

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

**A METAMODELING APPROACH FOR BIAS ESTIMATION OF
BIOLOGICAL REFERENCE POINTS**

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

DOCTOR OF PHILOSOPHY

in

STATISTICAL SCIENCE

by

Nicholas Grunloh

June 2024

The Dissertation of Nicholas Grunloh
is approved:

Professor Herbert Lee, Chair

Edward J. Dick Ph.D.

Professor Paul Parker

Professor Bruno Sanso

Dean Peter F. Biehl
Vice Provost and Dean of Graduate Studies

Copyright © by
Nicholas Grunloh
2024

Table of Contents

List of Figures	v
Abstract	vii
Dedication	viii
Acknowledgments	ix
1 Introduction	1
2 Pella-Tomlinson Model	6
2.1 Introduction	7
2.2 Methods	7
2.2.1 Model	7
2.2.2 Reference Points	8
2.2.3 Simulation	9
2.2.4 Design	10
2.2.5 Gaussian Process Metamodel	10
2.2.6 Catch	13
2.2.7 Two-Parameter Production Model Inference	15
2.2.8 Continuous model formulation	16
2.3 Results	19
2.3.1 An <i>MSY</i> -Optimal Catch History	19
2.4 Discussion	22
3 Schnute Model	26
3.1 Introduction	27
3.2 Methods	27
3.2.1 Model	27
3.2.2 Simulation	29
3.3 Discussion	29
4 Conclusion	33

A Inverting $\frac{B^*}{B(0)}$ and γ for the PT Model	34
Bibliography	36

List of Figures

1.1	<i>left</i> : An index of abundance data, catch per unit effort (CPUE), for Namibian Hake from 1965 to 1987. <i>right</i> : The associated catch data for Namibian Hake over the same time period.	3
2.1	The Pella-Tomlinson 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 regimes are shown in blue and red respectively. .	7
2.2	(<i>left</i>) Relative fishing with low, medium, and high contrast. (<i>right</i>) Population biomass and catch at each associated level of contrast.	15
2.3	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.	19

2.4	Joint bias direction for $(F^*, \frac{B^*}{B_0})$ estimates under the misspecified Schaefer Model. The intensity of color represents the excess bias relative to the shortest possible mapping. Results in the low contrast setting are shown <i>left</i> , and the high contrast setting is shown <i>right</i>	21
2.5	Bias in F^* under the Schaefer model when PT data are generated with increasing contrast so that F^* and $\frac{B^*}{B_0}$ are fixed at 0.699 and 0.201 respectively.	22
3.1	The Schnute production function plotted across a variety of parameter values. Regimes of similarly behaving curves are grouped by color.	27

Abstract

A Metamodeling Approach for Bias estimation of Biological Reference Points

by

Nicholas Grunloh

Stock assessments often assume a two-parameter functional form (e.g., Beverton-Holt or Ricker) for the expected recruitment produced by a given level of spawning output. Mangel et al. [14] and others have shown that biological reference points such as $\frac{F^*}{M}$ and $\frac{B^*}{B(0)}$ are largely determined by a single parameter (steepness) when using two-parameter relationships. These functions introduce strong correlations between reference points (RP) that are pre-determined by the functional form, rather than a biological characteristic of the stock. Mangel et al. note that use of a three-parameter stock-recruitment relationship allows for independent estimation of these reference points. This research seeks to understand the nature of biases in reference points resulting from fitting a two-parameter logistic functional form when the true relationship follows a three-parameter stock-recruitment relationship (SRR). This work demonstrates the useful limits of the misspecified Schaefer model, and the mechanisms of model failure which arise from mapping a three-dimensional parameter space into two dimensions.

To myself,
Perry H. Disdainful,
the only person worthy of my company.

Acknowledgments

I want to “thank” my committee, without whose ridiculous demands, I would have graduated so, so, very much faster.

Chapter 1

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 (RPs) are commonly derived from the ODE equilibrium equations and depend upon the parameterization of biomass production. Two-parameter forms of the production function have been shown to limit the theoretical domain of RPs [14]. The limited RP-space of two-parameter models makes these models vulnerable to model misspecification with respect to RPs, and thus the limiting structure of two-parameter models may in and of itself induce bias in RP estimation using these models. The behavior of RP estimation is not well understood and as a result patterns of bias in RP estimation may easily go unnoticed. A metamodeling approach is developed here to describe RP biases and explore mechanisms of model failure under the most common two-parameter models.

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 modelled population that is vulnerable to fishing. The index is often observed alongside a variety of other known quantities, but at a minimum, each index will be observed in the presence of some known catch for the period. Figure (1.1) shows the classic Namibian Hake dataset [15, 11, 13] exemplifying the form.

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

$$I_t = qB_t e^{\varepsilon} \quad \varepsilon \sim N(0, \sigma^2). \quad (1.1)$$

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.

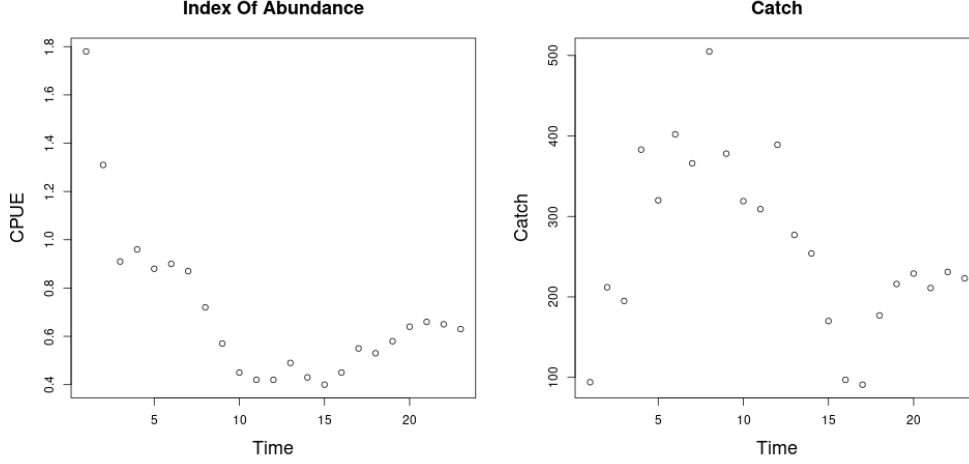


Figure 1.1: *left*: An index of abundance data, catch per unit effort (CPUE), for Namibian Hake from 1965 to 1987. *right*: The associated catch data for Namibian Hake over the same time period.

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

$$\frac{dB}{dt} = P(B(t); \theta) - Z(t)B(t). \quad (1.2)$$

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

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

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

39 From a management perspective a major goal of modeling is to accurately infer a

40 quantity known as *maximum sustainable yield* (MSY). One could maximize simple yield at a
 41 particular moment in time (and only for that moment) by fishing all available biomass in that
 42 moment. This strategy is penny-wise but pound-foolish (not to mention ecologically devastat-
 43 ing) since it doesn't leave biomass in the population to reproduce in the future. We seek to fish
 44 in a way that allows (or even encourages) future productivity in the population. This is accom-
 45 plished by maximizing the equilibrium level of catch over time. Equilibrium yield is considered
 46 by replacing the steady state biomass (\bar{B}) in the assumed form for catch, so that $\bar{Y} = F\bar{B}(F)$,
 47 where $\bar{\cdot}$ indicates a value at steady state. MSY is found by maximizing $\bar{Y}(F)$ with respect to
 48 F , and F^* is the fishing rate at MSY. Going forward let * decorate any value derived under the
 49 condition of MSY.

50 Fisheries are very often managed based upon reference points which serve as simpli-
 51 fied heuristic measures of population behavior. The mathematical form of RPs depends upon
 52 the model assumptions through the production function. While a number of different RPs exist
 53 which describe the population in different (but related) ways, the most common RPs revolve
 54 around the concept of MSY (or robust ways of measuring MSY [10, 16]). Here the focus is pri-
 55 marily on the RPs $\frac{B^*}{\bar{B}(0)}$ and F^* ($\frac{F^*}{M}$ when appropriate) for their pervasive use in modern fisheries
 56 [17].

