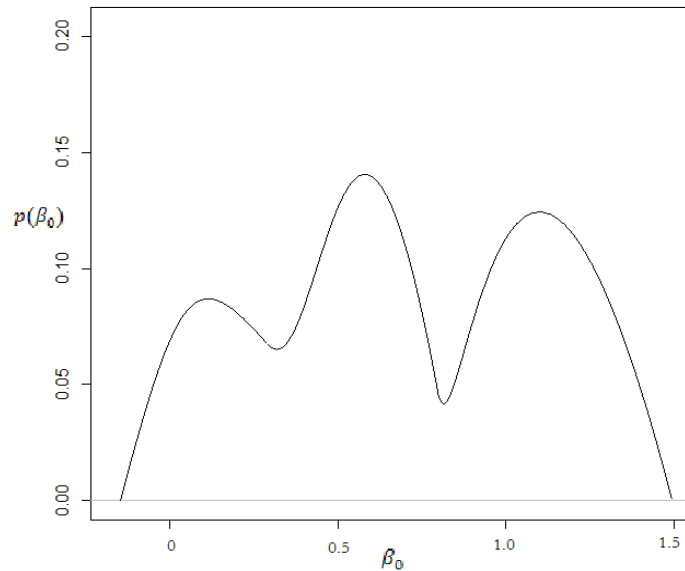
unimodal.

Figure 9. Hypothetical posterior distribution for

$\boldsymbol{\beta_0}$ being a mixture of Guassians.

If this were the case, it would mean that one or more of the models given in Eqs. (38)-(40) is

considered to be "true" for estimating rates of natural mortality, and that there are essentially

different populations with different mortality rate distributions. Instead, I will assume that

these three models are estimates of the overall model that is unknown. This assumption leads

us to assume that the posterior distributions for β_0 and β_1 ought to be unimodal, therefore, the

model format of the data will be:

$$Y = X\beta + \epsilon \tag{43}$$

where $\epsilon_i \sim N(0, \sigma^2)$. To obtain the posterior estimates for $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ and σ^2 I will first need to have a new dataset and select the prior distributions.

6.1 Data Overview and Analysis

The data used in the Bayesian regression model is taken from Lorenzen (1996). The data have several categories that are not going to be used here in this analysis, so after removing some of the categories the first four data points in the set are given in Table 3.

Table 3. The Lorenzen (1996) Dataset

Serial #	Weight $[g]$	M[y^{-1}]	Spec #	Species	Environment	System	Reference #
1	9	1.7	2	*Alestes sadleri*	Lake Victoria	Lakes	Getabu 1987
2	16	1.1	24	*Coregonus artedi*	Wisconsin	Lakes	Beverton & Holt 1959
3	28	1.2	24	*Coregonus artedi*	Wisconsin	Lakes	Beverton & Holt 1959
4	23	1	24	*Coregonus artedi*	Wisconsin	Lakes	Beverton & Holt 1959
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

The dataset has 634 measurements of weight and mortality rate for particular species in different systems including lakes, rivers, oceans, ponds, tanks, and cages. I also removed all of the data points that were taken from McGurk (1986) since that data is already included through the prior via the Peterson & Wroblewski (1986) model. After removing this data there are 541 data points remaining. The residual plot and Normal Q-Q plot are given in Figure 10.
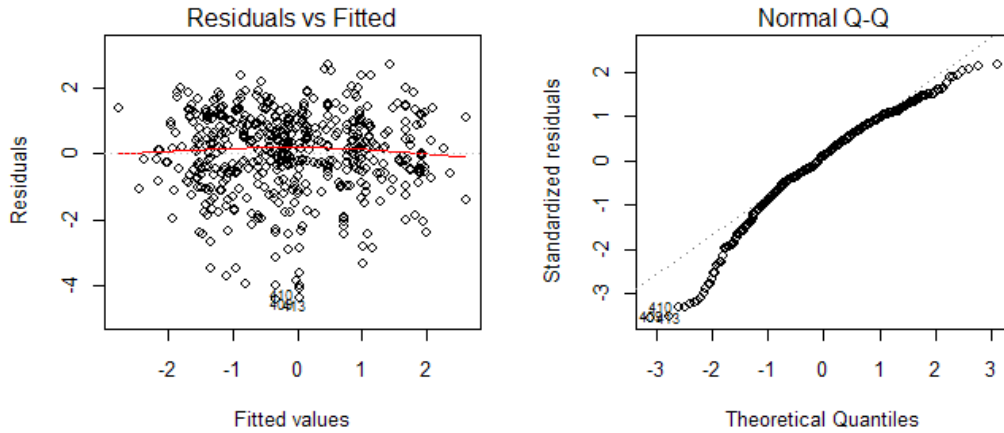
Figure 10. Residual and Normal Q-Q plot for

linear regression model from Lorenzen (1996) data, N=541.

The normality assumption is not being met here as the residual plot shows that there are

outliers that are clustered far outside the acceptable range.

Since I am mainly interested in mortality rates for marine species, I used the data found

in the ocean system as the base category for a pair-wise multiple comparisons of the means of

each system category. Using ANOVA analysis via the Tukey (1956) method, the output of the

SAS procedure is summarized in Table 4, where the system categories are: ocean=1, lakes=2,

rivers=3, ponds=4, tanks=5, and cages=6.

