Metamodeling for Bias Estimation of Biological Reference Points

Nicholas Grunloh

April 10, 2023

1 Introduction

Data for a typical surplus-production model comes in the form of an index of abundance through time which is assumed to be proportional to the reproducing biomass for the population of interest. The index is often observed alongside a variety of other known quantities, but at a minimum, each observed index will be observed in the presence of some known catch for the period.

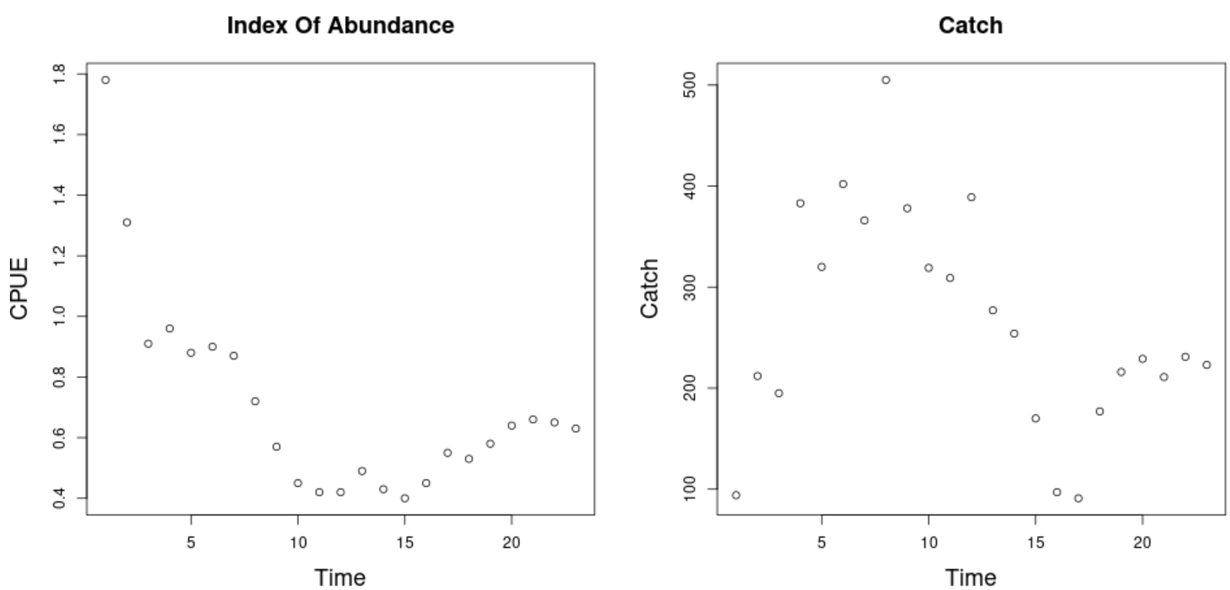


Figure 1: *left*: An observed series of index of abundance data for Namibian Hake from 1965 to 1987 (Hilborn & Mangel, 1997). *right*: The associated catch data for Namibian Hake over the same time period.

The observed indices are assumed to have multiplicative log-normal errors, and thus the following observation model arises naturally,

$$I_t = qB_te^{\epsilon} \quad \epsilon \sim N(0, \sigma^2). \tag{1}$$

Above q is often referred to as the "catchability parameter"; it serves as the proportionality constant mapping between the observed index of abundance and biomass. σ^2 models residual variation. Biologically speaking q and σ^2 are often treated as nuisance parameters with the "biological parameters" entering the model through a process model on biomass.

Biomass is assumed to evolve as an ODE; in this case I focus on the following form,

$$\frac{dB}{dt} = P(B(t); \boldsymbol{\theta}) - Z(t)B(t). \tag{2}$$

Here biomass is assumed to change in time by two processes, net production of biomass into the population, $P(B)$, and various sources of biomass removal, Z, from the population.

Firstly, the population grows through a production function, $P(B)$. Production in this setting is defined as the net biomass increase due to all reproduction and maturation processes. The production function is assumed to be a parametric (generally non-linear) function relating the current biomass of the population to an aggregate production of biomass.

Secondly, the population decreases as biomass is removed by various sources that are assumed to remove biomass linearly with biomass. Above, $Z(t)$, is an aggregate rate of removal. When the fishing rate, $F(t)$, is the only source of removal $Z(t) = F(t)$, however often models will also included other linear terms in $Z(t)$. Commonly the rate of "natural mortality", M, is also included as an additional term so that $Z(t) = M + F(t)$.

From a management perspective a major goal of modelling is to accurately infer a quantity known as *maximum sustainable yield* (MSY). One could maximize simple yield at a particular moment in time (and only for that moment) by fishing all available biomass in that moment. This strategy is penny-wise but pound-foolish (not to mention ecologically devastating) since it doesn't leave biomass in the population to reproduce in the future. We seek to fish in a way that allows (or even encourages) future productivity in the population. This is accomplished by maximizing the equilibrium level of catch over time. Equilibrium yield is considered by replacing the steady state biomass ($\bar{B}$) in the assumed form for catch, so that $\bar{Y} = F\bar{B}(F)$, where $^{-}$ indicates a value at steady state. The steady state biomass is a function of F. MSY is found by optimizing $\bar{Y}(F)$ with respect to F, and F^* is the fishing rate at MSY. Going forward let * decorate any value derived under the condition of MSY.

Fisheries are very often managed based upon reference points (RPs) which serve as simplified heuristic measures of population behavior. The mathematical form of RPs depends upon the model assumptions through the production function. While a number of different

RPs exist which describe the population in different (but related) ways, the most common RPs revolve around the concept of MSY (or robust ways of measuring MSY (Hilborn, 2010; Punt et al., 2016)). Here the focus is primarily on the RPs $\frac{B^*}{B(0)}$ and F^* ($\frac{F^*}{M}$ when appropriate) for their pervasive use in modern fisheries (Mangel et al., 2013; Punt & Cope, 2019).

F^* is the afore mentioned fishing rate which results in MSY. $\frac{B^*}{B(0)}$ is the depletion of the stock at MSY. That is to say $\frac{B^*}{B(0)}$ describes the fraction of the unfished population biomass that will remain in the equilibrium at MSY. In general $F^* \in \mathbb{R}^+$ and $\frac{B^*}{B(0)} \in (0,1)$, however under the under the assumption of a two parameter production function the model will be structurally unable to capture the full theoretical range of RPs (Mangel et al., 2013).

Many of the most commonly used production functions depend only on two parameters. For example, the Schaefer model (cite) depends only on the biological parameters r and K, and limits RP inference so that under the Schaefer model $\left(F^*, \frac{B^*}{B(0)}\right) \in \left(\mathbb{R}^+, \frac{1}{2}\right)$. Similarly the Beverton-Holt (Beverton & Holt, 1957, BH) and Ricker (Ricker, 1954) curves are also two parameter production functions that do not model the full theoretical space of RPs (Mangel et al., 2013).

The bias-variance trade-off (Ramasubramanian & Singh, 2017) makes it clear that the addition of a third parameter in the production function will necessarily reduce estimation bias. However the utility of this bias reduction is still under debate because the particular mechanisms and behavior (direction and magnitude) of these biases for key management quantities are not fully understood or described. Lee et al. (2012) provides some evidence that estimation of productivity parameters are dependent on biomass contrast as well as model specification. Conn et al. (2010) comes to similar conclusions via calibration modeling techniques. These studies indicate improtant factors that contribute to inferential failure, but they do not offer mechanisms of model failure, nor do they consider how different types of model misspecification interact with the information content of a given biomass series.

In this study I consider the behavior of inference when index data are simulated from three parameter PT and Schnute production models, but the simulated data are fit using intentionally misspecified two the parameter logistic or BH production models. The work begins with a derivation of RPs under the three parameter models. The parametric forms of RPs under the three parameter models are then inverted to develop a simulation setting for analyzing inference under the two parameter models. Finally a Gaussian Process (GP) metamodel (Gramacy, 2020) is constructed for exploration and analysis of RP biases.

A key insight of this approach is that bias is considered broadly across RP-space to uncover patterns and correlations between RPs. The GP metamodel is explicit about trade-offs between RPs so as to inform the full utility of reducing bias, as well as to suggest mechanisms for understanding what causes bias. Further, the effect of contrast on estimation is considered together with model misspecification.

2 Methods

2 .1 PT/Schaffer Model

The three parameter PT family has a convenient form that includes, among others (Fox Jr., 1970; Rankin & Lemos, 2015), the logistic production function as a special case. Pella-Tomlinson production function is parameterized so that $\boldsymbol{\theta} = [r, K, \gamma]$ and the family takes the following form,

$$P(B; [r, K, \gamma]) = \frac{rB}{\gamma - 1}\left(1 - \left(\frac{B}{K}\right)^{(\gamma-1)}\right). \quad (3)$$

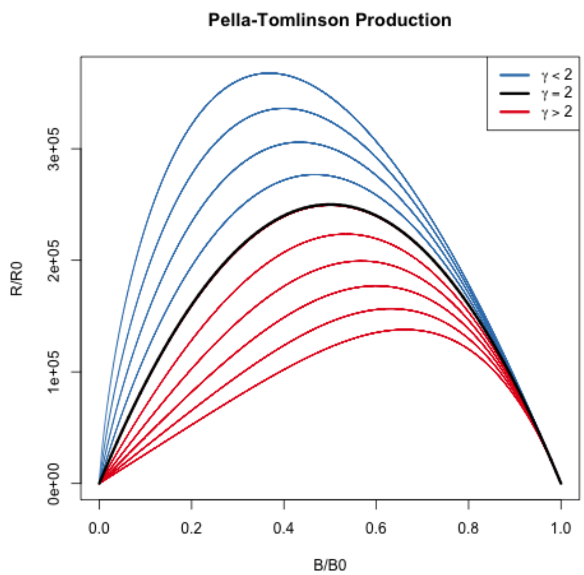


Figure 2: The PT production function plotted across a variety of parameter values. The special cases of Logistic production is shown in black, and the left-leaning and right-leaning regiems are shown in blue and red respectively.

γ is a parameter which breaks PT out of the restrictive symmetry of the logistic curve. In the special case of $\gamma = 2$ Eq (3) collapses back to the logistic curve, however in general $\gamma \in (1, \infty)$. The parameter r controls the maximum reproductive rate of the population in the absence of competition for resources (i.e. the slope of production function at the origin). K is the so called "carrying capacity" of the population. In this context the carrying capacity can be formally stated as steady state biomass in the absence of fishing (i.e. $\bar{B}(0) = K$). In Figure (2) PT recruitment is shown for a range of parameter values so as to demonstrate the various recruitment shapes that can be achieved by PT recruitment.

While the form of the PT curve produces some limitations (cite), importantly the introduction of a third parameter allows enough flexibility to fully describe the space of reference points used in management. To see this, the reference points are analytically derived for the PT model below.

2 .2 PT Reference Points

With $B(t)$ representing biomass at time t, under PT production, the dynamics of biomass are defined by the following ODE,

$$\frac{dB}{dt} = \frac{rB}{\gamma - 1}\left(1 - \left(\frac{B}{K}\right)^{\gamma-1}\right) - FB. \tag{4}$$

An expression for the equilibrium biomass is attained by setting Eq (4) equal to zero, and rearranging the resulting equation to solve for B. Thinking of the result as a function of F gives,

$$\bar{B}(F) = K\left(1 - \frac{F(\gamma-1)}{r}\right)^{\frac{1}{(\gamma-1)}}. \tag{5}$$

At this point it is convenient to notice that $\bar{B}(0) = K$. The expression for B^* is given by evaluating Eq (5) at F^*. To get an expression for F^*, the equilibrium yield is maximized with respect to F,

$$F^* = \underset{F}{\operatorname{argmax}}\, F\bar{B}(F). \tag{6}$$

In the case of PT production this maximization can be done analytically, by differentiating the equilibrium yield with respect to F as follows,

$$\frac{d\bar{Y}}{dF} = \bar{B}(F) + F\frac{d\bar{B}}{dF} \tag{7}$$

$$\frac{d\bar{B}}{dF} = -\frac{K}{r}\left(1 - \frac{F(\gamma-1)}{r}\right)^{\frac{1}{\gamma-1}-1}. \tag{8}$$

Setting Eq (7) equal to 0, substituting $\bar{B}(F)$ and $\frac{d\bar{B}}{dF}$ by Equations (5) and (8) respectively, and solving for F produces the following expression for the fishing rate required to produce MSY,

$$F^* = \frac{r}{\gamma} \tag{9}$$

Plugging the above expression for F^* back into Eq (5) gives the following expression for biomass at MSY,

$$B^* = K\left(\frac{1}{\gamma}\right)^{\frac{1}{\gamma-1}}. \tag{10}$$

The above derived expressions for $\bar{B}(0)$, B^*, and F^* can then be used to build a specific analytical form for the biological reference points in terms of only biological model parameters.

$$F^* = \frac{r}{\gamma} \qquad\qquad \frac{B^*}{\bar{B}(0)} = \left(\frac{1}{\gamma}\right)^{\frac{1}{\gamma-1}} \tag{11}$$

2 .3 Simulation

- introduce metamodeling idea

Indices of abundance are simulated from the three parameter PT production model broadly over the space of F^* and $\frac{B^*}{\bar{B}(0)}$ values. These PT data are then fit with a misspecified two parameter Schaefer model so as to observe the effect of productivity parameter model misspecification upon RP inference.

Generating simulated indices of abundance from the PT model requires inverting the relationship between $\left(F^*, \frac{B^*}{\bar{B}(0)}\right)$, and (r, γ). It is not generally possible to analytically invert this relationship for many three parameter production functions (Punt & Cope, 2019; J. T. Schnute & Richards, 1998). Most three parameter production functions lead to RPs that require expensive numerical methods to invert; more over the numerical inversion procedure can often be unstable. That said, for the case of PT this relationship is analytically invertible, and leads to the following relationship

$$r = \gamma F^* \qquad\qquad \gamma = \frac{W\left(\frac{B^*}{B(0)}\log\left(\frac{B^*}{B(0)}\right)\right)}{\log\left(\frac{B^*}{B(0)}\right)}. \tag{12}$$

Above W is the Lambert product logarithm function. More details about this derivation, and the Lambert product logarithm, are given in Appendix (4).

2 .3.1 Latin Hypercube Sampling

A Latin hypercube sample (LHS) of size n, in an m dimensional space, samples uniformly among uniform grids of size n in each dimension of the design space. By intersecting the grids of each dimension, n^m cells are produced, from which a total of n samples are taken. Crutially only one sample is taken from a given element of each grid in each dimension so as to reduce clumping of the n samples across the design space.

Letting $\mathcal{F}$ and $\mathcal{B}$ be equally spaced grids, of size n, on $F_{MSY} \in (0.1, 0.7)$ and $\frac{B_{MSY}}{B_0} \in (0.2, 0.6)$ respectively, a LHS samples 1 point in n of the n^2 cells produced by $\mathcal{F} \times \mathcal{B}$.

Each of the sampled LHS design locations represent a unique PT model with the sampled RP values. The productivity parameters of the PT, at each design location, are obtained by applying Eq. (12). Since K does not enter the RP calculation its value is fixed arbitrarily at 10000. The value of q is fixed at a typically small value of 0.0005. σ is fixed at the relatively small value of 0.01 to focus specifically on the behavior of population parameters. These parameters fully specify the PT model for the purposes of generating index data for each $\left(F^*, \frac{B^*}{B(0)}\right)$ pair.

2 .3.2 Catch

It is known that the behavior of catch can effect inference on the productivity parameters (Hilborn & Walters, 1992). In this setting contrast refers to changes in the long term trends of index data. Figure (3, *right*) demonstrates an example of biomass that includes contrast induced by catch. It is not well understood how contrast may factor into inferential failure induced by model misspecification. A variety of catches are investigated.

Catch is parameterized so that $F(t)$ can be controlled with respect to F^*. Recall that catch is assumed to be proportional to biomass, so that $C(t) = F(t)B(t)$. To control $F(t)$ with respect to F^*, $C(t)$ is specified by defining the quantity $\frac{F(t)}{F^*}$ as the relative fishing rate. $B(t)$ is defined by the solution of the ODE, and F^* is defined by the biological parameters of the model, see Eq (42). By defining $\frac{F(t)}{F^*}$, catch can then be written as $C(t) = F^* \left(\frac{F(t)}{F^*}\right) B(t)$.

Intuitively $\frac{F(t)}{F^*}$ describes the fraction of F^* that $F(t)$ is specified to for the current $B(t)$. When $\frac{F(t)}{F^*} = 1$, $F(t)$ will be held at F^*, and the solution of the ODE brings $B(t)$ into equilibrium at B^*. For constant $\frac{F(t)}{F^*}$ biomass comes to equilibrium as an exponential decay from K approaching B^*. When $\frac{F(t)}{F^*} < 1$, $F(t)$ is lower than F^* and $B(t)$ is pushed toward $\bar{B} > B^*$. Contrarily, when $\frac{F(t)}{F^*} > 1$, $F(t)$ is higher than F^* and $B(t)$ is pushed toward $\bar{B} < B^*$; the precise values of $\bar{B}$ can be calculated from Eq (15).

