UNIVERSITY OF CALIFORNIA
SANTA CRUZ

Biological Modeling in Data Poor Scenarios

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

MASTER OF SCIENCE

in

STATISTICS AND APPLIED MATHEMATICS

by

Daniel Hively

Applied Mathematics and Statistics and Center for Stock Assessment Research

November 2011

The Thesis of Daniel Hively
is approved:

Professor Marc Mangel, Chair

Stephen Ralston

Dr. Steve Ralston, Reader

Abstract

Many biological systems under study are characterized as data-poor. In order to enhance our analyses of these systems, we will need to develop modeling tools that can appropriately utilize the data, and determine if the data are being properly utilized. I examine this problem through two different data-poor systems: West Coast Rockfish (*Sebastes* spp.) fisheries, and mosquito (*Anopheles gambiae*) host attack behavior dynamics. Stock assessment models are used by fisheries scientists to estimate stock status, in order to help inform managers when they determine future catch policy. These models employ a large set of assumptions, which may not be appropriate for the species under consideration. When species are data-poor, it isn't fully known how uncertainty in stock assessments, and the catch policy based on them, truly effect the population. I implement a life history simulator to generate data with which to quantify the accuracy of the stock assessment models. Mosquito attack behavior is important to understand since they are vectors for several diseases, including malaria. By using a behavior model, I examine how mosquito attack behavior changes due to its state and its environment. I find that uncertainty has significant impacts on stock assessment accuracy and mosquito attack behavior, and I show that the models can be used to enhance the existing data and gain further insights into the systems.

Introduction

Many study systems in biology are typified by a lack of data. Yet several of these data poor systems have significant impacts on health, jobs, and the economy, which necessitates further study. New approaches are necessary to study systems under these conditions. A natural supplement to field work and data collection is a modeling approach, which can aid in several ways.

Some of the major impediments to data gathering are the costs and time commitments necessary for measurements and experiments in the systems. Characteristics under study may not be significant within the system. An important result from modeling may be an identification of which pieces of data would be most valuable to study, thus avoiding investments in research funds or manpower that could be put to better use in other aspects of the research. Thus incorporating modeling into a project can help to optimize research.

In cases where modeling approaches are used, it is important to understand whether they are appropriate, or that they are used appropriately, for that data scenario. Models introduce a suite of assumptions into an analysis that may be the source of additional unnecessary error. And within a data poor scenario, these assumptions may have yet unknown impacts.

In this document, I will examine two very different systems: West Coast rockfish (*Sebastes* spp.) and the malaria vector mosquito (*Anopheles gambiae*). Though the systems differ in several significant ways, they both share a lack of data. By introducing novel modeling approaches, current and future research can be significantly improved by identifying and addressing weaknesses with current approaches and methods. Coupling current data with modeling can help to improve future data gathering by finding the key drivers in the system.

West Coast rockfish are an important resource whose harvest is regulated through stock assessments to ensure continued economic viability and persistence of the stock. A stock assessment is a quantitative tool that uses available catch and survey data to make estimates about the state of the population, as well as the biological characteristics of the species. The estimates derived from the assessment are used to help guide fishing policy decisions for future catch. Many of the species are data poor, which results in increased uncertainty in the estimated population dynamics and life history traits. This is an even greater problem for *Sebastes* spp. since they tend to be long-lived, slow to mature (Love et al., 2002), and older fish contribute significantly more towards reproduction than younger fish (Ralston and MacFarlane, 2010). This leads to increased vulnerability to overfishing (Leaman, 1991). Therefore it is extremely important to better understand both their biological system, and the level of error generated for data poor stocks. I approach this problem with a life history simulator that I can use to generate data with which to examine and quantify the effects of data poor assessments. I begin with an illustrative example assessment, followed by use of the simulator with a currently used assessment model.

Malaria and its mosquito vector have far different impacts. Malaria continues to be an enormous public health concern despite massive efforts towards eradication of it and its mosquito vectors (World Health Organization, 2010). It also causes severe economic problems for the countries most affected due to the rate of infection and its associated virulence. In order to minimize and prevent future outbreaks, we need to learn more about the behavior of the vector, including mosquito attack behavior. By better understanding the mechanism of infection, we can better target vulnerabilities in the infection process to disrupt its spread. However behavioral data on mosquitoes, particularly *Anopheles gambiae*, are not common. To examine this problem I implement a model that I use to study the impacts of mosquito attack characteristics to environment and mosquito state.

The aim of this work is to demonstrate that utilization of appropriate mathematical techniques can be a powerful aid to handle situations where data are poor or non-existent. They can improve current methods and even help to inform novel techniques.

Section 1: Rockfish Life History Simulator to Examine Data-Poor Fisheries

Background

Stock assessments are widely applied throughout fisheries management and are vital for evaluating the status and projecting the future catch of a stock (Gulland, 1988). In particular, stock assessments produce estimates of abundance, which fisheries managers utilize in establishing policies such as catch limits to prevent overfishing and maintain sustainable fisheries (Rose and Cowan, 2003). In order to perform a stock assessment, various relevant data for the stock, including both fisheries data (landings, coded wire tags, hatcheries) and fisheries independent data (surveys), may be incorporated in the assessment. However, the inclusion of data that are not appropriate for the system being modeled can be detrimental to the accuracy of the results of the assessment; therefore only a subset of all available data may be used (Wang et al., 2009). The fisheries data may include the fishery type (commercial, recreational), total catch (biomass, numbers), geographic area of the catch, month and year of the catch, fishing effort, age composition of the catch, sex composition of the catch, and length or weight composition of the catch (Chen et al., 2003; Rose and Cowan, 2003). Alternatively, fisheries independent data may inform managers about the local movement and dispersal of the stock, migratory patterns, recapture rates, natural mortality estimates, growth rates, and other life history parameters (Gulland, 1988). Many stock assessments use these data as inputs to complex statistical analyses (e.g. Stock Synthesis 3 (SS3), see Methot (2005) for details; the Ecosystem Diagnosis and Treatment (EDT) model discussed in McElhany et al. (2010); Chen et al. (2003)). The results of this process determine the status of the stock and influence policies enacted regarding the stock (Figure 1).

Within a stock assessment assumptions must be made about the life history parameters, structure, and population dynamics of the stock. These assumptions are traditionally drawn from the data, borrowed from similar species, or come from convention (see Rose and Cowan (2003) regarding estimates of steepness (the fraction of maximum recruitment expected when spawning biomass is at 20% of its unfished biomass), also see Mangel et al. (2010)). For instance, accurately establishing the rate of natural mortality for a stock is quite difficult since this rate is not directly observable. Therefore, several equations estimating mortality have emerged. These include: constant mortality ($M = 0.2$), mortality depending on age ($M(a)$), and mortality depending on length at age ($M(L(a))$) (Jennings et al., 2001; Post and Evans, 1989; Lorenzen, 1996). When a stock assessment is performed, the fisheries scientist must choose appropriate equations and parameters (or a range of parameters) for the assessment, based upon available data about the stock, similar species, or convention. Although a range of parameters, structures, and equations governing the dynamics of the population may be used in the stock assessment, in order to make modeling the stock tractable, many scenarios must be excluded.

For example, consider the Namibian hake fishery (Hilborn and Mangel, 1997). An unregulated fishery developed on an unfished stock until a dramatic decrease in catch per unit effort (CPUE) was experienced. A regulatory agency was then established for the stock, leading to a reduction in catch. In order to model the fluctuating fish population dynamics, two models were considered, one less complex (4 free parameters) than the other (5 free parameters). Throughout the development of the models, whether stochastic or deterministic, Hilborn and Mangel made decisions regarding which parameters to derive from the data, which parameters to specify a priori, and what types of uncertainty to include. After the models were developed and the CPUE data were input into both models, the following step was to determine which model best fit the data. This was defined as the model that was the most consistent with the data and that used the fewest free parameters. Hilborn and Mangel found that for this data set the less “realistic” model (de Valpine and Hilborn, 2005) was a better fit (cf. Ludwig and Walters (1985)).

As more parameters and models are considered, this process can become even more computationally intensive, thus investigating every possible scenario is not feasible; some assumptions must be made, no matter what type of model is used. This results in a disparity between assumptions made in the stock assessment tool and the true dynamics of the stock, which may have significant consequences for the output of the assessment. Through simulation, I propose to test the robustness of stock assessment tools to these assumptions. Similar to a Management Strategy Evaluation (MSE), I use an operating model, which is more complex than the management tool, to evaluate the management tool. However, while the focus of an MSE is to identify the most robust management procedure (including decision rules and management actions) across a variety of criteria, my work focuses on the assumptions within an assessment tool and the accuracy of population estimates derived from the assessment tool. Specifically, I simulate complex age structure for the species within the operating model, while the assessment tool does not take this complexity into account. Additionally, I investigate the importance of assuming a constant natural mortality rate within the assessment tool while the operating model is constructed with a non-constant natural mortality rate (see He et al. (2011)); this allows me to determine the type and quantity of error that emerges when the assumptions within the management tool differ from the true biology of the organism.

Different quantities and types of data are collected for different stocks (Chen et al., 2003). The variability in the data depends on the agency collecting the data and their jurisdiction, historical catch records, funding and time constraints, the status of the stock, politics, and the importance of the stock both recreationally and commercially. There exists a gradient in data between stocks that varies from data poor to data rich (Figure 2). Data poor stocks are typically characterized by short time series and few if any records of catch composition, only total catch by numbers or biomass. For data rich stocks we may have long time series with diverse catch compositions such as catch by age, sex, length or weight, and location. Furthermore, data rich stocks may be described by data corresponding to movement, mortality and growth rates, effort, and other life history characteristics. Stocks that fall between the extremes of data poor and data rich are described by any subset of the data available for data rich stocks. Performing a stock assessment on a data poor stock adds another layer of complexity to the process since many of the life history parameters and assumptions about the structure and dynamics of the stock must be determined without data (Wetzel and Punt, 2011). Even for data rich stocks, selecting the relevant data for the model can be quite challenging. Another goal of my research is to test the effectiveness of the stock assessment tool when confronted with various data levels. For example, I input subsets of the total catch time series into the assessment to determine the influence of the length of catch time series on the accuracy of the estimates generated by the assessment tool.

There are several sources of uncertainty in the stock assessment process: 1) the functional forms selected to model the dynamics of the stock, since there are often several appropriate functions available; 2) observation error due to the data collection method; 3) assumptions made about the data, the system being examined, and the functions used during the modeling process, which all incur additional uncertainty (Wang et al., 2009). The main obstacle in quantifying the error in stock assessment estimates arises from the difficulty in tracking fish populations with any degree of certainty, so data poor fisheries are providing data that represent an even more incomplete picture of the true stock dynamics. It is not known how well different levels and types of data represent a stock in the assessment process. Yet it is important to evaluate the robustness of stock assessment models to ensure that managers have access to viable information when making policy decisions under these conditions. However no surveys exist which track a rockfish population completely to study how well data captures the population's dynamics, and how accurate assessment model estimates are in tracking the dynamics.

In order to approach this problem, I use a mathematical model to numerically simulate a *Sebastes* population, as a discrete-time dynamical system. The model is a life history simulator that generates catch data from the simulated population. Unlike most models, the simulator is structured by age, length and location. It is also population-based instead of individual-based. This allows for fine-scale dynamics to be tracked while avoiding unnecessary computational complexity. The system is computed numerically due to the complexity and stochasticity within the equations. This further allows me to study the population dynamics and the effects of changes to the system (such as the introduction of fishing or a change in the environment) with ease.

The simulator generates a population that is tracked by age, length, region and time. The population in turn is tracked across the number of individual fish, biomass of the population, fecundity of the population, and yield from catch in both number of fish and biomass. The level of composition can be consolidated by combining the data over the appropriate characteristic. For example, to obtain total yield in biomass with no composition for the time series, the time series for yield in biomass is combined across age, length, and region (Figure 3). This allows the simulator to generate data for many classes of models, not just ones with a matching set of state variables. Therefore the type of data (composition across variables), and the amount of data (length and position of time series), can be varied to reflect different fishery data scenarios. In use, the life history simulator will produce an appropriate time series of catch data for the stock assessment model under consideration. The stock assessment model uses the data to calculate estimates about the population. This can include time series of population biomass, or life history parameter values. The results from the assessment are compared to the full population information from the simulator to quantify stock assessment accuracy (Figure 4).

The life history simulator is a complex mathematical model, as are the stock assessment models. Therefore I begin with an illustrative "toy" assessment model in order to demonstrate the capability and use of the simulator. I then use the simulator with a true stock assessment model, Stock Synthesis 3. In both cases I describe the assessment model and the fishing scenarios under examination.

Life History Simulator Framework

The simulator is constructed as a discrete time dynamical system that tracks the age a, length l, and region r of the population through time t (maximum values are a_{max}, l_{max}, r_{max}, and t_{max} respectively). Age and time are tracked in years, length is in centimeters, and region is a numerical representation of different areas. The population size within a region changes as a result of mortality, recruitment and migration. I let $N(a, t, l, r)$ represent the number of individuals of age a at time t of length l in region r.

I use von Bertalanffy growth to model the relationship between age (yr) and length (cm) (von Bertalanffy, 1957). This assumes that length is asymptotic for older ages, resulting in smaller increments to length as age increases. Additionally, I assume that growth can vary spatially within populations (Boehlert and Kappenman, 1980; Haldorson and Love, 1990), due to variable amounts of resources and differing metabolic effects between regions. I let $L(a, r)$ denote the length of a fish of age a in region r. von Bertalanffy growth is described by asymptotic length $L_\infty(r)$, the maximal length for individuals in the population; growth rate $k(r)$, which determines how quickly individuals approach asymptotic length; $a_0(r)$, an offset parameter which compensates for the initial size of an individual.

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

I assume an allometric relationship between length and weight $W(l, r)$ (gm), where $c_1(r)$ denotes density and $c_2(r) \sim 3$ establishes an approximately cubic relationship between length and weight so that

$$W(l, r) = c_1(r) \cdot l^{c_2(r)} \tag{2}$$

Total mortality $Z(l, r) = M(l, r) + \mathcal{F}(l, r)$ (yr^{-1}) is determined by natural mortality, $M(l, r)$, and fishing mortality, $\mathcal{F}(l, r)$. I assume no senescence in this version of the simulator, i.e. that larger individuals are less prone to predation and are better able to overcome disease. Larger size should decrease the types and number of predators able to consume them, and should allow larger stores of energy with which to resist disease or parasitism. These factors will vary by location since the habitat and ecosystem are spatially heterogeneous, resulting in differing sources of mortality across the seascape. Natural mortality on length l in region r is represented by $M(l, r)$. I model the size selectivity of mortality by including size-independent $M_0(r)$ and size-dependent $M_1(r)$ components that depend on region r (Lorenzen, 1996).

$$M(l, r) = M_0(r) + \frac{M_1(r)}{l} \tag{3}$$

The example assessment models constant natural mortality. I calculated 5 different values for the pair of parameters $(M_0(r), M_1(r))$ associated with natural mortality, that generate the same survival to maximum age as constant mortality (dropping the region dependence for notation simplicity; here I denote these 5 pairs as $M^* = \{M^*(1), M^*(2), M^*(3), M^*(4), M^*(5)\}$ where each j entry contains an (M_0, M_1) pair such that $M^*(j) = (M_0(j), M_1(j))$, for more details see Appendix A). These cases describe the proportion of natural mortality that is size-dependent, i.e. the proportion that is size-dependent for $M^*(1)$ is 0, so $M_1(1) = 0$ (this corresponds to constant natural mortality), for $M^*(2)$ 25% of the natural mortality is size-dependent, etc., up to $M^*(5)$ where 100% of the natural mortality is size-dependent so $M_0(5) = 0$. These cases are implemented with 5 different simulation runs, one for each M^* pair.

I model gear selectivity with increased pressure on larger fish and implement it with a sigmoid function, where fishing mortality is bounded between 0 and maximum harvest mortality $\mathcal{F}_{max}(r)$. The size at which 50% maximum harvest strength occurs, $\mathcal{F}_{50}(r)$, is an inflection point where sizes below or above this value converge to zero or maximum harvest mortality, respectively. The slope at the point of inflection, $\mathcal{F}_s(r)$, determines the smoothness of the transition from low to high mortality. Each of these parameters vary by region to account for spatial variation in fishing fleet strategy.

$$\mathcal{F}(l,r) = \frac{\mathcal{F}_{max}(r)}{1 + exp\Big(\mathcal{F}_s(r) \cdot \big(\mathcal{F}_{50}(r) - l\big)\Big)} \tag{4}$$

Length accrues over time at different rates according to migratory patterns. Population transitions occur on an annual timescale and include migration between regions during each time step. I assume that migration occurs at the beginning of the year. As a result, fish will be subject to the environmental parameters, including growth and mortality, of the region to which they migrate.

I include spatially varying growth, so that each region r has its own von Bertalanffy growth parameters. I also include variable growth within regions by assuming that each length at age has a Gaussian distribution centered at the value calculated from the length equation. For example, a cohort at age a will follow a length composition according to the distribution $Normal\big(L(a,r), s^2(a,r)\big)$, and at age $a+1$ the cohort will follow the distribution

$Normal\big(L(a+1,r), s^2(a+1,r)\big)$, and so on. All fish within an age class are aggregated together and redistributed into a new length distribution when their age increases. The process is outlined in the following algorithm:

1. Calculate $N_L(a,t,r) = \sum_{l=0}^{l_{max}} N(a,t,l,r)$
2. Calculate $L(a+1,r)$
3. Calculate $s^2(a+1,r) = v_c \cdot L(a+1,r)$, where v_c = coefficient of variation
4. Generate the distribution $Normal\big(L(a+1,r), s^2(a+1,r)\big)$
5. Multiply $N_L(a,t,r)$ by the distribution to obtain the length distribution of age $a+1$ fish
6. Increment age
7. Return to step 1

This results in the method being memoryless, since the length distribution of an age class at year t is independent of the length distribution of that age class the following year $t+1$. So each year the population of fish for age a are collected together when calculating the length distribution for age $a+1$ and then redistributed to the Gaussian distribution for $L(a+1)$, regardless of any truncation of the length distribution the previous year (Figure 5). This implementation follows the usual implementations of Stock Synthesis 3 (Methot, 2005). Though it has the downside of not tracking dynamic length composition, it still broadly allows us to include variability in growth rates due to changing environmental or biological processes. This growth implementation will be referred to as the regenerative length scenario. A more sophisticated growth model is examined in a following section.