Table 4. Mean Comparison of System Categories via the Tukey Method

Comparison	Mean Difference	95% Confidence Int.	
1-2	-0.12	(-0.92	0.67)*
1-3	-0.37	(-1.20	0.46)*
1-4	0.70	(-0.08	1.49)*
1-5	1.30	(0.50	2.10)
1-6	1.81	(0.96	2.66)

This shows that oceans, rivers, lakes, and ponds are statistically similar having zero contained

the 95% confidence intervals of the mean comparisons. The system categories of tanks and

cages are not statistically similar to oceanic systems. This verifies what one might expect; that

natural environments behave similarly to one another with respect to rates of natural

mortality, and that artificial habitats do not have the same rate of mortality as natural habitats

do. Thus, I will only include the data from the environmental systems of oceans, lakes, rivers,

and ponds.

After cleaning the data I have 369 data points left that I will use in the Bayesian

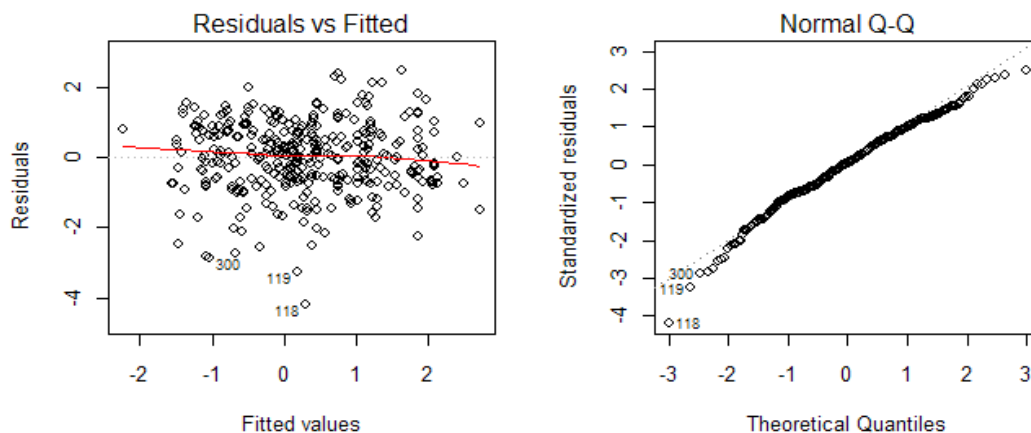regression analysis. Summary plots of the cleaned data are given in Figure 11.

Figure 11. Residual and Normal Q-Q plot for

linear regression model from cleaned Lorenzen (1996) data, N=369.

The data are now much closer to meeting the assumption of normality; however, there are still

three of the residuals that are flagged as being outliers. Based on Figure 11, observations 118,

119, and 300 are apparent outliers, but without having any reason to believe that these are

erroneous observations I will keep them in the analysis. To check for autocorrelation in the

residuals, I use the Durbin-Watson test statistic which yields a value of 1.149, which is within

the ideal value range being greater than 1 (Faraway, 2005).
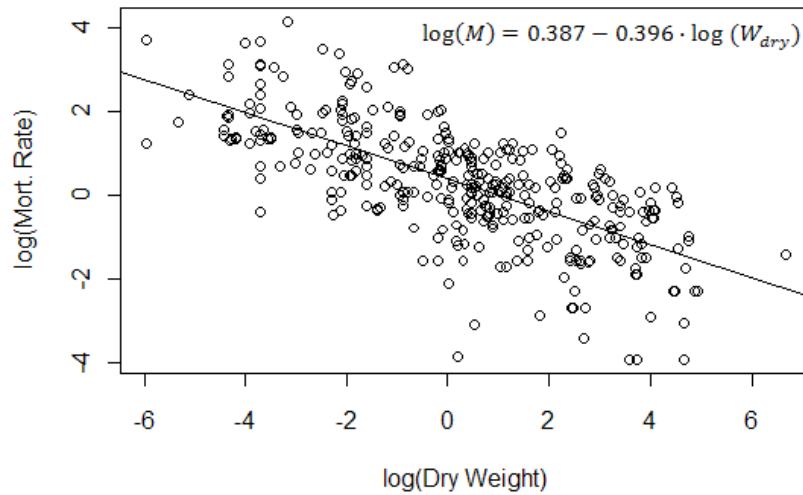
The least squares model fit is given in Figure 12.

Figure 12. Log-linear least squares regression fit line for cleaned Lorenzen (1996) data.

The least squares estimate for the slope is -0.396 which is a very different estimate from those given by Peterson & Wroblewski (1984) or McCoy & Gillooly (2008), who give estimates of -0.25 and -0.27. Lorenzen (1996) gives a slope estimate that as -0.29 for 'natural habitats', but that estimate includes the McGurk data which I have excluded here. This gives us reason to believe that these other models may not be providing an accurate model estimate for the slope term in the log-linear model. I will use Bayesian linear regression methods to achieve a more inclusive estimate for natural mortality rate that takes into account the Lorenzen data as well as the parameter estimates from previous models through an informative prior distribution.