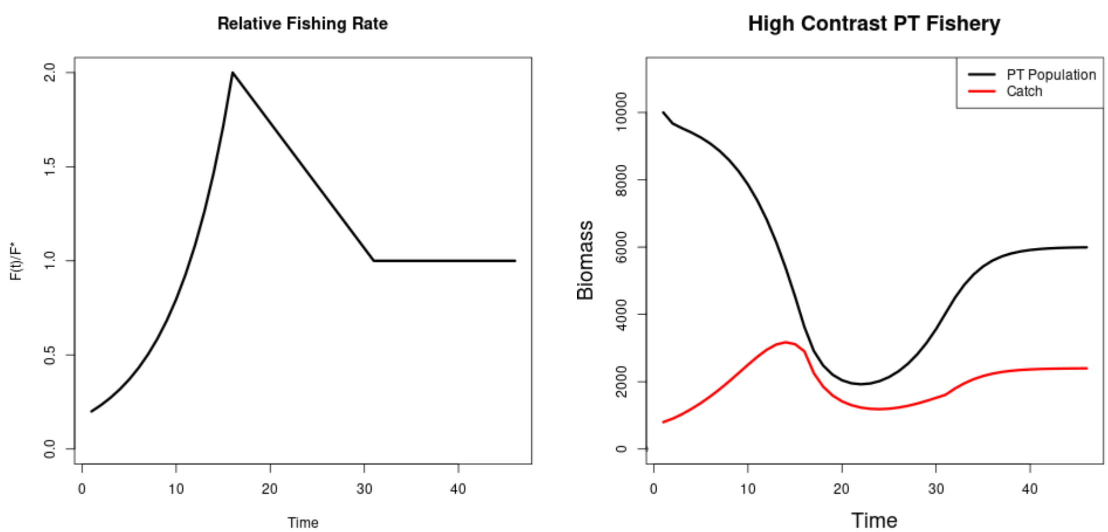


Figure 3: $(left)$ Relative fishing specified so as to induce contrast. $(right)$ Population biomass and catch demonstrating contrast in a PT population with $F^* = 0.4$ and $\frac{B^*}{B(0)} = 0.6$.

In practice, catch is determined by a series of observed, assumed known, catches. Catch observations are typically observed on a quarterly (or yearly) basis, so that the ODE may be discretized via Euler's method with integration step sizes to match the observation frequency of the modeled data. In this case, catch is sampled as would be done in practice however, the simulation can encounter a variety of issues working with the naively discretized ODE. As a result the ODE is integrated implicitly via the Livermore Solver (Radhakrishnan, 1993, lsode), and catch is linearly interpolated between sampled epochs.

- ?quantification of degrees of information? (avg curvature?)
- remake picture w/o PT references

2 .3.3 Continuous model formulation

a preface to regularity issues: identifiability, stiffness, and continuity.

An important (and often overlooked) implementation detail is the solution to the ODE which defines the progression of biomass through time (See Eq(33)). As a statistical model it is of paramount importance that this ODE not only have a solution, but also that the solution be unique. Of primary concern, uniqueness of the ODE solution is necessary for the identifiability of the statistical model.

If the form of $\frac{dB}{dt}$ is at least Lipschitz continuous, then the Cauchy-Lipschitz-Picard theorem provides local existence and uniqueness of $B(t)$.Recall from Eq(33) that $\frac{dB}{dt}$ is separated into a term for recruitment into the population, $R(B)$, and a term for removals via catch, C. For determining Lipschitz continuity of $\frac{dB}{dt}$, the smallest Lipschitz constant of $\frac{dB}{dt}$ will be the sum of the constants for each of the terms $R(B)$ and C separately. Typically any choice of $R(B)$ will be continuously differentiable, which implies Lipschitz continuity (since the set of continuous differentiable functions is a subset of the set of Lipschitz continuous functions). Thus, the assumed form of $R(B)$ does not typically introduce continuity concerns, unlike some potential assumptions for C.

In practice C is determined by a series of observed, assumed known, catches. Catch observations are typically observed on a quarterly basis, but in practice may not be complete for every quarter of the modeled period. It is overwhelmingly common to discretize the ODE via Euler's method with integration step sizes to match the observation frequency of the modeled data. This is often convenient but can present several issues. This strategy often pushes the assumption of catch continuity under the rug, but for identifiability of the statistical model an implicit assumption of continuity of the catches is required. While mechanistically at the finest scale fishers must only catch discrete packets of biomass (i.e. individual fish), it is sensible to consider catches at the quarterly (or yearly) scale as accruing in a continuous way. Furthermore any assumption of continuity will be required to be at least Lipschitz continuous for the required regularity of the model.

Here I assume catches accrue linearly between observed catches. This assumption defines the catch function as a piecewise linear function of time, with the smallest Lipschitz constant for the catch term defined by the steepest segment of the catch function. This assumption represents one of the simplest ways of handling catch, while retaining Lipschitz continuity overall. Furthermore linearly interpolated catch is adequately parsimonious for the typical handling of catches.

2 .3.4 Integration and Stiffness

As previously mentioned, the overwhelming majority of implementations of population dynamics models discretize the ODE using Euler's method with the integration step sized fixed so as to match the observation frequency. In this setting we explore model parameterizations that explore the full extent of biologically relevant reference points. This exercise produces some combinations of parameters that result in numerically stiff ODEs.

The concept of stiffness in ODEs is hard to precisely characterize (cite). Hairer and Wanner [5, p. 2] describe stiffness in the following pragmatic sense, "Stiff equations are problems for which explicit methods don't work". It is hard to make this definition more mathematically precise, but this is without a doubt a consistent issue for models parameterized so that ζ is greater than about $\frac{1}{2}$. Euler's method, as often implemented, is particularly poorly suited for these stiff regions of parameter space. In these stiff regions it is necessary to integrate the ODE with an implicate integration method.

Several of the most common implicate methods were tried including the Livermore Solver for ODEs (lsode), and the Variable Coefficient ODE Solver (vode) as implemented in the deSolve package of R (cite). The difference between implicate solvers is negligible, while most explicit methods result in wildly varying solutions to the ODE, and in still regions of parameter space explicate methods completely fail to represent the model as stated in the stiff regions of parameter space. Results shown here are computed using the lsode integration method since it runs relatively quickly and has a relatively smaller footprint in system memory.

2 .4 Schnute/BH Model

The Schnute production function is a three parameter generalization of many of the most common two parameter production functions (Deriso, 1980; J. Schnute, 1985). It can be written in the following form, with parameters α, β, and γ,

$$P_s(B;[\alpha,\beta,\gamma]) = \alpha B(1-\beta\gamma B)^{\frac{1}{\gamma}}. \quad (13)$$

The BH and Logistic production functions arise when γ is fixed to -1 or 1 respectively, and the Ricker model is a limiting case as $\gamma \to 0$.

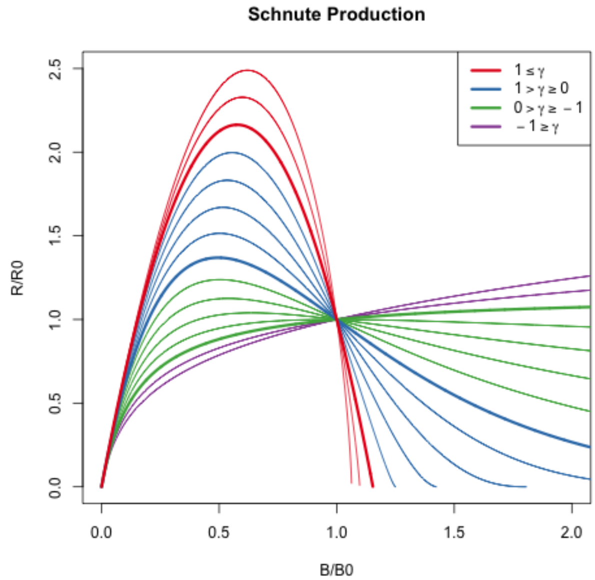


Figure 4: The Schnute production function plotted across a variety of parameter values. The special cases of BH, Ricker, and Logistic production are shown in green, blue, and red respectively.

Inference of BH productivity parameters under a wide variety of data is of particular interest due to the overwhelming popularity of the BH assumption in fisheries models. Since Schnute production models can represent a quantifiably wide varienty of possible productivity behaviours, they present an ideal simulation environment for inquiry of the relability of inference under the BH assumption.

Under Schnute production, biomass dynamics evolve according to the following ODE,

$$\frac{dB}{dt} = P_s(B;\theta) - (M+F)B. \quad (14)$$

This equation largely takes the same form as previously described, except that P_s is the Schnute production function and natural mortality, M, is modeled explicatly here. Natural mortality models the instantaneous rate of mortality from all causes outside of fishing. Explicatly modeling natural mortality in this way is not only a typical assumption of fisheries models, but is also key to the making RPs well defined over the relavant domain of γ.

The derivation of RPs under Eq. (14) follows a similar logic as under the PT model.

An expression for equilibrium biomass is attained by setting $\frac{dB}{dt} = 0$ and rearranging the resulting expression to solve for B

$$\bar{B}(F) = \frac{1}{\gamma\beta}\left(1 - \left(\frac{M+F}{\alpha}\right)^{\gamma}\right). \tag{15}$$

The above expression quickly yeilds B_0, B_{MSY} by evaluation at $F = 0$ and $F = F_{MSY}$ respectively,

$$B_0 = \frac{1}{\gamma\beta}\left(1 - \left(\frac{M}{\alpha}\right)^{\gamma}\right) \tag{16}$$

$$\frac{B_{MSY}}{B_0} = \frac{1 - \left(\frac{M+F_{MSY}}{\alpha}\right)^{\gamma}}{1 - \left(\frac{M}{\alpha}\right)^{\gamma}}. \tag{17}$$

Attaining an expression for F_{MSY} requires maximization of equilibrium yeild, $\bar{Y} = F\bar{B}(F)$, with respect to F. Analytically maximizing proceeds by differentiating $\bar{Y}$ to produce

$$\frac{d\bar{Y}}{dF} = \bar{B}(F) + F\frac{d\bar{B}}{dF} \tag{18}$$

$$\frac{d\bar{B}}{dF} = -\frac{1}{\beta}\left(\frac{\left(\frac{M+F}{\alpha}\right)^{\gamma}}{F+M}\right). \tag{19}$$

Setting $\frac{d\bar{Y}}{dF} = 0$, filling in the expressions for $\bar{B}(F)$ and $\frac{d\bar{B}}{dF}$, then rearranging to solve for F_{MSY} is less yeilding here than it was in the case of the PT model. This proceedure falls short of providing an analytical solution for F_{MSY} directly in terms of θ, but rather shows that F_{MSY} must respect the following expression,

$$0 = \frac{1}{\gamma} - \left(\frac{1}{\gamma} + \frac{F_{MSY}}{F_{MSY}+M}\right)\left(\frac{F_{MSY}+M}{\alpha}\right)^{\gamma}. \tag{20}$$

The lack of an analytical solution here is understood. J. T. Schnute and Richards (1998, pg. 519) specifically points out that F_{MSY} cannot be expressed analytically in terms of productivity parameters, but rather gives a partial analytical expression for the inverse relationship. Although parameterized slightly differently, J. T. Schnute and Richards (1998) derives expressions for α and β as a function of RPs and γ.

Since RPs are left without a closed form expression, computing RPs from productivity parameters amounts to numerically solving the system formed by collecting the expressions (20), (16), and (17).

2 .4.1 Simulation

For the purposed of simulation, it is not neccessary to completely know the precise relationships mapping RPs $\mapsto \theta$ or $\theta \mapsto$ RPs. Simulation only requires enough knowledge of these mappings to gather a list of (α, β, γ) tuples, for data generation under the Schnute model, and the corresponding RPs in some reasonable spacefilling design over RP space.

Similarly to J. T. Schnute and Richards (1998), expressions (20) and (16) are solved for α and β respectively, to produce the partial mapping $\big(F_{MSY}, B_0\big) \mapsto \big(\alpha(\cdot, \gamma),\ \beta(\cdot, \cdot, \gamma)\big)$ in terms of RPs and γ. By further working with Eq. (17), to identify γ, the following system is obtained,

$$\begin{aligned} \alpha &= (M + F_{MSY})\left(1 + \frac{\gamma F_{MSY}}{M + F_{MSY}}\right)^{1/\gamma} \\ \beta &= \frac{1}{\gamma B_0}\left(1 - \left(\frac{M}{\alpha}\right)^{\gamma}\right) \\ \frac{B_{MSY}}{B_0} &= \frac{1 - \left(\frac{M+F_{MSY}}{\alpha}\right)^{\gamma}}{1 - \left(\frac{M}{\alpha}\right)^{\gamma}}. \end{aligned} \tag{21}$$

For a population experienceing natural mortality M, by fixing F_{MSY}, B_0, and $\frac{B_{MSY}}{B_0}$ the above system can fully specify α and β for a given γ. Notice for a given γ a cascade of closed form solutions for α and β can be obtained. First $\alpha(\gamma)$ can be computed, and then $\beta(\alpha(\gamma), \gamma)$ can be computed. If $\alpha(\gamma)$ is filled back into the expression for $\frac{B_{MSY}}{B_0}$, the system collapses into a single onerous expression for $\frac{B_{MSY}}{B_0}(\alpha(\gamma), \gamma)$. For bevity, define the function $\zeta(\gamma) = \frac{B_{MSY}}{B_0}\big(\alpha(\gamma), \gamma, F_{MSY}, M\big)$ based on Eq. (17).

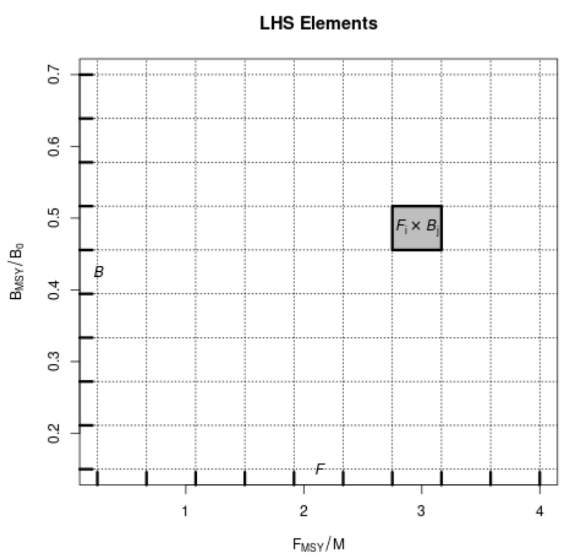
Inverting $\zeta(\gamma)$ for γ, and computing the cascade of $\alpha(\gamma)$, and then $\beta(\alpha(\gamma), \gamma)$, fully defines the Schnute model for a given $(\frac{F_{MSY}}{M}, \frac{B_{MSY}}{B_0})$. However inverting ζ accuratly is extremely difficult. Inverting ζ analytically is not feasible, and typical methods of numerically inverting ζ are unstable and expensive. Rather than numerically invert precise values of $\zeta(\gamma)$, γ is sampled so that the overall simulation design is space filling.

2 .4.2 Latin Hypercube Sampling

- a quick lit review of space filling designs

A Latin hypercube sample (LHS) of size n, in an m dimensional space, samples uniformly among uniform grids of size n in each dimension of the design space. By intersecting the grids of each dimension, n^m cells are produced, from which a total of n samples are taken. Crutially only one sample is taken from a given element of each grid in each dimension so as to reduce clumping of the n samples across the design space.

Letting $\mathcal{F}$ and $\mathcal{B}$ be equally spaced grids, of size n, on $\frac{F_{MSY}}{M} \in (0.25, 4)$ and $\frac{B_{MSY}}{B_0} \in (0.15, 0.7)$ respectively, a LHS samples 1 point in n of the n^2 cells produced by $\mathcal{F} \times \mathcal{B}$. Given the structured relationship between the RPs and productivity parameters α, β, and γ, obtaining a uniform LHS sample among $\mathcal{F} \times \mathcal{B}$ requires a tactful navigation of the system of equations seen in Eq. (21). The LHS grid setup and rough sampling strategy can be seen in Figure (5).



Given B_0, M, and F_{MSY}:

1) Draw $\gamma^* \sim \gamma | F_{MSY}, M$.
2) Compute $\frac{B_{MSY}}{B_0} = \zeta(\gamma^*)$
3) Compute $\alpha^* = \alpha(\gamma^*, F_{MSY}, M)$
4) Compute $\beta^* = \beta(\alpha^*, \gamma^*, M, B_0)$

Figure 5: $(left)$ LHS grids. Intersecting $\mathcal{F}$ and $\mathcal{B}$ produces n^2 cells; a particular cell $\mathcal{F}_i \times \mathcal{B}_j$ is shown in grey. $(right)$ An outline of the sampling proceedure for γ (and associated quantities) given B_0, M, and F_{MSY}.