$N(a,t,l,r)$ is comprised of individuals that stayed in region r between years and individuals that migrated in from other regions. The proportion of the population from a region

p that migrates to region r is γ_{pr}, where $\gamma_{rr} = 1 - \sum_{p \neq r}^{r_{max}} \gamma_{pr}$ specifies those that did not migrate between years. Fish begin the year at their current length l and following survival, they grow to length l_{new} according to the growth rules specified in the previous subsection. Migration occurs first in the time step, followed by mortality, then reproduction and growth.

$$N\big(a+1, t+1, l_{new}, r\big) = \sum_{p=1}^{r_{max}} \gamma_{pr} \cdot N(a,t,l,p) \cdot e^{-Z(l,r)} \tag{5}$$

I calculate the proportion of mature fish for a specified length and region with the sigmoid function $p_m(l,r)$. The output is bounded between 0 and 1, which specifies the fraction of fish capable of reproduction that year. The function is parameterized the same as harvest mortality with a length at inflection parameter, $p_{50}(r)$, and slope at inflection parameter, $p_s(r)$; both of which include spatial variation.

$$p_m(l,r) = \frac{1}{1 + exp\Big(p_s(r) \cdot \big(p_{50}(r) - l\big)\Big)} \tag{6}$$

The total number of eggs, $E_T(t,r)$, are calculated from the egg output of mature individuals over all lengths and ages. The fecundity relationship between length and number of eggs produced is determined by the parameters d_1 and d_2. Larger rockfish contribute significantly more to egg production than smaller fish (Dick, 2010), indicating that weight alone is not a sufficient proxy for calculating egg production, therefore by convention $d_2 > 3$ for several rockfish species. I model the total number of eggs as

$$E_T(t,r) = \sum_{l=1}^{l_{max}} \sum_{a=1}^{a_{max}} p_m(l,r) \cdot N(a,t,l,r) \cdot \big(d_1 \cdot l^{d_2}\big) \tag{7}$$

I use deterministic Beverton-Holt recruitment (Beverton and Holt, 1957) to calculate recruits, including spatial variability in recruitment success (Love et al., 1991; Field and Ralston, 2005). First I create a larval pool for each year by summing over regions. Larval dispersal occurs at the beginning of the time step, where the majority of the year is spent in the new region. I let ψ_{pr} denote the fraction of larval drift between regions p and r. The year following recruitment, larvae grow to juveniles with initial length $L(0,r)$. If $\alpha(p)$ represents maximum per capita reproduction in the region p in which spawning/release occurred prior to dispersal, and $\beta(r)$ characterizes density depending on the region r to which larvae settle, then the recruits in region r are

$$N\big(0,t,L(0,r),r\big) = \sum_{p=1}^{r_{max}} \psi_{pr} \cdot \frac{\alpha(p) \cdot E_T(t,p)}{1+\beta(r) \cdot E_T(t,p)} \tag{8}$$

Yield, $Y(a,t,l,r)$, is calculated as the fraction of individuals removed from the population that perished due to fishing, and assumes no incidental fishing mortality; the parameter ω converts the fished biomass from weight in grams to weight in metric tons:

$$Y(a,t,l,r) = \frac{\mathcal{F}(l,r)}{Z(l,r)} \cdot \big(1-e^{-Z(l,r)}\big) \cdot N(a,t,l,r) \cdot W(l,r) \cdot \omega \tag{9}$$

I use a bias function, $\phi(a,t,l,r)$, to quantify the inaccuracy of the assessment model estimates. The function measures the level of error present by computing the relative difference between the estimated quantity, $\hat{A}(a,t,l,r)$, and the true simulated quantity, $A(a,t,l,r)$. The quantity $A(a,t,l,r)$ is different between analyses, but the calculation remains the same.

In order to easily compare between the results of separate assessment scenarios in certain analyses, I calculate the absolute mean of the bias, $\bar{\phi}$, across all variables as a benchmark statistic.

$$\phi(a,t,l,r) = \frac{\hat{A}(a,t,l,r) - A(a,t,l,r)}{\hat{A}(a,t,l,r)} \tag{10}$$

$$\bar{\phi} = \sum_{a=1}^{a_{max}} \sum_{t=1}^{t_{max}} \sum_{l=1}^{l_{max}} \sum_{r=1}^{r_{max}} \frac{|\phi(a,t,l,r)|}{(a_{max} \cdot t_{max} \cdot l_{max} \cdot r_{max})} \tag{11}$$

I can rewrite the bias equation into the more intuitive Equation (12). This version allows us to easily relate the estimated value to the true value. For instance, a bias value of $\phi(a,t) = 0.2$ would result in $\hat{A}(a,t) = 1.25 \cdot A(a,t)$, signifying that the abundance estimate is 25% greater than the true value. I can extend this form to $\bar{\phi}$ to make statements about the percentage of overestimation or underestimation on average across age and time.

$$\hat{A}(a,t) = \frac{A(a,t)}{\big(1 - \phi(a,t)\big)} \tag{12}$$

All variables and parameters used in the rockfish simulator are specified in Table 1.

Example Assessment

The example assessment uses a Schaefer model (Schaefer, 1954) with observation error as outlined in Hilborn and Mangel (1997). This version has carrying capacity K, population growth rate r and catchability q as unknown parameters. The goal of this stock assessment model is to obtain maximum likelihood estimates for the three parameters, so that they can be used to estimate the biomass time series of the population over the twenty years following the onset of fishing. The likelihood estimates are calculated numerically with a wide range and fine granularity of values. Further details on the method are in Hilborn and Mangel (1997). In Table 2, I show the differences in assumptions between the simulator and the stock assessment model. The model is intentionally simple in order to provide an illustrative example of how the simulator can be used. Current work is being done on using the simulator with Stock Synthesis 3, a stock assessment tool currently used for many rockfish assessments.

For the simulator, I model a chilipepper (*Sebastes goodei*) population using life history parameter values from the literature (Field, 2009; Love et al., 2002), and one region. The time series for total annual yield in tons is calculated from the simulator without error as catch and input into the stock assessment as $C(t)$,

$$C(t) = \sum_{a=1}^{a_{max}} \sum_{l=1}^{l_{max}} \sum_{r=1}^{r_{max}} Y(a,t,l,r) \tag{13}$$

The time series has a length of 20 years, and begins with the onset of fishing in the simulator. The catch is generated from a constant value of $\mathcal{F}_{max}(r) = 0.1$ across all 20 years. The stock assessment additionally requires an index of abundance time series, which represents a measurement of total biomass. In order to generate this I compute the true biomass $B_T(t)$ from the population $N(a,t,l,r)$ by converting the units to tons and creating dependence only on time.

$$B_T(t) = \sum_{a=1}^{a_{max}} \sum_{l=1}^{l_{max}} \sum_{r=1}^{r_{max}} N(a,t,l,r) \cdot W(l,r) \cdot \omega \tag{14}$$

The observed index of abundance $I_{obs}(t)$ is calculated from the true biomass $B_T(t)$. I assume the index is an imperfect measurement of the abundance of the population subject to observation error, which I implement as a log-normal perturbation of the population biomass, where $X(t) \sim Normal(0, \sigma_x^2)$.

$$I_{obs}(t) = B_T(t) \cdot e^{X(t)} \tag{15}$$

The equation does not include the unbiasing term $-\frac{1}{2} \cdot \sigma_x^2$ in the exponent since runs performed with its inclusion showed no significant impacts. I assumed robust measurements and used the value $\sigma_x = 0.01$. Since this is a random time series, I generated 50 observed index of abundance time series for the stock assessment model to generate 50 estimates of abundance.

In addition, I ran the assessment under several scenarios to study the effects of biological assumptions and data availability. The assessment model uses total biomass without length composition, as a consequence, constant natural mortality (size-independent) must be in used in the assessment. I vary the level of size-dependence operating on natural mortality within the simulator as specified previously. This allows us to examine the error generated by a mismatch in natural mortality assumptions between the operating and assessment models.

In order to examine the impacts of data availability, I enter catch time series of varying lengths, from the same run, into the assessment model. The simulator uses constant natural mortality during these runs so that the error generated is attributable to the time series variation, and not disparity in natural mortality assumptions. The time series for catch across the entire 20 years is denoted $C_{1,20}(t)$, the time series for the first and last 15 years are $C_{1,15}(t)$ and $C_{6,20}(t)$, respectively, and the time series for the first, middle and last 10 years are $C_{1,10}(t)$, $C_{6,15}(t)$ and $C_{11,20}(t)$, respectively.

Example Assessment - Results

The abundance estimates for rockfish vary significantly between natural mortality scenarios. Under the first scenario, $M^*(1)$, where the simulator and assessment model both assume constant mortality (size-independent), the assessment generates $\bar{\phi} \approx 0.146$, which translates to an average overestimate of the true biomass by approximately 17% (Figure 6). This reflects the error independent of natural mortality assumptions, and can be associated with the lack of age-structure within the assessment model. When the simulator uses a natural mortality rate with 25% size-dependence, the absolute average bias nearly quadruples $\bar{\phi} \approx 0.535$, resulting in an average overestimate of roughly 115%. This increase in bias shows the amount of error associated with the increasing mismatch in natural mortality assumptions.

Under the remaining scenarios I obtain $\bar{\phi} \approx 0.7$, and the time series for the bias are nearly identical. This indicates that the most significant changes to average absolute bias between mortality scenarios occur between those with less size-dependence. In order to better understand the effect of the size-dependence of natural mortality on assessment accuracy, I perform additional simulations to fill in the gaps between $M^*(1)$ and $M^*(2)$, as well as $M^*(2)$ and $M^*(3)$. I use 5% increments between the 3 cases, generating 8 more scenarios. With the new estimates, a relationship emerges, between percentage of size-dependent natural mortality and absolute average bias, that is clearly asymptotic (Figure 7). We can also observe a clear monotonic increase in bias as the the simulator's natural mortality assumption diverges from a constant implementation.

When I vary the amount of data that is entered into the stock assessment model, I also observe a significant change in the accuracy of the estimates (Table 3). For time series that start on the same year ($C_{1,20}(t)$, $C_{1,15}(t)$, $C_{1,10}(t)$ and $C_{6,20}(t)$, $C_{6,15}(t)$), absolute average bias increases as the length of the time series decreases. Starting at the onset of fishing, when the time series decreases from 20 to 15 to 10 years, the average overestimate in the abundance increases from 17% to 43% to 111%, respectively. Similarly, when starting from year 6, when I truncate the time series length from 15 years to 10 years, I observe an increase from 71% to 112% in average overestimation. Additionally, the starting year of the time series considerably influences the accuracy of the abundance estimate. When I examine $C_{1,15}(t)$ and $C_{6,20}(t)$, there is a sharp increase in average overestimation from 43% to 71%.

Full Assessment

I now examine the stock assessment tool used for *Sebastes*, Stock Synthesis 3 (Methot, 2005), which is a sophisticated statistical model. It can take as input a large variety of information, such as time series for catch, indices of abundance, non-fishery survey counts, age and length composition samples, as well as parameter values for life history and fishery characteristics. The assessment tool then calculates estimates of time series such as population biomass, fecundity, and fishing effort, and updated estimates of the life history and fishery parameter values.

As in the example assessment, I study one of the biological assumptions inherent to SS3, namely how the length distribution of the population varies over time. The SS3 implementation has the length at age follow a Gaussian distribution, where the mean is $L(a)$ and the standard deviation is specified by the user (though a σ is typically selected that is approximately 10% of the mean). However this method is also memoryless, meaning that the length distribution of an age class at year t is independent of the length distribution of that age class the following year $t+1$. So each year the population of fish for age $a-1$ are collected together when calculating the length distribution for age a and then redistributed to the Gaussian distribution for $L(a)$, regardless of any truncation of the length distribution the previous year due to fishing. This regenerative process produces a strict bound around the $L(a)$ values and ignores all size-selective mortality acting on the population. This matches the implementation of growth specified in the simulator section.

I examine a new version of growth implementation which retains a memory of the population's length structure, and introduces variability to growth through the actual underlying mechanisms. Since new lengths are not generated each year, growth in the simulator will occur by accruing annual growth increments. I use the von-Bertalanffy size equation to calculate the change in length, $\Delta L(a)$, given a current length, $L(a)$.

$$\begin{aligned} L\big(a+\Delta a\big) &= \frac{q}{k}\cdot\Big(1-e^{-k\cdot\Delta a}\Big)+L\big(a\big)\cdot e^{-k\cdot\Delta a} \\ \implies L\big(a+\Delta a\big)-L\big(a\big) &= \Delta L\big(a\big)=\Big(\frac{q}{k}-L(a)\Big)\cdot\Big(1-e^{-k\cdot\Delta a}\Big) \end{aligned} \tag{16}$$

Now to introduce variation in $\Delta L(a)$, I want a method that addresses the source of the variation by addressing stochasticity in biological and environmental processes, instead of an ad-hoc distribution on the function for length. By examining the initial growth differential equation, $\frac{dL}{da}=q-k\cdot L$, we see that q and k represent gains and costs to growth. These processes are the true source of variation, since intrinsic and extrinsic factors drive the physiological response of the fish. I introduce distributions on these parameters, which in turn results in a distribution P for $\Delta L(a)$.

$$k\sim\Gamma\big(\alpha,\beta\big) \text{ and } q\sim N\big(\mu,\sigma^2\big) \implies \Delta L(a)\sim P \tag{17}$$

This above formulation results in $\Delta L\big(q,k,L(a)\big)$. Growth depends on draws from the distributions for q and k, as well as current length. However, for our particular goal, it may be more intuitive to think of the dependence as $\Delta L(l)$ with current length l, since length is not a strict function of age. This results in the desired effect of variability in growth tied to biological factors, with a growth model that retains a memory of current length.

The number of $\Delta L(l)$ distributions depends on the maximum size of the species, and the granularity in the implementation of length (i.e. centimeters, millimeters, etc.), since there will be a $\Delta L(l)$ distribution associated with each length l. The distributions will be non-standard due to the $\frac{q}{k}$ term in Equation (16). Consequently, each distribution will be generated by taking several draws from the distributions for q and k in order to generate a distribution for each $\Delta L(l)$ through numerical computation. So fish at current length l will grow at varying rates according to the growth increment distribution for that length, $\Delta L(l)$. I implement this by taking the fish of current length l, $N(a,t,l,r)$, and apportioning them by the weights of the distribution for ΔL to obtain $N(a+1,t+1,l+\Delta L(l),r)$ (Figure 8). This implementation is referred to as the non-regenerative length scenario, and is outlined in the following algorithm:

1. Generate the $\Delta L(l)$ distribution
    (a) Make a draw from $k \sim \Gamma\big(\alpha,\beta\big)$
    (b) Make a draw from $q \sim N\big(\mu,\sigma^2\big)$
    (c) Calculate a point in $\Delta L(l)$ (Equation 16)
    (d) Return to step 1a until $\Delta L(l)$ is sufficiently determined (10000 draws were used)
2. Obtain the shifted distribution $l+\Delta L(l)$ from current length l
3. Multiply $N(a,t,l,r)$ by the shifted distribution to obtain the new length distribution
4. Increment length
5. Return to step 1

With regenerative length distribution (outlined earlier) and non-regenerative length distribution implementations, I analyze the changes in accuracy to SS3's estimates between the assumptions, in the same way that I analyzed the impacts of length-dependence on natural mortality in the previous example assessment. I use Greenspotted rockfish (*Sebastes chlorostictus*) life history parameter values (Benet et al., 2009; Dick, 2011), and a single region. Within the implementations for regenerative and non-regenerative length distributions, I vary other inputs in the stock assessment to generate several fishery scenarios.

SS3 takes prior parameter estimates, and ranges of parameter values, when it generates posterior estimates of the parameter values. This is another source of data that I examine by varying the number of parameters estimated. The scenarios estimate the logarithm of initial recruitment, $log(R_0)$, as a parameter value. This parameter value is used in the generation of the Beverton-Holt parameter values, converting from a steepness parameterization. The prior estimate value fed in to SS3 was a slight perturbation of the true value, and a wide range of possible values was given to allow the assessment model to widely explore the parameter space. Other scenarios also estimated the fishing selectivity parameters, P_{50} and P_{95} (P_{50} corresponds to $\mathcal{F}_{50}$, P_{95} replaces $\mathcal{F}_s$ and is from an equivalent, reparameterized version of the selectivity function used in SS3). The initial estimates of these parameters input into SS3 were also slight perturbations of the true values, and both had wide ranges for parameter space. All other parameters were fixed at their true values.

The scenarios are also differentiated by the amount of length composition data fed into the stock assessment model. Since SS3 takes length samples as input, I calculated samples from the simulator. I implemented this by taking multinomial draws from the fish caught by harvest for each required year, where the weight of a draw from a particular length bin was based on the proportion of the total population that had that specific weight. This allowed the length sampling to accurately reflect the same proportions that length sampling dockside would encounter. The length sampling was varied by the number of years sampled, and the number of samples taken within a year. For a data poor scenario, I use 10 years with 50 samples each year. For the data rich scenario, I use 20 years with 200 samples each year. Additionally, since there are multinomial draws in the generation of the length data, they are stochastic and 50 runs were performed in each scenario to generate a distribution of estimates. This work covers eight scenarios with 3 varied factors: data poor and data rich, 1 and 3 parameters estimated, and regenerative vs. non-regenerative length distributions (Table 4). The simulator used both length distribution implementations, and the stock assessment model only used the regenerative length distribution assumption. All scenarios use the same $\mathcal{F}_{max}(r)$ time series, which is time-dependent and non-constant in this section

(Dick, 2011). Fishing mortality is commonly assumed to follow the relationship given by the equation $F = q \cdot E$, where F is fishing mortality, q is catchability, and E is fishing effort. When $q = 1$, effort and mortality are equivalent, which I assume in the following analyses with $\mathcal{F}_{max}(r)$. This time series will also be estimated in all the scenarios. No index of abundance data are used in the following analyses, only catch data.

Full Assessment - Results

Since there are many time series estimates generated by SS3, I discuss only depletion and fishing effort. Depletion is the fraction of the current population egg output compared to the unfished population egg output. Values will always be between 0 and 1, and the time series will always start at 1. Additionally, depletion can be directly compared between the regenerative and non-regenerative length scenarios since they are scaled into fractions. Fishing effort is the $\mathcal{F}_{max}$ parameter, assumed here to be time-dependent, and can also have its estimates compared between scenarios since they use the same values. I also examine the bias of the time series estimates using Equation 10. The accuracy of the estimates correspond to the proximation of the mean result to the true value. The width between quantiles, or the spread of the results, corresponds to the precision where a larger width signifies less precise results.

The depletion estimates are qualitatively similar to the true simulator depletion, but an offset exists which worsens over time (Figure 9). Additionally, the width between the 10% and 90% quantiles in the estimates get worse over time. An important observation is that even under the data rich scenario with less estimation and matching length regeneration assumptions, $S_{1,r}$, the offset exists. This is likely a result of a mismatch in other assumptions between the simulator and the stock assessment model, in particular the implementation of the timing of certain events within the year. SS3 uses a mix of continuous and discrete dynamics, while the simulator uses only discrete dynamics.

The mean depletion estimates under the regenerative length scenarios are more accurate than those under the non-regenerative length scenarios. However, width between the 10% and 90% quartiles is smaller under the non-regenerative length scenarios. When we examine the bias another interesting difference emerges (Figure 10). Under regenerative length, $S_{1,r}$ performs the best and $S_{3,p}$ performs the worst, as expected. In the remaining scenarios, the number of parameters estimated has a bigger impact than the amount of data, with $S_{1,p}$ having improved accuracy over $S_{3,r}$. This relationship changes under the non-regenerative length scenarios, with $S_{3,r}$ trading spaces with $S_{3,p}$. Additionally the depletion bias is significantly worse under non-regenerative length, with the underestimates worsening by as much as 20% between respective scenarios. Across all cases, the bias trends towards more negative values over time.