57 F^* is the afore mentioned fishing rate which results in MSY. $\frac{B^*}{\bar{B}(0)}$ is the depletion of
 58 the stock at MSY. That is to say $\frac{B^*}{\bar{B}(0)}$ describes the fraction of the unfished population biomass
 59 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
 60 the assumption of two-parameter production, models will be structurally unable to capture the
 61 full theoretical range of RPs.

62 Many of the most commonly used production functions depend only on two-parameters.
 63 For example, the Schaefer model depends only on the biological parameters r and K , and limits
 64 RP inference so that under the Schaefer model $\left(F^*, \frac{B^*}{\bar{B}(0)}\right) \in \left(\mathbb{R}^+, \frac{1}{2}\right)$. The two-parameter Fox
 65 model [6] limits $\left(F^*, \frac{B^*}{\bar{B}(0)}\right) \in \left(\mathbb{R}^+, \frac{1}{e}\right)$. Similarly the two-parameter Cushing [4], Beverton-
 66 Holt [1, BH] and Ricker [20] production functions do not model the full theoretical space of
 67 RPs [14, 26].

68 The bias-variance trade-off [18] makes it clear that the addition of a third parameter in

69 the production function will necessarily reduce estimation bias. However the utility of this bias
70 reduction is still under debate because the particular mechanisms and behavior (direction and
71 magnitude) of these biases for key management quantities are not fully understood or described.
72 Lee et al. [12] provides some evidence that estimation of productivity parameters are dependent
73 on changes in biomass trend through time (i.e. contrast) as well as model specification. Conn et
74 al. [2] comes to similar conclusions via calibration modeling techniques. These studies indicate
75 important factors that contribute to inferential failure. However they do not offer mechanisms
76 of model failure, nor do their experimental designs allow for the control of different types of
77 model misspecification.

78 In this study I consider the behavior of inference when index data are simulated from
79 three-parameter PT and Schnute production models, but the simulated data are fit using inten-
80 tionally misspecified two-parameter logistic or BH production models. The work begins with
81 a derivation of RPs under the three-parameter models. A method is then presented for gener-
82 ating simulation designs based on the parametric form of RPs which serves as a control on the
83 nature of simulated model misspecification. Finally a Gaussian Process (GP) metamodel [7] is
84 constructed for exploration and analysis of RP biases.

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

⁹⁰ **Chapter 2**

⁹¹ **Pella-Tomlinson Model**

92 2.1 Introduction

93 2.2 Methods

94 2.2.1 Model

The three-parameter Pella-Tomlinson (PT) family has a convenient form that includes, among others [6, 19], the logistic production function as a special case. PT production function is parameterized so that $\theta = [r, K, \gamma]$ and the family takes the following form,

$$P_p(B; [r, K, \gamma]) = \frac{rB}{\gamma-1} \left(1 - \left(\frac{B}{K} \right)^{(\gamma-1)} \right). \quad (2.1)$$

95 γ is a parameter which breaks PT out of
 96 the restrictive symmetry of the logistic curve. In
 97 general $\gamma \in (1, \infty)$, with the logistic model appear-
 98 ing in the special case of $\gamma = 2$, and the Fox model
 99 appearing as a limiting case as $\gamma \rightarrow 1$. The parame-
 100 ter r controls the maximum per-capita growth rate
 101 of the population in the absence of competition for
 102 resources (i.e. the slope of production function at
 103 the origin). K is the so called "carrying capaci-
 104 ty" of the population. In this context the carry-
 105 ing capacity can be formally stated as steady state
 106 biomass in the absence of fishing (i.e. $\bar{B}(0) = K$).
 107 In Figure (2.1) PT production is shown for a range
 108 of parameter values so as to demonstrate the vari-
 109 ous productivity shapes that can be achieved under
 110 PT.

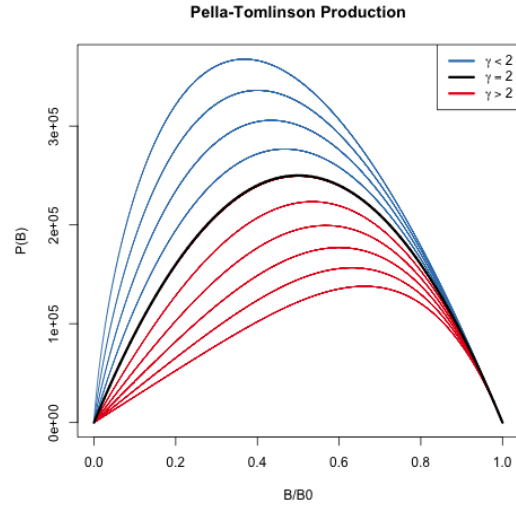


Figure 2.1: The Pella-Tomlinson production function plotted across a variety of parameter values. The special cases of Logistic production is shown in black, and the left-leaning and right-leaning regimes are shown in blue and red respectively.

111 While the form of the PT curve produces some limitations [5], importantly the intro-
 112 duction of a third parameter allows enough flexibility to fully describe the space of reference

113 points used in management. To see this, the reference points are analytically derived for the PT
 114 model below.

115 2.2.2 Reference Points

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

$$\frac{dB}{dt} = \frac{rB}{\gamma-1} \left(1 - \left(\frac{B}{K} \right)^{\gamma-1} \right) - FB. \quad (2.2)$$

An expression for the equilibrium biomass is attained by setting Eq (2.2) equal to zero, and rearranging the resulting equation to solve for B . Thinking of the result as a function of F gives,

$$\bar{B}(F) = K \left(1 - \frac{F(\gamma-1)}{r} \right)^{\frac{1}{\gamma-1}}. \quad (2.3)$$

118 At this point it is convenient to notice that $\bar{B}(0) = K$. The expression for B^* is given
 119 by evaluating Eq (2.3) at F^* . To get an expression for F^* , the equilibrium yield is maximized
 120 with respect to F ,

$$F^* = \operatorname{argmax}_F F \bar{B}(F). \quad (2.4)$$

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

$$\frac{d\bar{Y}}{dF} = \bar{B}(F) + F \frac{d\bar{B}}{dF} \quad (2.5)$$

$$\frac{d\bar{B}}{dF} = -\frac{K}{r} \left(1 - \frac{F(\gamma-1)}{r} \right)^{\frac{1}{\gamma-1}-1}. \quad (2.6)$$

Setting Eq (2.5) equal to 0, substituting $\bar{B}(F)$ and $\frac{d\bar{B}}{dF}$ by Equations (2.3) and (2.6) respectively, and solving for F produces the following expression for the fishing rate required

to produce MSY,

$$F^* = \frac{r}{\gamma} \quad (2.7)$$

Plugging the above expression for F^* back into Eq (2.3) gives the following expression for biomass at MSY,

$$B^* = K \left(\frac{1}{\gamma} \right)^{\frac{1}{\gamma-1}}. \quad (2.8)$$

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 productivity parameters.

$$F^* = \frac{r}{\gamma} \quad \frac{B^*}{\bar{B}(0)} = \left(\frac{1}{\gamma} \right)^{\frac{1}{\gamma-1}} \quad (2.9)$$

121 2.2.3 Simulation

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 [17, 21]. 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

$$r = \gamma F^* \quad \gamma = \frac{W \left(\frac{B^*}{\bar{B}(0)} \log \left(\frac{B^*}{\bar{B}(0)} \right) \right)}{\log \left(\frac{B^*}{\bar{B}(0)} \right)}. \quad (2.10)$$

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

124 Using Eq. (2.10) to obtain production parameters, a PT production model can be fully
125 defined for any combination of the RPs F^* and $\frac{B^*}{\bar{B}(0)}$. Since K does not enter the RP calculation
126 its value is fixed arbitrarily at 10000.

Indices of abundance are simulated from the three-parameter PT production model broadly over the space of F^* and $\frac{B^*}{B(0)}$ via a space filling design as described in Section (??). A small amount of residual variation, $\sigma = 0.01$, is added to the simulated index, and these data are then fit with a Schaefer model, at various degrees of misspecification, so as to observe the effect of productivity model misspecification upon RP inference.