Since it is not practicle to invert $\zeta(\gamma)$, a uniform sample in $\frac{B_{MSY}}{B_0}$ can be obtained by modeling γ as a random variable, with realization γ^*, and thinking of $\zeta(\gamma)$ as its cumulatice distribution function (CDF). The aim is to model γ as an easily sampled random varaible with a CDF that closely approximates ζ, so that $\zeta(\gamma^*)\dot{\sim}U(\zeta_{min}, 1)$ as closely as possible. There may be many good models for the distribution of γ, but in this setting the following distribution is very effective,

$$\gamma \sim \zeta_{min}\delta(\gamma_{min}) + t(\mu, \sigma, \nu)\mathbf{1}_{\gamma>\gamma_{min}}. \quad (22)$$

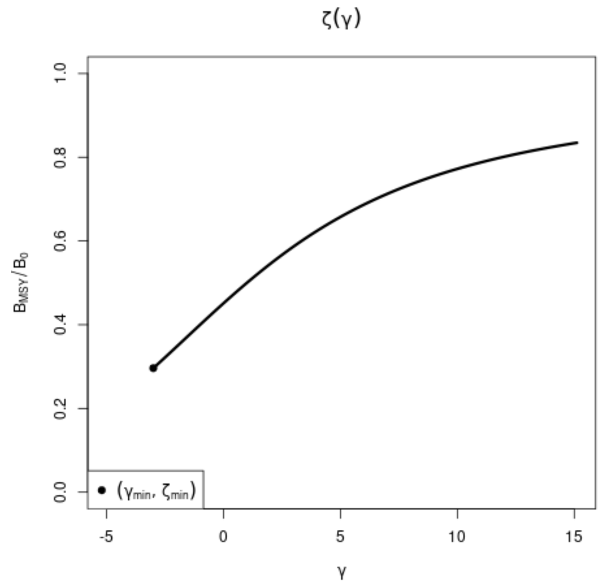


Figure 6: $\zeta(\gamma)$ Plotted for $F_{MSY} = 0.1$ and $M = 0.2$. The point $(\gamma_{min}, \zeta_{min})$ shows the lowest biologically meaningful value of γ; below which productivity is negative.

Above, t is the density of the three parameter location-scale family Student's t distribution with location μ, scale σ, and degrees of freedom ν. $\mathbf{1}_{\gamma>\gamma_{min}}$ is an indicator function that serves to truncate Student's t distribution at the lower bound γ_{min}. $\delta(\gamma_{min})$ is the dirac delta function evaluated at γ_{min}, which is scaled by the known value ζ_{min}; this places probability mass ζ_{min} at the point γ_{min}. Since sampling from Student's t distribution is readily doable, sampling from a truncated Student's t mixture only requires slight modification.

Let T be the CDF of the modeled distribution of γ. Since the point $(\gamma_{min}, \zeta_{min})$ is known from the dynamics of the Schnute model at a given RP, full specification of Eq. (22) only requires determining the values for μ, σ, and ν which make T best approximate $\zeta(\gamma)$. Thus, the values of μ, σ, and ν are chosen by minimizing the L^2 distance between $T(\gamma)$ and $\zeta(\gamma)$.

$$[\hat{\mu}, \hat{\sigma}, \hat{\nu}] = \underset{[\mu,\sigma,\nu]}{\arg\min} \int_\Gamma \big(T(\gamma; \mu, \sigma, \nu) - \zeta(\gamma)\big)^2 d\gamma \quad (23)$$

Fitting the distribution $T(\gamma|\hat{\mu},\hat{\sigma},\hat{\nu})$ for use generating γ^* values at a specific F_{MSY} and M releases the need to invert ζ. $T(\gamma|\hat{\mu},\hat{\sigma},\hat{\nu})$, together with the structure in Eq. (21), allows for the collection of an approximate LHS sample via the algorithm seen in Algorithm (1).

$\frac{F_{MSY}}{M}$ is drawn uniformly from $\mathcal{F}_i$. Conditioning on the sample of F_{MSY}, and M, $T(\gamma|\hat{\mu},\hat{\sigma},\hat{\nu})$ is fit and γ^* is sampled. ζ^* is then computed and placed into the appropriate grid element $\mathcal{B}_j$. Given γ^*, the cascade $\alpha(\gamma^*)$, and $\beta(\alpha(\gamma^*),\gamma^*)$, can be computed. The algorithm continues until all of the design elements, $(\frac{F_{MSY}}{M},\zeta^*) \Leftrightarrow (\alpha^*,\beta^*,\gamma^*)$, have been computed for all $i \in [1,...,n]$.

Algorithm 1 LHS of size n on rectangle R.

procedure $LHS_n(R)$
- Define n-grids $\mathcal{F},\mathcal{B} \in R$
- **for** each grid element i **do**
  - Draw $\frac{F_{MSY}}{M} \sim Unif(\mathcal{F}_i)$
  - Compute $[\hat{\mu},\hat{\sigma},\hat{\nu}]$ given F_{MSY} & M
  - **while** $\mathcal{B}_j$ not sampled **do**
    - Draw $\gamma^* \sim T(\gamma|\hat{\mu},\hat{\sigma},\hat{\nu})$
    - Compute $\zeta^* = \zeta(\gamma^*)$
    - Compute j such that $\zeta^* \in \mathcal{B}_j$
  - **end while**
  - Compute $\alpha^* = \alpha(\gamma^*, F_{MSY}, M)$
  - Compute $\beta^* = \beta(\alpha^*,\gamma^*, M, B_0)$
  - Save $(\frac{F_{MSY}}{M},\zeta^*) \Leftrightarrow (\alpha^*,\beta^*,\gamma^*)$ in $\mathcal{F}_i \times \mathcal{B}_j$
- **end for**

end procedure

2 .4.3 Design Refinement

Since the behavior of RP inference, under misspecified models, will vary in yet-unkown ways, the exact sampling design density may be hard to know a'priori. Several factors, including the particular level of observation uncertainty, high variance (i.e. hard to resolve) features of the response surface, or simply "gappy" instanciations of the initial LHS design may necessitate adaptive design refinement, to accuratly describe RP biases. Given the tempermental relationship between RPs and productivity parameters in the Schnute model, a recursive refinement algorithm, that makes use of the previously described LHS routine, is developed.

Holes in the existing design are identified based on maximin design principles. That is to say, new design points are collected based on areas of the RP design space which maximizes the minimum distance between all pairs of points in the current design.

$$d(\boldsymbol{x}, \boldsymbol{x'}) = \sqrt{(\boldsymbol{x} - \boldsymbol{x'})^T \boldsymbol{D}^{-1} (\boldsymbol{x} - \boldsymbol{x'})} \tag{24}$$
$$\boldsymbol{D} = \textbf{diag}\left[\big(max(\mathcal{F}) - min(\mathcal{F})\big)^2,\ \big(max(\mathcal{B}) - min(\mathcal{B})\big)^2\right]$$

Above, d is a scaled distance function that defines the distance between points in the differing scales of $\frac{B_{MSY}}{B_0}$ and $\frac{F_{MSY}}{M}$. $\boldsymbol{D}$ is a diagonal matrix that measures the squared size of the domain in each axis of so as to normalizing distances to a common scale.

If $\boldsymbol{X_n}$ is the initial design, computed on R_{full}, let $\boldsymbol{x_a}$ be the augmenting point which maximizes the minimum distance between all of the existing design points,

$$\boldsymbol{x_a} = \underset{\boldsymbol{x'}}{\text{argmax}} \ \ \min\{d(\boldsymbol{x_i}, \boldsymbol{x'}) : i = 1, ..., n\}. \tag{25}$$

The point $\boldsymbol{x_a}$ is used as an anchor for augmenting $\boldsymbol{X_n}$. An additional $LHS_{n'}$ (via Algorithm (1)) is collected, adding n' design points, centered around $\boldsymbol{x_a}$, to the overall design. The augmenting region, $R_{(x_a,d_a)}$, for collecting $LHS_{n'}$ is defined based on the square centered at $\boldsymbol{x_a}$ with side length $2d_a$, where $d_a = \min\{d(\boldsymbol{x_i}, \boldsymbol{x_a}) : i = 1, ..., n\}$, in the space defined by the metric d.

Due to the tendency of maximin sampling to cluster augmenting points on the edges of the design space, $R_{(x_a,d_a)}$ is truncated by the outer most limits of R_{full} so as to focus design augmentation within the specified domain of the simulation. Furthermore, since the design space has a nonlinear constraint, the calculation of x_a is further truncated based on a convex hull defined by the existing samples in the overall design.

In summary, an initial $X_n = LHS_n(R_{full})$ design is computed based on an overall simulated region of RPs R_{full}. The maximin augmenting point, x_a, is computed at a maximin distance of d_a from the existing samples. An augmenting design $X_{n'} = LHS_{n'}(R_{(x_a,d_a)})$ is collected and added to X_n. Design refinement carries on recursively collecting augmenting designs in this way until the desired maximin distance falls below the desired level.

2 .5 Gaussian Process Metamodel

- add metamodeling context

A GP is a stochastic process generalizing the multivariate normal distribution to an infinite dimensional analog. GPs are often specified primarily through the choice of a covariance (or correlation) function which defines the relationship between locations in an index set. Typically the index set is spatial for GPs, with points closely related in the index set resulting in correlated effects in the model. In this setting the model is over the space of reference points. A GP model implies an n dimensional multivariate normal distribution on the observations of the model with a correlated error structure defined by the modeled covariance function.

Each design location of the simulation produces an estimate of two productivity parameters under the restricted production model. Each of the fitted productivty parameter estimates are then modeled using independent instances of the following model in Eq (26).

Let $\mathbf{y}$ be a vector collecting the fitted log productivity parameter MLEs under the restricted two parameter production models. Furthermore, let $\boldsymbol{\omega}$ be a vector of the MLE standard error estimates, on the variance scale (via the inverted Fisher information of the production model log likelihood). $\boldsymbol{X}$ is the n x 2 LHS design matrix of RPs derived above for each respective three parameter data generating model.

$$
\begin{aligned}
&\mathbf{y} = \beta_0 + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v} + \boldsymbol{\epsilon} \\
&\boldsymbol{v} \sim N_n(\mathbf{0}, \tau^2 \boldsymbol{R}_{\boldsymbol{\ell}}) \\
&\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \boldsymbol{\omega}' \boldsymbol{I})
\end{aligned}
\tag{26}
$$

ϵ models independent normally distributed error, which provides an ideal mechanism for propagating uncertainty from inference in the simulation step into the metamodel. By matching each y_i with an observed ω_i variance term, ϵ serves to down weight the influence of each y_i in proportion to the inferred production model sampling distribution uncertainty. This has the effect of smoothing the GP model in a way similar to the nugget effect (Gramacy & Lee, 2012), although the application here models this effect heterogeniously.

The term, $\boldsymbol{v}$, contains spatially correlated GP effects. The correlation matrix, $\boldsymbol{R}_{\boldsymbol{\ell}}$ describes how RPs closer in the simulation design are more correlated. This spatial effect is modeled with a squared exponential correlation function,

$$R(\boldsymbol{x}, \tilde{\boldsymbol{x}}) = \exp\left(\sum_{i=1}^{2} \frac{-(x_i - \tilde{x}_i)^2}{2\ell_j^2}\right). \tag{27}$$

R has an anisotropic separable form which allows for differing length scales, ℓ_1 and ℓ_2, in the different RP axes. The flexibility to model correlations separately in the different RP axes is key due to the differences in the extent of the RP domains marginally. The metamodel parameters β_0, $\boldsymbol{\beta}$, τ^2, ℓ_1 and ℓ_2 are fit via MLE against the observations $\mathbf{y}$, $\boldsymbol{X}$, and $\boldsymbol{\omega}$ from simulation fits.

Predictive estimates are obtained via kriging (cite).

$$\hat{y}(\mathbf{x}) = \beta_0 + \mathbf{x}\boldsymbol{\beta} + \mathbf{r}(\mathbf{x})'\boldsymbol{R}_{\boldsymbol{\ell}}^{-1}\Big(\mathbf{y} - \big(\beta_0 + \boldsymbol{X}\boldsymbol{\beta}\big)\Big) \tag{28}$$

$\hat{y}(\mathbf{x})$ is a predicted value of the metamodel at the RP location $\mathbf{x}$. $\mathbf{r}(\mathbf{x})$ is defined as the vector of correlation function evaluations for the predictive location $\mathbf{x}$ against all observations in $\boldsymbol{X}$ (i.e. $\mathbf{r}(\mathbf{x}) = \boldsymbol{R}(\mathbf{x}, \boldsymbol{x}_i)\ \forall\ \boldsymbol{x}_i \in \boldsymbol{X}$).

Maybe a bit of hinting at uses.

3 Results

While metamodeling occurs on the inferred productivity parameters of the restricted production model, the metamodel can also be used to build metamodeled estimates of major biological RPs. The relavant transformations are given Eqs. (11, 17, 20) with γ fixed to the restricting case.

Applying the metamodel predictive surfaces to the RP estimate allows for the quantification of RP bias induced by model misspecification of two parameter production functions. Below RP bias is quantified by the following relative measure of bias, similar to a percent error calculation.

$$\text{Relative Bias} = \frac{\hat{RP} - RP}{RP} \tag{29}$$

Above RP is a stand in for the the true value of any of the biological reference points under the three parameter data generating production model, and $\hat{RP}$ refers to the metamodel estimate of each RP quantity under the two parameter restricted cases.

3 .1 PT/Schaffer

region of chaotic fits due to model misspecification

- RP break out plots with mechanism of failure

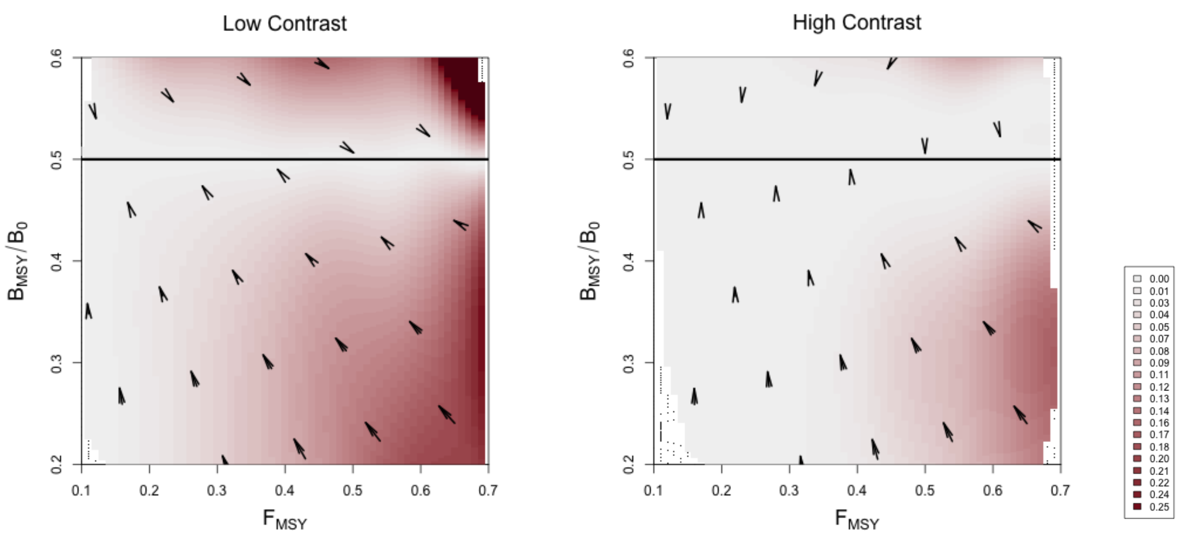


Figure 7: Joint bias direction for $(F_{MSY}, \frac{B_{MSY}}{B_0})$ estimates under the misspecified Schaefer Model. The intesity of color represents the excess bias relative to the shortest possible mapping.

- Mechanism plot on SRR LHS Grid plot
- some detail views in B_{MSY}, B_0, F_{MSY}, $?F_{SPR?}$, MSY

Figure (8) shows four of the most misspecified example production function fits as compared to the true data generating PT production functions. In the rug plots below each set of curves the observed biomasses demonstrate the exponential decay from K to B^* in each case. In particular, notice how only biomasses greater than the PT B^* are observed.

Due to the leaning of the true PT curves, and the symmetry of the logistic parabola, the logistic curve only observes information about its slope at the origin from data observed on the right portion of the PT curves. Above the Schaefer line PT is steeper on the right of B^* than it is on the left, and so the the logistic curve over-estimates r, and thus F^*, for data generated above the Schaefer line. Below the Schaefer line the vice versa phenomena occurs. Below the Schaefer line PT is shallower to the right of B^* than it is on the left and so the logistic parabola estimate tends to under estimate F^*.