The stock assessment estimates for fishing effort exhibit different patterns. The estimates are again qualitatively similar with some offset (Figure 11). The accuracy of the mean estimate changes over time. When fishing effort is low, the optimal scenario, $S_{1,r}$, does not perform the best (Figure 12). This is another example of the effect of the mismatched assumptions. However as fishing effort starts to increase, the bias increases over all scenarios resulting in $S_{1,r}$ performing best over the years with greatest fishing. During the years of increased fishing, non-regenerative length scenarios perform worse, with overestimation increasing by as much as 46% between scenarios with regenerative length distributions. The width between quantiles of the estimates increases with less data, and increases further under the non-regenerative length scenarios.

The parameter estimates were also strongly effected by data quality and model assumptions. The regenerative length distribution scenarios (Figure 13) generated worse estimates of $log(R_0)$ than the non-regenerative scenarios (Figure 14). This is also likely a result of the offset seen in previous results. In the most data rich scenarios, $S_{1,r}$ and $S^N_{1,r}$, the quantile width is smallest, while for the most data poor scenarios, $S_{3,p}$ and $S^N_{3,p}$, it is largest. However the non-regenerative results exhibited smaller quantile widths than the regenerative results.

In contrast, the scenarios with regenerative length distributions had more accurate selectivity parameter estimates (Figure 15) than those with non-regenerative length distributions (Figure 16). There was also no significant relationship between model assumption and the quantile width of the selectivity parameter estimates. However, data rich scenarios had smaller quantile widths than data poor ones.

Discussion

Performing a stock assessment involves making assumptions about the implemented model and the stock, based upon available data. When the assumptions within a stock assessment do not match the true dynamics of a species, or too little data are available, an increasing disparity between the abundance estimated by the assessment and the true abundance is expected. I have presented a tool for quantifying the robustness of stock assessment estimates by identifying these disparities, over a range of available data. Through simulation and implementation of two stock assessments, I have demonstrated that a significant disparity can arise for rockfish stocks.

The disparity is present in the sample assessment even when natural mortality is assumed constant in the simulator and the assessment model, which indicates further sources of error in the estimates. The simulator contains important dynamics that are not present in the assessment. The rockfish life history simulator includes size-dependent life history and environmental characteristics, while the sample assessment tool treats each unit of biomass the same. There is no capability in the example assessment model to handle selectivity in any form. Additionally, certain life history characteristics are simplified in the example assessment, for example egg production and Beverton-Holt stock recruitment in the rockfish version of the simulator are approximated in the example assessment model through population growth and carrying capacity parameters. The additional dynamics within the simulator create complexity that is not captured within the sample assessment tool. Similar to the Namibian hake fishery mentioned in the introduction (Hilborn and Mangel, 1997), the choice of assessment model imposed restrictions on which parameters could be specified a priori, which data could be included, and thus what types of uncertainty were present in the results. Additionally, the length of the time series results in differences in abundance of nearly 100%. This indicates that the length of the time series available can strongly influence the accuracy of the assessment results.

The Stock Synthesis 3 assessments results also showed disparities. In this case, mismatch is due to differing assumptions in the life history of the species, outside of the population length distributions, and not simplified dynamics. This is another illustration that careful consideration is needed when building models, since model assumptions consequently lead to biological assumptions in the life history or population dynamics. Yet the difference in results between regenerative and non-regenerative growth also show that how length distributions in the population are treated is an important consideration in stock assessments, and can result in model misspecification. Considering that regenerative length distributions are an unlikely assumption in the actual population, treating them as such in the assessment model appears to be inappropriate in light of these results. However, when the population has a non-regenerative length distribution and the assessment model assumes a regenerative length distribution, the assessment model compensates by increasing fishing and decreasing depletion. This results in a more conservative population estimate, which could subsequently result in more risk-averse fishing policy decisions.

I have shown that careful consideration must be made when collecting and using data for analyses, since the type of data is just as important as the amount of data. Improving data poor fisheries will involve more than just expanding the length of the time series of data. The type of data needs to be appropriate for the life history considerations of that species, and must capture any consequential impacts on the population. The example assessment revealed that the initial years of fishing are important to include since they have a dramatic impact on the state of the population. The full assessment showed that having more information on life history parameters can strongly impact assessment accuracy as well. The lack of appropriate data can lead to the use of inappropriate models, as well as generating and maintaining a large degree of uncertainty in the model results, which both lead to inaccurate and unstable estimates. Decreasing uncertainty will involve obtaining improved data, which will allow for more applicable models to be used.

In the analyses I examined simplified and complex stock assessment models. Both exhibited changes to accuracy when the amount and quality of data were varied, and when simulator assumptions were altered. The accuracy of the simplified model was much more sensitive to input changes than the complex model, and was more inaccurate and imprecise. However the complex model required significantly more parameters as input. This illustrates another issue in constructing models: how much realism to include in the model. Complex models have the potential to be more accurate since they are likely to include more factors impacting the system. Yet the complexity brings significant costs. Including more factors means that more data and information are necessary, for example SS3 requires several life history parameter in addition to data time series. This leads to a need for more experiments or fieldwork to produce adequate parameter values. Complex models are also more computationally intensive, requiring longer run-times or faster computers. Furthermore, there is no guarantee that the added complexity captures the driving factors in the system, so there is the risk that the additional costs produce no added benefit. Yet simple models have a different set of problems. Their estimates are likely to be more inaccurate and more sensitive to assumptions and data availability, leading to potentially less valuable results. Both kinds of models have several potential issues, illustrating again the importance of careful model construction. The ideal balance between parsimony and realism will vary between systems and problems, but it must also be considered to ensure effective models.

Although previous studies have considered the importance of different types and amounts of data for assessments (Brooks et al., 2010; Wang et al., 2009), few studies have been done to examine the relationship between stock assessment robustness and data richness through a life history simulator (He et al., 2011). With further work I anticipate the ability to discover species-specific thresholds of data that ensure reasonable assessment accuracy. Additionally, this work examined only a single region, so a natural extension would be to analyze a stock with multiple regions to study spatial impacts. This also has the potential to assist in data collection by specifying which types of data are most valuable for specific stocks. The simulator can also be applied to other fisheries of interest with reparameterization.

These results will help to identify the weaknesses of current methods in order to inform the implementation of future stock assessment methods, which will lead to improved policy decisions for fisheries management.

Section 2 - Mosquito Attack Behavior and Its Adaptions to Disturbance

Background

Malaria continues to be an enormous public health concern despite massive efforts towards eradication of it and its mosquito vectors (World Health Organization, 2010). It is still a huge source of mortality worldwide, devastating communities in both the short and long term. In order to minimize and prevent future outbreaks, we need to learn more about the behavior of the infection through the behavior of the vector. Mosquitoes can be difficult to study, especially in the field, due to their size and behavior. For example, *Anopheles gambiae*, the major malaria vector and the species I will examine, seeks blood at night, primarily from sleeping human hosts, and is smaller and more delicate than many other species, such as *Aedes aegypti*. Additionally, *A. gambiae* prefers humid environments and can have sophisticated host detection and attack techniques. The mechanisms that drive the mosquito response to assorted stimuli are poorly understood. While research is being done on *Anopheles* feeding behavior (Stone et al., 2011; Roitberg et al., 2010), there is still much to learn.

As such, the interaction between *Anopheles gambiae* and humans can be characterized as a data poor system. This necessitates a method to supplement experimental work, both in aiding in the understanding of the interactions between host and parasite, and in optimizing data collection, saving both cost and time. By examining certain behaviors and life history characteristics, the important factors driving the system can be identified. For example, identifying a factor as having no impact on mosquito behavior can help inform future experiments on what measurements to perform to optimize research funds.

The most important part of the interaction is the method of transmission, the mosquito's feeding behavior. Gaining more information on what drives this interaction, and specifically how she handles the feeding scenario under different conditions, would be a great boon to curbing further infections. By better understanding the mechanism of infection, we can better target vulnerabilities in the infection process to disrupt its spread. A practical application is to use this information to make additional inferences about how the mosquitoes will respond to current or future anti-malaria methods, such as the modeling approach used to analyze mosquito response to chemical treatments and bednetting (Roitberg and Mangel, 2009).

The natural aid to a data poor system, as illustrated in the previous section, is a modeling tool. In generating a model, it is important to understand the life history of *Anopheles gambiae*. Female mosquitoes will generally be sexually mature within 2 days following the pupal stage. A single mating encounter will provide her with all future fertile eggs. Following mating she will seek her first blood-meal and begin her first gonotrophic cycle, which refers to the period where she seeks and obtains a blood meal, and then oviposits. She seeks blood-meals at night in order to help develop her eggs, from (typically sleeping) human hosts. When her somatic reserves are low she will seek nectar, as she derives little energy for herself from the blood-meal (van Handel, 1965). As such nectar is generally used for energy, while blood is used for egg development (Foster, 1995). Once she has obtained an appropriate amount of blood, she will find a nearby surface and spend the next hour removing water from the blood-meal. She then seeks a safer location to spend the following days (generally 2) digesting the blood to produce a batch of eggs. She will lay her developed eggs in a nearby source of water, before beginning a new gonotrophic cycle. Females rarely have more than a few successful gonotrophic cycles in their lifetime, and rarely live longer than a few weeks (Spielman and D'Antonio, 2001). In this section, I propose a model which incorporates the important characteristics of the adult female mosquito life history in order to supplement the experimental data-gathering techniques already in use. Furthermore, it can be used to help inform existing models.

I use a model to examine how mosquito, in particular the major malaria vector species *Anopheles gambiae*, feeding behavior adapts to host disturbance. When a mosquito lands on a host seeking a blood meal, she chances disturbing the host and incurring a defensive response. Through the initial probing phase and the subsequent feeding process she may be felt by brushing a hair or as a result of sensitive flesh. Once disturbed, the host may ignore the mosquito or attempt to dislodge or kill her. Additionally some species prefer to feed on sleeping hosts, minimizing the possibility of the host defensively responding to the mosquito following a disturbance. She must decide whether to risk continue feeding on the host following these disturbances, or leave and find a new host (Figure 17). Her decisions will be driven by her current energy reserves (Ma and Roitberg, 2008; Tsunoda et al., 2010).

Behaviorial Modeling Framework

The process is modeled through stochastic dynamic programming (SDP), a common optimization approach for determining optimal animal behavior (Mangel and Clark, 1988; Clark and Mangel, 2000). The model contains a suite of decisions available to the animal at each time step, except the final one, where each decision is associated with a fitness consequence which is maximized over the behaviors to select an optimal action. Here the measure of fitness is egg or offspring survival. The decisions in the model are not true decisions made by the animal, but reflect optimal evolved responses, given a current state of the animal or system.

The final time step is often defined to be the end of a breeding season, or the end of the animal's life, therefore no decisions may be selected in the final time step. Instead a function is defined which converts the state variables into fitness, specified as the terminal fitness function. With the fitness values defined at the end of the time horizon, the model moves backward one time step. Current fitness is then calculated from a discounted (by mortality and other factors) terminal fitness. This process is iterated backward in time until the start of the season, across all state variables for each time step. This creates an array of optimal behaviors for each combination of state variables.

I use SDP to create a model to determine the attack behaviors of a female mosquito that optimize her maximum expected reproductive fitness, measured in successful egg production. The model contains the following state variables: blood reserves (b) which range from 0 to b_{max} and tracks the amount of blood the mosquito has fed on recently; energy reserves (e) which range from 0 to e_{max} and tracks the amount of energy the mosquito has to fuel her search for a blood meal and maintain soma; host awareness (a) which indicates whether the mosquito has disturbed the host and takes the value 0 if the host hasn't been disturbed, and 1 if the host has been disturbed; time within the modeled night is (t), which begins at 0 and ends at T.

The variables determine the fitness function, which I define as $R(b, e, a, t)$ = maximum accumulated reproductive success when blood reserves are b, energy reserves are e, host awareness is a, and the current time step is t. The decisions available to the mosquito (and the associated function) are to seek a new host ($V_s(b, e, a, t)$), to feed on the current host ($V_f(b, e, a, t)$), or to leave and oviposit ($V_o(b, e, a, t)$). These functions are defined in the following pages. Then

$$R(b, e, a, t) = max\left[V_s(b, e, a, t), V_f(b, e, a, t), V_o(b, e, a, t)\right] \tag{18}$$

When the host defends itself the mosquito is forced to flee, and she will have only a subset of decisions available to her: Seek a new host, $V_s(b, e, a, t)$, and leave to oviposit, $V_o(b, e, a, t)$. In these cases I use the fitness function $L(b, e, a, t)$ = maximum accumulated reproductive success when blood reserves are b, energy reserves are e, host awareness is a, and the current time step is t.

$$L(b,e,a,t) = max\left[V_s(b,e,a,t), V_o(b,e,a,t)\right] \tag{19}$$

At the final time $t = T$ she has no time left to seek a blood meal, so she will attempt to oviposit with what she has, so that $R(b,e,a,T) =$ is determined by the eggs she can produce. The fitness benefit she gets is given by the function $r_b(b,e)$, determined by her blood and energy reserves, so that $R(b,e,a,t) = r_b(b,e)$. Blood meal size has a large impact as that directly relates to the amount of eggs produced (Briegel, 1990). Both variables must meet a minimum threshold to account for the mosquito needing a base amount of energy to survive and a base amount of blood to even begin to produce eggs successfully, which is implemented through the indicator function $I_{\{e \geq 10, b \geq 10\}}$. Indicator functions take a value of 1 when the arguments in the subscript are true, and 0 when they are false. The parameter s_c converts from volume of blood reserves into number of eggs. Then

$$R(b,e,a,T) = r_b(b,e) = \frac{b}{s_c} \cdot I_{\{e \geq 10, b \geq 10\}} \tag{20}$$

The probability that the host becomes aware of the mosquito is P_a. Once detected, the probability that the host subsequently defends against the mosquito is P_d. I assume that the mosquito knows when it has disturbed the host, due to the host tensing or shifting at the disturbance. Though these probabilities will likely vary by host (and hence follow some distribution), for this version of the model they are fixed.

When the host does defend itself, the mosquito's survival varies both by the amount of blood she's carrying and whether she chose to stay or flee following the disturbance. The mosquito's speed and ability to evade an attack decrease as she carries more blood due to the increased mass (Roitberg et al., 2003). Additionally, if the mosquito chooses to flee she has more time to react since she has already begun to withdraw her proboscis. I incorporate this by using separate mortality rates associated with the mosquito staying on the host, $\mu_1(b)$, or fleeing the host, $\mu_2(b)$, when the host defends itself. I implement these as linear functions, where $\mu_1(b)$ has a greater slope and intercept than $\mu_2(b)$, so that $m_1 \geq m_2$ and $k_1 \geq k_2$.

$$\mu_1(b) = \frac{m_1}{b_{max}} \cdot b + k_1 \tag{21}$$

$$\mu_2(b) = \frac{m_2}{b_{max}} \cdot b + k_2 \tag{22}$$

As the mosquito feeds on the host, there will be diminishing returns to the blood reserves, where the increment to blood reserves is specified as $g(b)$. She will fill up with blood, and begin to excrete some of the liquid in order to keep more of the protein that she needs for egg development. The parameter k_g determines the amount of blood she gets once she becomes full and the diminishing returns are in full effect. The g_1 and g_2 parameters control how quickly the diminishing returns act given her current blood reserves. Then

$$g(b) = k_g + \frac{g_1}{1 + g_2 \cdot b} \tag{23}$$

The mosquito makes decisions according to her current situation (determined by the current variable values). With this decision she is either leaving another host that became aware of her ($a = 1$), or she is beginning the night ($a = 0$). This assumption is valid since she will have no reason to leave a host that does not sense her, so she may obtain a full blood meal. If she is leaving another host, she must survive the mortality associated with the host defending itself given that she decided to leave before the host's action $\mu_2(b)$, expending e_d

units of energy in the process. Her ability to survive this source of mortality is dependent on her current blood reserves, due to the hindrance of the additional mass. With both of these scenarios she has to survive some background mortality μ over time t_s to find a new host. She expends e_h units of energy seeking the host. Then the fitness value of seeking a new host is

$$V_s(b,e,a,t) = \begin{cases} \text{if } a = 1 \\ \qquad P_d \cdot e^{-\mu_2(b)-\mu} \cdot \Big[R(b, e - e_h - e_d, 0, t + t_s)\Big] \\ \qquad + \big(1 - P_d\big) \cdot e^{-\mu} \cdot \Big[R(b, e - e_h, 0, t + t_s)\Big] \\ \\ \text{if } a = 0 \\ \qquad e^{-t_s \cdot \mu} \cdot \Big[R(b, e - e_h, 0, t + t_s)\Big] \end{cases} \tag{24}$$

When the mosquito continues to feed, V_f, there is the possibility that the host is aware of the mosquito ($a = 1$) or isn't aware ($a = 0$). If the host is aware, then the host may defend itself with probability P_d. The mosquito's ability to survive this is determined by the amount of blood she's carrying, since carrying more blood will impede her escape velocity and agility. The mortality associated with host defense given that she did not decide to leave is $\mu_1(b)$. Leaving the host takes t_a time and uses e_d units of energy. If there is no disturbance then the mosquito continues to feed for a single unit of time, using e_f energy and gaining $g(b)$ units of blood. The fitness function $L(b, e, a, t)$ is used in the case where the host defends itself, since she will always flee in that situation and doesn't have the option to continue feeding.

$$
V_f(b,e,a,t) = \begin{cases} \text{if } a = 1 \\ \qquad P_d \cdot e^{-\mu_1(b)} \cdot L(b, e - e_d, 0, t + t_a) \\ \qquad\quad + \big(1 - P_d\big) \cdot \Big(P_a \cdot R(b + g(b), e - e_f, 1, t + 1) \\ \qquad\quad + \big(1 - P_a\big) \cdot R(b + g(b), e - e_f, 0, t + 1)\Big) \\ \\ \text{if } a = 0 \\ \qquad P_a \cdot R(b + g(b), e - e_f, 1, t + 1) \\ \qquad\quad + \big(1 - P_a\big) \cdot R(b + g(b), e - e_f, 0, t + 1) \end{cases} \tag{25}
$$

When the mosquito decides to leave and oviposit, V_o, there is again the possibility of the host being aware ($a = 1$) or unaware ($a = 0$). If she is fleeing due to the disturbance associated with the host's awareness and if the host defends, she must survive the associated mortality $\mu_2(b)$. Otherwise she leaves to oviposit. She will choose this decision at some point within the modeled night, though her blood reserves may be from feeding on multiple hosts within the domicile if she was disturbed. If she survives the following two nights of background mortality $2M$ while she finishes egg development, she will oviposit. When she makes this decision, the model moves to the end of the time horizon and ends.

$$
V_o(b,e,a,t) = \begin{cases} \text{if } a = 1 \\ \qquad P_d \cdot e^{-\mu_2(b) - 2M} \cdot R(b, e, a, T) \\ \qquad\quad + \big(1 - P_d\big) \cdot e^{-2M} \cdot R(b, e, a, T) \\ \\ \text{if } a = 0 \\ \qquad e^{-2M} \cdot R(b, e, a, T) \end{cases} \tag{26}
$$

All variables, parameters, and functions used in the mosquito behavior model are specified in Table 5, along with the values used in the analysis. The final values chosen are illustrative, though a sensitivity analysis was run across a range of values.