6.2 Informative Prior Selection and Calculations

There are many priors that can be used to analyze any given data set. The two main choices are to use either informative priors, or vague/non-informative priors. I investigate both types of priors in my analysis of the Lorenzen (1996) dataset, but I will first show a method of

obtaining an informative prior that is based upon the three size-dependent models given in section 5.

I set a multivariate normal prior on $\boldsymbol{\beta}$

$$p(\boldsymbol{\beta}|\boldsymbol{\mu}, \Sigma_{\boldsymbol{\beta}}) \sim MVN(\boldsymbol{\mu}, \Sigma_{\boldsymbol{\beta}}) \tag{44}$$

I also set hyper- prior distributions on the mean and variance terms $\left(\boldsymbol{\mu} = \begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix} \text{ and } \Sigma_{\boldsymbol{\beta}} = \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right)$ as conjugate normal and Inverse-Wishart distributions (Gelman et al, 2003). The prior means for μ_0 and μ_1 are chosen to be weighted averages of the three model means for the slope and intercept terms, m_{ji} (where $j = 0,1$ and $i = 1,2,3$). Similarly, the prior standard deviations for μ_0 and μ_1 are chosen to be weighted averages of the three SEMs of the model estimates, s_{ji}

$$\mu_0 \sim N\left(\sum_i \pi_i m_{0i} , \sum_i \pi_i (s_{0i})^2\right) \tag{45}$$

$$\mu_1 \sim N\left(\sum_i \pi_i m_{1i} , \sum_i \pi_i (s_{1i})^2\right) \tag{46}$$

The values of m_{ji} and s_{ji} are based on the three models given in Eqs. (38)-(40). For example, the intercept term means in the three models are $[1.44 \, log(L_\infty) + log(k) - 2.28]$, 1.17, and 1.92, so I take $\boldsymbol{m_0} = ([1.44log(L\_\infty) + log(k) - 2.28], 1.17, 1.92)$. Similarly, m_1 is be based on the three slope term means from Eqs. (38)-(40), so I take $\boldsymbol{m_1} = (-0.54, -0.27, -0.25)$.

To address the uncertainty in $\log(k)$ and $\log(L_\infty)$ in m_{01}, I place hyper-priors on these terms as normals with the means and standard deviation values to be equal to those found by Pauly (1984).

$$\log(k) \sim N(-0.49, 0.39^2) \tag{47}$$

$$\log(L_\infty) \sim N(1.60, 0.39^2) \tag{48}$$

Using these distributions I can calculate the distribution of m_{01} utilizing the additive properties of normal random variables

$$m_{01} = 1.44\, log(L_\infty) + log(k) - 2.28 \tag{49}$$

$$= 1.44 N(-0.49, 0.39^2) + N(1.60, 0.39^2) - 2.28$$

$$= N([1.44 \cdot (-0.49) + 1.60 - 2.28], [(1.44 \cdot 0.39)^2 + 0.39^2])$$

I then obtain

$$m_{01} \sim N(-1.39, 0.68^2) \tag{50}$$

I base the estimate for the standard deviation term s_{ij} on SEM of the model parameter estimates in Models 1, 2, and 3. In Table 5, I show the fitted values for Models 2 and 3 with the corresponding SEMs for the slope and intercept term estimates.

Table 5. Parameter estimates with SEMs for Models 2 and 3

Model	Fitted Intercept β_0	Fitted Slope β_1
Model 2: McCoy & Gillooly (2008)	1.17 (SEM = 0.189)	-0.27 (SEM = 0.041)
Model 3: Peterson & Wroblewski (1984)	1.92 (SEM = 0.099)	-0.25 (SEM = 0.055)

For Model 3 the sample standard deviations are approximated using the graph of the data and a program called Data Thief (http://datathief.org/) to extract the data and infer the SEM. I will use the SEM for each term as the chosen values for $s_{.2}$ and $s_{.3}$. For Model 1, I obtain the SEMs for the slope and intercept terms by summing all of the SEMs for the estimated coefficients given in Table 2. Thus $s_{01} = 0.437$, $s_{02} = 0.189$, $s_{03} = 0.099$, $s_{11} = 0.046$, $s_{12} = 0.041$, and $s_{13} = 0.055$. I use the SEM in my prior distributions instead of the standard deviation of the data because I want to incorporate the uncertainty in the estimated value of the fitted slope and intercept terms, not the uncertainty of the overall data used to make those estimations.

For the model weights given as π_i in Eq. (44), I might assume that all models are equally likely giving that $\pi_i = \frac{1}{3}$ for $i = 1,2,3$; however, a more appealing option is to set the probabilities so that they are proportional to the amount of data in each model. That is, $\pi_i = \frac{n_i}{N}$; where n_i is the number of data points in the ith model, and where N is the total number of data points among all three models. This method seems to be preferable in that it will give weight to each model that is proportional to the amount of information that is supporting that particular model. I have that $n_1 = 168$, $n_2 = 168$, $n_3 = 90$, and $N = 426$, so the probabilities I give for the three models are $\pi_1 = 0.39$, $\pi_2 = 0.39$, and $\pi_3 = 0.22$.