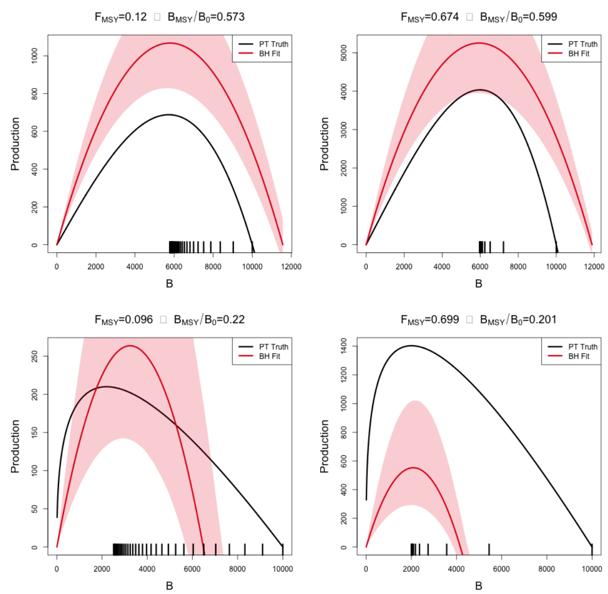


Figure 8: A comparison of the true PT production function (in black) and the estimated logistic curve (in red) with 95% CI shown. The examples shown represent the four corners of maximum model misspecification in the simulated RP-space. Observed biomasses are plotted in the rug plots below the curves.

Figure (8) indicates that the individual biases of B^* and K may behave quite differently. B^* appears to be estimated fairly accurately while K does not.

Figure (17) also gives some examples of the relative behavior of B^* and K. In Figure (16) it is clear that the bias behavior of $\frac{B^*}{B(0)}$ is locked in a fixed pattern under the Schaefer model. Figure (17) indicates that the individual biases of B^* and K may behave quite differently. B^* appears to be estimated fairly accurately while K does not.

3 .2 Schnute/BH

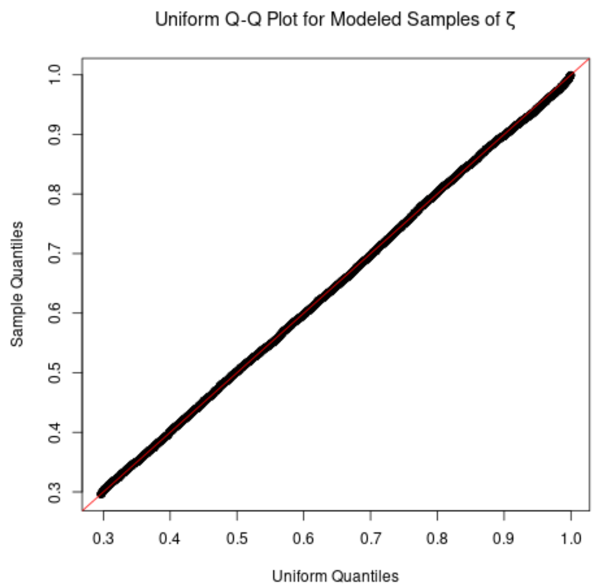


Figure 9: Uniform Q-Q plot for sampled ζ against theoretical uniform quantiles.

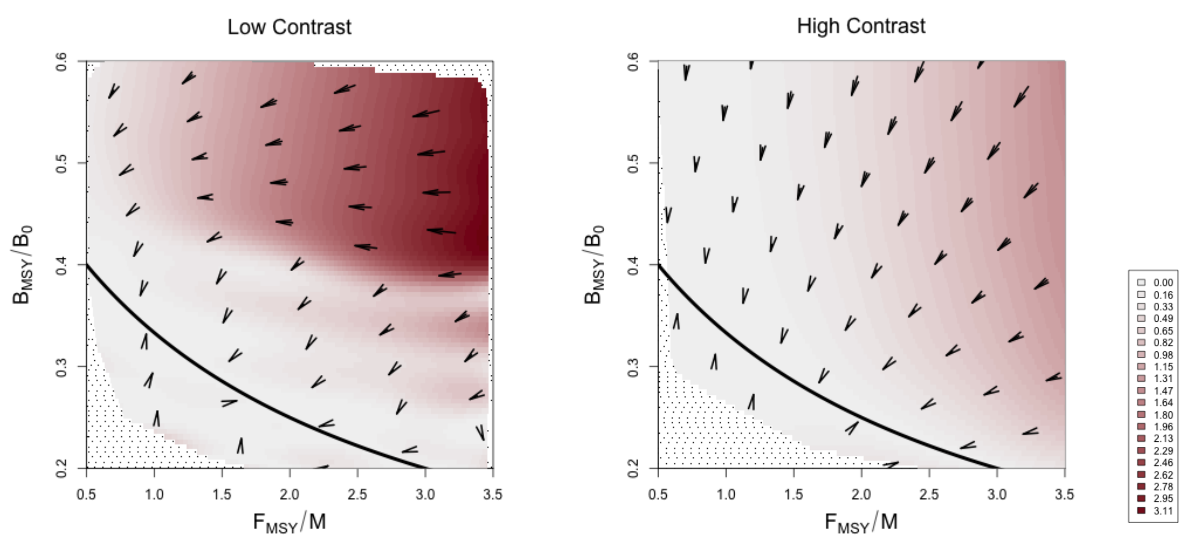


Figure 10: Joint bias direction for $\left(\frac{F_{MSY}}{M}, \frac{B_{MSY}}{B_0}\right)$ estimates under the misspecified BH Model. The intesity of color represents the excess bias relative to the shortest possible mapping.

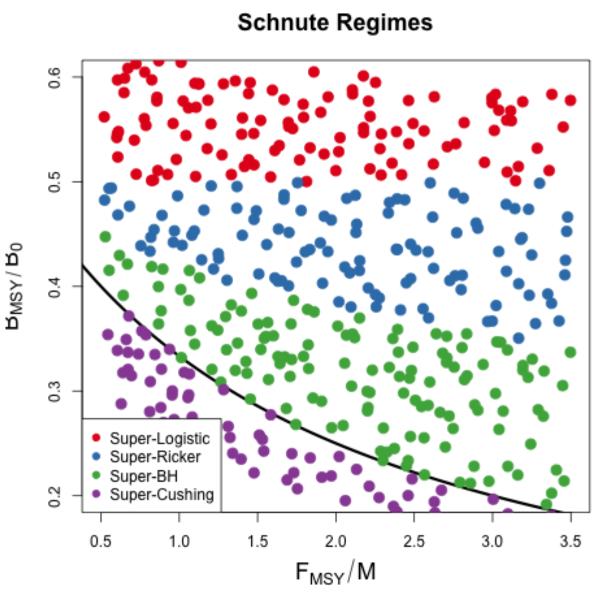


Figure 11: grid plot of production functions.)

4 Appendix: Inverting $\frac{B^*}{B(0)}$ and γ for the PT Model

For bevity let $\zeta = \frac{B^*}{B(0)}$.

$$\zeta = \left(\frac{1}{\gamma}\right)^{\frac{1}{\gamma-1}}$$
$$\zeta = \gamma\zeta^{\gamma}$$
$$\zeta = \gamma e^{\gamma\log(\zeta)}$$
$$\zeta\log(\zeta) = \gamma\log(\zeta)e^{\gamma\log(\zeta)}$$

The Lambert product logarithm, W, is defined as the inverse function of $z = xe^x$ such that $x = W(z)$. Applying this definition allows for the isolation of γ.

$$\begin{aligned}\gamma\log(\zeta) &= W\left(\zeta\log(\zeta)\right)\\ \gamma &= \frac{W\left(\zeta\log(\zeta)\right)}{\log(\zeta)}\end{aligned} \tag{30}$$

The Lambert product logarithm is a mulivalued function with a branch point at $-\frac{1}{e}$. The principal branch, $W_0(z)$, is defined on $z \in \left(-\frac{1}{e}, \infty\right)$, and the lower branch, $W_{-1}(z)$, is defined on $z \in \left(-\frac{1}{e}, 0\right)$. Taken individually, each respective branch is analytic, but cannot be expressed in terms of elementary functions.

When $\zeta \in \left(0, \frac{1}{e}\right)$ the solution of interest in Eq. (12) comes from W_0. When $\zeta \to \frac{1}{e}$, the Fox Model emerges as $\gamma \to 1$. When $\zeta \in \left(\frac{1}{e}, 1\right)$ the solution of interest comes from W_{-1}. For the use case presented here, Eq. (12) is to be interpreted as,

$$\gamma = \begin{cases} \frac{W_0(\zeta\log(\zeta))}{\log(\zeta)} & \zeta \in \left(0, \frac{1}{e}\right) \\ \frac{W_{-1}(\zeta\log(\zeta))}{\log(\zeta)} & \zeta \in \left(\frac{1}{e}, 1\right) \end{cases}. \tag{31}$$

Prager 2002, Figure(2).

https://math.stackexchange.com/questions/3004835/is-the-lambert-w-function-analytic-if-not-everywhere-then-on-what-set-is-it-ana https://researchportal.bath.ac.uk/en/publications/algebraic-properties-of-the-lambert-w-function-from-a-result-of-r

https://cs.uwaterloo.ca/research/tr/1993/03/W.pdf

5 Introduction

The most fundamental model in modern fisheries management is the surplus-production model. These models focus on modeling population growth via nonlinear parametric ordinary differential equations (ODE). Key management quantities called reference points (RP) are commonly derived from the ODE equilibrium equations and depend upon the parameterization of biomass production. Two-parameter parameterizations of the production function have been shown to limit the theoretical domain of RPs (Mangel et al., 2013). The limited RP-space of two parameter models are a major source of model misspecification for RPs and thus induce bias in RP estimation. The behavior of RP estimation bias is not well understood and as a result often underappreciated. A metamodeling approach is developed here to describe RP biases and explore mechanisms of model failure in the Schaefer model.

Data for a typical surplus-production model comes in the form of an index of abundance through time which is assumed to be proportional to the reproducing biomass for the population of interest. The index is often observed alongside a variety of other known quantities, but at a minimum, each observed index will be observed in the presence of some known catch for the period.

The observed indices are assumed to have multiplicative log-normal errors, and thus the following observation model arises naturally,

$$I_t = qB_te^{\epsilon} \quad \epsilon \sim N(0, \sigma^2). \tag{32}$$

Above q is often referred to as the "catchability parameter"; it serves as the proportionality constant mapping between the observed index of abundance and biomass. σ^2 models residual variation. Biologically speaking q and σ^2 are often treated as nuisance parameters with the "biological parameters" entering the model through a process model on biomass.

Biomass is assumed to evolve as an ODE; in this case I focus on the following form,

$$\frac{dB}{dt} = P(B(t); \boldsymbol{\theta}) - C(t). \tag{33}$$

Here biomass is assumed to change in time by two processes, net production of biomass into

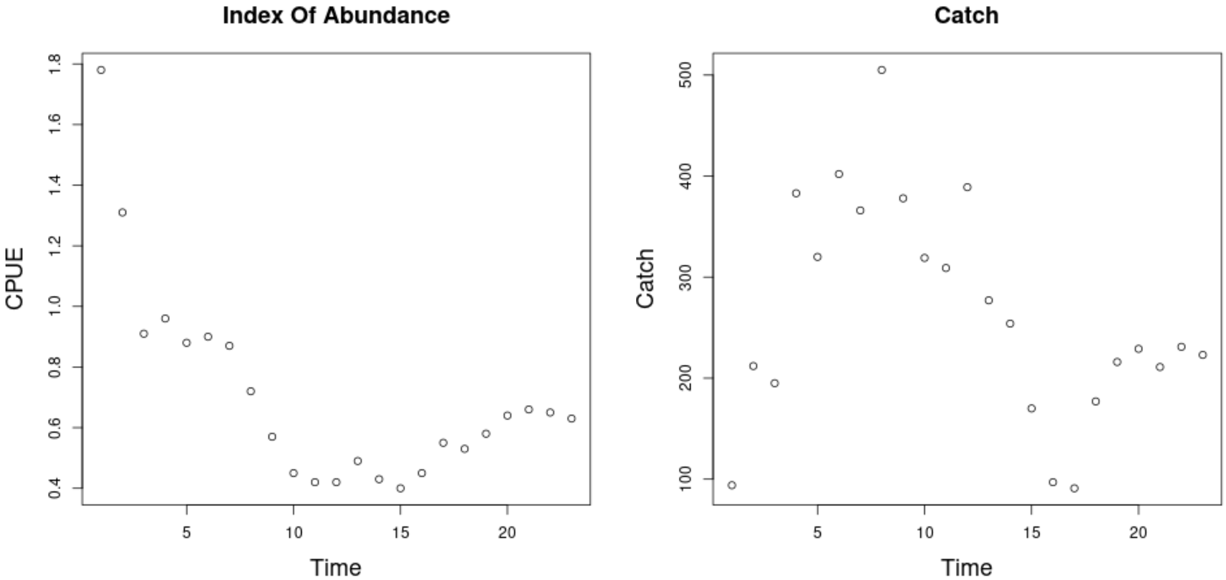


Figure 12: *left*: An observed series of index of abundance data for Namibian Hake from 1965 to 1987 (Hilborn & Mangel, 1997). *right*: The associated catch data for Namibian Hake over the same time period.

the population, and catches removing biomass from the population.

Firstly, the population grows through a production function, $P(B)$. Production in this setting is defined as the net biomass increase due to all reproduction and maturation processes accounting for all naturally occurring sources of mortality other than the recorded fishing from humans. The production function is assumed to be a parametric function that relates the current biomass of the population to an aggregate production of biomass.

Secondly, the population decreases as biomass is removed due to catch, $C(t)$. While catches (aka yields) are observable quantities (Pearson & Erwin, 1997), the model assumes that catch is proportional to biomass with the proportionality constant representing the fishing rate, $F(t)$, so that $C(t) = F(t)B(t)$. From a management perspective a major goal of the model is to accurately infer a quantity known as *maximum sustainable yield* (MSY). One could maximize simple yield at a particular moment in time (and only for that moment) by fishing all available biomass in that moment. This strategy is penny-wise but pound-foolish (not to mention ecologically devastating) since it doesn't leave biomass in the population to reproduce for future time periods. We seek to fish in a way that allows (or even encourages) future productivity in the population. This is accomplished by maximizing the equilibrium level of catch over time. Equilibrium yield is considered by replacing the steady state biomass ($\bar{B}$) in the assumed form for catch, so that $\bar{C} = F\bar{B}(F)$, where $\bar{}$ indicates a value at steady state. Naturally the steady state biomass is a function of F; we will see a specific example of this in Section (6 .2). MSY is found by optimizing $\bar{C}(F)$ with respect to F, and F^* is the fishing rate at MSY. Going forward let * decorate any value derived under the condition of MSY.

The canonical production model in fisheries is the Schaefer model. The Schaefer model is formed by choosing P to be logistic growth (Mangel, 2006) parameterized by $\boldsymbol{\theta} = [r, K]$ so that the family of production functions takes the following form,

$$P(B; [r, K]) = rB\left(1 - \frac{B}{K}\right). \tag{34}$$

r is a parameter controlling the maximum reproductive rate of the population in the absence of competition for resources (i.e. the slope of production function at the origin). K is the so called "carrying capacity" of the population. In this context the carrying capacity can be formally stated as steady state biomass in the absence of fishing (i.e. $\bar{B}(0) = K$).

The logistic production function produces idealized parabolic recruitment with equilibrium quantities taking very simple forms that can be easily understood from the graphical construction seen in Figure (13). Positive recruitment is observed when $B \in (0, K)$. Due to the parabolic shape of the logistic production function it is straightforward to see that yield is maximized by fishing the stock down to B^*, where the stock attains its peak productivity. By symmetry it is clear that this peak occurs at $B^* = \frac{K}{2}$. The fishing rate required to hold the stock at MSY is $F^* = \frac{r}{2}$, which is half of the stock's maximum reproductive rate. MSY is then the product of F^* and B^* so that $MSY = \frac{rK}{4}$.

