Metamodeling for Bias Estimation of Biological Reference Points

Nicholas Grunloh

May 8, 2023

1 Introduction

set the context for my question

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.

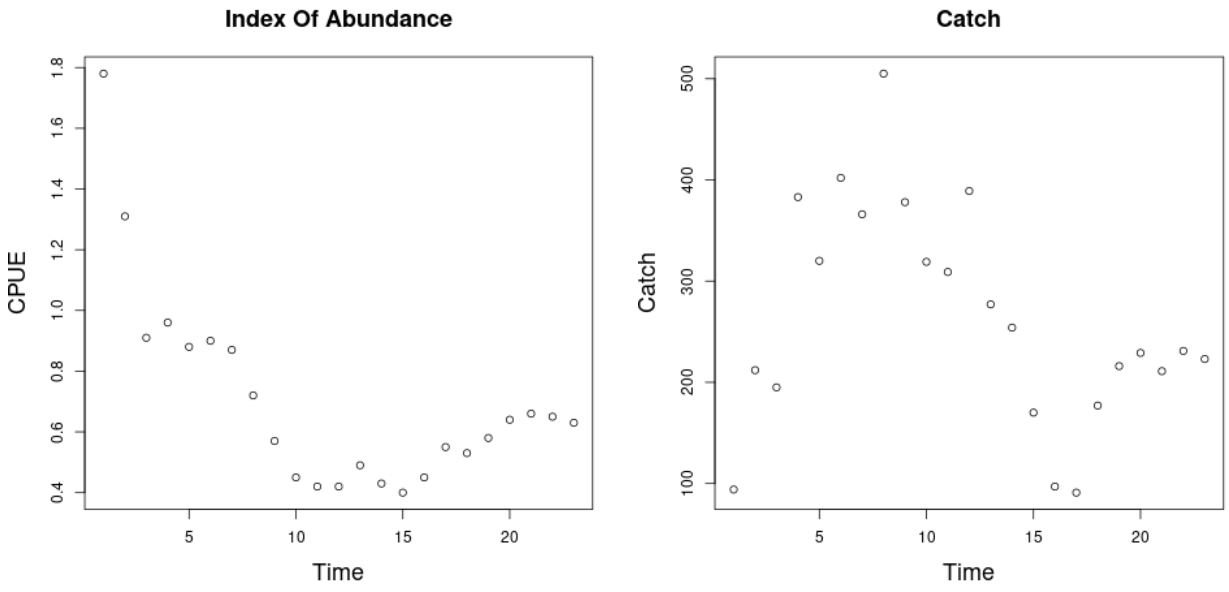

Figure 1: *left*: An observed series of index of abundance data for Namibian Hake from 1965 to 1987 (Hilborn & Mangel, 1997). *right*: The associated catch data for Namibian Hake over the same time period.

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

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

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 modeling 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 $\bar{}$ indicates a value at steady state. MSY is found by maximizing $\bar{Y}(F)$ 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 production models 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 important 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 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 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,

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

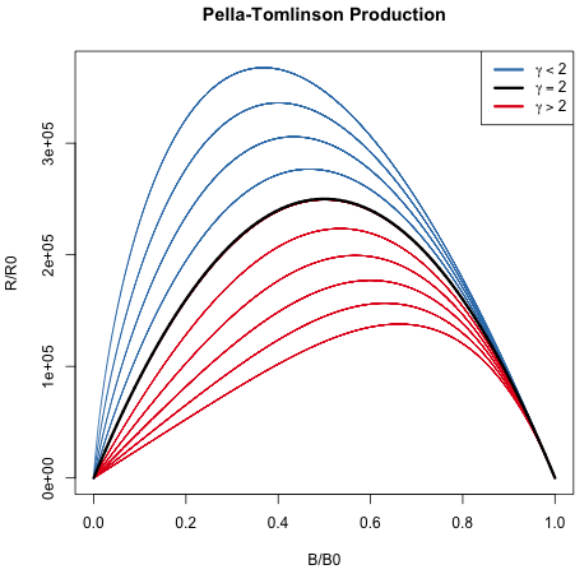

Figure 2: The PT production function plotted across a variety of parameter values. The special cases of Logistic production is shown in black, and the left-leaning and right-leaning regimes are shown in blue and red respectively.