So for μ_0 in Eq.(45), I rewrite it as

$$p(\mu_0|m_0^*) = N(m_0^*, s_0^2) \tag{51}$$

where $m_0^* = 0.39 \cdot \mu_{01} + 0.39 \cdot 1.17 + 0.22 \cdot 1.92 = 0.39\mu_{01} + 0.88$. This is just the weighted average of the intercept term estimates from the three models given in Eqs. (38)-(40). Since I have hyper-priors on μ_{01} I can use the properties of the normal distribution to obtain

$$m_0^* \sim N(0.39 \cdot (-1.39) + 0.88, (0.39 \cdot 0.68)^2)$$

$$\sim N(0.34, 0.27^2)$$

For s_0^2, I just take the weighted average of the slope term SEMs. Thus

$$s_0^2 = 0.39 \cdot (0.437)^2 + 0.39 \cdot (0.189)^2 + 0.22 \cdot (0.099)^2 = 0.30^2$$

By the properties of the normal distribution I have that

$$p(\mu_0 | m_0^*) = N(0.34, 0.30^2 + 0.27^2) = N(0.34, 0.40^2) \tag{52}$$

For μ_1 in Eq.(46) I rewrite it as

$$p(\mu_1) = N(m_1^*, s_1^2) \tag{53}$$

thus $m_1^* = 0.39 \cdot (-0.52) + 0.39 \cdot (-0.27) + 0.22 \cdot (-0.25) = -0.36$

and $s_1^{*2} = 0.39 \cdot (0.046)^2 + 0.39 \cdot (0.041)^2 + 0.22 \cdot (0.055)^2 = 0.46^2$

This results in the hierarchical model with informative conjugate priors and hyper-priors

$$\mathbf{W} \sim W_2(\nu, \mathbf{I}) \tag{54}$$

$$p(\mathbf{\Sigma}_\beta) \sim \mathbf{W}^{-1} \tag{55}$$

$$p(\boldsymbol{\mu}) \sim N\left(\begin{pmatrix} 0.34 \\ -0.36 \end{pmatrix}, \begin{pmatrix} 0.40^2 & 0 \\ 0 & 0.46^2 \end{pmatrix} \right) \tag{56}$$

$$p(\boldsymbol{\beta}|\boldsymbol{\mu}, \Sigma_{\boldsymbol{\beta}}) \sim N_2(\boldsymbol{\mu}, \Sigma_{\boldsymbol{\beta}}) \tag{57}$$

$$p(\sigma^2) \sim \Gamma^{-1}(1000, 1000) \tag{58}$$

$$\boldsymbol{y} \sim \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{59}$$

where $\boldsymbol{W}$ has the Wishart distribution, ν is the degrees of freedom parameter, and

$\epsilon_i \sim N(0, \sigma^2 \mathbf{I})$. I can set ν to be any number greater than 1, but for my initial model run I will set

$\nu = 500$ which will give only a small amount of variability to the prior distribution for $\boldsymbol{\beta}$, thus

drawing the posterior closer to the prior mean. I investigate how the selection of ν effects the

posterior estimates in section 7.3.

6.3 Vague prior Selection

I also consider a model using a vague prior and compare it with the model that uses the

informative prior. I use the vague priors as follows

$$\mathbf{W} \sim W_2(500, \mathbf{I}) \tag{60}$$

$$p(\Sigma_{\beta}) \sim \mathbf{W}^{-1} \tag{61}$$

$$p(\boldsymbol{\mu}) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1000 & 0 \\ 0 & 1000 \end{pmatrix}\right) \tag{62}$$

$$p(\boldsymbol{\beta}|\boldsymbol{\mu}, \Sigma_{\beta}) \sim N_2(\boldsymbol{\mu}, \Sigma_{\beta}) \tag{63}$$

$$p(\sigma^2) \sim \Gamma^{-1}(1000, 1000) \tag{64}$$

6.4 Posterior Distributions via Gibbs Sampler

I use WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs) to sample from the posterior

distributions for $\boldsymbol{\beta}, \Sigma_{\beta}$, and σ^2. The setup for the sampling in WinBUGS is a Gibbs sampler that

samples from the complete conditional distributions. The full conditional which is used to

obtain the complete conditionals for Gibbs sampling is

$$p(\boldsymbol{\beta}, \boldsymbol{\mu}, \sigma^2, \boldsymbol{\Sigma_\beta} | Y) = p(\sigma^2) \cdot p(\boldsymbol{\mu}) \cdot p(\boldsymbol{\Sigma_\beta}) \cdot p(\boldsymbol{\beta} | \boldsymbol{\mu}, \boldsymbol{\Sigma_\beta}) \cdot p(Y | \boldsymbol{\beta}, \boldsymbol{\mu}, \sigma^2, \boldsymbol{\Sigma_\beta})$$
(65)