2.2.4 Design

Letting \mathcal{F} and \mathcal{B} be regular grids, of size $n = 100$, on $F^* \in (0.1, 0.7)$ and $\frac{B^*}{B_0} \in (0.2, 0.6)$ respectively, a LHS design of size 100 is collected among the 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. Since the relationship mapping RPs analytically to productivity parameters can be found for the PT model, LHS designs the the PT model are computed directly in RP space and Eq. (2.10) is used to map the sampled RP design locations to PT productivity parameters.

2.2.5 Gaussian Process Metamodel

At its core, a metamodel is simply a model of some mapping of inputs to outputs (the mapping itself is typically defined by a computer model). By modeling the mapping with a statistical model (that explicitly defines the relevant features of the mapping) a metamodel defines a specific ontology for the mapping. By simulating examples of the mapping, the inferential infrastructure of the statistical model is used to empirically learn an effective emulation of the mapping within the ontology defined by the statistical model. The predictive infrastructure of the statistical model is then useful as an approximate abstraction of the system itself to better understand the system through further data collection, cheap approximation of the mapping, and/or study of the mapping itself.

In this setting, the aim of metamodeling is to study how well RPs are inferred when typical two-parameter models of productivity (Logistic and BH) are misspecified for populations that are actually driven by more complicated dynamics. The simulation design, \mathbf{X} , provides a sample of different population dynamics that are driven by three-parameter production

153 functions broadly in RP space. By simulating index of abundance data from the three parameter
 154 model, and fitting those data with the two-parameter production model, we observe particu-
 155 lar instances of how well RPs are inferred at the given misspecification of the two-parameter
 156 model relative to the true three-parameter production model. By gathering all of the simulated
 157 instances of how RPs are inferred (under the two-parameter model), we form a set of example
 158 mappings to train a metamodel which represents the mapping of true RPs (under the three-
 159 parameter model) to estimates of RPs under the misspecified two-parameter production model.
 160 The metamodel is essentially a surrogate for inference under the misspecified two-parameter
 161 production model that controls for the specific degree of model misspecification.

162 A flexible GP model is assumed for the structure of the metamodel to describe the
 163 mapping of RPs under misspecified two-parameter models of productivity. A GP is a stochastic
 164 process generalizing the multivariate normal distribution to an infinite dimensional analog. GP
 165 models are often specified primarily through the choice of a covariance (or correlation) func-
 166 tion which defines the relationship between locations in the input space. Typically correlation
 167 functions are specified so that points closely related in space result in correlated effects in the
 168 model. In this setting the inputs to the GP metamodel are the space of reference points which
 169 define the simulated three-parameter production models.

While index of abundance data are generated from three-parameter models, at each
 design location of the simulation, fitting the restricted two-parameter model results in a maxi-
 mum likelihood estimate (MLE; and associated estimation uncertainty) of each of the produc-
 tivity parameters (i.e. Schaefer: $[\log(r), \log(K)]$, BH: $[\log(\alpha), \log(\beta)]$). To simplify the speci-
 fication of the metamodel, let \mathbf{y} be a vector collecting the fitted MLEs for one of the productivity
 parameters, and let $\boldsymbol{\omega}$ be a vector of estimates of the estimator variances (via the inverted Fisher
 information) at each \mathbf{y} . Each of the fitted productivity parameter estimates are then modeled

using independent instances of the following GP metamodel.

$$\begin{aligned}\mathbf{y} &= \beta_0 + \mathbf{X}\beta + \mathbf{v} + \epsilon \\ \mathbf{v} &\sim N_n(\mathbf{0}, \tau^2 \mathbf{R}_\ell) \\ \epsilon &\sim N_n(\mathbf{0}, \omega' \mathbf{I})\end{aligned}\tag{2.11}$$

\mathbf{X} is the $n \times 2$ LHS design matrix of RPs for each simulated three-parameter data generating model as described in Section (??). 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 [8], although the application here models this effect heterogeneously.

The term, \mathbf{v} , contains spatially correlated GP effects. The correlation matrix, \mathbf{R}_ℓ describes how RPs close together in the simulation design are more correlated than those that are far away. This spatial effect is modeled with a squared exponential correlation function,

$$R(\mathbf{x}, \tilde{\mathbf{x}}) = \exp\left(\sum_{i=1}^2 \frac{-(x_i - \tilde{x}_i)^2}{2\ell_j^2}\right).\tag{2.12}$$

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 , β , τ^2 , ℓ_1 and ℓ_2 are fit via MLE against the observations \mathbf{y} , \mathbf{X} , and ω from simulation fits.

Fitting the metamodel allows for a full predictive description of inference under the misspecified restricted models. Predictive estimates are obtained via kriging [3]

$$\hat{y}(\mathbf{x}) = \beta_0 + \mathbf{x}\beta + \mathbf{r}(\mathbf{x})' \mathbf{R}_\ell^{-1} \left(\mathbf{y} - (\beta_0 + \mathbf{X}\beta) \right)\tag{2.13}$$

184 $\hat{y}(\mathbf{x})$ is the predicted value of the modeled productivity parameter MLE under the
 185 two-parameter production model, when the index of abundance is generated from the three-
 186 parameter production model at RP location \mathbf{x} . $\mathbf{r}(\mathbf{x})$ is a vector-valued function of correlation
 187 function evaluations for the predictive location \mathbf{x} against all observations in \mathbf{X} (i.e. $\mathbf{r}(\mathbf{x}) =$
 188 $\mathbf{R}(\mathbf{x}, x_i) \forall x_i \in \mathbf{X}$).

189 While metamodeling occurs on the inferred productivity parameters of the restricted
 190 production model, the metamodel can also be used to build estimates of major biological RPs.
 191 For the BH model the relevant transformations for relating productivity parameters with RPs are
 192 given in Eqs. (3.5, 3.8) with γ fixed to -1; for the Schaefer model $\hat{B}^* = \frac{\hat{K}}{2}$ and $\hat{F}^* = \frac{\hat{r}}{2}$. Applying
 193 the metamodel predictive surfaces on the scale of RP estimates allows for the quantification
 194 of estimation bias that is induced by fitting a misspecified two-parameter production model to
 195 indices of abundance generated under three-parameter productivity.

196 2.2.6 Catch

197 It is known that contrast in the observed index and catch time series can effect infer-
 198 ence on the productivity parameters [9]. In this setting contrast refers to changes in the long
 199 term trends of index data. Figure (2.2, *right*) demonstrates an example of biomass that includes
 200 contrast induced by catch. It is not well understood how contrast may factor into inferential
 201 failure induced by model misspecification. Thus catch is parameterized so as to allow for a
 202 spectrum of possible contrast simulation settings.

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

208 Intuitively $\frac{F(t)}{F^*}$ describes the fraction of F^* that $F(t)$ is specified to for the current
 209 $B(t)$. When $\frac{F(t)}{F^*} = 1$, $F(t)$ will be held at F^* , and the solution of the ODE brings $B(t)$ into
 210 equilibrium at B^* . When $\frac{F(t)}{F^*}$ is held constant in time biomass comes to equilibrium as an
 211 exponential decay from K approaching B^* . When $\frac{F(t)}{F^*} < 1$, $F(t)$ is lower than F^* and $B(t)$ is

212 pushed toward $\bar{B} > B^*$. Contrarily, when $\frac{F(t)}{F^*} > 1$, $F(t)$ is higher than F^* and $B(t)$ is pushed
 213 toward $\bar{B} < B^*$; the precise values of \bar{B} can be calculated from the steady state biomass equations
 214 provided above and depend upon the specific form of the production function.