Results

The behavior of the mosquito is relatively independent of time until near the end of the night. As such, the following results are for a limited time horizon ($t_{max} = 100$) since the results are nearly identical, except for when the transition in optimal behavior occurs. The mosquito has stationary optimal behavior before $t = 85$ (Figure 18), where it changes due to there not being enough time left to seek a new host (Figure 19), and then once more at the end where she must leave to oviposit.

Her decisions strongly depend on her state. Prior to $t = 85$, if there is no threat of host awareness, she will choose to continue feeding on the current host if either her blood or energy reserves are below a threshold, otherwise she leaves to oviposit. She only chooses to seek a new host when the current one has become aware of her, and then only when her somatic reserves are above a threshold, but her blood reserves are low. Otherwise she oviposits if above an intermediate energy and blood reserve value, and continues to feed otherwise. When energy reserves are low, she doesn't have enough energy to leave the host and will thus continue feeding on it even when her crop gets full.

The fitness values are roughly similar for the unaware host cases (Figure 20). When the host is aware the fitness dynamics change significantly (Figure 21). Remaining on the host becomes significantly less optimal, and seeking a new host becomes less optimal for small energy reserve values.

Between $t = 80$ and $t = 84$, the region in which the mosquito will seek a new host, provided that her current host is aware, narrows until it is gone at $t = 85$. So as time increases, the blood reserve level necessary for her to risk seeking a new host decreases.

When $t = 85$, her behavior on an unaware host is unchanged. However when she is at an aware host, she will always leave to oviposit unless her energy or blood reserves are low. Seeking a new host is no longer an option due to time constraints. At the final time step her only decision is to leave and oviposit.

Varying the parameters revealed additional behavior dynamics. Increasing the rate of blood intake for $g_b(b)$ results in the mosquito being more likely to seek a new host, since she is able to fill her crop with ease, and thus only oviposit when she has a larger blood meal. When blood intake is diminished, she becomes less willing to seek a new host and more likely to leave and oviposit with her current blood reserves due to the rapid diminishing returns.

Increasing mortality results in the mosquito being more likely to oviposit with a lesser blood meal, since the risk of death is too high to try to increase her blood reserves. Decreasing mortality has the opposite effect, making her more willing to persist since there is little risk to her. These results are true for when all forms of mortality are varied, including host defense and background mortality. They are also true when host awareness is varied, since it changes the likelihood of host defense indirectly.

Changing energy costs and travel time yield intuitive results. Higher energy costs or travel times result in more conservative mosquito behavior, she will not seek out new hosts as often; and lower energy costs or travel time result in the mosquito being more willing to seek new hosts since her energy and time are not as constrained.

Discussion

There are several mechanisms driving mosquito behavior. Her state is an important factor in her optimal decisions, since she must balance obtaining as much of a blood meal as possible without using up her energy reserves. She must also be aware of her surroundings and especially must be aware of the state of her host. Despite all those states playing a significant role in her feeding behavior, the time state variable did not. Additionally, the hosts are unlikely to be aware of her presence and asleep, given that *Anopheles gambiae* typically feed on blood at night. Ingesting an entire uninterrupted blood meal lasts only a few minutes within the span of an entire night, so it would be logical for her behavior to not change drastically due to time under these conditions.

In addition to her state, the mosquito's behavior strongly responds to mortality. Macdonald (1957) indicated that adult female mosquito survivorship was a significant factor in the spread of malaria. I found that mosquitoes were much more likely to give up on feeding earlier in the evening if there was significant mortality present. Thus decreased survivorship resulted in a more timid mosquito that was less willing to continue attacking her current host or to even seek a new one. As a consequence, this would lead to less opportunities for parasite transmission from vector to host.

She reacted similarly when host awareness and defense was increased. She was more likely to leave to oviposit when the host was more often aware of her, and also willing to defend itself. The disturbance of the host becoming aware of her (a sleeping person shifting or someone scratching or swatting at the bite area) also results in inhibited bite behavior, since the host disturbance has an assumed threat of mortality as well. This indicates that not just mortality, but threat of mortality through a disturbance resulted in more conservative mosquito behavior.

Additionally the mosquito changed her behavior significantly based on her efficiency in taking in a blood meal. She became more willing to seek new hosts since she was easily able to increase her blood reserves. This could potentially lead to an increased number of transmissions if she moves from host to host with the malaria parasite.

These results show which environmental conditions or mosquito states result in more aggressive or conservative mosquito attack behavior. This can help inform future work on ways to inhibit the attack process by taking advantage of expected mosquito responses. Additionally, this model illustrates several important areas in the attack cycle that appear to be important factors in mosquito response, and would provide fruitful topics for future research.

Conclusions

Though the two systems and modeling approaches are vastly different, they share many similarities in their results, and many shared insights may be gained from them. Uncertainty in model specification can lead to significant impacts on estimate results. Varying model and parameter specification can lead to big changes in the population's or individual's behavior. The rockfish simulator showed that changing natural mortality from constant to length-dependent had a huge effect on the accuracy of the sample assessment, and modifying the assumption on the population's length distribution had an impact on the accuracy of SS3. Varying the parameters in the mosquito model resulted in divergent mosquito behavior, such as modifying the mortality or host parameters to cause an inhibition in her attack behavior.

Pairing these tools with data will help expand on what the data tells us. For example, changes in rockfish catch trends can be examined within the simulator to determine the causes. For mosquitoes, observed behavior can be matched to decision tables to make inference into mosquito state, or how to effect mosquito state. By combining experimental and fieldwork with models, the downsides of each can be addressed. Experiments and fieldwork are performed to gather data, but restarting the work or obtaining more data can be incredibly costly and time-consuming. Models, without calibration with data, are far less informative. Some foundational set of knowledge is necessary to properly construct a model which accurately reflects the system it is modeled after. By combining empirical and modeling approaches, each becomes more powerful and useful.

The model can help to inform future experiments or data by determining which data are useful. With the simulator, the amount of data necessary for a reasonable level of accuracy for a stock assessment can be calculated. For example, finding the amount of length composition data needed to obtain sufficient accuracy for managers. For mosquitoes, determining parameters or conditions that have an effect on the system, such as mortality or blood meal accessibility being such strong drivers in a mosquito's likelihood of feeding on more hosts.

Data can also inform models. The rockfish life history simulator can be fed known parameters and data to generate simulated data for other models or analyses, or to add to existing data. The mosquito model can use mosquito data to generate additional decision matrices for further behavior inference.

These two models demonstrate the strengths that modeling brings to biological analyses. We can address the pitfalls of data poor systems by using novel, thoughtful approaches. The utility and power of data can be improved with additional use of modeling, since it allows us to examine problems from new angles.

Appendix A: Mortality Parameters

In the example assessment, I investigate the impact of assuming length-independence on natural mortality. I do so by calculating natural mortality parameter values that produce populations with complete length-independence and complete length-dependence, in addition to parameterizations between the two extremes. I begin by assuming a pair of values for length-independent (constant) natural mortality, $M^*(1)$, to be entered into the survival equation, $S(a_s)$, where a_s is the age of survival. Using $a_s = a_{max}$ and constant natural mortality, the survival to maximum age is calculated.

$$S(a_s) = \prod_{a=0}^{a_s} exp\Big(- M_0 - \frac{M_1}{L(a)}\Big) \tag{27}$$

In order to generate the parameter values for increasing length dependence, I assume the same survival to maximum age in all scenarios. Thus $S(a_{max})$ is fixed. M_0 is obtained by taking a percentage of $S(a_{max})$, where c_p is a fraction denoting the percentage of length-dependence of natural mortality in the scenario. For example, $c_p = 1$ for total length-independence, $M^*(1)$, and $c_p = 0.5$ for half length-dependence, $M^*(3)$.

$$M_0 = S(a_{max}) * c_p \tag{28}$$

I compute M_1 by entering the calculated M_0 value and the fixed $S(a_{max})$ value into Equation (27), and solve for M_1.

$$M_1 = \frac{log\Big(S(a_{max})\Big) - a_{max} \cdot M_0}{\sum_{a=0}^{a_{max}} \Big(\frac{1}{L(a)}\Big)} \tag{29}$$

Parameter values were obtained for the five pairs of values in M^*.

References

Benet, D. L., Dick, E. and Pearson, D. E. (2009), 'Life history aspects of greenspotted rockfish (*Sebastes chlorostictus*) from central california', *Southwest Fisheries Science Center* .

Beverton, R. J. H. and Holt, S. J. (1957), *On the Dynamics of Exploited Fish Populations*, Chapman & Hall, London, UK.

Boehlert, G. W. and Kappenman, R. F. (1980), 'Variation of growth with latitude in two species of rockfish (sebastes pinniger and s. diploproa) from the northeast Pacific ocean', *Marine Ecology - Progress Series* **3**, 1–10.

Briegel, H. (1990), 'Fecundity, metabolism and body size in *Anophleles* (diptera, culiciaade), vectors of malaria', *J. Med. Entomol.* **27**, 839–850.

Brooks, E. N., Powers, J. E. and Cortés, E. (2010), 'Analytical reference points for age-structured models: Application to data-poor fisheries', *ICES Journal of Marine Science* **67**(1), 165–175.

Chen, Y., Chen, L. and Stergious, K. I. (2003), 'Impacts of data quantity on fisheries stock assessment', *Aquatic Sciences* **65**(1), 92–98.

Clark, C. W. and Mangel, M. (2000), *Dynamic State Variable Models in Ecology: Methods and Applications*, Oxford University Press, Inc.

de Valpine, P. and Hilborn, R. (2005), 'State-space likelihoods for nonlinear fisheries time-series', *Can. J. Fish. Aquat. Sci.* **62**(9), 1937–1952.

Dick, E. (2010), Modeling the Reproductive Potential of Rockfishes (Sebastes spp.), PhD thesis, Department of Ocean Sciences, University of California, Santa Cruz.

Dick, E. J. (2011), personal communication.

Field, J. C. (2009), 'Status of the chilipepper rockfish, sebastes goodei, in 2007', *Southwest Fisheries Science Center* .

Field, J. C. and Ralston, S. (2005), 'Spatial variability in rockfish (sebastes spp.) recruitment events in the California current system', *Can. J. Fish. Aquat. Sci.* **62**, 2199–2210.

Foster, W. A. (1995), 'Mosquito sugar feeding and reproductive energetics', *Annu. Rev. Entomol.* **40**, 443–474.

Gulland, J., ed. (1988), *Fish Population Dynamics: The Implications for Management*, 2nd edn, John Wiley & Sons Ltd., London, UK.

Haldorson, L. and Love, M. (1990), 'Maturity and fecundity in the rockfishes, sebastes spp., a review', *Marine Fisheries Review* **53**(2), 25–31.

He, X., Ralston, S. and MacCall, A. D. (2011), 'Interactions of age-dependent mortality and selectivity functions in age-based stock assessment models', *Fishery Bulletin* **109**(2), 198–216.

Hilborn, R. and Mangel, M. (1997), *The Ecological Detective: Confronting Models with Data*, Princeton University Press, Princeton, New Jersey, USA.

Jennings, S., Kaiser, M. J. and Reynolds, J. D. (2001), *Marine Fisheries Ecology*, Blackwell Science, Oxford, London.

Leaman, B. M. (1991), 'Reproductive styles and life history variables relative to exploitation and management of *Sebastes* stocks', *Environmental Biology of Fishes* **30**, 253–271.

Lorenzen, K. (1996), 'The relationship between body weight and natural mortality in juvenile and adult fish: A comparison of natural ecosystems and aquaculture', *Journal of Fish Biology* **49**, 627–647.

Love, M. S., Carr, M. H. and Haldorson, L. J. (1991), 'The ecology of substrate-associated juveniles of the genus sebastes', *Environmental Biology of Fishes* **53**, 225–243.

Love, M. S., Yoklavich, M. and Thorsteinson, L. (2002), *The Rockfishes of the Northeast Pacific*, University of California Press, Berkeley and Los Angeles, California, USA.

Ludwig, D. and Walters, C. J. (1985), 'Are age-structured models appropriate for catch-effort data?', *Can. J. Fish. Aquat. Sci.* **42**(6), 1066–1072.

Ma, B. O. and Roitberg, B. D. (2008), 'The role of resource availability and state-dependence in the foraging strategy of blood-feeding mosquitoes', *Evolutionary Ecology Research* **10**, 1111–1130.

Macdonald, G. (1957), *The epidemiology and control of malaria*, Oxford Univ. Press.

Mangel, M., Brodziak, J. and DiNardo, G. (2010), 'Reproductive ecology and scientific inference of steepness: A fundamental metric of population dynamics and strategic fisheries management', *Fish and Fisheries* **11**(1), 89–104.

Mangel, M. and Clark, C. W. (1988), *Dynamic Modeling in Behavioral Ecology*, Princeton University Press.

McElhany, P., Steel, E. A., Avery, K., Yoder, N., Busack, C. and Thompson, B. (2010), 'Dealing with uncertainty in ecosystem models: Lessons from a complex salmon model', *Ecological Applications* **20**(2), 465–482.

Methot, R. D. (2005), *Technical description of the stock synthesis II assessment program*, NOAA Tech. Memo. SEDAR 16-AW-04.

Post, J. R. and Evans, D. O. (1989), 'Size-dependent overwinter mortality of young-of-the-year yellow perch (perca flavescens): Laboratory, in situ enclosure, and field experiments', *Can. J. Fish. Aquat. Sci.* **46**, 1958–1968.

Ralston, S. and MacFarlane, B. R. (2010), 'Population estimation of bocaccio (*Sebastes paucispinis*) based on larval production', *Canadian Journal of Fisheries and Aquatic Sciences* **67**, 1005–1020.

Roitberg, B. D., Keiser, S. and Hoffmeister, T. (2010), 'State-dependent attacks in a mosquito', *Physiological Entomology* **35**, 46–51.

Roitberg, B. D. and Mangel, M. (2009), 'Mosquito biting rate as an emergent community property and the implications for malarial interventions', *Israel Journal of Ecology and Evolution* (in press).

Roitberg, B. D., Mondor, E. B. and Tyerman, J. G. A. (2003), 'Pouncing spider, flying mosquito: blood acquisition increases predation risk in mosquitoes', *Behav. Ecol.* **14**, 736–740.

Rose, K. A. and Cowan, Jr., J. H. (2003), 'Data, models, and decisions in U.S. marine fisheries management: Lessons for ecologists', *Annu. Rev. Ecol. Evol. Syst.* **34**, 127–151.

Schaefer, M. B. (1954), 'Some aspects of the dynamics of populations, important for the management of the commercial marine fisheries', *Inter-American Tropical Tuna Commission* **1**(2), 27–56.

Spielman, A. and D'Antonio, M. (2001), *Mosquito: A natural history of our most persistent and deadly foe*, Hyperion, New York.

Stone, C., Hamilton, I. and Foster, W. (2011), 'A survival and reproduction trade-off is resolved in accordance with resource availability by virgin female mosquitoes', *Animal Behaviour* **81**, 765–774.

Tsunoda, T., Fukuchi, A., Nanbara, S. and Takagi, M. (2010), 'Effect of body size and sugar meals on oviposition of the yellow fever mosquito, aedes aegypti (diptera: Culicidae)', *Journal of Vector Ecology* **35**(1), 56–60.

van Handel, E. (1965), 'The obese mosquito', *Journal of Physiology* **181**, 478–486.

von Bertalanffy, L. (1957), 'Quantitative laws in metabolism and growth', *Quarterly Review of Biology* **32**(3), 217–231.

Wang, S.-P., Maunder, M. N. and Aires-da-Silva, A. (2009), 'Implications of model and data assumptions: An illustration including data for the Taiwanese longline fishery into the eastern Pacific Ocean bigeye tuna (*Thunnus obesus*) stock assessment', *Fisheries Research* **97**(1-2), 118–126.

Wetzel, C. R. and Punt, A. E. (2011), 'Model performance for the determination of appropriate harvest levels in the case of data-poor stocks', *Fisheries Research* **110**(2), 342–355.

World Health Organization (2010), 'World malaria report'. http://www.who.int/malaria/world_malaria_report_2010/en/index.html.

Tables & Figures

Table 1: Rockfish Parameters

Variables	Interpretation
$N(a, t, l, r)$	Number of individuals in the population at age a, time t, length l, region r
$Y(a, t, l, r)$	Yield from individuals at age a, time t, length l, region r
$E_T(t, r)$	Total eggs at time t in region r
ψ_{ij}	Proportion of larvae from region i that settle as recruits in region j
γ_{ij}	Proportion of migrants from region i to region j
$\alpha(r), \beta(r)$	Beverton-Holt Recruitment parameters in region r
$L_\infty(r), k(r), \tau_0(r)$	Von Bertalanffy Growth parameters in region r
$W(l, r)$	Length-Weight Allometry for fish of length l in region r
c_1, c_2	Weight allometry parameters
d_1, d_2	Fecundity parameters
ω	Conversion parameter for translating units from grams to tons
$M_0(r), M_1(r)$	Natural mortality parameters in region r
$\mathcal{F}(l, r)$	Size selective fishing mortality for fish of length l in region r
$\mathcal{F}_{max}(r), \mathcal{F}_s(r), \mathcal{F}_{50}(r)$	Fishing mortality parameters in region r
$p_m(l, r)$	Maturity in region r
$p_s(r), p_{50}(r)$	Maturity parameters in region r
a, a_{max}	age, and maximum age
t, t_{max}	time, and maximum time
l, l_{max}	length, and maximum length
r, r_{max}	region, and maximum number of regions

Table 2: Rockfish Model Assumptions

Simulator	Stock Assessment
Age Structured Model	Production Model
Beverton-Holt recruitment	Logistic production
Size-dependent natural mortality	Constant natural mortality
Size-dependent fishing mortality	Constant fishing mortality
Biomass tracked by age, length, region, and time	Biomass tracked by time

Table 3: Bias in Abundance Estimates from Data Variability

Catch Time Series	$\bar{\phi}$
$C_{1,20}(t)$	0.146
$C_{1,15}(t)$	0.299
$C_{6,20}(t)$	0.416
$C_{1,10}(t)$	0.528
$C_{6,15}(t)$	0.529
$C_{11,20}(t)$	0.550

Table 4: Scenarios Used In Full Assessment

Reg. Length Distribution	10 years with 50 length samples	20 years with 200 length samples
1 parameter estimated	$S_{1,p}$	$S_{1,r}$
3 parameters estimated	$S_{3,p}$	$S_{3,r}$

Non-Reg. Length Distribution	10 years with 50 length samples	20 years with 200 length samples
1 parameter estimated	$S^N_{1,p}$	$S^N_{1,r}$
3 parameters estimated	$S^N_{3,p}$	$S^N_{3,r}$