Fisheries are very often managed based upon reference points which serve as simplified heuristic measures of population behavior. The mathematical form of RPs depends upon the model assumptions through the production function. While a number of different RPs exist which describe the population in different (but related) ways, the most common RPs revolve around the concept of MSY (or robust ways of measuring MSY (Hilborn, 2010; Punt

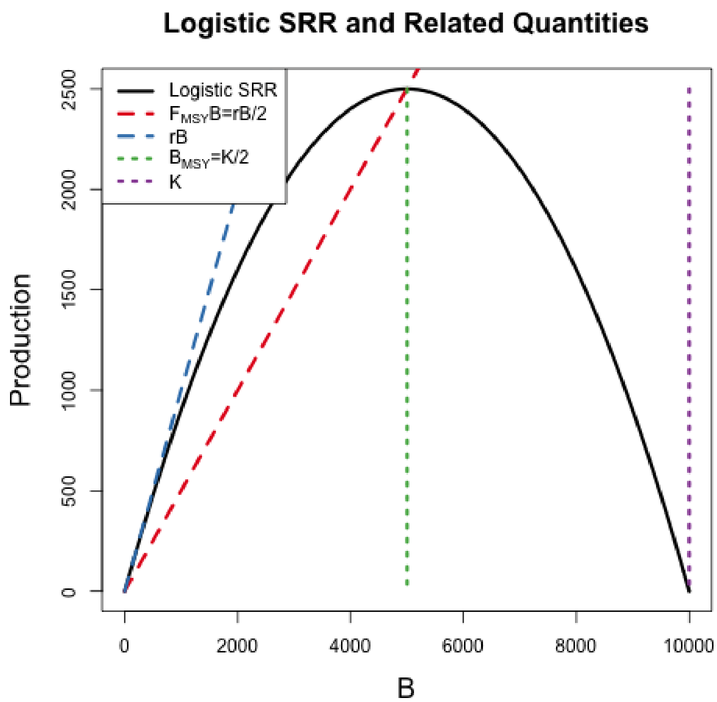


Figure 13:
The logistic production function in black plotted next to depictions of the key biological parameters and reference points. The slope at the origin (and thus r) is shown in blue, catch resulting in MSY in red, biomass at MSY in green, and K in purple at the right x-intercept. MSY is seen at the peak of the parabola, and is attained with a fishing rate of $\frac{r}{2}$ and biomass equilibrating to $\frac{K}{2}$.

et al., 2016)). Here the focus is primarily on the RPs F^* and $\frac{B^*}{\bar{B}(0)}$ for their pervasive use in modern fisheries (Mangel et al., 2013; Punt & Cope, 2019).

F^* is the afore mentioned fishing rate which results in MSY. $\frac{B^*}{\bar{B}(0)}$ is the depletion of the stock at MSY. That is to say $\frac{B^*}{\bar{B}(0)}$ describes the fraction of the unfished population biomass that will remain in the equilibrium at MSY. In general $F^* \in \mathbb{R}^+$ and $\frac{B^*}{\bar{B}(0)} \in (0, 1)$, however under the under the assumption of logistic production these quantities take the following form,

$$F^* = \frac{r}{2} \qquad\qquad \frac{B^*}{\bar{B}(0)} = \frac{1}{2} \tag{35}$$

so that $\left(F^*, \frac{B^*}{\bar{B}(0)}\right) \in \left(\mathbb{R}^+, \frac{1}{2}\right)$.

In current practice, production functions are typically chosen to depend only on two parameters. The Schaefer model as presented depends only on the biological parameters r and K, but other common two parameter choices of the production function are the Beverton-Holt (Beverton & Holt, 1957, BH) and Ricker (Ricker, 1954) curves. All of these two parameter production functions struggle similarly to model the full theoretical space of

RPs (Mangel et al., 2013).

The basis of the Schaefer model is ripe with debate (Kingsland, 1982), and the debate continues within modern fisheries modeling (Prager, 2002). On the one hand, Maunder (2003) argues that the Schaefer model is insufficient in large part due to the restriction it places on $\frac{B^*}{B(0)}$, at $\frac{1}{2}$, and further argues that the three parameter Pella-Tomlinson (PT) model (Pella & Tomlinson, 1969) should replace the Schaefer model to avoid biased parameter estimates. On the other hand, while Prager (2003) appreciates the limitations of the Schaefer model, he argues its usefulness as a well understood and simple model that has the ability to reasonably approximate dynamics in many data poor stocks.

The bias-variance trade-off (Ramasubramanian & Singh, 2017) makes it clear that the addition of a third parameter in the production function will necessarily reduce estimation bias. However the utility of this bias reduction is still under debate because the particular mechanisms and behavior (direction and magnitude) of these biases for key management quantities are not fully understood or described. Lee et al. (2012) provides some evidence that estimation of productivity parameters, and thus RPs via (Mangel et al., 2013), are dependent on biomass contrast as well as model specification. Conn et al. (2010) comes to similar conclusions via calibration modeling techniques. Despite this understanding of productivity estimation, the implications have not been extended to a joint description of biases on the scale of management RPs.

Together the general behavior of the PT model and the simplicity of the Schaefer model make the PT/Schaefer pair an ideal setting for beginning to understand the consequences of model misspecification on the production function. In this study I consider the behavior of inference when data are simulated from the three parameter PT production model but fit with the two parameter Schaefer model.

The work begins with a derivation of RPs under the three parameter PT model. The parametric forms of RPs under the PT model are then inverted to develop a simulation setting for analyzing inference under the two parameter Schaefer model. Finally a Gaussian Process (GP) metamodel (Gramacy, 2020) is constructed for exploration and analysis of RP biases.

A key insight of this approach is that bias is considered broadly across RP-space to uncover patterns and correlations between RPs. The GP metamodel is explicit about trade-offs between RPs so as to inform the full utility of reducing bias, as well as to suggest mechanisms for understanding what causes bias. Further, the effect of contrast on estimation is considered together with model misspecification.

6 Methods

6 .1 PT Model

The three parameter PT family has a convenient form that includes, among others (Fox Jr., 1970; Rankin & Lemos, 2015), the logistic production function as a special case to form the Schaefer model. The Pella-Tomlinson production function is parameterized so that $\boldsymbol{\theta} = [r, K, \gamma]$ and the family takes the following form,

$$P(B; [r, K, \gamma]) = \frac{rB}{\gamma - 1}\left(1 - \frac{B}{K}\right)^{\gamma-1}. \tag{36}$$

γ is a parameter which breaks PT out of the restrictive symmetry of the logistic curve. In the special case of $\gamma = 2$ Eq (36) collapses back to the logistic curve, however in general $\gamma \in (1, \infty)$. The parameters r and K maintain the same interpretation as they do in the logistic production function. In Figure (14) PT recruitment is shown for a range of parameter values so as to demonstrate the various recruitment shapes that can be achieved by PT recruitment.

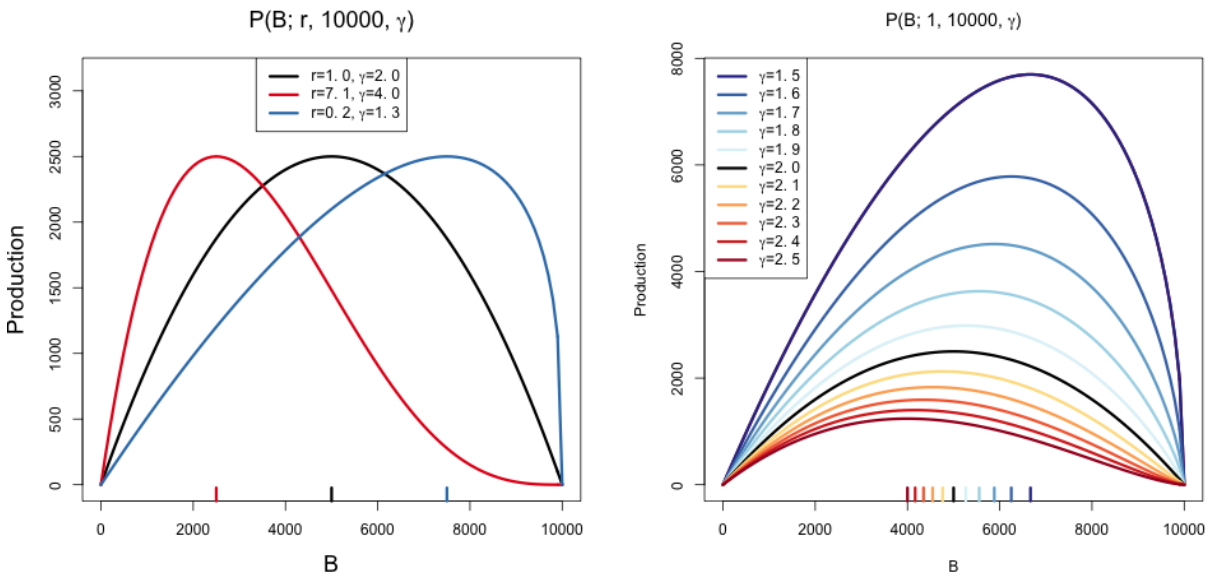


Figure 14: (*left*) PT production functions with parameters chosen so that MSY is consistent, but $\frac{B^*}{B(0)}$ is less than $\frac{1}{2}$ (in red), greater than $\frac{1}{2}$ (in blue), or equal to $\frac{1}{2}$ (in black; logistic production function). (*right*) PT production functions over a range of γ values with the values of r and K fixed at 1 and 10,000 respectively.

While the particular form of how γ appears in PT still produces some limitations to the form of the production function, importantly the introduction of a third parameter allows enough flexibility to fully describe the space of reference points used in management. To see this, the reference points are analytically derived for the PT model below.

6 .2 PT Reference Points

With $B(t)$ representing biomass at time t, under PT production, the dynamics of biomass are defined by the following ODE,

$$\frac{dB}{dt} = \frac{rB}{\gamma-1}\left(1-\frac{B}{K}\right)^{\gamma-1} - FB. \tag{37}$$

An expression for the equilibrium biomass is attained by setting Eq (37) equal to zero, and rearranging the resulting equation to solve for B. Thinking of the result as a function of F gives,

$$\bar{B}(F) = K\left(1-\left(\frac{F(\gamma-1)}{r}\right)^{\frac{1}{(\gamma-1)}}\right). \tag{38}$$

At this point it is convenient to notice that $\bar{B}(0) = K$. The expression for B^* is given by evaluating Eq (38) at F^*.

To get an expression for F^*, the equilibrium yield is maximized with respect to F,

$$F^* = \underset{F}{\operatorname{argmax}}\, F\bar{B}(F). \tag{39}$$

In the case of PT production this maximization can be done analytically (however many three parameter production functions do not result in tractable analytical solutions). In this case maximization can proceed by differentiating the equilibrium yield with respect to F as follows,

$$\frac{d\bar{Y}}{dF} = \bar{B}(F) + F\frac{d\bar{B}}{dF} \tag{40}$$

$$\frac{d\bar{B}}{dF} = -\frac{K}{F(\gamma-1)}\left(\frac{F(\gamma-1)}{r}\right)^{\frac{1}{\gamma-1}}. \tag{41}$$

Setting Eq (40) equal to 0, substituting $\bar{B}(F)$ and $\frac{d\bar{B}}{dF}$ by Equations (38) and (41) respectively, and then solving for F produces the following expression for the fishing rate required to produce MSY,

$$F^* = \frac{r}{\gamma-1}\left(\frac{\gamma-1}{\gamma}\right)^{\gamma-1}. \tag{42}$$

Plugging the above expression for F^* back into Eq (38) gives the following expression for biomass at MSY,

$$B^* = \frac{K}{\gamma}. \tag{43}$$

The above derived expressions for $\bar{B}(0)$, B^*, and F^* can then be used to build a specific analytical form for the biological reference points in terms of only biological model parameters.

$$F^* = \frac{r}{(\gamma-1)}\left(\frac{\gamma-1}{\gamma}\right)^{\gamma-1} \qquad \frac{B^*}{\bar{B}(0)} = \frac{1}{\gamma} \tag{44}$$

6 .3 Simulation Study

Indices of abundance are simulated from the three parameter PT production model over a grid of F^* and $\frac{B^*}{\bar{B}(0)}$ values. These PT data are then fit with a two parameter Schaefer model.

Generating simulated indices of abundance from the PT model requires inverting the relationship between $\left(F^*, \frac{B^*}{\bar{B}(0)}\right)$, and (r, γ). It is not generally possible to analytically invert this relationship for many three parameter production functions (Punt & Cope, 2019; J. T. Schnute & Richards, 1998). Most three parameter production functions lead to RPs that require expensive numerical methods to invert; more over the numerical inversion procedure can often be unstable. That said, for the case of PT this relationship is analytically invertible, and leads to the following relationship

$$r = F^* \left(\frac{1-\frac{B^*}{\bar{B}(0)}}{\frac{B^*}{\bar{B}(0)}}\right)\left(1-\frac{B^*}{\bar{B}(0)}\right)^{\left(\frac{\frac{B^*}{\bar{B}(0)}-1}{\frac{B^*}{\bar{B}(0)}}\right)} \qquad \gamma = \frac{1}{\frac{B^*}{\bar{B}(0)}}. \tag{45}$$

Indices are generated under the following conditions. Data are simulated at each point on the grid $\mathcal{F} \times \mathcal{B}$, with $F^* \in \mathcal{F}$ and $\frac{B^*}{\bar{B}(0)} \in \mathcal{B}$, where $\mathcal{F} = \{0.1, 0.2, ..., 0.7\}$ and $\mathcal{B} = \{0.2, 0.3, ..., 0.6\}$ as seen in Figure (15). These ranges of values for F^* and $\frac{B^*}{\bar{B}(0)}$ are selected to include a wide range of values thought to reflect many commonly assessed fisheries. The red X's in Figure (15) show four simulation locations where the Schaefer model is misspecified to a large degree and will be considered in more detail in Section(7 .1). For each $\left(F^*, \frac{B^*}{\bar{B}(0)}\right)$, the associated pair (r, γ) are computed from Eq (45). Since K does not enter the RP calculation its value is fixed arbitrarily at 10000. The value of q is fixed at a typically small value of 0.0005. σ is fixed at the relatively small value of 0.01 to focus specifically on the behavior of population parameters. These parameters fully specify the PT model for the purposes of generating index data for each $\left(F^*, \frac{B^*}{\bar{B}(0)}\right)$ pair.

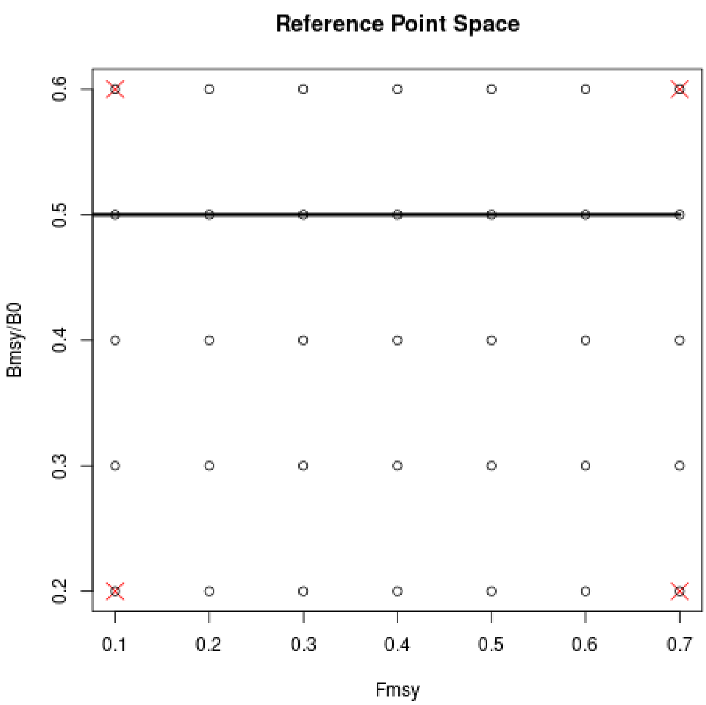


Figure 15:
Open circles show the location of the simulation grid $\mathcal{F} \times \mathcal{B}$. The horizontal line shows the constrained space of RPs for the Schaefer model. The red X's indicated 4 simulation locations where the Schaefer model is particularly misspecified.

6 .4 Catch

It is known that the behavior of catch can effect inference on the biological parameters (Hilborn & Walters, 1992). In particular it is thought that catch can induce "contrast" in index data so as to better inform r. In this setting contrast refers to changes in the long term trends of index data. Figure (19, *right*) demonstrates an example of biomass that includes contrast induced by catch. It is not well understood how contrast may factor into biases induced by model misspecification. To investigate this a variety of catches are investigated.

Catch is parameterized so that $F(t)$ can be controlled with respect to F^*. Recall that catch is assumed to be proportional to biomass with the proportionality constant amounting to the fishing rate, so that $C(t) = F(t)B(t)$. To control $F(t)$ with respect to F^*, $C(t)$ is specified by defining the quantity $\frac{F(t)}{F^*}$ as the relative fishing rate. $B(t)$ is defined by the solution of the ODE, and F^* is defined by the biological parameters of the model, see Eq (42). Thus by defining $\frac{F(t)}{F^*}$, catch can then be written as $C(t) = F^*\left(\frac{F(t)}{F^*}\right)B(t)$.

Intuitively $\frac{F(t)}{F^*}$ describes the fraction of F^* that $F(t)$ is specified to for the current $B(t)$. When $\frac{F(t)}{F^*} = 1$, $F(t)$ will be held at F^*, and the solution of the ODE brings $B(t)$ into equilibrium at B^*. For constant $\frac{F(t)}{F^*}$ the Schaefer model comes to equilibrium as an exponential decay from K approaching B^*. The relative fishing rate is defined on $[0, \infty)$; when $\frac{F(t)}{F^*} < 1$, $F(t)$ is lower than F^* and $B(t)$ is pushed toward $\bar{B} > B^*$. Contrarily, when $\frac{F(t)}{F^*} > 1$, $F(t)$ is higher than F^* and $B(t)$ is pushed toward $\bar{B} < B^*$; the precise values of $\bar{B}$ can be calculated from Eq (38).