For the simulations presented here, a family of fishing behaviors are considered where the fishing rate accelerates as technology and fishing techniques improve rapidly until management practices are applied, which ultimately brings fishing into equilibrium at F^* . This is parameterized as three distinct phases, over a total of 45 units of time, with each phase lasting 15 time units. The specific form is given below.

$$\frac{F(t)}{F^*} = ae^{bt}\mathbf{1}_{0 \leq t < 15} + (d - ct)\mathbf{1}_{15 \leq t < 30} + \mathbf{1}_{30 \leq t \leq 45} \quad (2.14)$$

The first term of Eq (2.14) is an exponential increase in fishing, the second term is a linear decline in relative fishing as initial management practices are applied, and the third term, $\mathbf{1}_{30 \leq t \leq 45}$, simply holds the fishing rate at F^* there after. These three phases are controlled by the four parameters a , b , c , and d . By enforcing that the interface of the phases meet at χ_{max} and 1 respectively the relative fishing series is reduced to a two-parameter family.

$$a = e^{\log(\chi_{max}) - 15b} \quad b = \frac{1}{t - 15} \log \left(\frac{\chi_{min}}{\chi_{max}} \right) \quad (2.15)$$

$$c = \frac{\chi_{max} - 1}{15 - 1} \quad d = 15c + \chi_{max} \quad (2.16)$$

215 By further specifying $\chi_{max} = 1.6^\chi$ and $\chi_{min} = 0.4^\chi$ the two-parameters χ_{max} , and χ_{min} can be
 216 reduced to the single parameter χ . The tuning parameter χ then singularly controls contrast that
 217 appears in time series data.

218 When $\chi = 0$, the relative fishing rate is a constant at 1 to create a low contrast simu-
 219 lation environment. As χ increases Eq (2.14) induces more and more contrast in the observed
 220 index and catch time series until $\chi = 1$ which produces a high contrast simulation environment.
 221 Figure (2.2) demonstrates a spectrum of contrast simulation environments as well as the time
 222 series data they induce in the solution of the production model ODE.

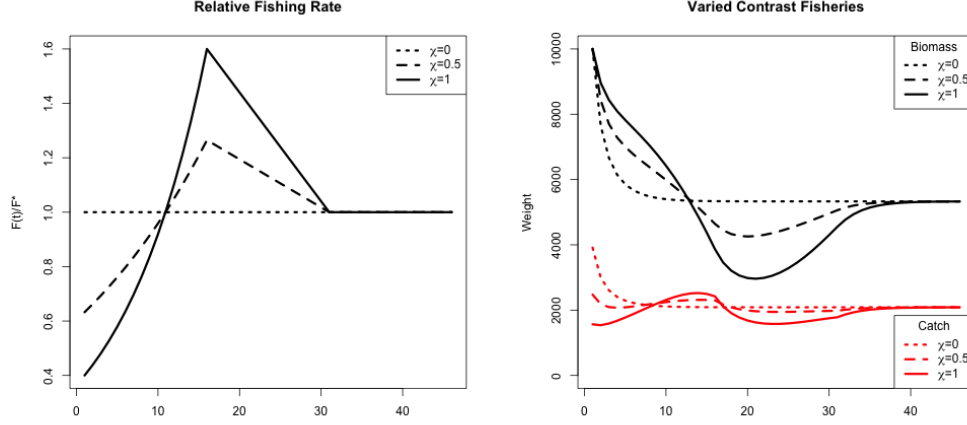


Figure 2.2: (left) Relative fishing with low, medium, and high contrast. (right) Population biomass and catch at each associated level of contrast.

2.2.7 Two-Parameter Production Model Inference

The simulated mapping results from fitting an intentionally misspecified two parameter production model to index of abundance data that are generated from a more complex three-parameter model of productivity. Thus, let I_t be an index of abundance simulated from the three-parameter PT or Schnute production models at time $t \in \{1, 2, 3, \dots, T\}$. However the fitted model is specified to be intentionally misspecified so that the fitted model is driven by a two-parameter Schaefer, or BH production model respectively.

The observation model for the fitted model is log-normal such that,

$$I_t | q, \sigma^2, \theta \sim LN(qB_t(\theta), \sigma^2). \quad (2.17)$$

$B_t(\theta)$ is defined by the solution of the ODEs defined by the Schaefer, or BH models. For the Schaefer model $\theta = [r, K]$, and for the BH model $\theta = [\alpha, \beta]$. From the perspective of the fitted model, the observed I_t are assumed independent conditional on q, σ^2, r, K and the two-parameter ODE model for biomass. Thus the log likelihood can be written as

$$\log \mathcal{L}(q, \sigma^2, \theta; I) = -\frac{T}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_t \log \left(\frac{I_t}{qB_t(\theta)} \right)^2. \quad (2.18)$$

In this setting, q is fixed at 0.0005 and M is fixed at 0.2, to focus on the inferential

231 effects of model misspecification on biological parameters. σ^2 and θ are reparameterized to the
 232 log scale and fit via MLE. Reparameterizing the parameters to the log scale improves the reli-
 233 ability of optimization, in addition to facilitating the use of Hessian information for estimating
 234 MLE standard errors.

235 Given that the biological parameters enter the likelihood via a nonlinear ODE, and
 236 further the parameters themselves are related to each other nonlinearly, the likelihood function
 237 can often be difficult to optimize. A hybrid optimization scheme is used to maximize the log
 238 likelihood to ensure that a global MLE solution is found. The R package GA [22, 23] is used to
 239 run a genetic algorithm to explore parameter space globally. Optimization periodically jumps
 240 into the L-BFGS-B local optimizer to refine optima within a local mode. The scheme functions
 241 by searching globally, with the genetic algorithm, across many initial values for starting the local
 242 gradient-based optimizer. The genetic algorithm serves to iteratively improve hot starts for the
 243 local gradient-based optimizer. Additionally, optimization is only considered to be converged
 244 when the optimum results in an invertible Hessian at the found MLE.

245 **2.2.8 Continuous model formulation**

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

250 If the form of $\frac{dB}{dt}$ is at least Lipschitz continuous, then the Cauchy-Lipschitz-Picard
 251 theorem provides local existence and uniqueness of $B(t)$. Recall from Eq(1.2) that $\frac{dB}{dt}$ is sepa-
 252 rated into a term for biomass production, $P(B)$, and a term for removals, $Z(t)B(t)$. For deter-
 253 mining Lipschitz continuity of $\frac{dB}{dt}$, the smallest Lipschitz constant of $\frac{dB}{dt}$ will be the sum of the
 254 constants for each of the terms $P(B)$ and $Z(t)B(t)$ separately. Typically any choice of $P(B)$ will
 255 be continuously differentiable, which implies Lipschitz continuity. At a minimum $Z(t)$ typically
 256 contains fishing mortality as a function of time $F(t)$ to model catch in time as $C(t) = F(t)B(t)$.
 257 $Z(t)$ may or may not contain M , but typically M is modeled as stationary in time and does not
 258 pose a continuity issue, unlike some potential assumptions for $C(t)$.

259 In practice $C(t)$ is determined by a series of observed, assumed known, catches. Catch
260 observations are typically observed on a quarterly basis, but in practice may not be complete for
261 every quarter (or year) of the modeled period. It is overwhelmingly common to discretize the
262 ODE in time via Euler's method with integration step sizes to match the observation frequency
263 of the modeled data. This is often computationally convenient when the underlying species
264 dynamics are reasonably well behaved, however when the dynamics model is used as a statistical
265 model, with the goal of inferring the behavior of the underlying species dynamics, the regularity
266 of the dynamics are not guaranteed. An implicit assumption of continuity of catch in time
267 provides the necessary regularity for the statistical model. Furthermore a continuous handling
268 of the dynamics provides improved accuracy in evaluating the ODE, particularly when inferring
269 productivity parameters which largely control the regularity of the dynamics.

270 While there are many ways to handle catch continuity, here I assume that catches
271 accrue linearly between observed catches. This assumption defines the catch function as a
272 piecewise linear function of time, with the smallest Lipschitz constant for the catch term defined
273 by the steepest time segment of the catch function. This assumption represents one of the
274 simplest ways of handling catch, while retaining Lipschitz continuity overall. Furthermore
275 linearly interpolated catch is adequately parsimonious for the typical handling of catches.