Table 5: Mosquito Parameters

Variables	Interpretation	Values/Functions
$R(b, e, a, t)$	Reproductive fitness	-
b	Blood reserves	$b_{max} = 100$
e	Energy reserves	$e_{max} = 100$
a	Host awareness	$a_{max} = 2$
t	Time in current night	$T = 100$
$D\big(V_s, V_f, V_o\big)$	Optimal decision	-
$V_s(b, e, a, t)$	Seek new host	-
$V_f(b, e, a, t)$	Feed on current host	-
$V_o(b, e, a, t)$	Oviposit	-
$L(V_s, V_o)$	Optimal decision when leaving current host	-
$r_b(b, e)$	Reproductive fitness benefit from laying eggs	$\frac{b}{0.01} \cdot I_{\{e \geq 10, b \geq 10\}}$
P_a	Prob. of host becoming aware	0.4
P_d	Prob. of host attacking mosquito after becoming aware	0.5
$g(b)$	Blood reserve increment	$1 + \frac{3}{1+0.1 \cdot b}$
$\mu_1(b)$	Mortality from host attack if mosquito stays	$\frac{0.15}{100} \cdot b + 0.2$
$\mu_2(b)$	Mortality from host attack if mosquito flees	$\frac{0.05}{100} \cdot b + 0.1$
t_s	Time spent finding new host	12
t_a	Time spent leaving host	2
μ	Background mortality rate at t time scale	0.005
M	Background mortality rate at daily time scale	0.15
e_f	Energy spent feeding per t time	1
e_h	Energy spent flying to or from host	10
e_d	Energy spent fleeing host attack	2

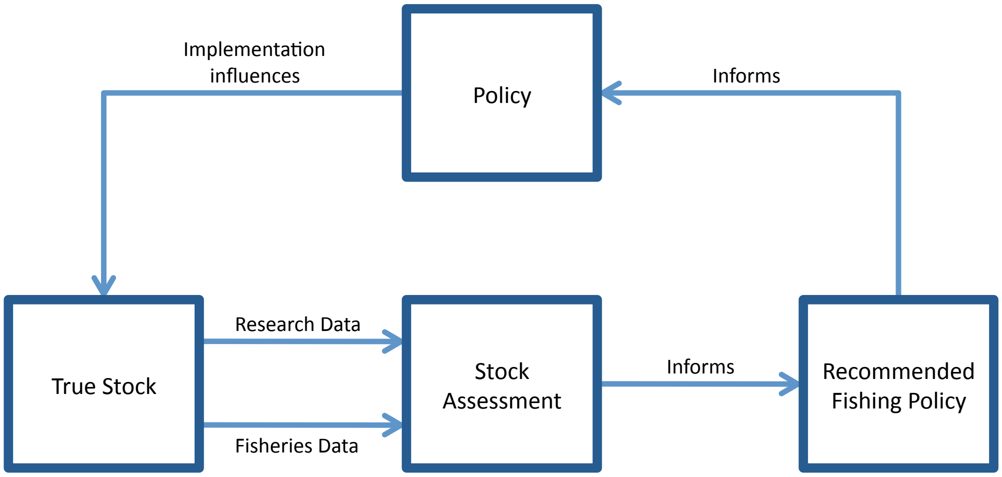


Figure 1: Both research (fisheries independent) and fisheries data may be collected for a stock and all or a subset of these data are input into a stock assessment. The results of the stock assessment can include an estimate of the abundance of the stock and are used to determine fishing regulations, which are recommended to committees, regulatory agencies, and policy makers. A policy is enacted, which in turn can influence the dynamics and status of the stock.

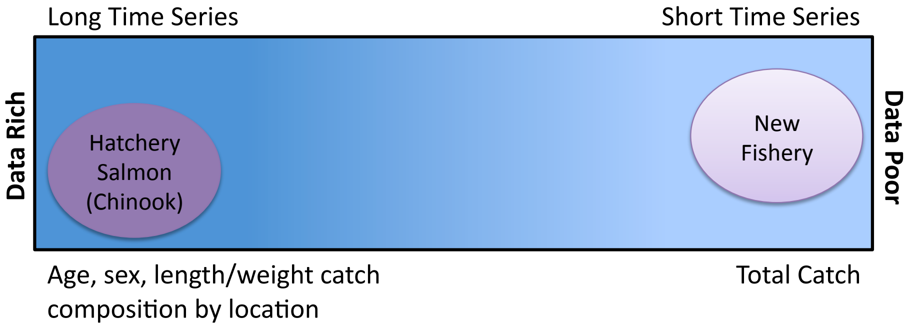
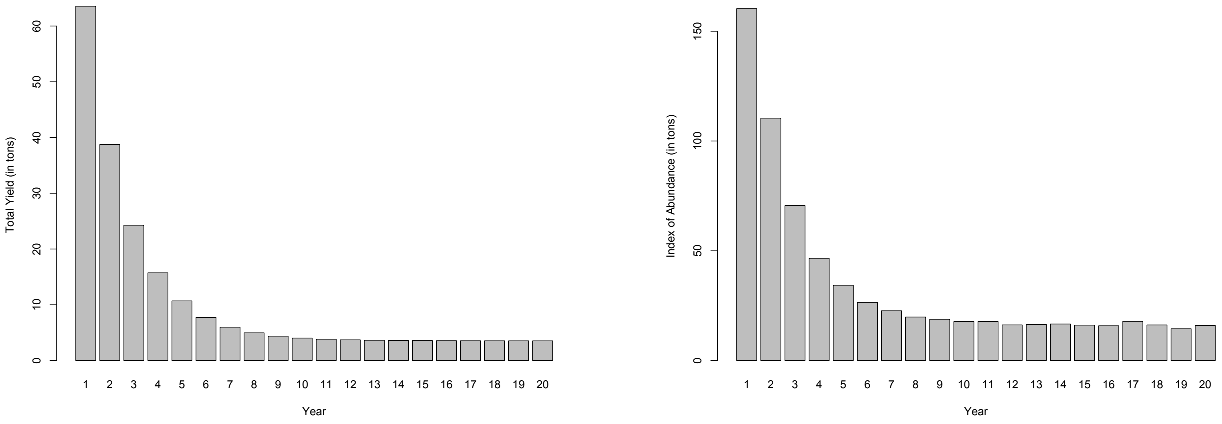
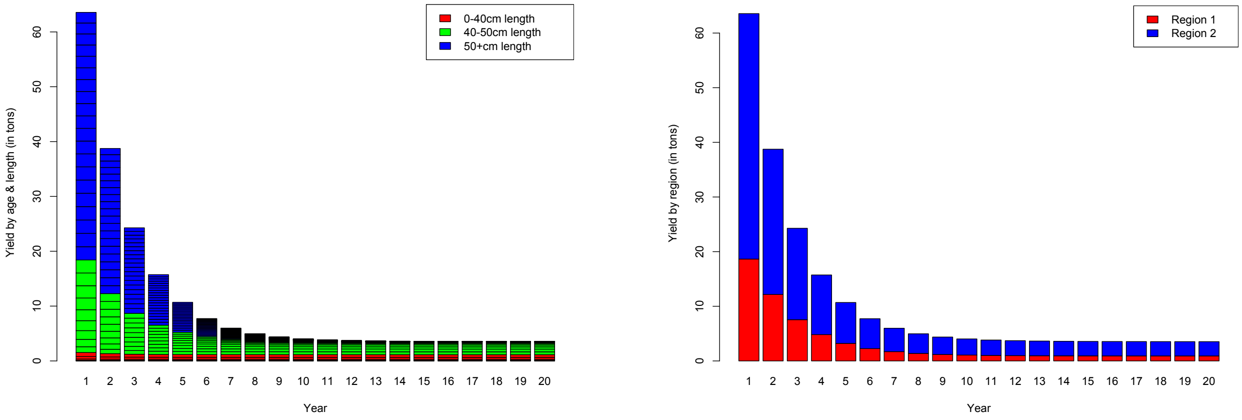


Figure 2: Data poor stocks are typically characterized by the simplest records of catch data, such as total catch in numbers or biomass over a short time series. A new fishery falls into the data poor end of the spectrum since it will necessarily have a short time series, and typically no catch composition. Alternatively, data rich stocks are often described by long time series and diverse catch compositions such as catch by age, sex, length or weight, and location. Additionally, data rich stocks can contain information pertaining to movement, mortality and growth rate, effort, and other life history parameters. Hatchery salmon, such as chinook, are typically considered data rich since recovered coded-wire-tags implanted in a proportion of these fish make it possible to reconstruct cohorts, determine distribution patterns, fishery impacts, and survival rates. Intermediate data level stocks may be described by any subset of the data present for data rich stocks.



(a) Total Yield

(b) Index of Abundance



(c) Composition of yield by age and length

(d) Composition of yield by region caught

Figure 3: Different levels of catch data generated from the life history simulator: (a) Total yield from catch per year. (b) Index of abundance per year. (c) Yield by age and length, per year. Each box represents an age class, and each color represents a length class. (d) Yield by region. Each color represents a catch region.

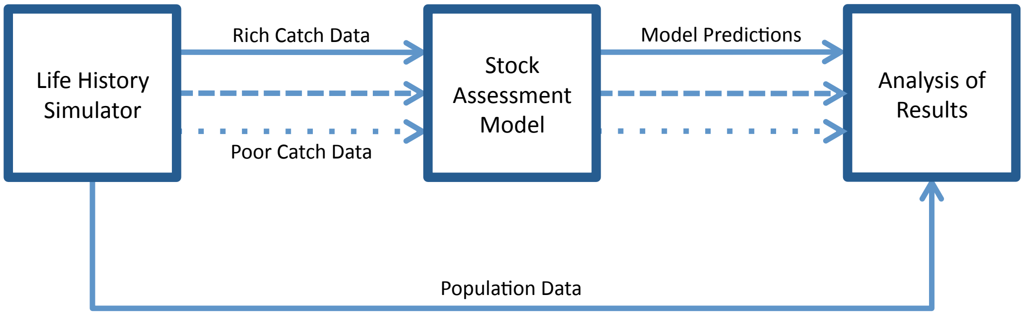


Figure 4: The life history simulator generates different levels of catch data. These are input into the stock assessment model to make predictions about the population. The results of the stock assessment for each level of data richness are compared to the complete population data from the life history simulator to determine model accuracy across increasing levels of available data.

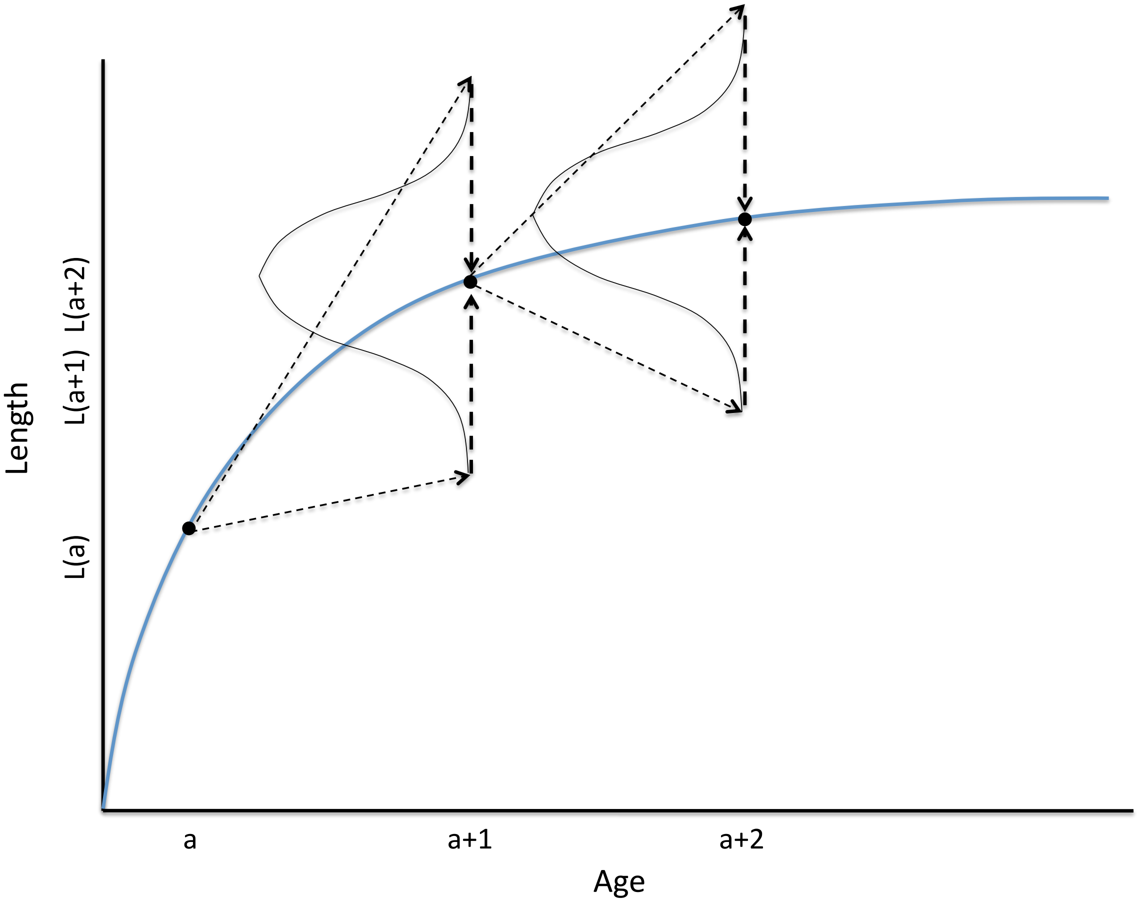


Figure 5: For rockfish, each length at age has a Gaussian distribution centered at the von Bertalanffy length for that age. The following year the individuals are combined and redistributed according to a new Normal distribution with a mean at the new von Bertalanffy length for the updated age.

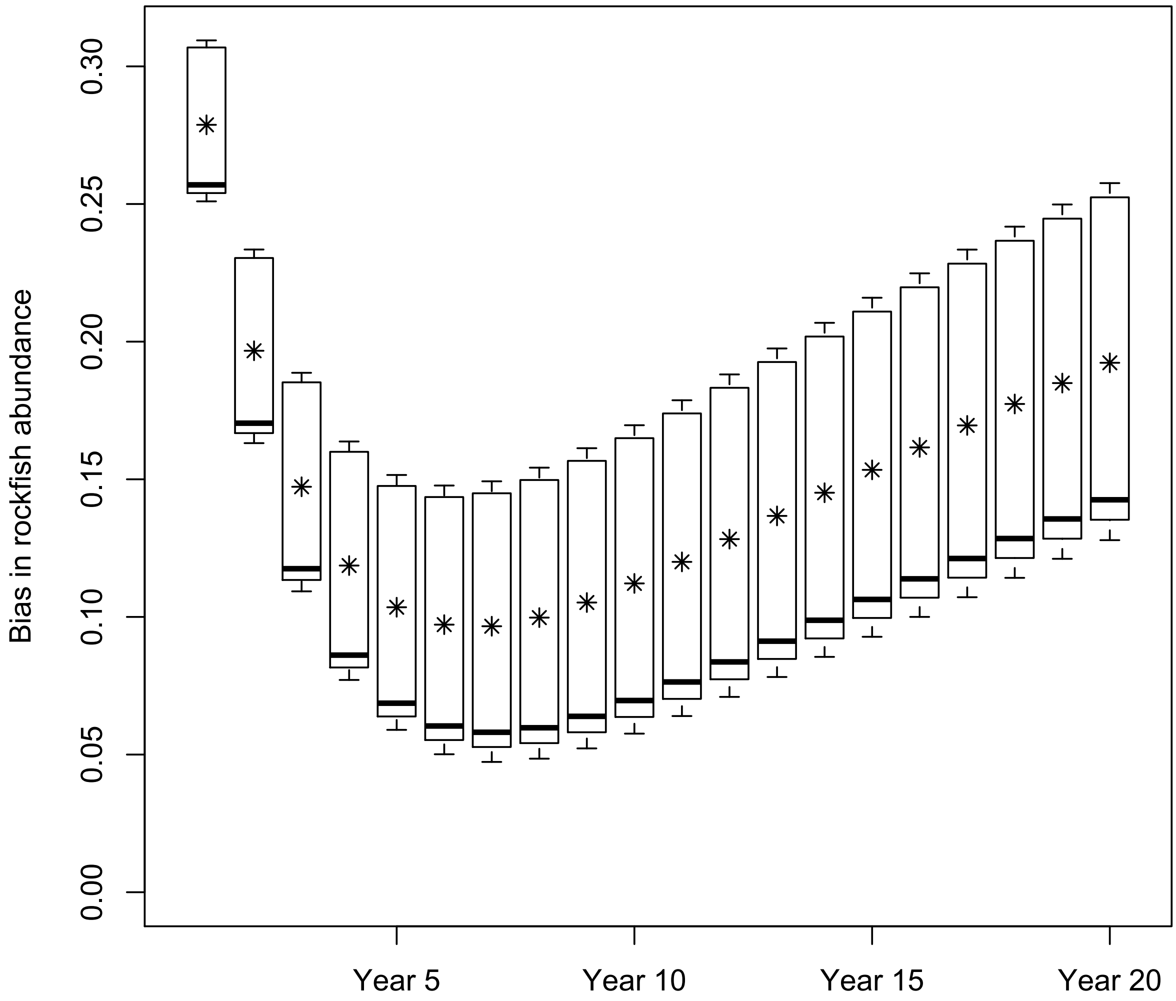


Figure 6: Bias in rockfish biomass under the $M^*(1)$ natural mortality scenario for each year, where the simulator and assessment model both assume constant (size-independent) natural mortality. The stars are the means, the thick bands in the box plot are the medians, the boxes are the interquartile range, and the bands are the range. Positive bias values indicate overestimation in abundance and negative values show underestimation.