In practice, catch is determined by a series of observed, assumed known, catches. Catch observations are typically observed on a quarterly (or yearly) basis, so that the ODE may be discretized via Euler's method with integration step sizes to match the observation frequency of the modeled data. In this case, catch is sampled as would be done in practice however, the simulation can encounter a variate of issues working with the naively discretized ODE. As a result the ODE is integrated implicitly via the Livermore Solver (Radhakrishnan, 1993, lsode), and catch is linearly interpolated between sampled epochs.

6 .5 Model Fitting

The goal of model fitting is to assess how the biological parameters of the two parameter Schaefer model behave under MLE inference when fit to PT data. Thus, let I_t be an observation of PT index data at time $t \in \{1, 2, 3, ..., T\}$. The observation model is log-normal such that,

$$I_t|q, \sigma^2, \boldsymbol{\theta} \sim LN(qB_t(\boldsymbol{\theta}), \sigma^2). \tag{46}$$

For the Schaefer model $\boldsymbol{\theta} = [r, K]$, and $B_t(\boldsymbol{\theta})$ is defined by the solution of the following ODE

$$\frac{dB}{dt} = rB\left(1 - \frac{B}{K}\right) - FB. \tag{47}$$

The I_t are assumed independent conditional on q, σ^2, r, K and the ODE model for biomass. Thus the log likelihood can be written as

$$\log \mathcal{L}(q, \sigma^2, \boldsymbol{\theta}; I) = -\frac{T}{2} \log\big(\sigma^2\big) - \frac{1}{2\sigma^2} \sum_t \log\left(\frac{I_t}{qB_t(\boldsymbol{\theta})}\right)^2. \tag{48}$$

In this setting, q is fixed at the true value of 0.0005 to focus on the inferential effects of model misspecification on biological parameters. σ^2, r, and K are reparameterized into the log scale as $log(\sigma^2)$, $log(r)$, and $log(K)$ and fit via MLE. σ^2 is allowed to be fit to assess overall model fit. Reparameterization of the parameters into the log scale improves the reliability of optimization in addition to facilitating the use of Hessian information for parameter estimate standard errors.

Given that the biological parameters enter the likelihood via a nonlinear ODE, and further the parameters themselves are related to each other nonlinearly, the likelihood function can often be difficult to optimize. A hybrid optimization scheme is used to maximize the log likelihood to ensure that a global MLE solution is found. The R package GA (Scrucca, 2013, 2017) is used to run a genetic algorithm to explore parameter space globally. Optimization occasionally jumps into the L-BFGS-B local optimizer to refine optima within a local mode.

The scheme functions by searching globally to iteratively improve hot starts for the local optimizer.

In Appendix A a profile likelihood method for estimating all of the parameters of the model is derived. The profile likelihood technique greatly improves the reliability of local optimizers when fitting the biological parameters alongside additional nuisance parameters. The catchability parameter q has the effect of rescaling biomass which can often function similarly to the role of the carrying capacity parameter K. Thus, the structure of the likelihood may confound q and K, and for some data these parameters may only be weakly identifiable. Posing the model in a Bayesian context provides a convenient mechanism for managing these weak identifiability issues. In a tactful Bayesian formulation q and σ^2 may then be marginalized out of the joint posterior to yield fast and reliable inference (Walters & Ludwig, 1994).

6 .6 Gaussian Process Metamodel

For assessing biological parameters over the simulated grid, as seen in Figure (15), a GP model is used as a flexible, stochastic interpolator over RP space. As previously established, in Section (6 .5), the biological parameters of interest are the Schaefer model's $log(r)$ and $log(K)$ parameters. Since the estimates of these parameters are random variables, with variances given by the inverse of the observed fisher information, interpolation of MLEs requires paying additional attention to propagating estimates of uncertainty into the metamodel.

A GP is a stochastic process generalizing the normal distribution to an infinite dimensional analog. GPs are often specified primarily through the choice of a covariance function which defines the relationship between locations in an index set. Typically the index set is spatial for GPs, and in this setting the model is across the reference point space, $\left(F^*, \frac{B^*}{B(0)}\right)$, of the three parameter PT data generating model. A GP model implies an n dimensional multivariate normal distribution on the observations of the model and the covariance function fills out the covariance matrix for the observations.

Modeling the estimates of $log(r)$ and $log(K)$ with independent GP models is used to extend analysis of all major biological RP over the simulated grid. Let $\hat{\mu}$ be the maximum likelihood estimate (MLE) of either $log(r)$ or $log(K)$. Additionally let $\hat{\omega}$ be the inverted

Hessian information of the log likelihood evaluated at $\hat{\mu}$.

Each grid location of the simulation produces a single fitted $\hat{\mu}_i$ at an associate $\left(F^*, \frac{B^*}{\bar{B}(0)}\right)$ location with $i \in \{1, ..., n\}$. $\hat{\boldsymbol{\mu}}$ is jointly modeled over the space of reference points as the following GP,

$$\begin{aligned}
\boldsymbol{x} &= \left(F^*, \frac{B^*}{\bar{B}(0)}\right) \\
\hat{\mu} &= \beta_0 + \boldsymbol{\beta}'\boldsymbol{x} + f(\boldsymbol{x}) + \boldsymbol{\epsilon} \\
f(\boldsymbol{x}) &\sim \text{GP}(0, \tau^2 R(\boldsymbol{x}, \boldsymbol{x}')) \\
\epsilon_i &\sim \text{N}(0, \hat{\omega}_i).
\end{aligned} \tag{49}$$

The GP residual variation provides an ideal mechanism for propagating uncertainty from inference in the simulation step into the metamodel. $\hat{\omega}_i$ is the observed residual variation for the inferred value, $\hat{\mu}_i$. This mechanism down weights the influence of each $\hat{\mu}_i$ in proportion to the inferred sampling distribution uncertainty. This has the effect of smoothing the GP model in a way similar to the nugget effect (Gramacy & Lee, 2012).

Here R is the squared exponential correlation function.

$$R(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(\sum_{j=1}^{2} \frac{-(x_j - x_j')^2}{2\ell_j^2}\right) \tag{50}$$

R has an anisotropic separable form to allow for differing length scales in the F^* and $\frac{B^*}{\bar{B}(0)}$ axes. The flexibility to model correlations separately in the different RP axes is key due to the differences in the extent of the RP domains marginally. ℓ_1 and ℓ_2 model the length scales for F^* and $\frac{B^*}{\bar{B}(0)}$ respectively. The metamodel parameters β_0, $\boldsymbol{\beta}$, τ^2, ℓ_1 and ℓ_2 are fit via MLE against the observations of $\hat{\mu}$ and $\hat{\omega}$ from simulation fits.

Predictive estimates of modeled quantities are obtained via kriging over intermediate values over RP space. Let $\check{}$ decorate any quantity that is derived for metamodel interpolation.

$$\check{\mu}(\check{s}) = \beta_0 + \boldsymbol{x}(\check{s})\boldsymbol{\beta} + R_{\boldsymbol{\ell}}(\check{s}, s)R_{\boldsymbol{\ell}}^{-1}(s, s)\Big(\hat{\mu}(s) - \big(\beta_0 + \boldsymbol{x}(s)\boldsymbol{\beta}\big)\Big) \tag{51}$$

7 Results

- chaotic regions
- best hits from Schaffer
- Schnute
- arrow plot
- RP specific plots

While interpolation occurs in the space of either $\log(r)$ or $\log(K)$, these interpolated values are used to build interpolated estimates of major biological reference points. Using the interpolated values $\check{\log(r)}$ and $\check{\log(K)}$ the following transformation are applied to interpolate RP quantities under the Schaefer model,

$$\check{B}^* = \frac{\check{K}}{2} \qquad \check{F}^* = \frac{\check{r}}{2}. \tag{52}$$

Using these interpolated RP quantities, the bias induced by model misspecification is quantified by the following relative measure of bias, similar to a percent error calculation.

$$\text{Relative Bias} = \frac{\check{RP} - RP}{RP} \tag{53}$$

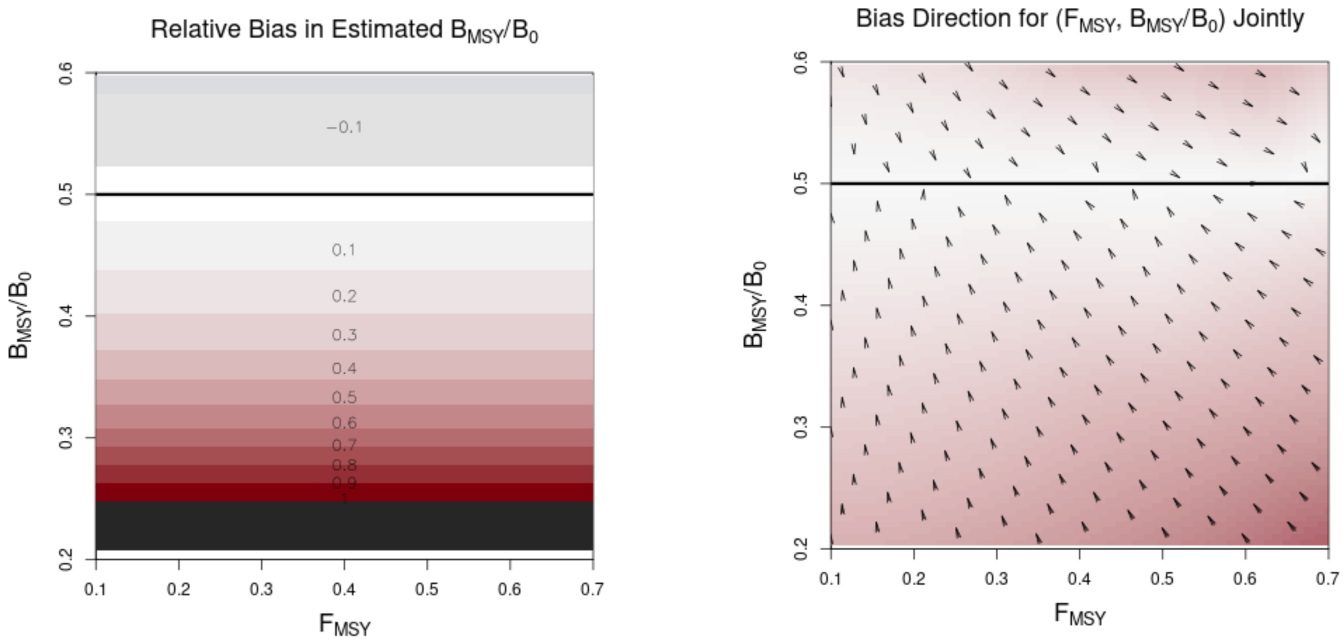
Above RP is a stand in for the the true value of any of the biological reference points under PT data generation, and $\check{RP}$ refers to the interpolated estimated RP quantity under the Schaefer model.

7 .1 An MSY-Optimal Catch History

When $F(t)$ is held constant at F^*, $B(t)$ comes to equilibrium as an exponential decay from K to B^*. Understanding model misspecification bias is simplified in this setting due to the relative simplicity of $B(t)$. However this simplicity is known to poorly inform estimates of r, and thus F^*, due to the limited range of the production function that is observed (Hilborn & Walters, 1992). This example is a "low contrast" setting.

Figure (16) shows the biases in F^* and $\frac{B^*}{B(0)}$ over the space of simulated RPs. The (*top-right*) panel of Figure (16) shows how data generated across a broad space of RPs are mapped onto the limited space of the Schaefer line. Below the Schaefer line, RP estimates are biased by over-estimating $\frac{B^*}{B(0)}$ and under-estimating F^*. Above the Schaefer line the vice-versa is true; $\frac{B^*}{B(0)}$ is under-estimated and F^* is over-estimated. In the (*left*) and (*bottom*) panels of Figure (16) the bias in $\frac{B^*}{B(0)}$ and F^* are shown component-wise; each panel showing the same patterns, but focusing on only one component of the bias at a time. In these panels red coloring indicates over-estimation of the RP and blue indicates under-estimation. Notice that the region of RPs near the Schaefer line enjoy relatively low bias since model misspecification is minor in this region.

Notice that under the Schaefer model B^* is necessarily half of K. Since $\frac{B^*}{B(0)}$ is always $\frac{1}{2}$ under the Schaefer model, the bias in $\frac{B^*}{B(0)}$ (as seen in Figure (16)) simply measures the distance from the data generating location vertically to the Schaefer line.



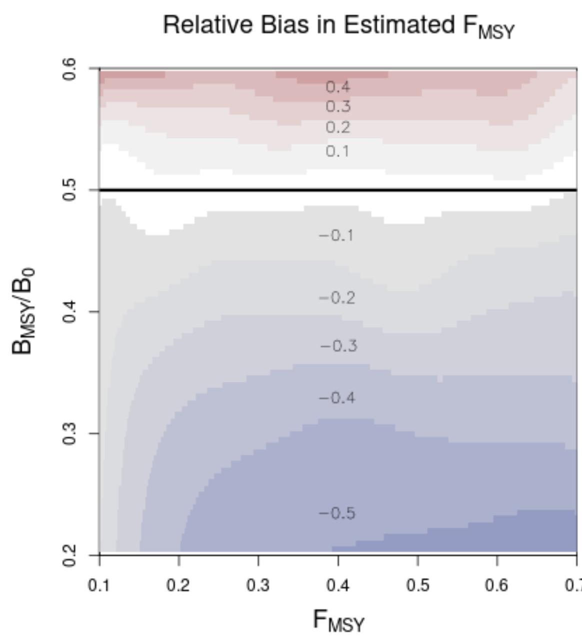


Figure 16: Heatplots showing the bias in RP estimation induced by model misspecification. In all cases the restricted RP-space of the Schaefer model is shown as a horizontal black line at $\frac{B^*}{B(0)} = 0.5$. (*left*) Relative bias in $\frac{B^*}{B(0)}$. (*top-right*) Bias in RP-space shown directionally. Arrows point from the location where data is generated, toward the location in on the Schaefer line where MLE projects. The intensity of color shows the absolute error as a distance in RP-space. (*bottom*) Relative bias in F^*.

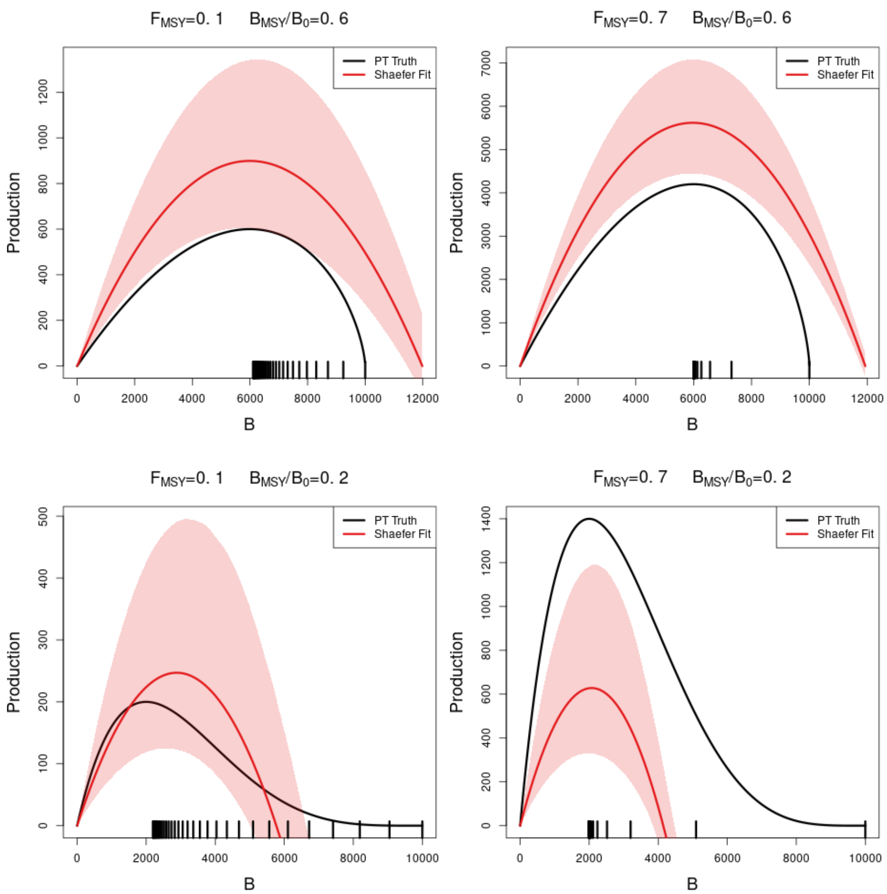


Figure 17: A comparison of the true PT production function (in black) and the estimated logistic curve (in red) with 95% CI shown. The examples shown represent the four corners of maximum model misspecification in the simulated RP-space. Observed biomasses are plotted in the rug plots below the curves.