So the complete conditionals are:

$$p(\boldsymbol{\mu} | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\Sigma_\beta}, Y) = c \cdot p(\boldsymbol{\mu}) \cdot p(\boldsymbol{\beta} | \boldsymbol{\mu}, \boldsymbol{\Sigma_\beta})$$
(66)

$$p(\boldsymbol{\Sigma_\beta} | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, Y) = c \cdot p(\boldsymbol{\Sigma_\beta}) \cdot p(\boldsymbol{\beta} | \boldsymbol{\mu}, \boldsymbol{\Sigma_\beta})$$
(67)

$$p(\sigma^2 | \boldsymbol{\mu}, \boldsymbol{\beta}, Y) = c \cdot p(\sigma^2) \cdot p(Y | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\Sigma_\beta})$$
(68)

$$p(\boldsymbol{\beta} | \boldsymbol{\mu}, \sigma^2, \boldsymbol{\Sigma_\beta}, Y) = c \cdot p(\boldsymbol{\beta} | \sigma^2, \boldsymbol{\mu}, \boldsymbol{\Sigma_\beta}) \cdot p(Y | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\mu}, \boldsymbol{\Sigma_\beta})$$
(69)

The Gibbs sampling updating scheme is as follows:

1. Set initial values of $\sigma^2, and\ \boldsymbol{\beta}$.

2. Sample $\boldsymbol{\mu}_{j+1}$ from $p(\boldsymbol{\mu}_j) \cdot p\left(\boldsymbol{\beta}_j \middle| \boldsymbol{\mu}_j, \boldsymbol{\Sigma_\beta}_j\right)$

3. Sample $\boldsymbol{\Sigma_\beta}_{j+1}$ from $p\left(\boldsymbol{\Sigma_\beta}_j\right) \cdot p\left(\boldsymbol{\beta}_j \middle| \boldsymbol{\mu}_{j+1}, \boldsymbol{\Sigma_\beta}_j\right)$

4. Sample σ^2_{j+1} from $p(\sigma^2_j) \cdot p\left(Y \middle| \boldsymbol{\beta}_j, \sigma^2_j, \boldsymbol{\mu}_{j+1}, \boldsymbol{\Sigma_\beta}_{j+1}\right)$

5. Sample β_{j+1} from $p\left(\boldsymbol{\beta}_j \middle| \boldsymbol{\mu}_{j+1}, \boldsymbol{\Sigma_\beta}_{j+1}\right) \cdot p\left(Y \middle| \boldsymbol{\beta}_j, \sigma^2_{j+1}, \boldsymbol{\mu}_{j+1} \boldsymbol{\Sigma_\beta}_{j+1}\right)$

6. Repeat steps $2 - 5$ for 101000 iterations and discard first 1000 runs as the burn-in

   period.

7. COMPARISON OF TWO BAYESIAN LINEAR REGRESSION MODELS

In this section I give posterior estimates and analysis for both hierarchical Bayesian regression models using the Lorenzen data; the first model employs the informative priors and the second model employs the vague priors as laid out in sections 6.2 and 6.3. I compare both models with each other and to the traditional least squares method. For each model, I also give sensitivity, posterior predictive, and residual analysis. I also make use of a hypothetical dataset to demonstrate that the informative prior model draws the posterior closer to the prior as the number of data points (n) is decreased.

7.1 Posterior Estimates Using Informative Priors

Using the informative priors as given in Eqs. (54)-(58), I run the model in WinBUGS and obtain the following posterior estimates along with the 95% credible intervals for each estimate shown in Table 6.

Table 6. Posterior Estimates Bayesian Regression Using Informative Priors

	Posterior Quantiles		
Parameter	2.5%	Mean	97.5%
β_0	0.284	**0.386**	0.488
β_1	-0.437	**-0.395**	-0.354
σ^2	0.087	**1.005**	1.162

The pair-wise correlation between β_0 and β_1 was -0.069, which is reasonably small.

I present the posteriors distributions for the slope and intercept terms in Figure 13.
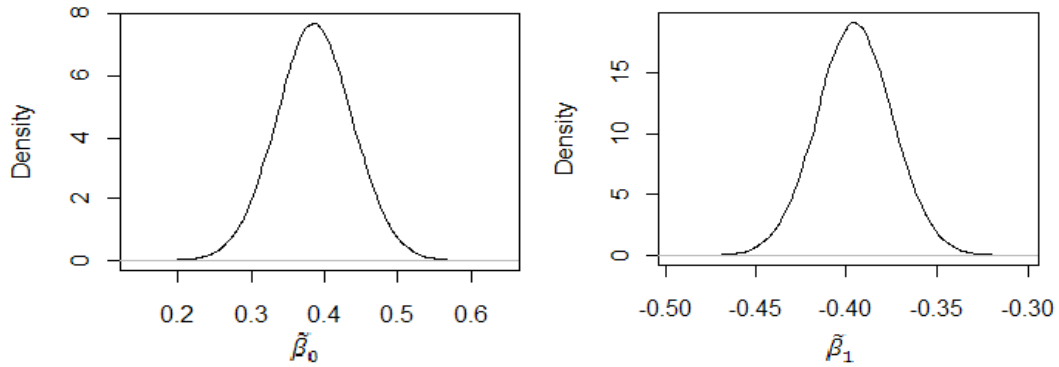
Figure 13. Posterior distributions for the slope and intercept terms using the informative priors.