276 **2.2.8.1 Integration and Stiffness**

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

282 The concept of stiffness in ODEs is hard to precisely characterize. Hairer and Wanner
283 [25, p.2] describe stiffness in the following pragmatic sense, "Stiff equations are problems for
284 which explicit methods don't work". It is hard to make this definition more mathematically
285 precise, but this a consistent issue for models of very productive species in the low contrast
286 simulation. Euler's method, as often implemented, is particularly poorly suited for these stiff

287 regions of parameter space. In these stiff regions it is necessary to integrate the ODE with an
288 implicit integration method.

289 Several of the most common implicit methods were tried including the Livermore
290 Solver for ODEs (lsode), and the Variable Coefficient ODE Solver (vode) as implemented in
291 the deSolve package of R [24]. The difference between implicit solvers is negligible, while
292 explicit methods result in wildly varying solutions to the ODE in stiff regions of parameter
293 space. Results shown here are computed using the lsode integration since it runs relatively
294 quickly and has a relatively smaller footprint in system memory.

2.3 Results

2.3.1 An *MSY*-Optimal Catch History

When $F(t)$ is held constant at F^* , as it is in the “low contrast” simulation setting, $B(t)$ 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 that this induces in $B(t)$. 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 [9].

Figure (2.3) shows four of the most misspecified example production function fits as compared to the true data generating PT production functions. The rug plots below each set of curves show how the observed biomasses decay exponentially from K to B^* in each case. In particular, notice how observations only exist where the PT biomass is greater than B^* . Due to the leaning of the true PT curves, and the symmetry of the logistic

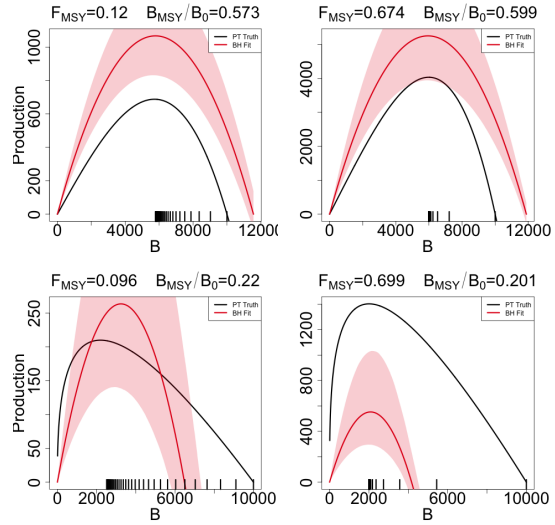


Figure 2.3: A comparison of the true PT production function (in black) and the estimated logistic function (in red) with 95% CI shown. The examples shown represent the four corners of maximum model misspecification in the simulated RP-space. Observed biomasses are plotted in the rug plots below the curves.

parabola, the logistic curve only observes information about its slope at the origin from data observed on the right portion of the PT curves. The top two panels of Figure (2.3) shows PT data generated such that $\frac{B^*}{B(0)} > 0.5$; in these cases PT is steeper to the right of B^* than it is on the left, and so the the logistic curve over-estimates r , and consequently also over-estimates F^* . The bottom two panels of Figure (2.3) show PT data generated with $\frac{B^*}{B(0)} < 0.5$ and where the vice versa phenomena occurs. 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^* .

2.3.1.1 Metamodeled Trends

Each point in the space of the RPs F^* and $\frac{B^*}{B(0)}$ uniquely identifies a complete PT model with different combinations of parameters values. Recall that when $\gamma = 2$ for the PT model, the PT curve becomes a parabola and is equivalent to the logistic curve of the Schaefer model. Since the logistic curve is symmetric about B^* , the Schaefer model must fix the value of $\frac{B^*}{B(0)}$ at the constant 0.5 for any value of F^* . So the line through RP space defined by $\frac{B^*}{B(0)} = 0.5 \ \forall \ F^*$, defines the subset of RP space where $\gamma = 2$ and where the PT model is equivalent to the Schaefer model. For brevity this subset of RP where $\frac{B^*}{B(0)} = 0.5$ will be referred to as the “Schaefer set”. Thus simulated data that are generated along the Schaefer set will be the only data that are not misspecified relative to the Schaefer model; as PT data are simulated farther and farther away from this line at $\frac{B^*}{B(0)} = 0.5$ model misspecification of the Schaefer model becomes worse and worse.

While Figure (2.3) demonstrates a real trend in simulation results, individual simulation runs will at best show jittery trends due to the stochastic nature of statistical inference. The GP process metamodel accounts for this stochasticity to focus analysis on the signal in the simulation results. Recall that metamodeling occurs on the scale of the inferred productivity parameters of the restricted production model, by transforming metamodel predictions via Eq. (2.9), metamodeled predictions are obtained for Schaefer RPs. By further subtracting the true data generating PT RPs from the predicted Schaefer RPs at each point in RP space a pattern of inferential RP bias, induced by model misspecification of the Schaefer model, can be seen.

Figure (2.4) shows the pattern of biases the Schaefer model creates when fit to PT data generated at each point of RP space. An equivalent way to think of Figure (2.4) is that since the Schaefer model must estimate RPs in the Schaefer set, the metamodel arrows indicate the mapping that is created by inferring RPs under a misspecified Schaefer model fit to PT data generated at each point over the pictured region.

Since $\frac{B^*}{B_0}$ must be 0.5 under the Schaefer model, biases in the $\frac{B^*}{B_0}$ direction must simply map vertically onto the Schaefer set. Due to this simplified RP geometry under the Schaefer model, the degree of bias in $\frac{B^*}{B_0}$ estimation is defined solely by the degree of model misspeci-

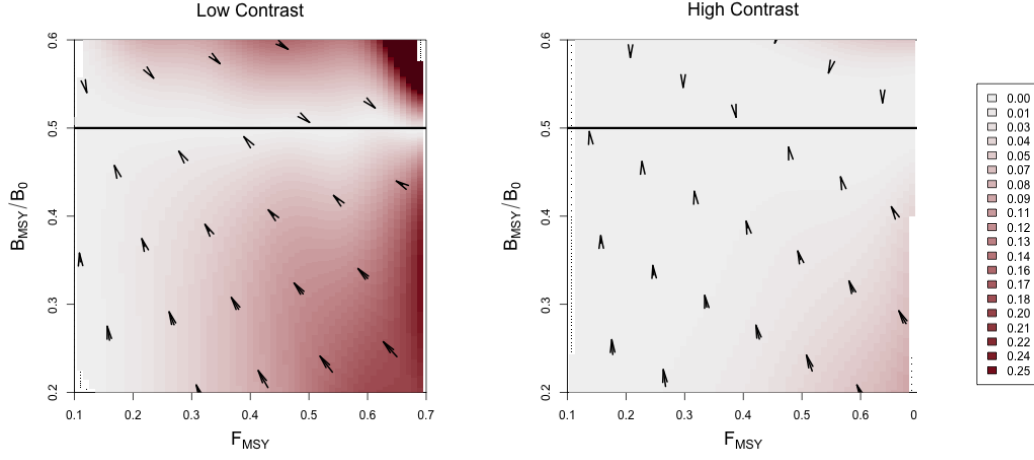


Figure 2.4: Joint bias direction for $(F^*, \frac{B^*}{B_0})$ estimates under the misspecified Schaefer Model. The intensity of color represents the excess bias relative to the shortest possible mapping. Results in the low contrast setting are shown *left*, and the high contrast setting is shown *right*.

350 fication irrespective of F^* . Furthermore, the closest possible point along the Schaefer set that
 351 Schaefer model inference could map RPs would be the perfectly vertical mapping. This pattern
 352 only contains the strictly necessary bias present in $\frac{B^*}{B_0}$, and zero bias in F^* . Any deviation from
 353 this minimal bias pattern is necessarily due to added bias in F^* .