Figure (17) shows four of the most misspecified example production function fits as compared to the true data generating PT production functions. In the rug plots below each set of curves the observed biomasses demonstrate the exponential decay from K to B^* in each case. In particular, notice how only biomasses greater than the PT B^* are observed. Due to the leaning of the true PT curves, and the symmetry of the logistic parabola, the logistic curve only observes information about its slope at the origin from data observed on the right portion of the PT curves. Above the Schaefer line PT is steeper on the right of B^* than it is on the left, and so the the logistic curve over-estimates r, and thus F^*, for data generated above the Schaefer line. Below the Schaefer line the vice versa phenomena occurs. Below the Schaefer line PT is shallower to the right of B^* than it is on the left and so the logistic parabola estimate tends to under estimate F^*.

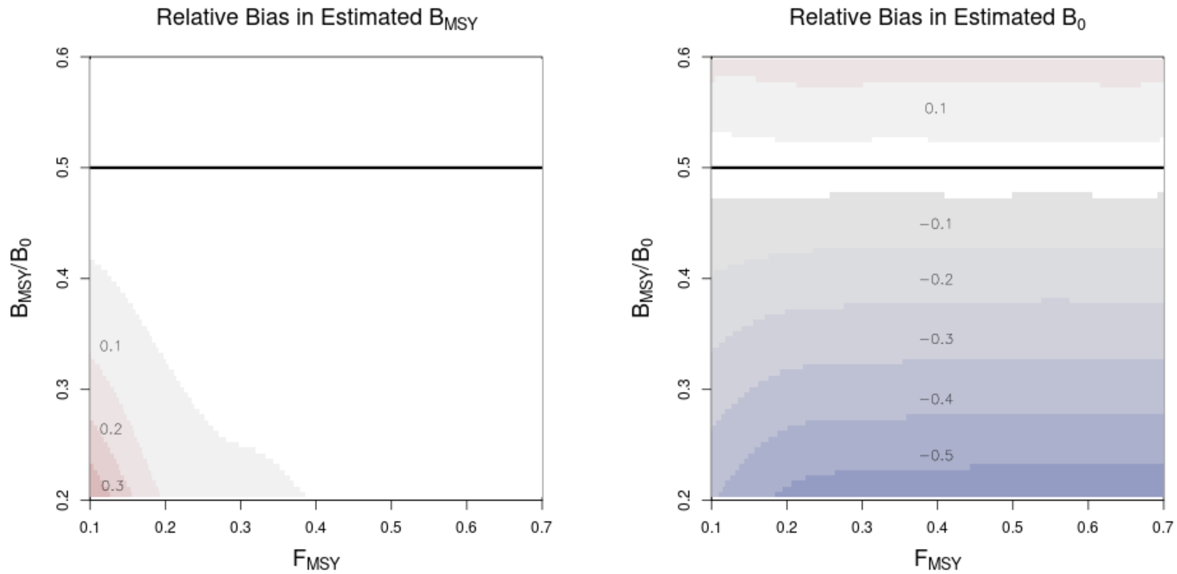


Figure 18: MLE Bias surfaces for B^* (*left*) and K (*right*) individually.

Figure (17) also gives some examples of the relative behavior of B^* and K. In Figure (16) it is clear that the bias behavior of $\frac{B^*}{\bar{B}(0)}$ is locked in a fixed pattern under the Schaefer model. Figure (17) indicates that the individual biases of B^* and K may behave quite differently. B^* appears to be estimated fairly accurately while K does not.

Figure (18) teases apart $\frac{B^*}{\bar{B}(0)}$ into individual bias surfaces for B^* and K respectively. Interestingly B^* enjoys a large region of RP-space with relatively low bias. Given that B^* has relatively consistently low bias, K maintains the expected inverse relationship with $\frac{B^*}{\bar{B}(0)}$ bias. Since the parabolic structure of the logistic function ties the ratio of B^* and $\bar{B}(0)$ to $\frac{1}{2}$, their is only one degree of freedom shared between B^* and $\bar{B}(0)$ so that their ratio is maintained at $\frac{1}{2}$. In this setting it appears that B^* estimation is largely conserved at the cost of K.

7 .2 More Informative Catch Histories

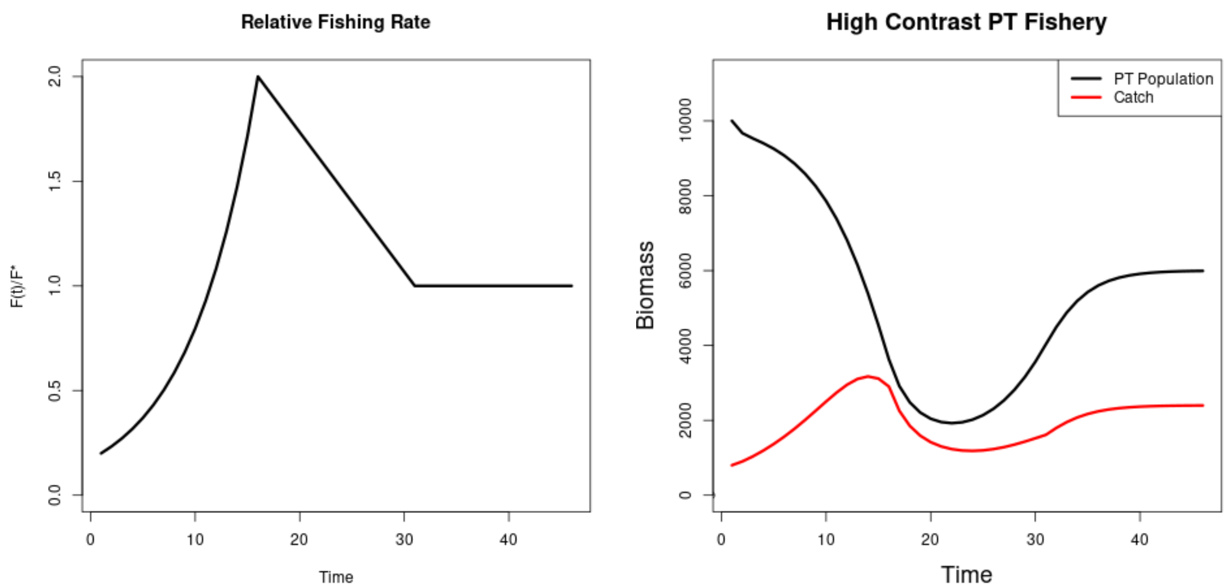


Figure 19: $(left)$ Relative fishing specified so as to induce contrast. $(right)$ Population biomass and catch demonstrating contrast in a PT population with $F^* = 0.4$ and $\frac{B^*}{B(0)} = 0.6$.

The setting of constant relative fishing rate is a useful simplification for building understanding of the dynamics that induce bias, but in practice constant fishing rate is a somewhat oversimplified setting. Consider a hypothetical stock where fishing rate accelerates as technology and fishing techniques improve rapidly until management practices are applied. Figure (19) demonstrates this more realistic, while still idyllic, fishing behavior. This population is exposed to a variety of fishing rates, which induce contrast in the generated indices and allows the fitting model to observe a decrease in the population followed by a rebuild of the stock. This represents a "high contrast" setting that is widely thought to better inform growth rate parameters, such as r.

Figure (20) shows the relative bias surfaces for B^* and F^* under 45 epochs of data in the high contrast setting. On the one hand, notice the relative lack of bias in F^* over a large swath of RPs far from the Schaefer line. On the other hand, notice that bias in B^* increases here relative to the low contrast setting. The pattern of bias in B^* maintains a similar pattern, and overall scale, as the low contrast setting seen in Figure (18), however

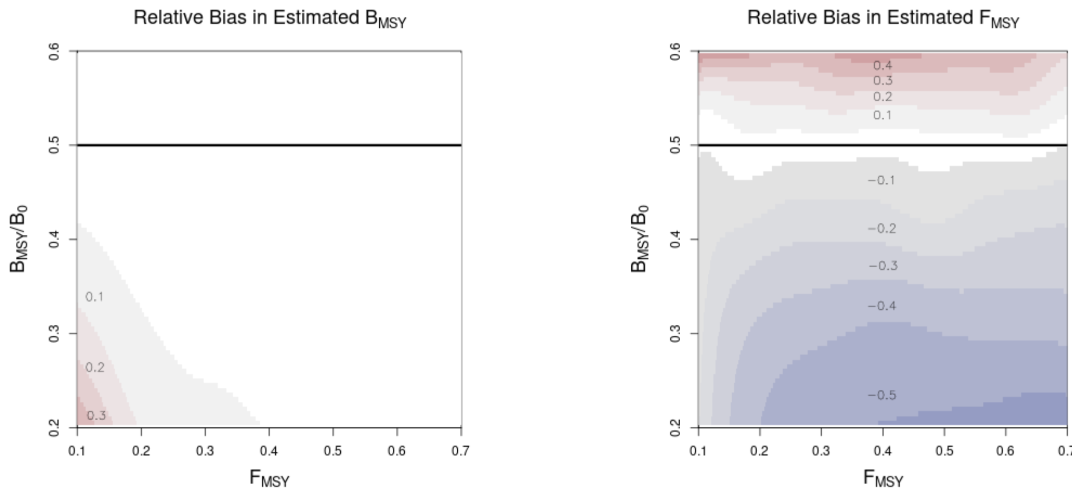


Figure 20: MLE Bias surfaces for B^* (*left*) and F^* (*right*) with relative fishing rate as specified in Figure (19)

a smaller region of RP-space enjoys low bias here. Due to the expanded pattern of B^* bias here, as compared with the low contrast setting, and the constrained relationship with $\frac{B^*}{B(0)}$, the bias surface for K maintains the same general inverse relationship with $\frac{B^*}{B(0)}$.

If the data are augmented so that the fishing rate is held at F^* for an additional 45 time epochs (90 epochs total), so that slower growing stocks may observe more data near B^*, Figure (21) shows the updated bias surfaces. The scale of bias in B^* is reduced, but the general patterns of bias remains similar for both RPs. While the bias behavior of B^* estimates are diminished, F^* biases are generally magnified.

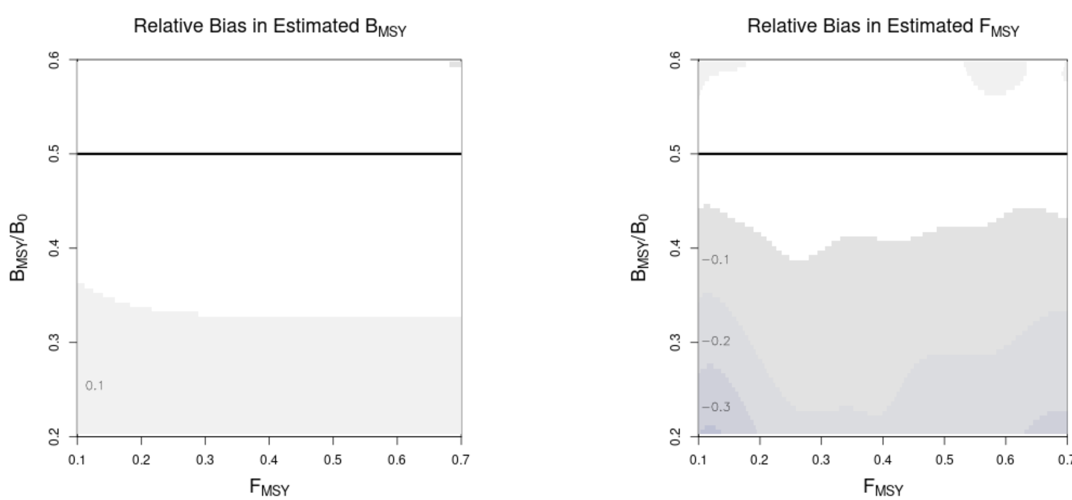


Figure 21: MLE Bias surfaces for B^* (*left*) and F^* (*right*) with relative fishing rate augmented with additional observations near equilibrium.

8 Discussion

Results presented here generally agree with what is known about estimating growth rate parameters (Lee et al., 2012; Conn et al., 2010; Magnusson & Hilborn, 2007), in this case r, and thus F^*. In the presence of contrast F^* estimation can enjoy very low bias even for a wide range of poorly specified models; conversely in the absence of contrast F^* estimation can suffer very large bias even for slightly misspecified models. In all cases when model misspecification is removed, even with weakly informative data, F^* estimation is unbiased. Model misspecification is thus a necessary but not sufficient condition for inducing bias.

While it is established that growth rate parameters require contrast to estimate, the implications of these biases jointly across a variety of RPs have not received as much attention. When considering B^* alongside F^* in varying contrast environments, it becomes clear that different data informs different parts of the production function differently. In low contrast environments B^* is estimation is remarkably unbiased across all but the most challenging instances of model misspecification. However in the presence of contrast, while F^* enjoys better estimation, B^* estimation experiences substantial bias for only modestly misspecified models. Further, by augmenting contrasting data with an additional period of low contrast data this pattern begins to reverse with B^* bias receding toward more poorly specified models and F^* bias encroaching toward only modestly misspecified models as seen in Figure (21).

The behavior of bias in estimating B^* and F^* suggests that the limited parameter space of the Schaefer model induces a trade off in estimating these parameters. In practice, when the true model is not known and the Schaefer model is unlikely to be correctly specified, one should at best expect to only estimate either B^* or F^* correctly depending on the particular degree of model misspecification. The observed contrast then serves to distribute the available information among B^* and F^*. Increasing the flexibility of the production function by moving toward curves with additional parameters could release these structural limitations (Mangel et al., 2013). Punt and Cope (2019) considers a suite of possible three parameter curves which could be used instead of current two parameter curves.

This study only explores the compatibility of the possible productivity shapes exhibited by the PT and Schaefer models. While the PT and Schaefer models are instructive for a variety of dome shaped production behaviors, it is possible that under different modeling assumptions, for example BH production or age structured models, different bias patterns will emerge. Extending this work to be able to make claims in those settings is necessary for developing more generally extensible claims.

Given the role that catch plays in understanding where the production function is informed, it is clear that good estimates of catch are important for contextualizing modeling inferences. While the production model treats catches as known without uncertainty, upon inspection of Figure (12, *right*) this assumption is clearly suspect. Results presented here only consider very deterministic catch histories. More work is needed to understand how jittery catch may affect RP estimation. A smoothing model of catch may be preferable for estimation, but results of this study suggest that even improvements to the contextual understanding of catch will be important for interpreting model inferences correctly.

A Proposal: Productivity & Growth Extensions

The Deriso production function presents a convenient three parameter form that is capable of representing many of the most common two parameter production functions as special cases (Deriso, 1980). The BH and Logistic production functions arise when γ is fixed to -1 or 1 respectively, and the Ricker model is a limiting case as $\gamma \to 0$ (J. Schnute, 1985).

$$\frac{dB}{dt} = P(B;\theta) - (M+F)B$$

$$P(B;[\alpha,\beta,\gamma]) = \alpha B(1-\beta\gamma B)^{\frac{1}{\gamma}}$$

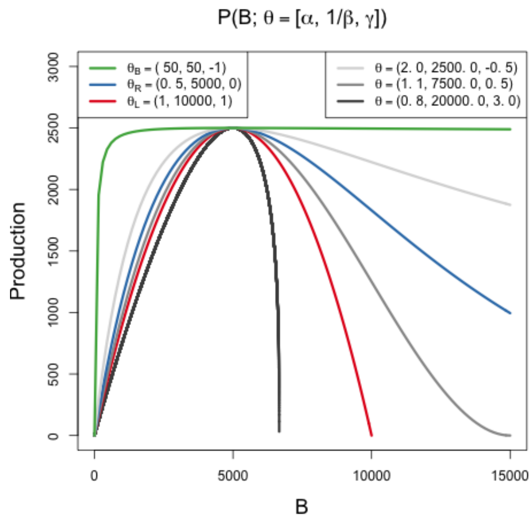


Figure 22: The Deriso production function plotted across a variety of parameter values. The special cases of BH, Ricker, and Logistic production are shown in green, blue, and red respectively.

Using the Deriso model as a data generating model across a wide range of RP-space, similarly as described in Section (6 .3), presents an ideal setting for extending the above study of RP biases across a broad range of productivity assumptions. Under the Deriso model inverting the relationship between RPs and model parameters is not fully analytically possible (J. T. Schnute & Richards, 1998). Numerical inversion of the nonlinear system seen in Eq (54) is required for determining parameter values for data generation. Notice for a given γ value, α and β can be solved analytically.

$$\begin{aligned}
\frac{B^*}{\bar{B}(0)} &= \frac{\left(\frac{\alpha}{M+F^*}\right)^{\frac{1}{\gamma}} - 1}{\left(\frac{\alpha}{M}\right)^{\frac{1}{\gamma}} - 1} \\
\alpha &= (M+F^*)\left[1 - \frac{1}{\gamma}\left(\frac{F^*}{M+F^*}\right)\right]^{-\gamma} \\
\beta &= \frac{1}{\gamma\bar{B}(0)}\left(1 - \left(\frac{\alpha}{M}\right)^{\frac{1}{\gamma}}\right)
\end{aligned} \tag{54}$$

$$P_{\mathrm{BH}}(B; [\alpha, \beta, -1]) = \frac{\alpha B}{(1 + \beta B)}$$

$$\frac{B^*}{\bar{B}(0)} = \frac{1}{\frac{F^*}{M} + 2}$$

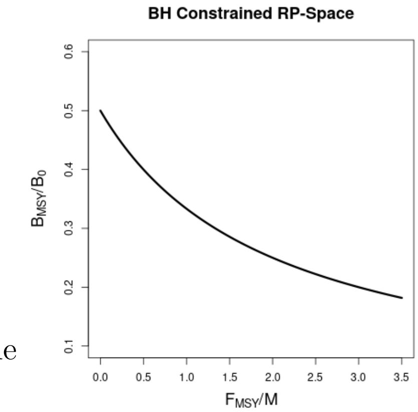


Figure 23: The restricted RP-space under the BH production function.