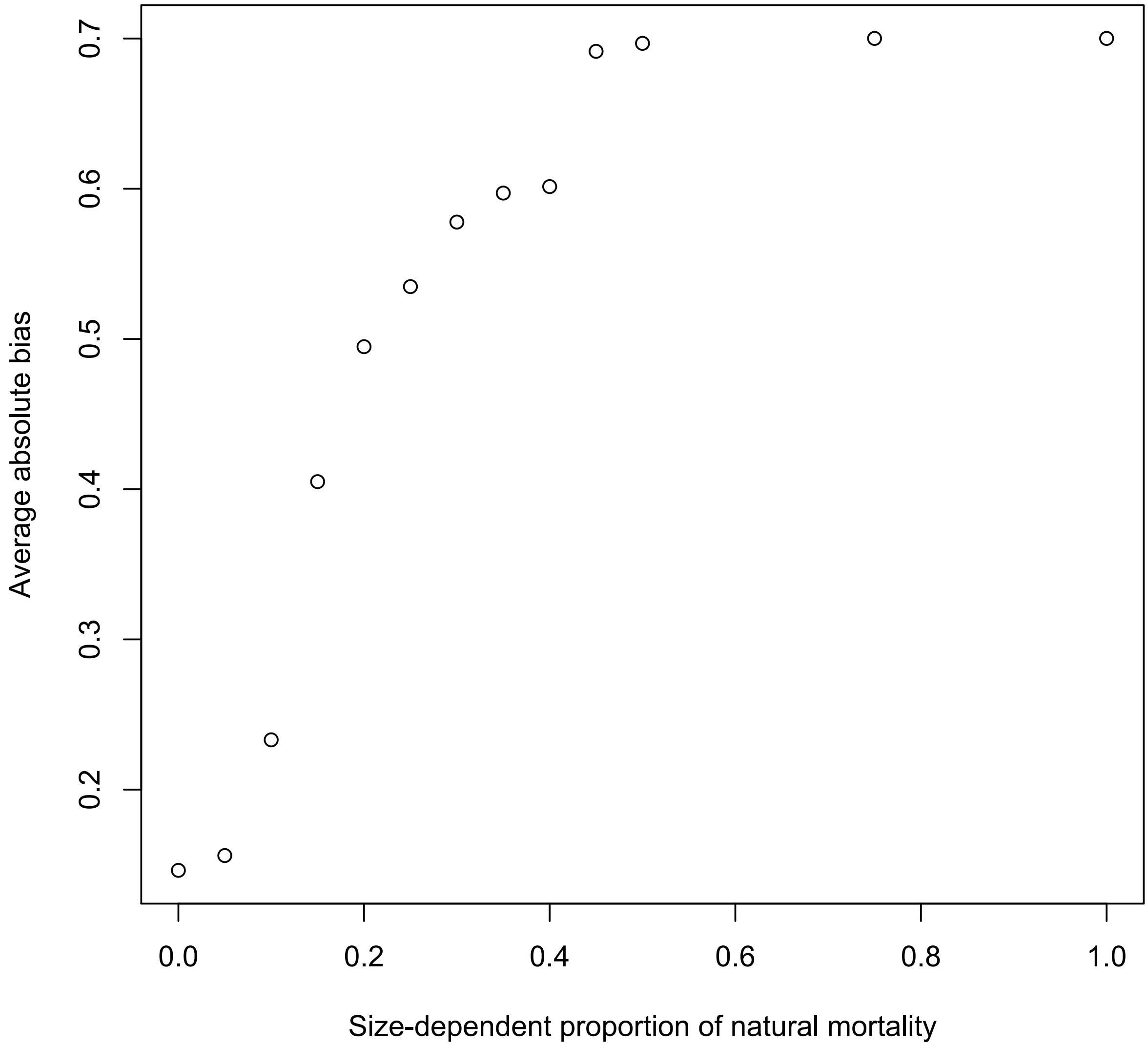


Figure 7: Absolute average bias, $\bar{\phi}$, over the initial natural mortality scenarios $M^{*}(1), M^{*}(2), M^{*}(3), M^{*}(4), M^{*}(5)$, and the additional runs performed to examine the bias at low size-dependence.

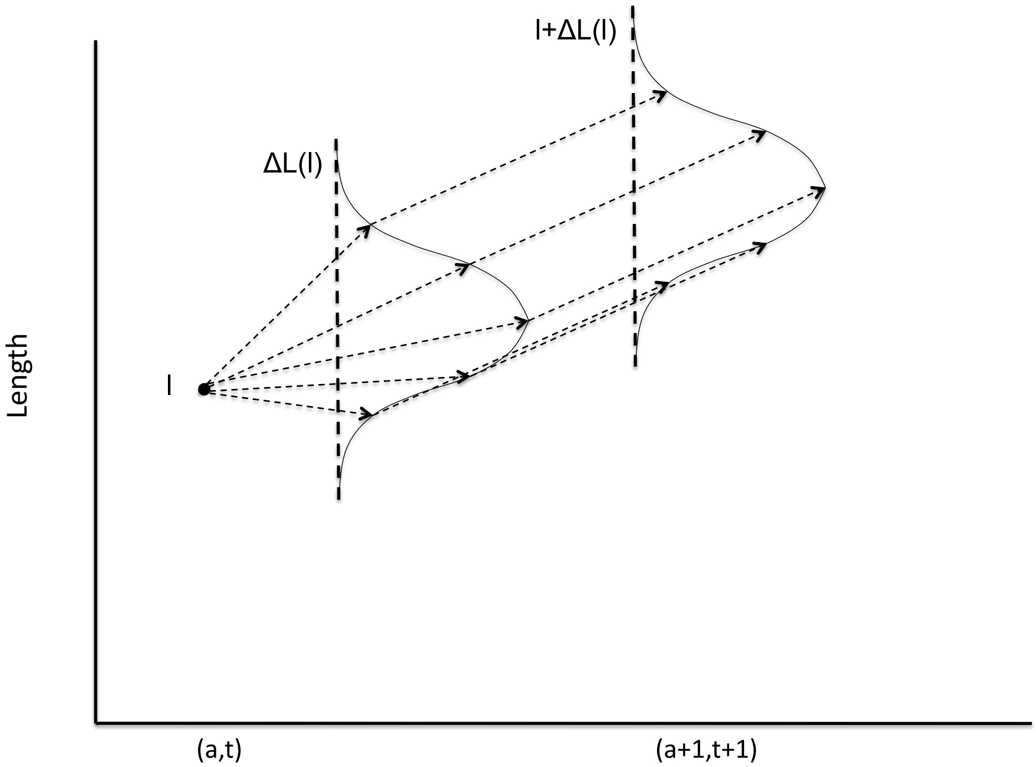


Figure 8: Fish of length l grow at variable rates. Their change in size is calculated from the distribution for the increment to growth, $\Delta L(l)$, and added to their current length to determine their size for the following year.

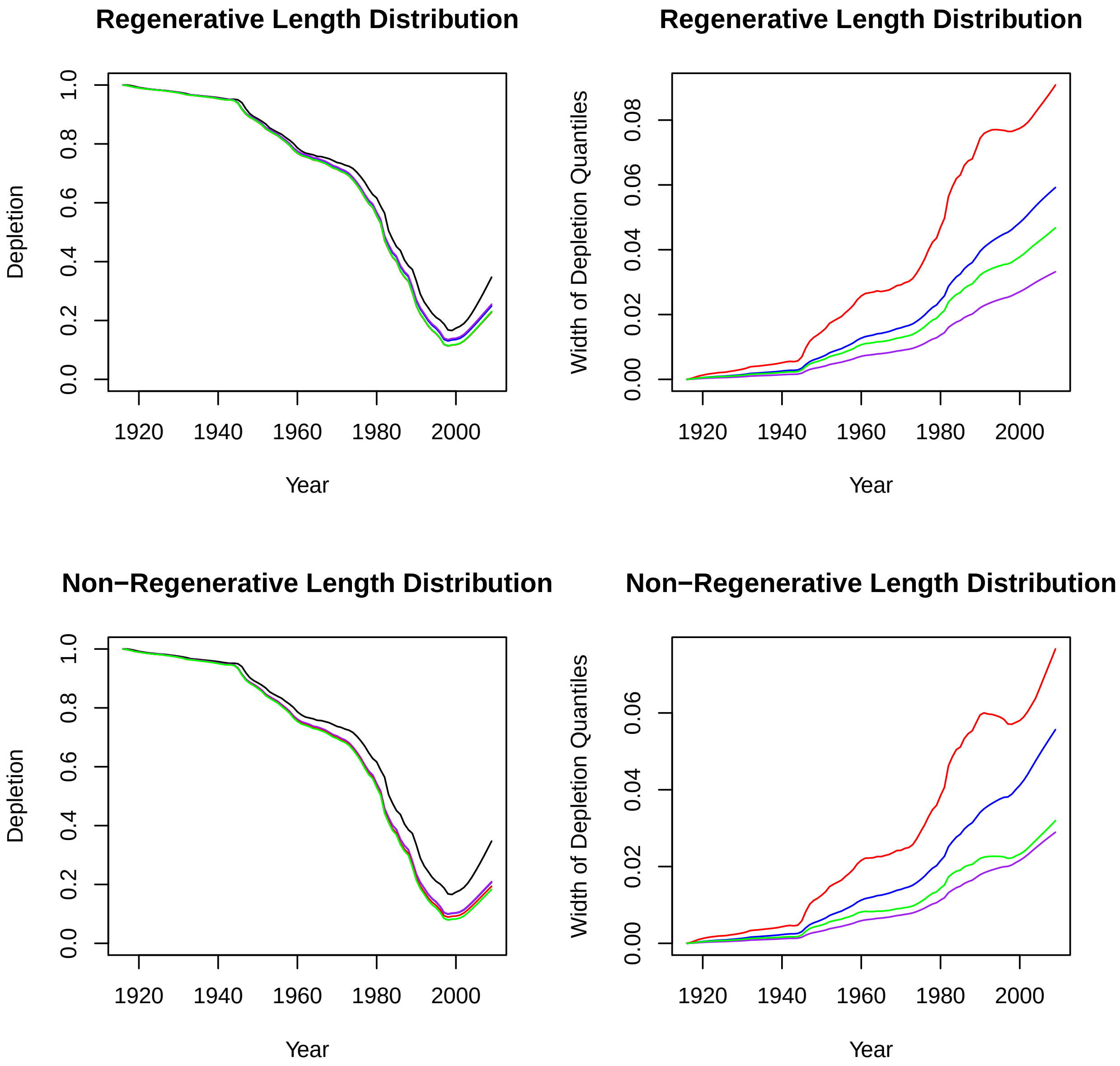


Figure 9: Time series of mean depletion estimates (left figures) and the width of the quantiles (right figures), where the black line is the simulated depletion, the purple line is data rich with 1 parameter estimated, the green line is data rich with 3 parameters estimated, the blue line is data poor with 1 parameter estimated, and the red line is data poor with 3 parameters estimated.

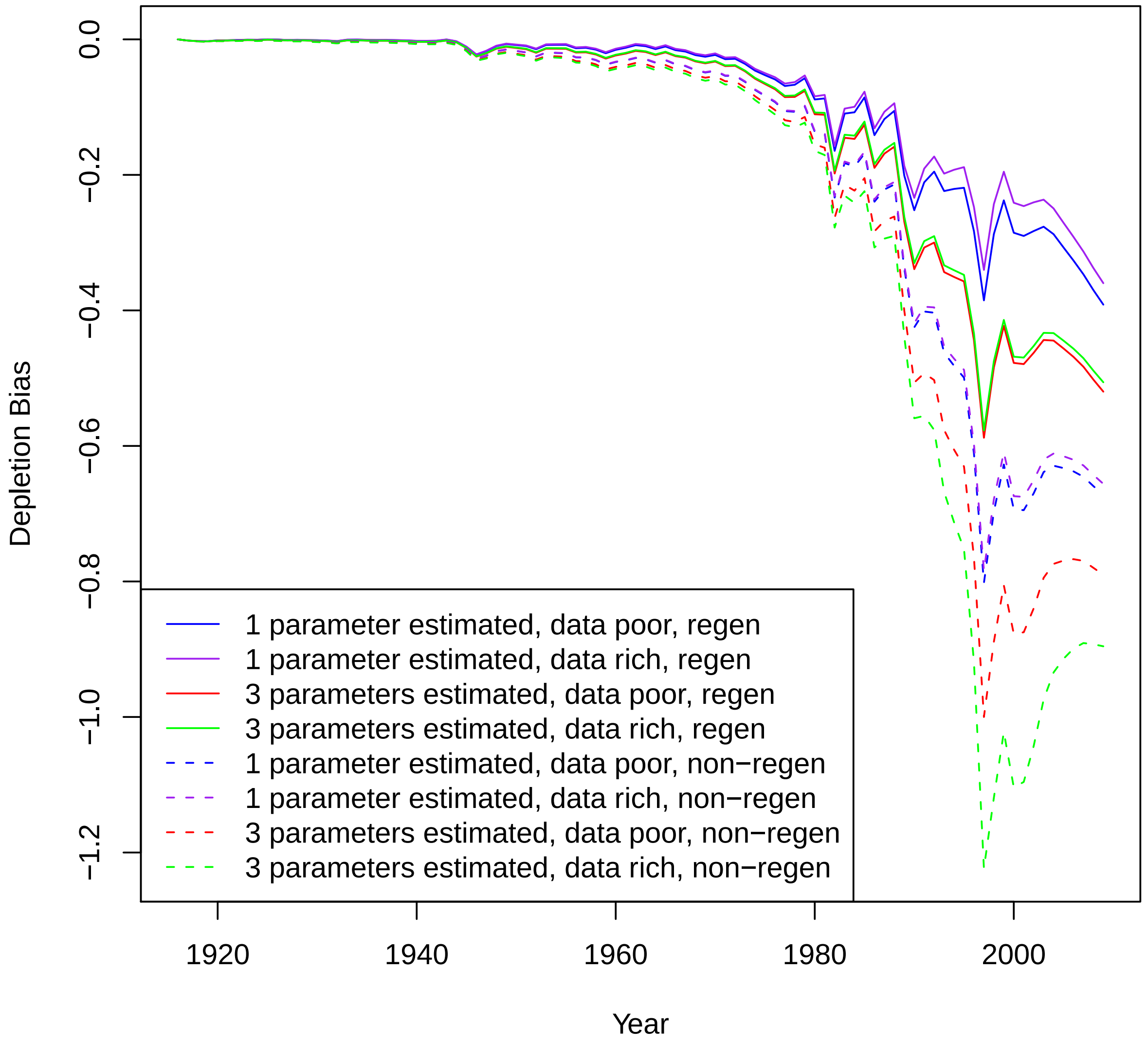


Figure 10: The time series of depletion bias where the purple lines are data rich with 1 parameter estimated, green lines are data rich with 3 parameters estimated, blue lines are data poor with 1 parameter estimated, and red lines are data poor with 3 parameters estimated. Solid lines are regenerative length and non-regenerative length distribution assumptions respectively.

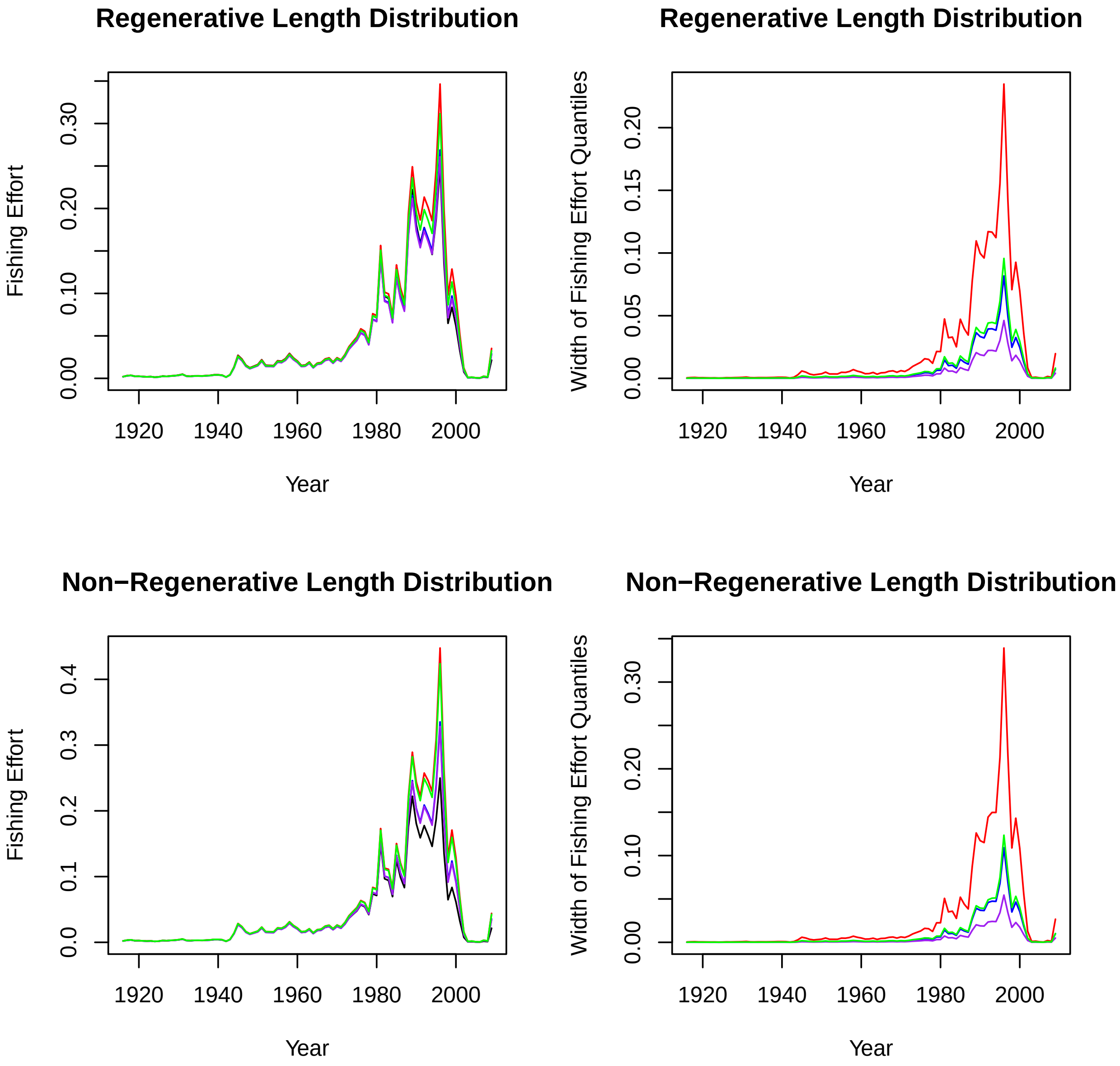


Figure 11: Time series of mean fishing effort estimates (left figures) and the width of the quantiles (right figures), where the black line is the simulated depletion, the purple line is data rich with 1 parameter estimated, the green line is data rich with 3 parameters estimated, the blue line is data poor with 1 parameter estimated, and the red line is data poor with 3 parameters estimated.