354 The two simulation settings shown in Figure (2.4) are identical except for the amount
 355 of contrast present in the simulated index. The left panel of Figure (2.4) shows RP biases in the
 356 low contrast setting, while the right panel shows the high contrast setting. Notice that in the low
 357 contrast setting the RP bias pattern is far from the minimum distance mapping, however when
 358 contrast is added the mapping becomes much closer to a minimal vertical bias mapping. In the
 359 low contrast setting the observed bias is consistent with the pattern and mechanism described
 360 in Figure (2.3), where F^* is underestimated for data generated below the Schaefer line and
 361 overestimated above the Schaefer set. In the high contrast simulation the mapping is nearly
 362 minimal distance with the exception of PT data generated with simultaneously low $\frac{B^*}{B_0}$ and high
 363 F^* .

364 Figure (2.5) demonstrates how bias in F^* estimation decreases as contrast is added to
 365 PT data as generated in the low $\frac{B^*}{B_0}$ and high F^* regime. By including additional contrast F^*

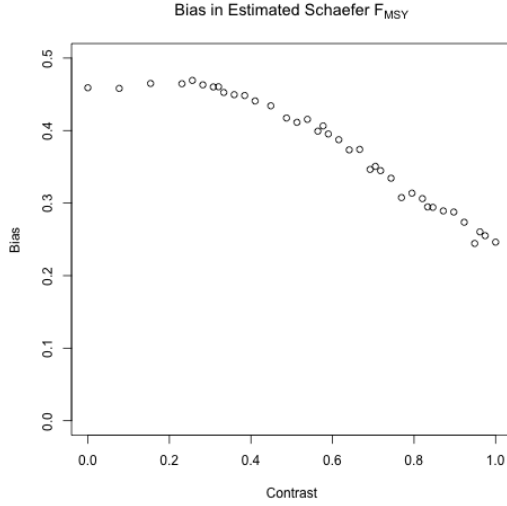


Figure 2.5: Bias in F^* under the Schaefer model when PT data are generated with increasing contrast so that F^* and $\frac{B^*}{B_0}$ are fixed at 0.699 and 0.201 respectively.

366 bias is decreased, however parameterizing contrast so as to fully extinguish F^* bias may require
 367 a more complex model of fishing.

368 2.4 Discussion

369 *Tease Out BH*

370
 371 Results presented here generally agree with what is known about estimating pop-
 372 ulation growth rate parameters [12, 2, ?]. These studies appreciate the role of contrast for
 373 estimating growth rates, however they struggle to make generally extensible conclusions since
 374 they focus only on a handful of stocks that fall short of forming a random sample of the greater
 375 population of possible stock behaviors. The LHS design methods presented here are designed
 376 specifically to simulate a representative sample of stocks broadly across the space of possible
 377 RPs. Furthermore, the simulation design, taken together with the GP metamodel of productivity
 378 parameter estimates, allows this study to control the degree of model misspecification and gen-
 379 eralize conclusions about the behavior of productivity estimation within the production model
 380 setting presented.

381 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. This pattern is particularly true for low-contrast inference under the Schaefer model where the geometry of the restricted RP set isolates estimation failure of F^* from $\frac{B^*}{B(0)}$. While contrast has a similar impact on F^* estimation under the BH model, the geometry of the BH RP set correlates estimation bias of F^* and $\frac{B^*}{B(0)}$. The GP metamodeling approach reveals a more general pattern that highly informative data sets (high contrast) produces a nearly minimal distance mapping of RPs onto the constrained RP set.

In all cases when model misspecification is removed, even with weakly informative data, RP estimation is unbiased and well estimated. Thus contrast alone is not the only factor leading to inferential failure. Model misspecification is a necessary but not sufficient condition for inducing RP estimation bias. The particular RP bias present depends on the RP geometry of the fitted model and how that geometry is misspecified relative to the data. The RP mapping is then oriented to the RP geometry of the fitted model.

While the relative fishing rate parameterized in Section (2.2.6) captures a usefully broad spectrum of relevant fishing behaviors, it is still limiting in the amount of information that it can induce. Improved methods for quantifying contrast in fisheries data, and/or methods of discovering more informative fishing behavior, could improve this analysis. In the absence of a maximally informative dataset simulation methods will not fully describe how inference fails, but the methods presented here tell the most complete picture yet, with explicit control of the degree model misspecification, contrast, and a simulation design that allows for uniform representative data generation across biologically meaningful stocks. The results presented here suggest the conjecture that under a maximally informative dataset, RP inference with a two parameter production function will be biased in the direction a shortest distance map from the true RPs onto restricted set of RPs under the two-parameter model.

Given the potential for model misspecification of RPs, a minimal distance mapping of RPs represents a best-case scenario where the total bias of RPs, when measured jointly, is minimized. That said, without recognizing the geometry of how two-parameter models of productivity limit RP space this may lead to unintuitive implications in RP estimation. For example, due to the shape of the BH RP set a minimal distance mapping ensures that if there

411 is bias in one of $\frac{B^*}{B_0}$ or F^* , there will necessarily be bias in the other RP. However under the
 412 Schaefer model, since the RP set is a constant in $\frac{B^*}{B_0}$, bias in F^* is not adulterated in the same way
 413 by bias in $\frac{B^*}{B_0}$ estimation. While models with constant RPs, such as the logistic model $\frac{B^*}{B_0} = \frac{1}{2}$ or
 414 the Fox model $\frac{B^*}{B_0} = \frac{1}{e}$, are extremely limited, they can be valuable tools for developing intuition
 415 precisely because they isolate RP estimation in their free RPs from the correlated RP biases
 416 present in models like the BH or Ricker model.

417 When one considers the implications of RP bias, overestimation of RPs carries the
 418 severe implication of management recommendations potentially leading to overfishing, while
 419 underestimation of RP leads to overly conservative management. In this sense, when the true
 420 model is not known, the geometry of the BH set together with the metamodeled bias trends
 421 makes the BH model a naturally conservative estimator of RPs for most stocks. For most non-
 422 BH populations the BH model is likely to make conservative errors in its estimates of F^* and $\frac{B^*}{B_0}$.
 423 The one notable exception to the conservatism of the BH model stands for data generated in the
 424 Cushing-like regime of Schnute RPs. In this regime the BH model tends to be fairly unbiased
 425 overall, however the bias that is present for these populations tends to be overestimation in both
 426 RPs, leading to much more severe management consequences for those populations.

427 The RP bias trends of the Schaefer model demonstrate much less conservatism than
 428 the BH overall. For any population with $\frac{B^*}{B_0} < 0.5$, $\frac{B^*}{B_0}$ will be overestimated. When the popula-
 429 tion comes from the regime where $\frac{B^*}{B_0} > 0.5$, $\frac{B^*}{B_0}$ will be under estimated, but F^* is likely to be
 430 overestimated depending on the degree of contrast present in the data. So while the Schaefer
 431 model is an intuitive model, it tends to lead to much less conservative RP estimation.

432 While it is important to recognize these limitations of two-parameter models of pro-
 433 ductivity, we should not solely accept conservatism as a rational of choosing a BH model
 434 of productivity. Increasing the flexibility of the production function by moving toward three-
 435 parameter models would release the underlying structural limitations [14] that cause these RP
 436 biases in the first place. Punt & Cope [17] considers a suite of possible three-parameter curves
 437 which could be used instead of current two-parameter curves. For all of their benefits, three
 438 parameter production functions have their own complicating factors, and the structure present
 439 in the Schnute model explored here makes it an intuitive bridge model for developing three-

440 parameter models going forward.

441 **Chapter 3**

442 **Schnute Model**

3.1 Introduction

3.2 Methods

3.2.1 Model

The Schnute production function is a three-parameter generalization of many of the most common two-parameter production functions [?, ?]. It can be written in the following form, with parameters α , β , and γ ,

$$P_s(B; [\alpha, \beta, \gamma]) = \alpha B(1 - \beta\gamma B)^{\frac{1}{\gamma}}. \quad (3.1)$$

The BH and Logistic production functions arise when γ is fixed to -1 or 1 respectively. The Ricker model is a limiting case as $\gamma \rightarrow 0$. For $\gamma < -1$ a family of strictly increasing Cushing-like curves arise, culminating in linear production as $\gamma \rightarrow -\infty$. These special cases form natural regimes of similarly behaving production functions as seen in Figure (3.1).