Inference under the BH model is of particular interest due to its overwhelming popularity in stock assessment. Similar to the limited RP-space of the Schaefer model the two parameter BH model also has a limited RP-space as shown in Figure (23). While the BH constrained RP space is more complicated than the Schaefer model, analogy to the results obtained under the PT-Schaefer simulation setting, and the flexible GP metamodel, should expedite the analysis of BH inference.

Individual Growth

Models that include individual growth and maturity dynamics are another important practical setting for extending the understanding of how productivity model misspecification can bias RPs estimation. The Deriso-Schnute delay-difference (DD) model provides a compact representation of simple age-structured dynamics (Deriso, 1980; J. Schnute, 1985, 1987). While various modeling strategies may be considered for including effects of age-structure in the population, the Deriso-Schnute DD model presents an ideal model for the simulation setting presented here. The compact representation of the Deriso-Schnute DD model via delay-differential equations accounts for the effects of individual growth and maturity while maintaining relatively fast computation.

The DD model is derived directly from an assumption of Von Bertalanffy growth (Von Bertalanffy, 1938) in weight, as seen in Eq (58). In this setting Von Bertalanffy growth relates individual age to individual weight by assuming linear instantaneous growth (as parameterized by the growth parameters κ and w_∞). The DD model expands the idea of biomass production into the processes of recruitment, individual growth, and maturity. This formulation separates the number of individuals in the population (N) from the biomass of the population (B). The dynamics of N, as seen in Eq (56), are very similar to that of the Deriso production model presented above, however the role of the production function is now filled by a "recruitment" function which describes how new individuals are added to the numbers equation. The B dynamics, can then be seen to describe biomass by an account of 1) biomass of new recruits, 2) the net growth of existing biomass, and 3) biomass lost due to mortality. The model accounts for maturity as knife-edge maturity at the instant an individual reaches age a_0.

$$\frac{dB}{dt} = \overbrace{w(a_0)R(B;\theta)}^{\text{Recruitment Biomass}} + \overbrace{\kappa\left[w_\infty N - B\right]}^{\text{Net Growth}} - \overbrace{(M+F)B}^{\text{Mortality}} \tag{55}$$

$$\frac{dN}{dt} = R(B;\theta) - (M+F)N \tag{56}$$

$$R(B;[\alpha,\beta,\gamma]) = \alpha B(t-a_0)(1-\beta\gamma B(t-a_0))^{\frac{1}{\gamma}} \tag{57}$$

$$w(a) = w_\infty(1-e^{-\kappa a}) \tag{58}$$

For the purpose of inference the parameters κ, w_∞, a_0 and M are typically fixed at values determined by the population of study. Thus the primary inferential goal of the DD model is again focused on learning the recruitment parameters. Using the Deriso-Schnute DD model as a data generating model across a wide range of RP-space, and fitting those data under a BH restriction of the DD model, further extends the simulation study of RP bias to include the effects of individual growth.

Summary

My dissertation will include extensions of the metamodeling analysis of RP bias, as presented in Sections (5 - 8), in the context of the Deriso-BH simulation setting. Additionally my dissertation will include further extensions into the analysis of RP bias under model misspecification of the DD model's recruitment function. The Deriso production/recruitment function presents numerical challenges that require careful numerics and a different handling of the simulation design, but allows for extensions into all of the most widely used models of productivity. Furthermore, the DD model allows for an efficient extension of results into to context of maturity and growth dynamics.

References

Beverton, R. J., & Holt, S. J. (1957). *On the dynamics of exploited fish populations* (Vol. 11). Springer Science & Business Media.

Conn, P. B., Williams, E. H., & Shertzer, K. W. (2010). When can we reliably estimate the productivity of fish stocks? *Canadian Journal of Fisheries and Aquatic Sciences*, *67*(3), 511–523.

Deriso, R. B. (1980, February). Harvesting Strategies and Parameter Estimation for an Age-Structured Model. *Canadian Journal of Fisheries and Aquatic Sciences*, *37*(2), 268–282. Retrieved 2020-05-13, from https://www.nrcresearchpress.com/doi/abs/10.1139/f80-034 doi: 10.1139/f80-034

DeYoreo, M. (2012). *Integrating catchability out of the likelihood.*

Fox Jr., W. W. (1970). An Exponential Surplus-Yield Model for Optimizing Exploited Fish Populations. *Transactions of the American Fisheries Society*, *99*(1), 80–88. Retrieved 2022-02-17, from https://onlinelibrary.wiley.com/doi/abs/10.1577/1548-8659%281970%2999%3C80%3AAESMFO%3E2.0.CO%3B2 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1577/1548-8659%281970%2999%3C80%3AAESMFO%3E2.0.CO%3B2) doi: 10.1577/1548-8659(1970)99<80:AESMFO>2.0.CO;2

Gramacy, R. B. (2020). *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences.* Chapman and Hall/CRC.

Gramacy, R. B., & Lee, H. K. (2012). Cases for the nugget in modeling computer experiments. *Statistics and Computing*, *22*(3), 713–722. (Publisher: Springer)

Hilborn, R. (2010). Pretty good yield and exploited fishes. *Marine Policy*, *34*(1), 193–196. (Publisher: Elsevier)

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

Hilborn, R., & Walters, C. J. (1992). Quantitative Fisheries, Stock Assessment: Choice Dynamics, and Uncertainty Chapman and Hall. *New York*.

Kingsland, S. (1982). The refractory model: the logistic curve and the history of population ecology. *The Quarterly Review of Biology*, *57*(1), 29–52. (Publisher: Stony Brook Foundation, Inc.)

Lee, H.-H., Maunder, M. N., Piner, K. R., & Methot, R. D. (2012, August). Can steepness of the stock–recruitment relationship be estimated in fishery stock assessment models? *Fisheries Research*, *125-126*, 254–261. Retrieved 2022-01-29, from https://linkinghub.elsevier.com/retrieve/pii/S0165783612001099 doi: 10.1016/j.fishres.2012.03.001

Magnusson, A., & Hilborn, R. (2007). What makes fisheries data informative? *Fish and Fisheries*, *8*(4), 337–358. (Publisher: Wiley Online Library)

Mangel, M. (2006). The Theoretical Biologist's Toolbox: Quantitative Methods for Ecology and Evolutionary Biology..

Mangel, M., MacCall, A. D., Brodziak, J., Dick, E., Forrest, R. E., Pourzand, R., & Ralston, S. (2013, April). A perspective on steepness, reference points, and stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*(6), 930–940. Retrieved 2019-07-03, from https://www.nrcresearchpress.com/doi/10.1139/cjfas-2012-0372 doi: 10.1139/cjfas-2012-0372

Maunder, M. N. (2003). Is it time to discard the Schaefer model from the stock assessment scientist's toolbox? *Fisheries Research*, *61*(1-3), 145–149.

Pearson, D. E., & Erwin, B. (1997). Documentation of California's commercial market sampling data entry and expansion programs.

Pella, J. J., & Tomlinson, P. K. (1969). A generalized stock production model. *Inter-American Tropical Tuna Commission Bulletin*, *13*(3), 416–497.

Prager, M. H. (2002). Comparison of logistic and generalized surplus-production models applied to swordfish, Xiphias gladius, in the north Atlantic Ocean. *Fisheries Research*, *58*(1), 41–57. (Publisher: Elsevier)

Prager, M. H. (2003, March). Reply to the Letter to the Editor by Maunder. *Fisheries Research*, *61*(1), 151–154. Retrieved 2022-01-30, from https://www.sciencedirect.com/science/article/pii/S0165783602002746 doi: 10.1016/S0165-7836(02)00274-6

Punt, A. E., Butterworth, D. S., Moor, C. L. d., Oliveira, J. A. A. D., & Haddon, M. (2016).

Management strategy evaluation: best practices. *Fish and Fisheries*, *17*(2), 303–334. Retrieved 2018-12-13, from https://onlinelibrary.wiley.com/doi/abs/10.1111/faf.12104 doi: 10.1111/faf.12104

Punt, A. E., & Cope, J. M. (2019, September). Extending integrated stock assessment models to use non-depensatory three-parameter stock-recruitment relationships. *Fisheries Research*, *217*, 46–57. Retrieved 2019-07-19, from http://www.sciencedirect.com/science/article/pii/S0165783617301819 doi: 10.1016/j.fishres.2017.07.007

Radhakrishnan, K. (1993). Description and Use of LSODE, the Livermore Solver for Ordinary Differential Equations. , 124.

Ramasubramanian, K., & Singh, A. (2017). *Machine learning using R* (No. 1). Springer.

Rankin, P. S., & Lemos, R. T. (2015, October). An alternative surplus production model. *Ecological Modelling*, *313*, 109–126. Retrieved 2022-02-11, from https://www.sciencedirect.com/science/article/pii/S0304380015002732 doi: 10.1016/j.ecolmodel.2015.06.024

Ricker, W. E. (1954). Stock and recruitment. *Journal of the Fisheries Board of Canada*, *11*(5), 559–623. (Publisher: NRC Research Press Ottawa, Canada)

Schnute, J. (1985, March). A General Theory for Analysis of Catch and Effort Data. *Canadian Journal of Fisheries and Aquatic Sciences*, *42*(3), 414–429. Retrieved 2020-05-13, from https://www.nrcresearchpress.com/doi/abs/10.1139/f85-057 doi: 10.1139/f85-057

Schnute, J. (1987). A general fishery model for a size-structured fish population. *Canadian Journal of Fisheries and Aquatic Sciences*, *44*(5), 924–940. (Publisher: NRC Research Press Ottawa, Canada)

Schnute, J. T., & Richards, L. J. (1998, February). Analytical models for fishery reference points. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*(2), 515–528. Retrieved 2020-01-14, from https://www.nrcresearchpress.com/doi/abs/10.1139/f97-212 doi: 10.1139/f97-212

Scrucca, L. (2013, April). GA: A Package for Genetic Algorithms in R. *Journal of Statistical Software*, *53*, 1–37. Retrieved 2022-01-17, from https://doi.org/10.18637/jss.v053.i04 doi: 10.18637/jss.v053.i04

Scrucca, L. (2017). On Some Extensions to GA Package: Hybrid Optimisation, Parallelisation and Islands EvolutionOn some extensions to GA package: hybrid optimisation, parallelisation and islands evolution. *The R Journal*, *9*(1), 187–206. Retrieved 2022-01-17, from https://journal.r-project.org/archive/2017/RJ-2017-008/index.html

Von Bertalanffy, L. (1938). A quantitative theory of organic growth (inquiries on growth laws. II). *Human biology*, *10*(2), 181–213. (Publisher: JSTOR)

Walters, C., & Ludwig, D. (1994). Calculation of Bayes posterior probability distributions for key population parameters. *Canadian Journal of Fisheries and Aquatic Sciences*, *51*(3), 713–722. (Publisher: NRC Research Press Ottawa, Canada)

Appendix A: Profile Likelihood MLE

Given that q has the effect of rescaling the mean function, a naive handling of q has the potential to interfere with the inference on $\boldsymbol{\theta}$. While the parameter q is typically identifiable, it can introduce lesser modes which complicate naive inference.

Below I outline a profile likelihood method for MLE inference on q and σ^2. However if posed in a Bayesian context, q and σ^2 may be marginalized out of the joint posterior to yield a direct sampling scheme for q and σ^2 which factors the posterior into the form $p(q,\sigma^2,\boldsymbol{\theta}|I) = N(\log(q)|\sigma^2,\boldsymbol{\theta},I)IG(\sigma^2|\boldsymbol{\theta},I)p(\boldsymbol{\theta}|I)$ (Walters & Ludwig, 1994; DeYoreo, 2012)

The joint likelihood on the log scale can be written as,

$$\log\mathcal{L}(q,\sigma^2,\boldsymbol{\theta};I) = -\frac{T}{2}\log\big(\sigma^2\big) - \frac{1}{2\sigma^2}\sum_t \log\left(\frac{I_t}{qB_t(\boldsymbol{\theta})}\right)^2. \tag{59}$$

First Eq (59) is maximized with respect to q by partial differentiation of Eq (59) with respect to q,

$$\frac{\partial \log\mathcal{L}}{\partial q} = -\frac{1}{q\sigma^2}\left(\sum_t \log\left(\frac{I_t}{B_t(\boldsymbol{\theta})}\right) - T\log(q)\right) \tag{60}$$

The maximum of the likelihood in the q direction is attained when $\frac{\partial \log\mathcal{L}}{\partial q} = 0$. By setting $\frac{\partial \log\mathcal{L}}{\partial q}$ to 0 and solving for q, the MLE of q in terms of $\boldsymbol{\theta}$ can be written as

$$q(\boldsymbol{\theta}) = e^{\frac{1}{T}\sum_t \log\left(\frac{I_t}{B_t(\boldsymbol{\theta})}\right)} = \left(\prod_t \frac{I_t}{B_t(\boldsymbol{\theta})}\right)^{\frac{1}{T}}. \tag{61}$$

Notice that $\hat{q}(\boldsymbol{\theta})$ is the geometric mean of the empirical scaling factors between the observed index and modeled biomass at each time. This form is emblematic of the interpretation of the q parameter as the proportionality constant between the observed index and the modeled biomass. Additionally notice that $\hat{q}$ is a function of $\boldsymbol{\theta}$, so that achieving the global maximum of the likelihood function still requires maximization over $\boldsymbol{\theta}$. Furthermore, $\hat{q}(\boldsymbol{\theta})$ is only a function of $\boldsymbol{\theta}$ and that σ^2 does not enter the expression. This will be helpful in further maximization of the likelihood with respect to σ^2.

Now to maximize in the σ^2 direction Eq (59) is differentiated with respect to σ^2,

$$\frac{\partial \log \mathcal{L}}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_t \log\left(\frac{I_t}{qB_t(\boldsymbol{\theta})}\right)^2. \tag{62}$$

The maximum of the likelihood in the σ^2 direction is attained when $\frac{\partial \log \mathcal{L}}{\partial \sigma^2} = 0$. Setting $\frac{\partial \log \mathcal{L}}{\partial \sigma^2}$ to 0 and solving for σ^2 produces the following MLE as a function of $\boldsymbol{\theta}$,

$$\sigma^2(\boldsymbol{\theta}) = \frac{1}{T} \sum_t \log\left(\frac{I_t}{q(\boldsymbol{\theta})B_t(\boldsymbol{\theta})}\right)^2 \tag{63}$$

Notice that the conditionally MLE of σ^2 is not only a function of $\boldsymbol{\theta}$ but also a function of q. As previously noted, $q(\boldsymbol{\theta})$ is only a function of $\boldsymbol{\theta}$, and so to achieve a global maximum of the joint likelihood, $\sigma^2(\boldsymbol{\theta})$ is written entirely in terms of $\boldsymbol{\theta}$ by replacing q by $q(\boldsymbol{\theta})$ as seen above.

By combining Eq (61) and Eq (63) the MLEs of q and σ^2 can be written entirely in terms of $\boldsymbol{\theta}$. Furthermore, this realization allows the joint maximization of the likelihood to be reduced to the following profile log-likelihood,

$$\log \mathcal{L}(\boldsymbol{\theta}; I) = -\frac{T}{2} \log\big(\sigma^2(\boldsymbol{\theta})\big) - \frac{1}{2\sigma^2(\boldsymbol{\theta})} \sum_t \log\left(\frac{I_t}{q(\boldsymbol{\theta})B_t(\boldsymbol{\theta})}\right)^2. \tag{64}$$

This profile log-likelihood is maximized numerically over $\boldsymbol{\theta}$, and the estimates for q and σ^2 are given by evaluating Equations (61) and (63) at $\hat{\boldsymbol{\theta}}$.

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \log \mathcal{L}(\boldsymbol{\theta}; I) \tag{65}$$

$$\hat{\sigma}^2 = \sigma^2(\hat{\boldsymbol{\theta}}) \tag{66}$$

$$\hat{q} = q(\hat{\boldsymbol{\theta}}) \tag{67}$$

This profile formulation via $\hat{q}(\boldsymbol{\theta})$ and $\hat{\sigma}^2(\boldsymbol{\theta})$ reduces the computational complexity of this numerical optimization, while also avoiding the multimodality issues induced by q.