To test the fit of the model with the data that is observed I examine the posterior predictive distributions of the minimum, median, and maximum values in the Lorenzen (1996) dataset. I obtain 1000 simulated draws (y^{rep}) from each predictive distribution and plot them against the actual observed value in Figure 14 (Gelman, 2003; Albert, 2009).
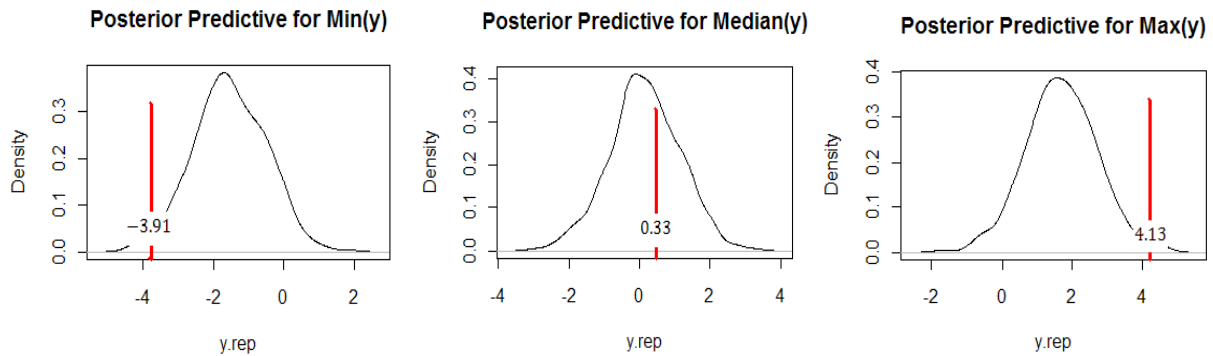

Figure 14. Posterior predictive distribution samples for minimum, median, and maximum y values in the Lorenzen (1996) dataset with actual observed values indicated.

Here I see that perhaps there are some outliers in the dataset that do not quite fit the model. In looking at the p-values I see that the minimum and maximum values are 0.99 and 0.016. There were roughly 20 data points that had posterior predictive p-values that were unacceptable at

the 5% level, this is less than 5% of the total dataset. This is somewhat high but not extreme so I

can conclude that the model has reasonable predictive abilities (Gelman et al, 2003).

In looking at the standardized residuals of the model (Figure 15), I can see that there are

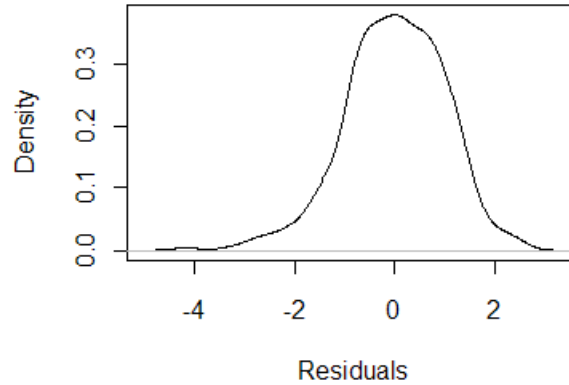indeed some outliers in the model particularly on the left side of the residual density plot

Figure 15. Residual density plot of the Bayesian regression model using informative priors

The outlier residuals here might have been avoided if the outliers seen in the cleaned Lorenzen

(1996) data set had been removed. The residual density is not ideal but it is not extreme

consider I included some outliers in m analysis.

7.2 Posterior Estimates and Analysis Using Vague priors

Running the model with the vague priors as given in Eqs. (60)-(64) I obtain the posterior

estimates given in Table 7.

Table 7. Posterior estimates using Bayesian regression with vague priors

| Parameter | Posterior Quantiles | | |
	2.5%	Mean	97.5%
β_0	0.284	**0.387**	0.490
β_1	-0.437	**-0.396**	-0.354
σ^2	0.087	**1.005**	1.162

38

The pair-wise correlation between β_0 and β_1 was found to be -0.070 which is basically the same as seen before. In fact there are essentially no differences between the posterior estimates obtained between the informative and the vague prior models. The small differences can be explained by MCMC error, and since the dataset is so large, having 369 data points, the posteriors are both essentially equivalent to each other as well as to the least squares estimates for $\boldsymbol{\beta}$.

The posterior predictive analysis for the model using the vague prior yielded essentially the same results as were seen in the analysis of the informative prior model. This is to be expected since there is enough data in the model to have the likelihood dominate the prior.