The behavior of RP inference under the BH model 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 variety of possible productivity behaviors, they present an ideal simulation environment for inquiry of the reliability of inference under the BH assumption.

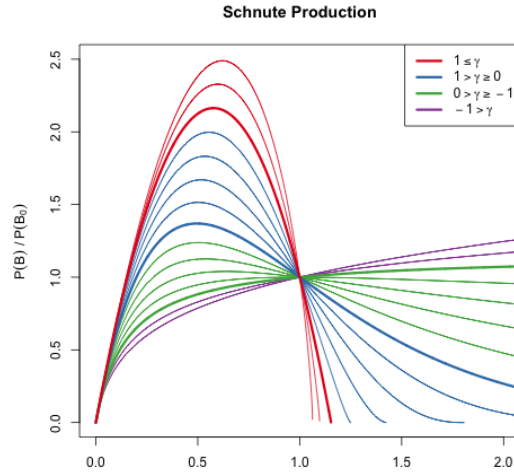


Figure 3.1: The Schnute production function plotted across a variety of parameter values. Regimes of similarly behaving curves are grouped by color.

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

$$\frac{dB}{dt} = P_s(B; \theta) - (M + F)B. \quad (3.2)$$

462 This equation largely takes the same form as previously described, except that P_s is the Schnute
 463 production function and natural mortality, M , is modeled explicitly here. Natural mortality
 464 models the instantaneous rate of mortality from all causes outside of fishing. While Eq. (3.2)
 465 models M explicitly, natural mortality is implicit to the structure of the previously decribed
 466 Schaefer, Fox, and PT production models. Explicitly modeling natural mortality allows for the
 467 production function not to approach (or intersect) 0 for large biomasses (e.g. BH production).
 468 In turn, the Schunte model requires the addition of the term $-MB$ to form an interpretable yeild
 469 curve and make RPs well defined over the relevant domain of γ .

The derivation of RPs under Eq. (3.2) 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

$$\bar{B}(F) = \frac{1}{\gamma\beta} \left(1 - \left(\frac{M+F}{\alpha} \right)^\gamma \right). \quad (3.3)$$

The above expression quickly yields B_0 , B^* by evaluation at $F = 0$ and F^* respectively,

$$B_0 = \frac{1}{\gamma\beta} \left(1 - \left(\frac{M}{\alpha} \right)^\gamma \right) \quad (3.4)$$

$$\frac{B^*}{B_0} = \frac{1 - \left(\frac{M+F^*}{\alpha} \right)^\gamma}{1 - \left(\frac{M}{\alpha} \right)^\gamma}. \quad (3.5)$$

Attaining an expression for F^* requires maximization of equilibrium yield, $\bar{Y} = F\bar{B}(F)$,

with respect to F . Analytically maximizing proceeds by differentiating \bar{Y} to produce

$$\frac{d\bar{Y}}{dF} = \bar{B}(F) + F \frac{d\bar{B}}{dF} \quad (3.6)$$

$$\frac{d\bar{B}}{dF} = -\frac{1}{\beta} \left(\frac{\left(\frac{M+F}{\alpha}\right)^\gamma}{F+M} \right). \quad (3.7)$$

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^* is less yielding here than it was in the case of the PT model. This procedure falls short of providing an analytical solution for F^* directly in terms of θ , but rather shows that F^* must respect the following expression,

$$0 = \frac{1}{\gamma} - \left(\frac{1}{\gamma} + \frac{F^*}{F^* + M} \right) \left(\frac{F^* + M}{\alpha} \right)^\gamma. \quad (3.8)$$

470 The lack of an analytical solution here is understood. Schnute & Richards [21, pg.
471 519] specifically point out that F^* cannot be expressed analytically in terms of productivity
472 parameters, but rather gives a partial analytical expression for the inverse relationship. Although
473 parameterized slightly differently, Schnute & Richards derive expressions for α and β as a
474 function of RPs and γ .

475 Since RPs are left without a closed form expression, computing RPs from productivity
476 parameters amounts to numerically solving the system formed by collecting the expressions
477 (3.8), (3.4), and (3.5).

478 3.2.2 Simulation

479 3.3 Discussion

480 *Tease Out BH*

481

482 Results presented here generally agree with what is known about estimating pop-
483 ulation growth rate parameters [12, 2, ?]. These studies appreciate the role of contrast for
484 estimating growth rates, however they struggle to make generally extensible conclusions since

they focus only on a handful of stocks that fall short of forming a random sample of the greater population of possible stock behaviors. The LHS design methods presented here are designed specifically to simulate a representative sample of stocks broadly across the space of possible RPs. Furthermore, the simulation design, taken together with the GP metamodel of productivity parameter estimates, allows this study to control the degree of model misspecification and generalize conclusions about the behavior of productivity estimation within the production model setting presented.

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. This pattern is particularly true for low-contrast inference under the Schaefer model where the geometry of the restricted RP set isolates estimation failure of F^* from $\frac{B^*}{B(0)}$. While contrast has a similar impact on F^* estimation under the BH model, the geometry of the BH RP set correlates estimation bias of F^* and $\frac{B^*}{B(0)}$. The GP metamodeling approach reveals a more general pattern that highly informative data sets (high contrast) produces a nearly minimal distance mapping of RPs onto the constrained RP set.

In all cases when model misspecification is removed, even with weakly informative data, RP estimation is unbiased and well estimated. Thus contrast alone is not the only factor leading to inferential failure. Model misspecification is a necessary but not sufficient condition for inducing RP estimation bias. The particular RP bias present depends on the RP geometry of the fitted model and how that geometry is misspecified relative to the data. The RP mapping is then oriented to the RP geometry of the fitted model.

While the relative fishing rate parameterized in Section (2.2.6) captures a usefully broad spectrum of relevant fishing behaviors, it is still limiting in the amount of information that it can induce. Improved methods for quantifying contrast in fisheries data, and/or methods discovering more informative fishing behavior, could improve this analysis. In the absence of maximally informative dataset simulation methods will not fully describe how inference fails, but the methods presented here tell the most complete picture yet, with explicit control of the degree model misspecification, contrast, and a simulation design that allows for uniform representative data generation across biologically meaningful stocks. The results presented here

514 suggest the conjecture that under a maximally informative dataset, RP inference with a two
515 parameter production function will be biased in the direction a shortest distance map from the
516 true RPs onto restricted set of RPs under the two-parameter model.

517 Given the potential for model misspecification of RPs, a minimal distance mapping
518 of RPs represents a best-case scenario where the total bias of RPs, when measured jointly,
519 is minimized. That said, without recognizing the geometry of how two-parameter models of
520 productivity limit RP space this may lead to unintuitive implications in RP estimation. For
521 example, due to the shape of the BH RP set a minimal distance mapping ensures that if there
522 is bias in one of $\frac{B^*}{B_0}$ or F^* , there will necessarily be bias in the other RP. However under the
523 Schaefer model, since the RP set is a constant in $\frac{B^*}{B_0}$, bias in F^* is not adulterated in the same way
524 by bias in $\frac{B^*}{B_0}$ estimation. While models with constant RPs, such as the logistic model $\frac{B^*}{B_0} = \frac{1}{2}$ or
525 the Fox model $\frac{B^*}{B_0} = \frac{1}{e}$, are extremely limited, they can be valuable tools for developing intuition
526 precisely because they isolate RP estimation in their free RPs from the correlated RP biases
527 present in models like the BH or Ricker model.

528 When one considers the implications of RP bias, overestimation of RPs carries the
529 severe implication of management recommendations potentially leading to overfishing, while
530 underestimation of RP leads to overly conservative management. In this sense, when the true
531 model is not known, the geometry of the BH set together with the metamodeled bias trends
532 makes the BH model a naturally conservative estimator of RPs for most stocks. For most non-
533 BH populations the BH model is likely to make conservative errors in its estimates of F^* and $\frac{B^*}{B_0}$.
534 The one notable exception to the conservatism of the BH model stands for data generated in the
535 Cushing-like regime of Schnute RPs. In this regime the BH model tends to be fairly unbiased
536 overall, however the bias that is present for these populations tends to be overestimation in both
537 RPs, leading to much more severe management consequences for those populations.