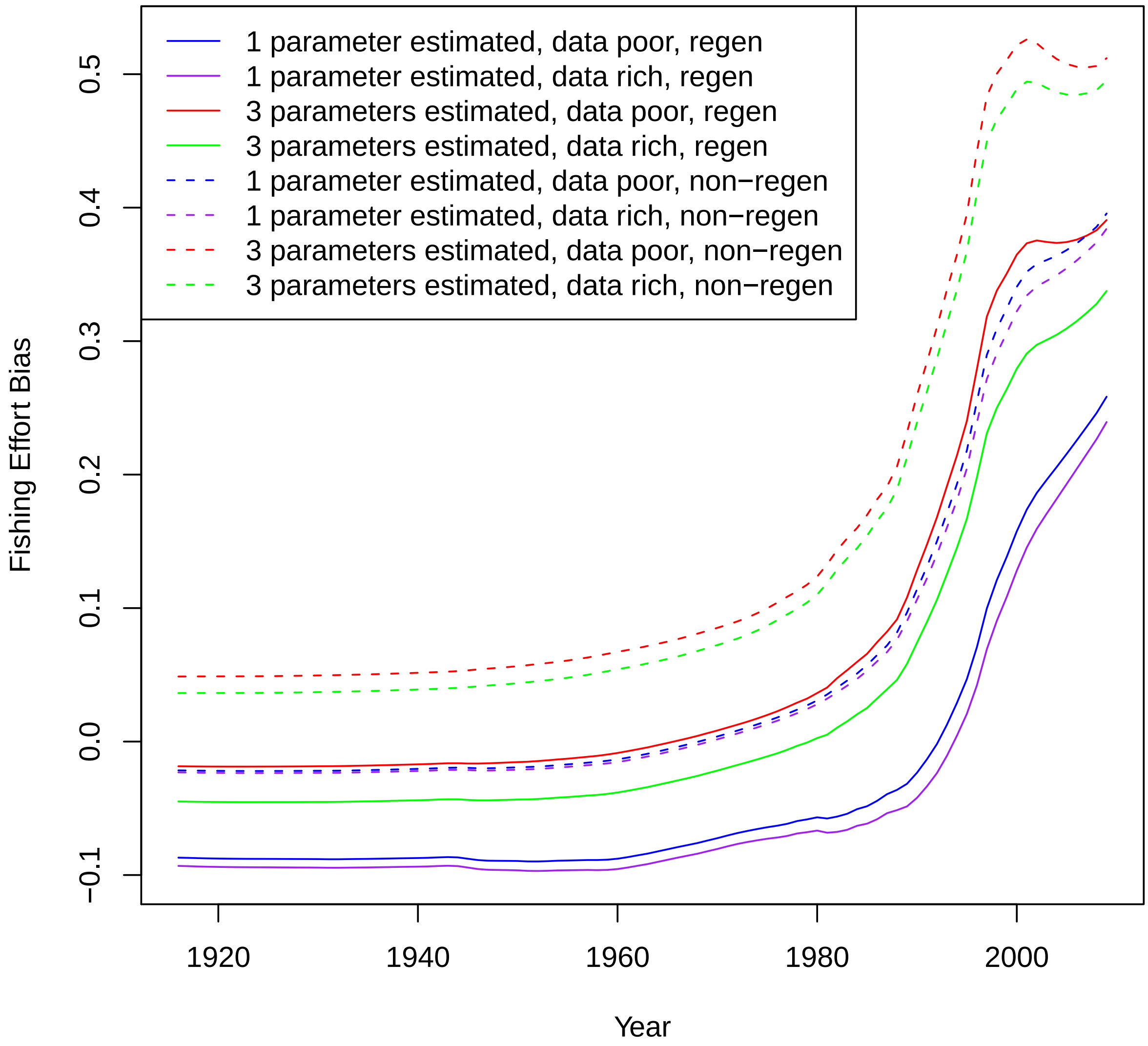


Figure 12: The time series of fishing effort bias where the purple lines are data rich with 1 parameter estimated, green lines are data rich with 3 parameters estimated, blue lines are data poor with 1 parameter estimated, and red lines are data poor with 3 parameters estimated. Solid lines are regenerative length and non-regenerative length distribution assumptions respectively.

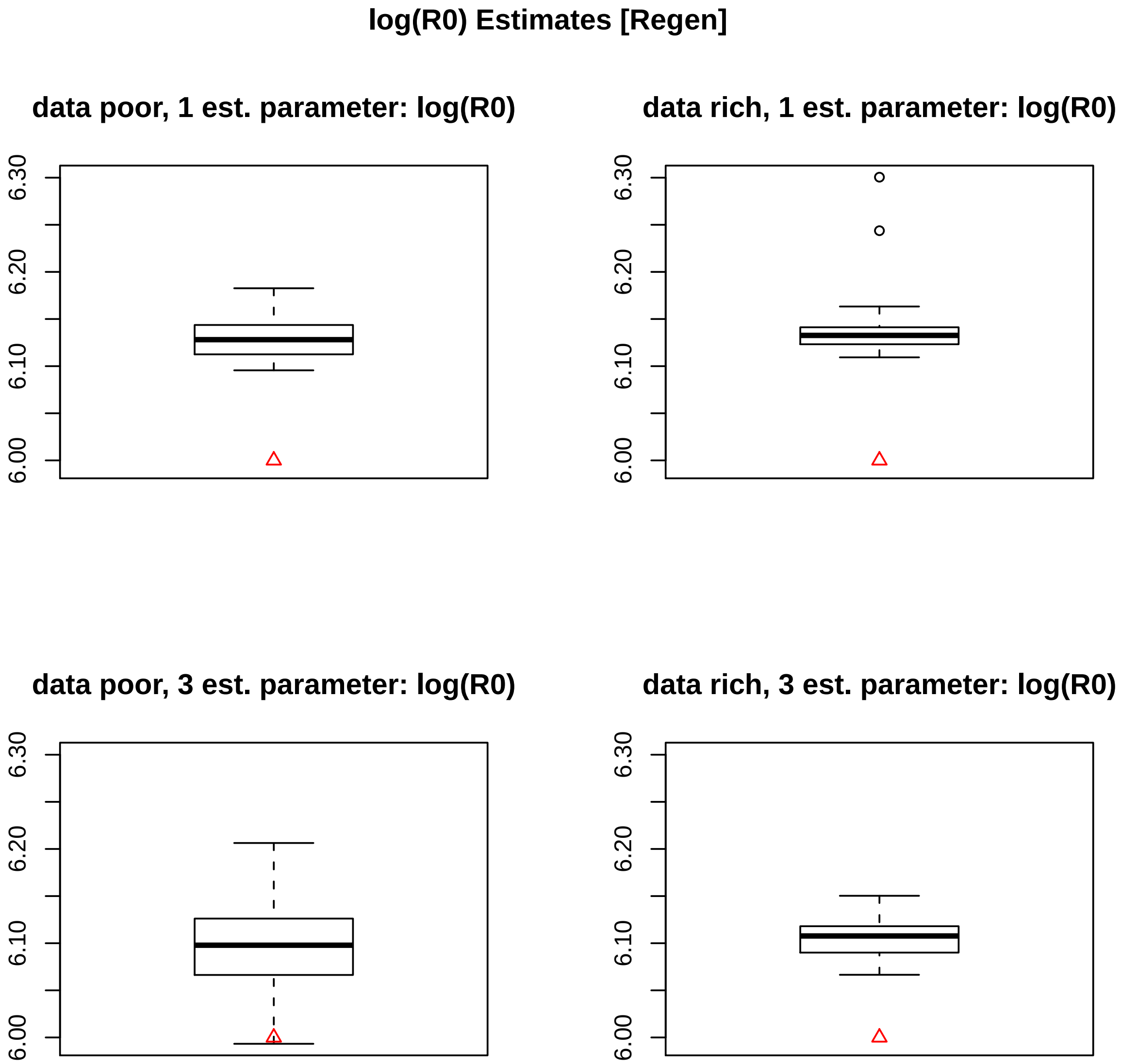


Figure 13: Box plots of the parameter estimate for $log(R_0)$, where the red triangle is the true value, under the regenerative length distribution assumption.

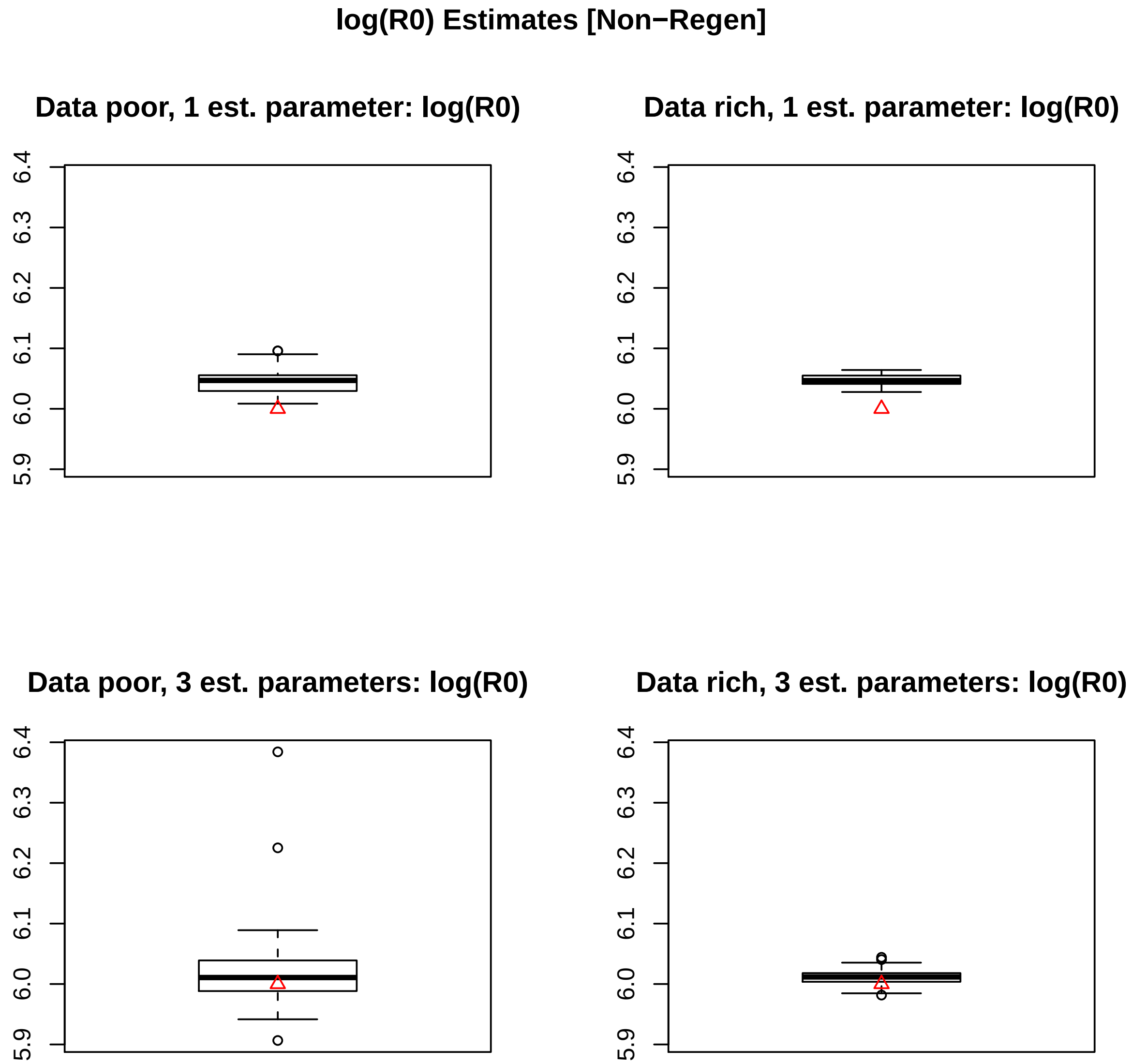


Figure 14: Box plots of the parameter estimate for $log(R_0)$, where the red triangle is the true value, under the non-regenerative length distribution assumption.

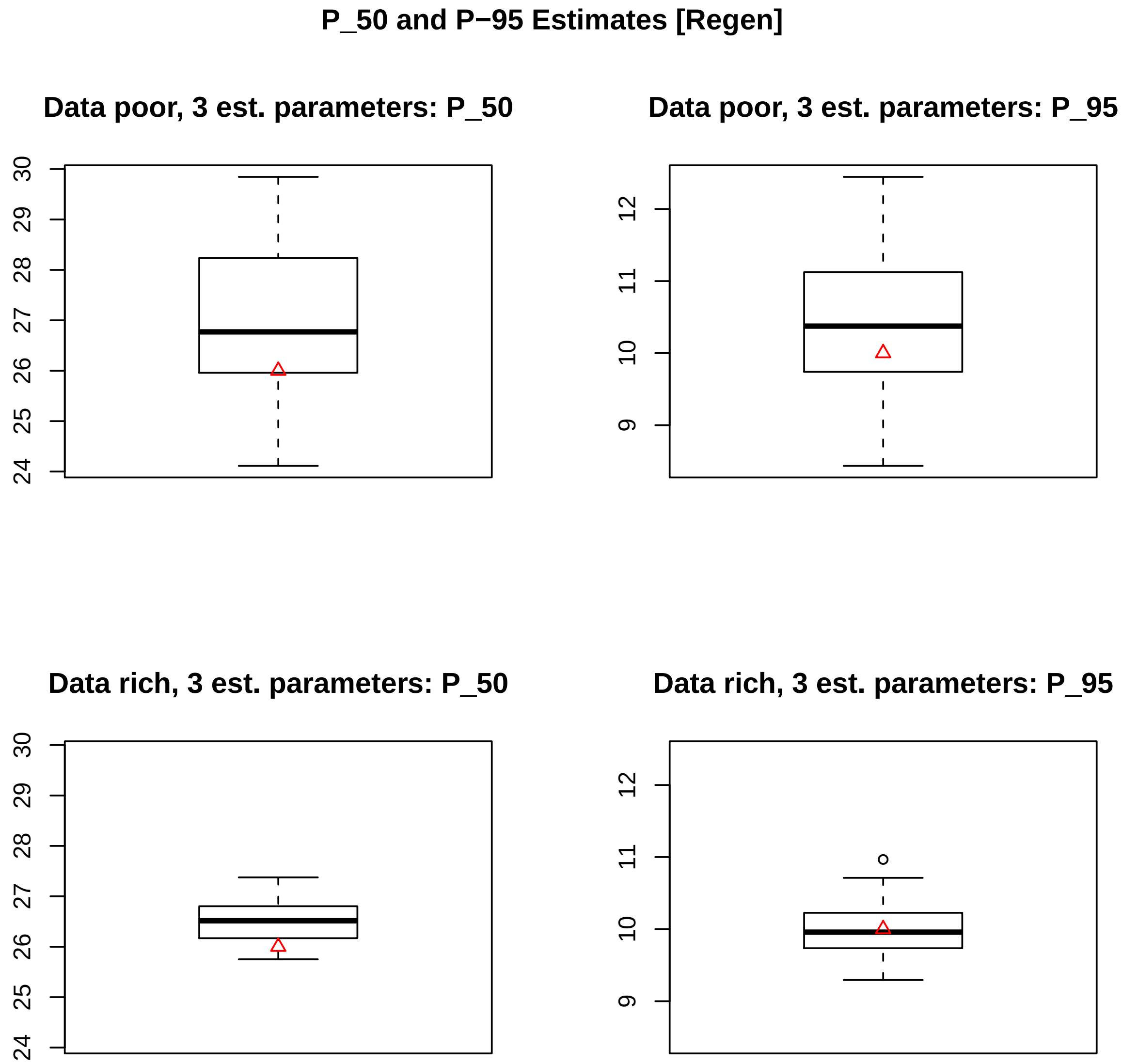


Figure 15: Box plots of the estimates for the selectivity parameters, where the red triangle is the true value, under the regenerative length distribution assumption.

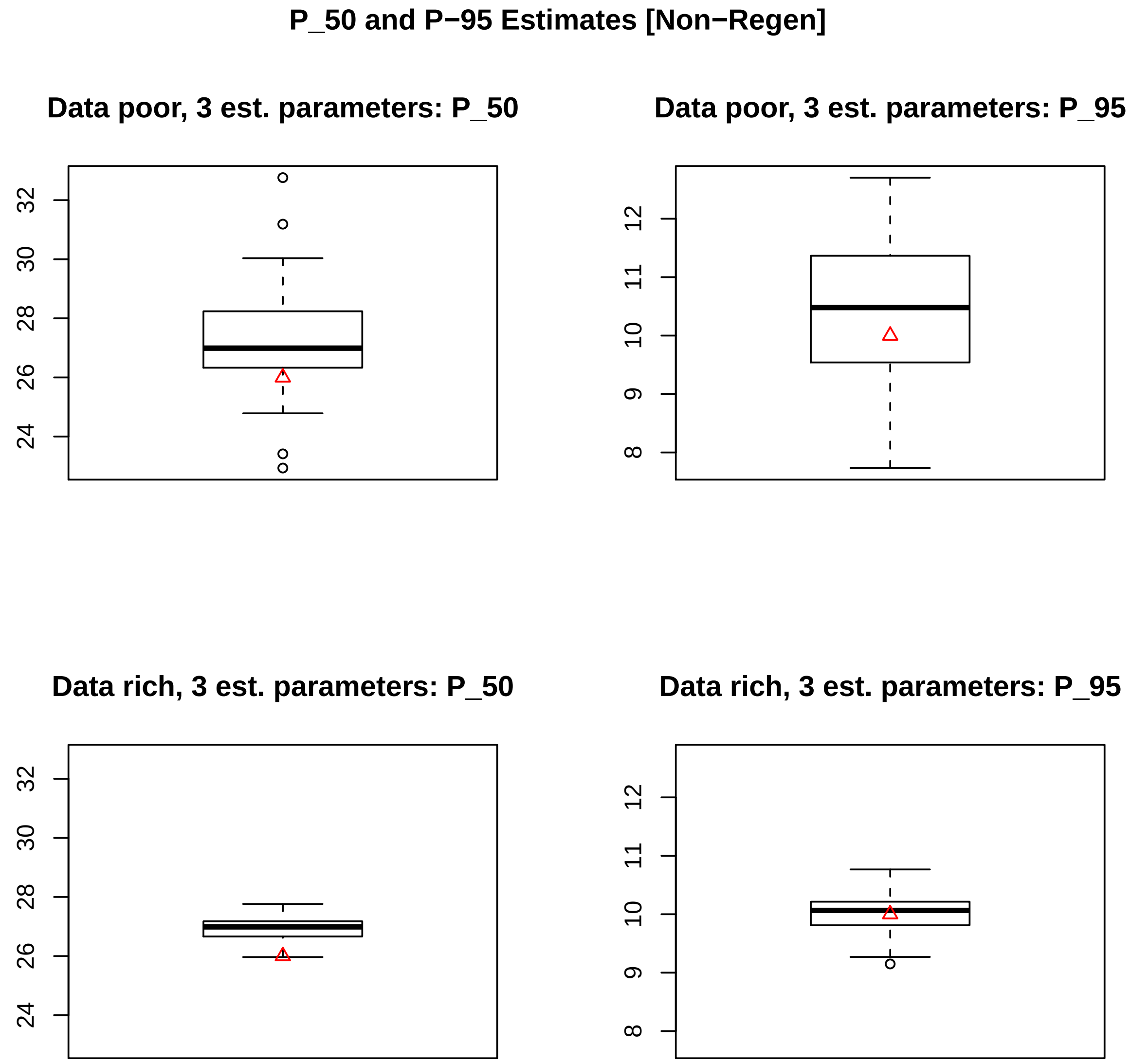


Figure 16: Box plots of the estimates for the selectivity parameters, where the red triangle is the true value, under the non-regenerative length distribution assumption.