7.3 Sensitivity Analysis

In Bayesian analysis it is important to investigate the sensitivity of the model to differences in the data, likelihood, and the prior distributions (Gelman et al, 2003). I first look at how changes in the amount of data used in the model effects how much the prior distribution has an influence on the posterior. To illustrate this I generated a hypothetical set of data points that has 40 observations. The actual values of these data points lie within the ranges seen in the Lorenzen (1996) dataset. Using the least squares estimates as the basis for the comparison, I look at how the informative and vague prior models are affected by the number of data points by running both models using $5, 10, 20, 30$, and the full set of 40 data points from this hypothetical data set (Table 8).

n	Parameter	Posterior Estimates		
		Informative Prior Model	Vague prior Model	Least Squares Estimates
5	β_0	0.219 (0.235)	0.067 (0.354)	0.084 (0.260)
	β_1	-0.607 (0.277)	-0.811 (0.392)	-0.789 (0.310)
10	β_0	0.267 (0.126)	0.262 (0.126)	0.258 (0.105)
	β_1	-0.560 (0.130)	-0.576 (0.129)	-0.574 (0.110)
20	β_0	0.231 (0.084)	0.226 (0.084)	0.226 (0.079)
	β_1	-0.418 (0.079)	-0.419 (0.078)	-0.420 (0.073)
30	β_0	0.245 (0.073)	0.242 (0.075)	0.242 (0.072)
	β_1	-0.463 (0.071)	-0.466 (0.071)	-0.466 (0.068)
40	β_0	0.157 (0.072)	0.151 (0.073)	0.151 (0.071)
	β_1	-0.385 (0.069)	-0.386 (0.070)	-0.385 (0.068)

In Table 8 I see that the smaller n becomes, the more the posterior is drawn towards the prior means of $\beta_0 = 0.34$ and $\beta_1 = -0.36$. The vague prior model is less affected by changes in the amount of data and is nearly identical to the least squares estimate from $n = 10$ and higher. The informative prior is very influential up until n is equal to 40. Even though there is only a difference of around 0.006 in the estimate for β_0 when $n = 40$, this is significant since the estimate is on the log-scale which would be a difference of 1 unit on the actual scale value.

Using the hypothetical dataset with 20 data points I look at how the informative prior model behaves in its predictive abilities when n is relatively small. Looking at 1000 samples (y^{rep}) from the posterior predictive distributions for the minimum, median, and maximum values in the hypothetical dataset (Figure 16)
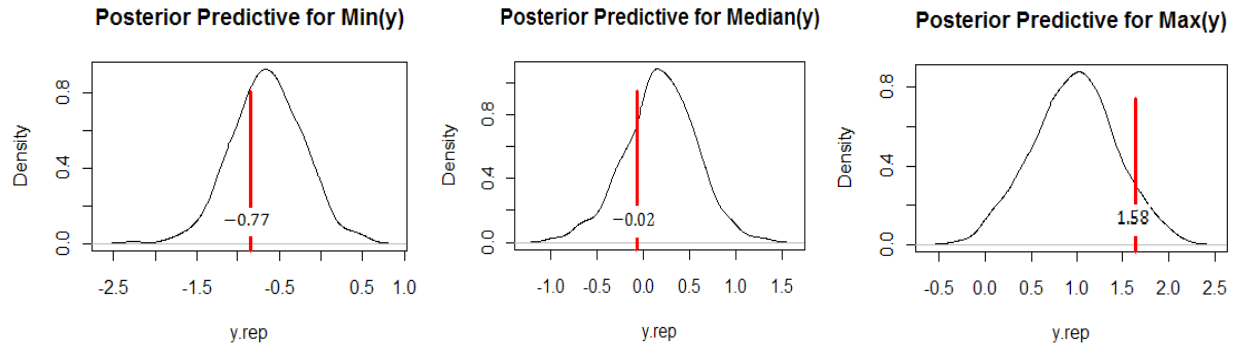
Posterior Predictive for Min(y) **Posterior Predictive for Median(y)** **Posterior Predictive for Max(y)**

Figure 16. Posterior predictive distributions for minimum, median, and maximum y values hypothetical dataset with actual observed values indicated.

I observe that there are no apparent issues with the predictive abilities of the model. All p-values were within the acceptable ranges at the 5% level.

Investigating how changes in the prior distributions affect the posterior distribution for both models, I see that the selection of the degrees of freedom in W, where $\Sigma_\beta \sim W^{-1}$, greatly affected the outcome of the posterior for Σ_β in both models, for example selecting $v = 2$ with $n = 20$ data points would yield a posterior for $\Sigma_{\beta_{11}} = 7$, whereas if $v = 500$ with n=20, $\Sigma_{\beta_{11}} = 0.002$.

The amount of data had an effect on the sensitivity of the model to the prior selection as well. Even though the posteriors for Σ_β change drastically depending on the prior selection, the posterior for β and σ^2 were not greatly affected when n was large. Also for larger data sets the posterior for Σ_β would tend to shrink as well as the standard deviation of the estimates for β and σ^2. This is to be expected since as more data are included in the model the prior has less of an influence on the posterior distribution, and the posterior will tend to equal the likelihood

as n is greater than 40. Changing the prior values for $\boldsymbol{\beta}$ or σ^2 had little to no effect on the

posteriors for both the informative prior and the vague prior models.

DISCUSSION

Having looked at two hierarchical Bayesian linear regression models, I found that for

large datasets (larger than 40 observations) the informative prior model gives essentially

equivalent posterior estimates as to the model that utilizes the vague priors. Given smaller

datasets, the informative prior is significantly different from the least squares model, but as the

dataset gets larger, at around 40 data points, the informative prior model yields equivalent

results to the vague prior model. The vague prior model estimates are nearly identical to the

least squares model estimates; however, the uncertainty in those estimates is more

conservative, or wider, in the vague prior Bayesian model.