538 The RP bias trends of the Schaefer model demonstrate much less **conservatism** than
539 For any population with $\frac{B^*}{B_0} < 0.5$, $\frac{B^*}{B_0}$ will be overestimated. When the population comes from
540 the regime where $\frac{B^*}{B_0} > 0.5$, $\frac{B^*}{B_0}$ will be under estimated, but F^* is likely to be overestimated
541 depending on the degree of contrast present in the data. So while the Schaefer model is an
542 intuitive model, it tends to lead to much less conservative RP estimation.

543 While it is important to recognize these limitations of two-parameter models of pro-
544 ductivity, we should not solely accept conservatism as a rational of choosing a BH model
545 of productivity. Increasing the flexibility of the production function by moving toward three-
546 parameter models would release the underlying structural limitations [14] that cause these RP
547 biases in the first place. Punt & Cope [17] considers a suite of possible three-parameter curves
548 which could be used instead of current two-parameter curves. For all of their benefits, three
549 parameter production functions have their own complicating factors, and the structure present
550 in the Schnute model explored here makes it an intuitive bridge model for developing three-
551 parameter models going forward.

552 **Chapter 4**

553 **Conclusion**

554 **Appendix A**

555 **Inverting $\frac{B^*}{\bar{B}(0)}$ and γ for the PT Model**

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

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

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 .

$$\begin{aligned}\gamma \log(\zeta) &= W(\zeta \log(\zeta)) \\ \gamma &= \frac{W(\zeta \log(\zeta))}{\log(\zeta)}\end{aligned}\tag{A.1}$$

556 The Lambert product logarithm is a multivalued function with a branch point at $-\frac{1}{e}$. The prin-
557 cipal branch, $W_0(z)$, is defined on $z \in (-\frac{1}{e}, \infty)$, and the lower branch, $W_{-1}(z)$, is defined on
558 $z \in (-\frac{1}{e}, 0)$. Taken individually, each respective branch is analytic, but cannot be expressed in
559 terms of elementary functions.

560 When $\zeta \in (0, \frac{1}{e})$ the solution of interest in Eq. (2.10) comes from W_0 . When $\zeta \rightarrow \frac{1}{e}$,
561 the Fox Model emerges as $\gamma \rightarrow 1$. When $\zeta \in (\frac{1}{e}, 1)$ the solution of interest comes from W_{-1} .
562 For the use case presented here, Eq. (2.10) is to be interpreted as,

$$\gamma = \begin{cases} \frac{W_0(\zeta \log(\zeta))}{\log(\zeta)} & \zeta \in (0, \frac{1}{e}) \\ \frac{W_{-1}(\zeta \log(\zeta))}{\log(\zeta)} & \zeta \in (\frac{1}{e}, 1) \end{cases}.\tag{A.2}$$

Bibliography

- [1] Raymond JH Beverton and Sidney J. Holt. *On the dynamics of exploited fish populations*, volume 11. Springer Science & Business Media, 1957.
- [2] Paul B. Conn, Erik H. Williams, and Kyle W. Shertzer. When can we reliably estimate the productivity of fish stocks? *Canadian Journal of Fisheries and Aquatic Sciences*, 67(3):511–523, 2010.
- [3] Noel Cressie. *Statistics for spatial data*. John Wiley & Sons, 2015.
- [4] D. H. Cushing. The Dependence of Recruitment on Parent Stock in Different Groups of Fishes. *ICES Journal of Marine Science*, 33(3):340–362, May 1971.
- [5] R. I. Fletcher. On the restructuring of the Pella-Tomlinson system. *Fish. Bull.*, 76(3):515–521, 1978.
- [6] William W. Fox Jr. An Exponential Surplus-Yield Model for Optimizing Exploited Fish Populations. *Transactions of the American Fisheries Society*, 99(1):80–88, 1970. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1577/1548-8659%281970%2999%3C80%3AAESMFO%3E2.0.CO%3B2>.
- [7] Robert B. Gramacy. *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*. Chapman and Hall/CRC, 2020.
- [8] Robert B. Gramacy and Herbert KH Lee. Cases for the nugget in modeling computer experiments. *Statistics and Computing*, 22(3):713–722, 2012. Publisher: Springer.
- [9] R. Hilborn and C. J. Walters. Quantitative Fisheries, Stock Assessment: Choice Dynamics, and Uncertainty Chapman and Hall. *New York*, 1992.
- [10] Ray Hilborn. Pretty good yield and exploited fishes. *Marine Policy*, 34(1):193–196, 2010. Publisher: Elsevier.
- [11] Ray Hilborn and Marc Mangel. *The Ecological Detective: Confronting Models with Data*. Princeton University Press, March 1997.
- [12] Hui-Hua Lee, Mark N. Maunder, Kevin R. Piner, and Richard D. Methot. Can steepness of the stock–recruitment relationship be estimated in fishery stock assessment models? *Fisheries Research*, 125-126:254–261, August 2012.

- 591 [13] Marc Mangel. The Theoretical Biologist’s Toolbox: Quantitative Methods for Ecology
592 and Evolutionary Biology. 2006.
- 593 [14] Marc Mangel, Alec D. MacCall, Jon Brodziak, E.j. Dick, Robyn E. Forrest, Roxanna
594 Pourzand, and Stephen Ralston. A perspective on steepness, reference points, and stock
595 assessment. *Canadian Journal of Fisheries and Aquatic Sciences*, 70(6):930–940, April
596 2013.
- 597 [15] André E. Punt. *Model selection for the dynamics of southern African hake resources*. PhD
598 Thesis, University of Cape Town, 1988.
- 599 [16] André E. Punt, Doug S. Butterworth, Carryn L. de Moor, José A. A. De Oliveira, and
600 Malcolm Haddon. Management strategy evaluation: best practices. *Fish and Fisheries*,
601 17(2):303–334, 2016.
- 602 [17] André E. Punt and Jason M. Cope. Extending integrated stock assessment models to
603 use non-dependant three-parameter stock-recruitment relationships. *Fisheries Research*,
604 217:46–57, September 2019.
- 605 [18] Karthik Ramasubramanian and Abhishek Singh. *Machine learning using R*. Number 1.
606 Springer, 2017.
- 607 [19] Peter Sheldon Rankin and Ricardo T. Lemos. An alternative surplus production model.
608 *Ecological Modelling*, 313:109–126, October 2015.
- 609 [20] William Edwin Ricker. Stock and recruitment. *Journal of the Fisheries Board of Canada*,
610 11(5):559–623, 1954. Publisher: NRC Research Press Ottawa, Canada.
- 611 [21] Jon T Schnute and Laura J Richards. Analytical models for fishery reference points.
612 *Canadian Journal of Fisheries and Aquatic Sciences*, 55(2):515–528, February 1998.
- 613 [22] Luca Scrucca. GA: A Package for Genetic Algorithms in R. *Journal of Statistical Soft-*
614 *ware*, 53:1–37, April 2013.
- 615 [23] Luca Scrucca. On Some Extensions to GA Package: Hybrid Optimisation, Parallelisation
616 and Islands EvolutionOn some extensions to GA package: hybrid optimisation, paralleli-
617 sation and islands evolution. *The R Journal*, 9(1):187–206, 2017.
- 618 [24] Karline Soetaert, Thomas Petzoldt, and R. Woodrow Setzer. Solving Differential Equa-
619 tions in R: Package deSolve. *Journal of Statistical Software*, 33:1–25, February 2010.
- 620 [25] Gerhard Wanner and Ernst Hairer. *Solving ordinary differential equations II*, volume 375.
621 Springer Berlin Heidelberg, 1996.
- 622 [26] Justin D. Yeakel and Marc Mangel. A generalized perturbation approach for exploring
623 stock recruitment relationships. *Theoretical Ecology*, 8(1):1–13, February 2015.