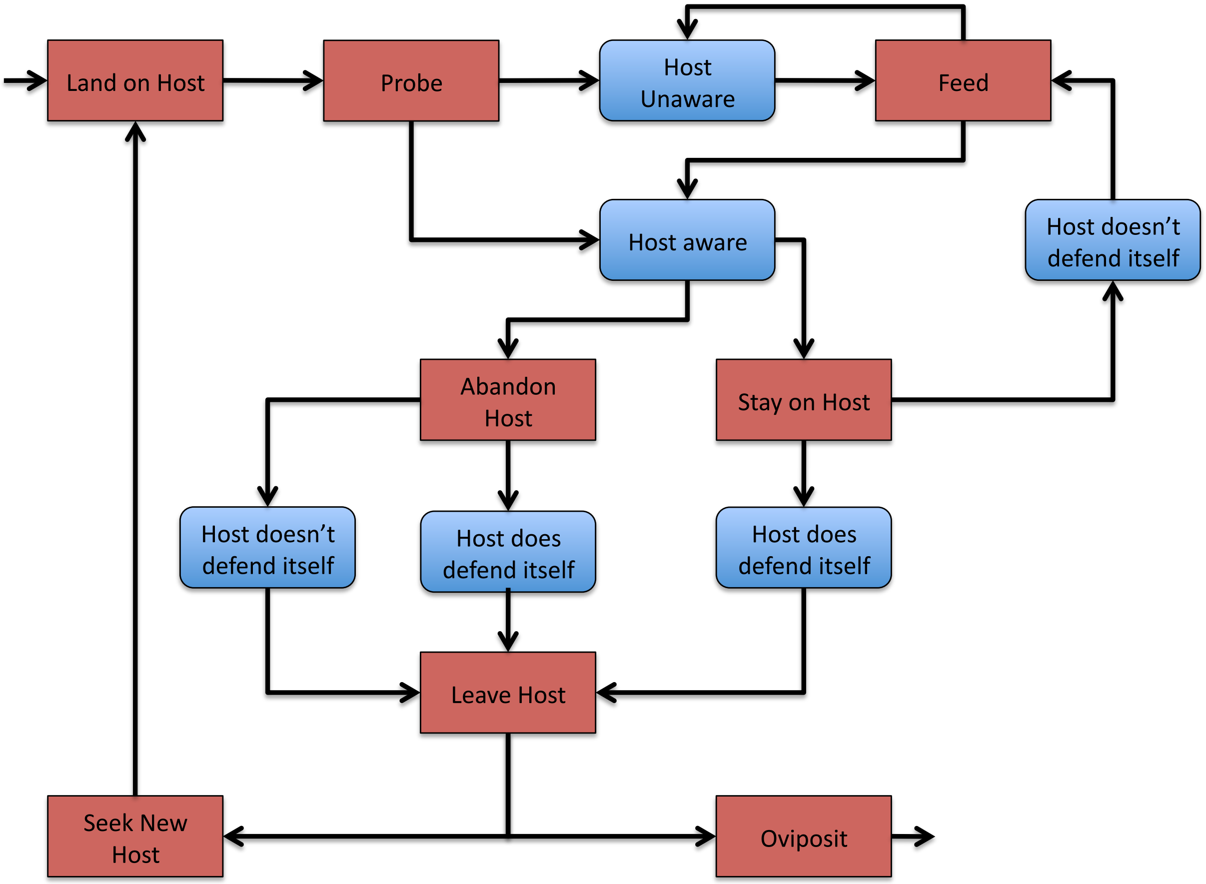


Figure 17: Modeled mosquito behavior process

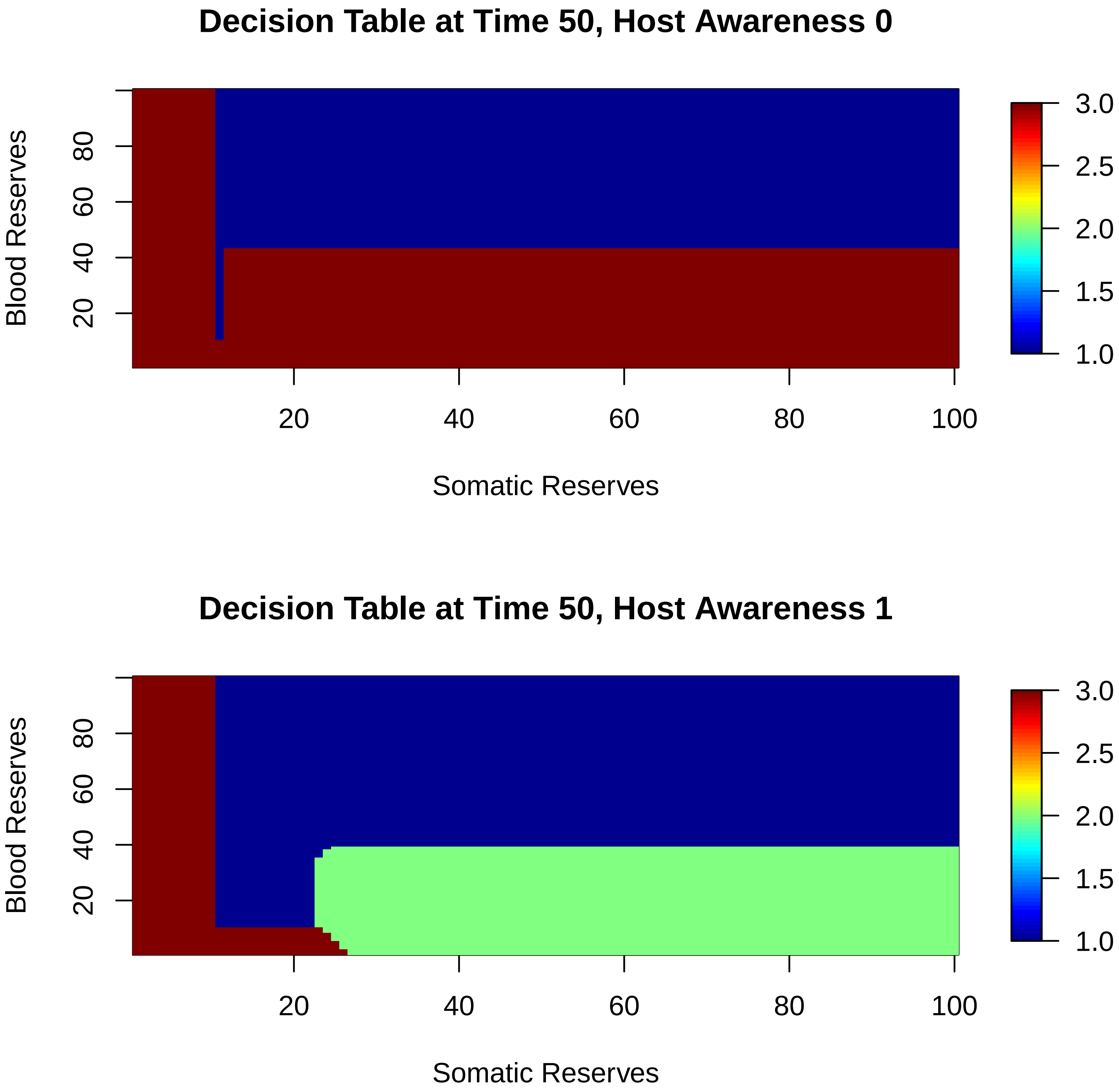


Figure 18: Decision rules for current blood and somatic reserve values, with both states of awareness, at time 50 (though the figure is generally equivalent to the decision tables between time 0 and time 84). Red (3) signifies the decision to continue feeding on the current host, teal (2) signifies the decision to seek a new host, and blue (1) is when the mosquito decides to leave to oviposit.

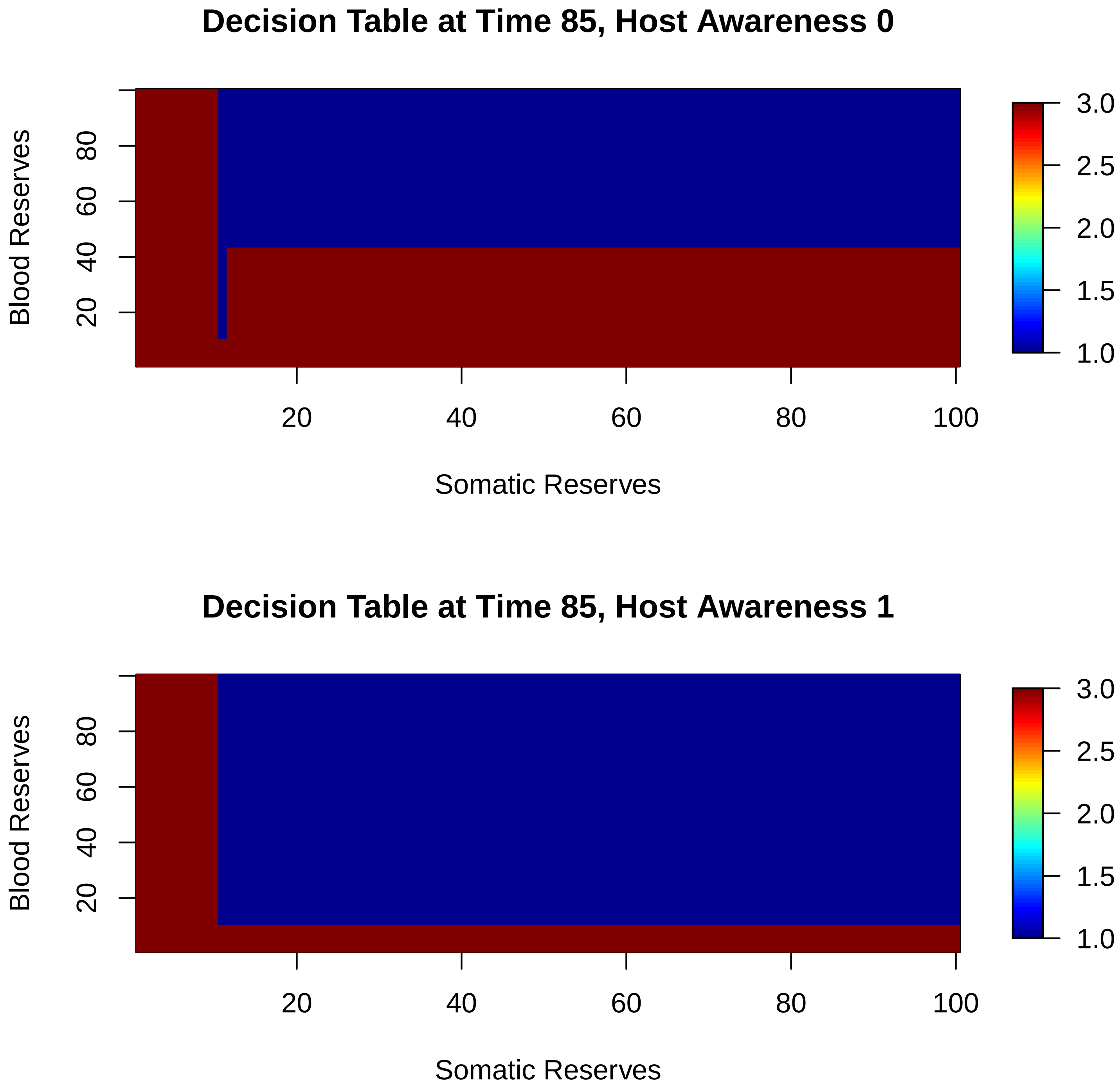
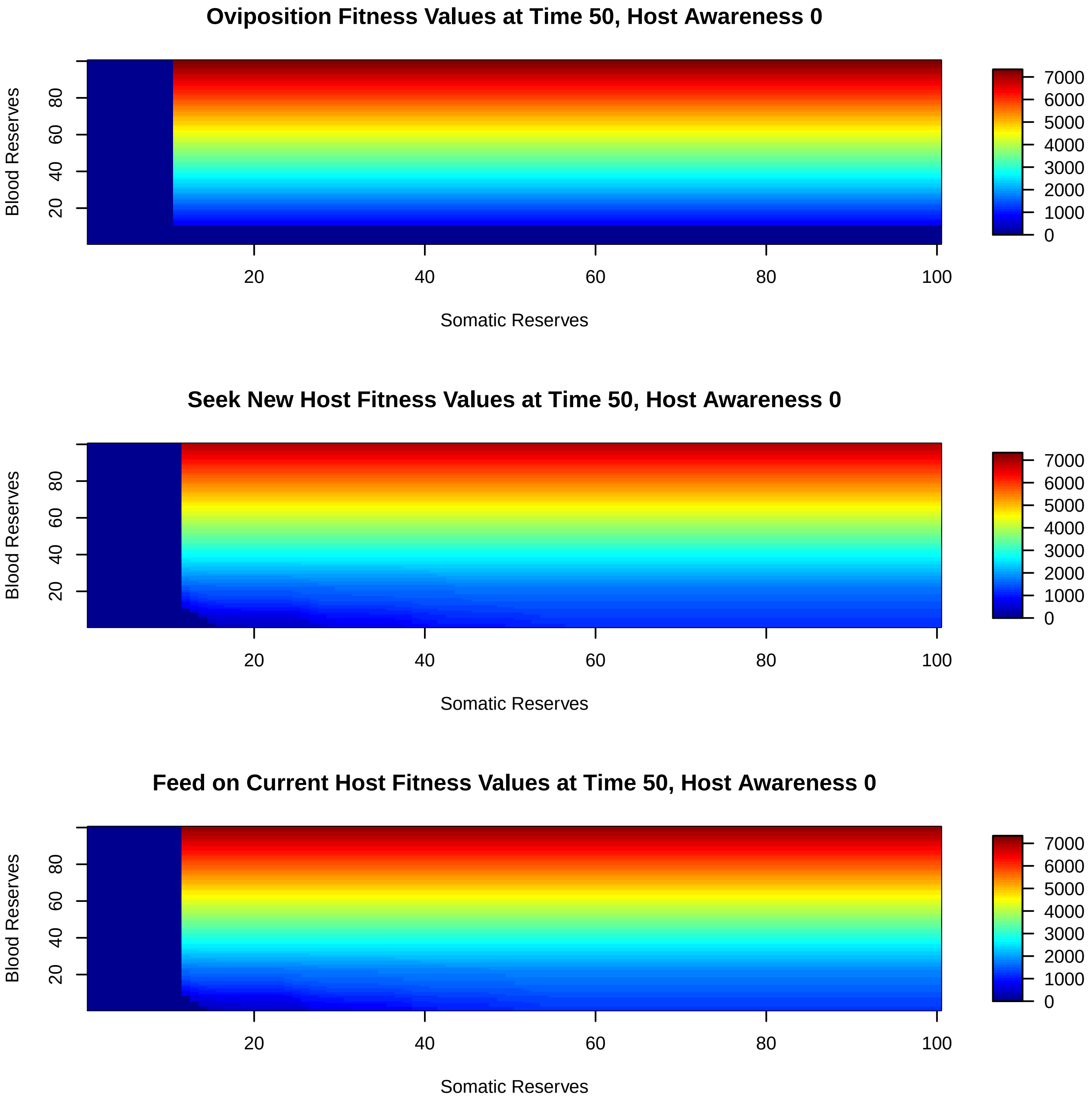


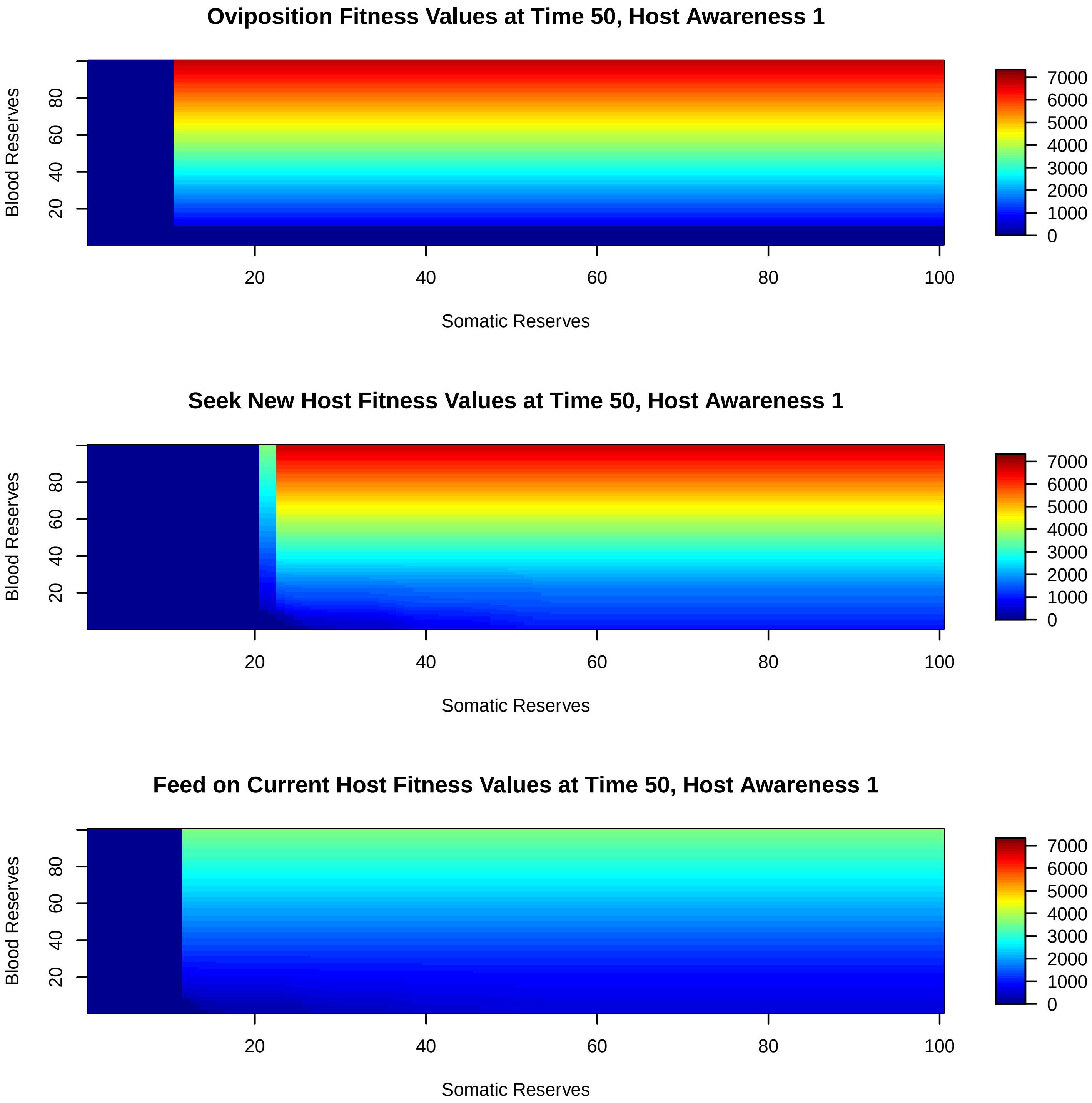
Figure 19: Decision rules for current blood and somatic reserve values, with both states of awareness, at time 85 (though the figure is generally equivalent to the decision tables between time 85 and time 97). Red (3) signifies the decision to continue feeding on the current host, teal (2) signifies the decision to seek a new host, and blue (1) is when the mosquito decides to leave to oviposit.


Oviposition Fitness Values at Time 50, Host Awareness 0

Seek New Host Fitness Values at Time 50, Host Awareness 0

Feed on Current Host Fitness Values at Time 50, Host Awareness 0

Figure 20: Fitness values for current blood and somatic reserve values, at time 50 and with a non-disturbed host. Warmer colors (red) indicate higher values, while colder colors (blue) indicate lower values.



Oviposition Fitness Values at Time 50, Host Awareness 1

Seek New Host Fitness Values at Time 50, Host Awareness 1

Feed on Current Host Fitness Values at Time 50, Host Awareness 1

Figure 21: Fitness values for current blood and somatic reserve values, at time 50 and with a disturbed host. Warmer colors (red) indicate higher values, while colder colors (blue) indicate lower values.