For non-Bayesians, the prior model in section 6.2 can be used to give a more

comprehensive estimate for M over those models that have been given by Peterson &

Wroblewski (1984), Gislason et al (2010), and McCoy & Gillooly (2008). The prior model by itself

is

$$\log{(M)} = \beta_0^* - \beta_1^*\big(\log\big(W_{dry}\big)\big) \tag{70}$$

with $\beta_0^* \sim N(0.34, 0.40^2)$ and $\beta_1^* \sim N(-0.36, 0.46^2)$. These have fairly large variance terms giving

more informative and conservative estimates for M.

Here I have looked at using size-dependent methods for estimating rates of natural

mortality. Incorporating other studies of size-dependent natural mortality rates could be useful

here such as Pauly (1980) who gives a model that is dependent on the maximum weight W_∞. A conversion could be made between W_∞ and W_{dry} which would allow it to be compared to these other models I have utilized here. The posterior distributions I obtained in Section 7 can also be updated as new data becomes available.

References

Albert J., 2009. Bayesian Computation with R, Second Edition. Springer, Seattle, WA.

Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances. Philosophical Transactions, Royal Society of London 53: 370-418.

Beverton, R.J.H. and Holt, S.J., 1957. On the dynamics of exploited fish populations. Ministry of Agriculture, Fisheries and Food. Fishery Investigations, London, 2, 19–533.

Cushing, D. H., 1974. The possible density-dependence of larval mortality and adult mortality in fishes, p. 103-111 . In J.H.S. Blaxter [ed.] The early Life history of fish. Springer-Verlag, New York, NY.

Faraway, J. J., 2005. Linear Models with R. Chapman & Hall/CRC, Boca Raton, FL.

Fraley, C. and Raftery A. E., 2007. Bayesian regularization for normal mixture estimation and model-based clustering. Journal of Classification, 24:155:181.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B., (2003). Bayesian Data Analysis, 2nd ed. CRC Press, London.

Gislason, H., Daan N. , Rice J. , and Pope J., 2010. Size, growth, temperature and the natural mortality rate of marine fish. Fish and Fisheries 11:149-158.

Gunderson, D.R. and Dygert, P.H., 1988. Reproductive effort as a predictor of natural mortality rate. Journal du Conseil International pour l' Exploration de la Mer 44, 200–209.

Hewitt, D. A., and Hoenig J. M., 2005. Comparison of two approaches for estimating natural mortality based on longevity. Fishery Bulletin 103:433-437.

Jeffreys, H., 1939. Theory of probability. 1st edition. Clarendon Press, Oxford, United Kingdom.

Jennings, S., M. J. Kaiser, and Reynolds J. D., 2001. Marine Fisheries Ecology. Blackwell Science Ltd., London.

Jensen, A. L., 1996. Beverton and Holt life history invariants result from optimal trade-off of reproduction and survival. Canadian Journal of Fisheries and Aquatic Sciences 53:820-822.

Letourneur Y., 1998. Length-weight relationship of some marine fish species in Reunion Island, Indian Ocean. Naga 21:37–4.

Lorenzen K., 1996. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. J Fish Biol 49:627–647.

Mangel M., 2006. The theoretical biologist's toolbox. Cambridge University Press, Cambridge.

McCoy, M.W. and Gillooly, J.F., 2008. Predicting the natural mortality rates of plants and animals. Ecology Letters 11, 710–716.

McGurk, M.D., 1986. Natural mortality of marine pelagic fish eggs and larvae: role of spatial patchiness. Marine Ecology Progress Series 34, 227–242.

Pauly, D. (1980) On the interrelationships between natural mortality, growth parameters, and mean environmental temperature in 175 fish stocks. Journal du Conseil International pour l' Exploration de la Mer 39, 175–192.

Pearcy, W. G., 1962. Ecology of an estuarine population of winter flounder,Psedopbeumrrecoes ameaicaeanus (Walbaum). Part 1-4. Bull. Bingham Oceanogr. Coll. 18: 16-37.

Peterson, I., and Wroblewski J. S., 1984. Mortality Rates of Fishes in the Pelagic Ecosystem. Canadian Journal of Fisheries and Aquatic Sciences 41:1117-1120.

Press WH. 1997. Understanding data better with Bayesian and global statistical methods. In Unsolved Problems in Astrophysics, ed. JN Bahcall, JP Ostriker, pp. 49–60. Princeton, NJ: Princeton Univ. Press

Ursin, E., 1967. A mathematical model of some aspects of fish growth , respiration, and mortality. J. Fish Res. Board Can. 24:2355-2452.

Vetter E.F., 1988. Estimation of natural mortality in fish stocks: A review. Fish Bull., U.S. 86, 25–43.

Wade, P. R., 2000. Bayesian methods in conservation biology. Conservation Biology 14:1308-1316.

Ware, D. M., 1975. Relation between egg size, growth and natural mortality of larval fish. J. Fish Res. Board Can. 32:2503-2512.

Zhang, Z., K. L. Chan, Y. Wu, and C. Chen (2004). Learning a multivariate Normal mixture model with the reversible jump MCMC algorithm. Statistics and Computing 14, 343:355.