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 .1.1 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 productivity parameters.

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

2 .1.2 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 (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

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

Above W is the Lambert product logarithm function. More details about this derivation, and the Lambert product logarithm, are given in Appendix (5).

Using Eq. (12) to obtain production parameters, a PT production model can be fully

defined for any combination of the RPs F^* and $\frac{B^*}{B(0)}$. Since K does not enter the RP calculation 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 (2 .3). 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 Schnute 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,

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

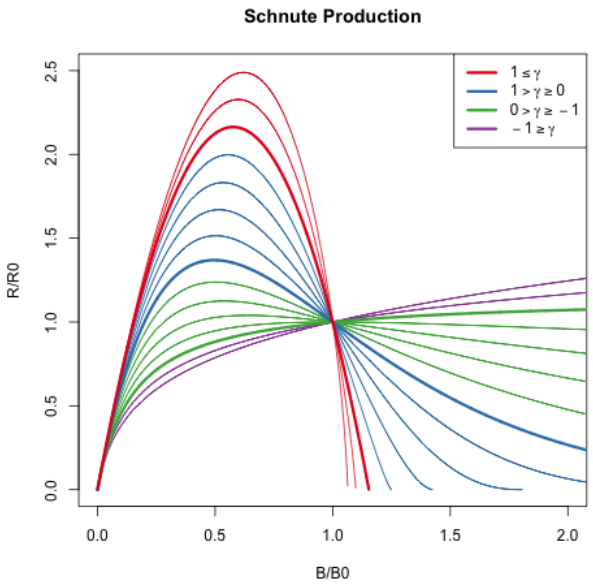

Figure 3: The Schnute production function plotted across a variety of parameter values. The special cases of BH, Ricker, and Logistic production are shown in green, blue, and red respectively.

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 $\gamma \to 0$.

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.

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

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

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 explicitly here. Natural mortality models the instantaneous rate of mortality from all causes outside of fishing. Explicitly modeling natural mortality is not only a typical assumption of fisheries models, but is also key to the making RPs well defined over the relevant 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 yields 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 yield, $\bar{Y} = F\bar{B}(F)$, with respect to F. Analytically maximizing proceeds by differentiating $\bar{Y}$ to produce

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

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

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 yielding here than it was in the case of the PT model. This procedure 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 .2.1 Simulation

For the purposed of simulation, it is not necessary 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 space-filling design over RP space.

Similarly to J. T. Schnute and Richards (1998), expressions (20) and (16) are solved for α and β respectively. This leads to the partial mapping $\left(F_{MSY}, B_0\right) \mapsto \left(\alpha(\cdot, \gamma),\ \beta(\cdot, \cdot, \gamma)\right)$ 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 experiencing 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 brevity, 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 accurately is extremely difficult. Inverting ζ analytically is not feasible, and numerical methods for inverting ζ are unstable and can be computationally expensive. Rather than numerically invert precise values of $\zeta(\gamma)$, γ is sampled so that the overall simulation design is space filling as described in Section (2 .3.2).

Each design location defines a complete Schnute production model with the given RP values. Indices of abundance are simulated from the Schnute model at each design location, a small amount of residual variation, $\sigma = 0.01$, is added to the simulated index, and the data are then fit with a misspecified BH production model. The design at large captures various degrees of model misspecification relative to the BH model, so as to observe the effect of productivity model misspecification upon RP inference.

2 .3 Latin Hypercube Sampling

- a quick lit review of space filling designs

A Latin hypercube sample (LHS) of size n, in the 2 dimensional space defined by RPs, distributes samples so as to spread points across a design region in a broadly representative way. A LHS design extends the notion of a univariate random uniform sample across multiple dimensions so that each margin of the design space enjoys a uniform distribution.

LHS designs achieve this notion of uniformity by first partitioning each dimension of the design space into regular grids of size

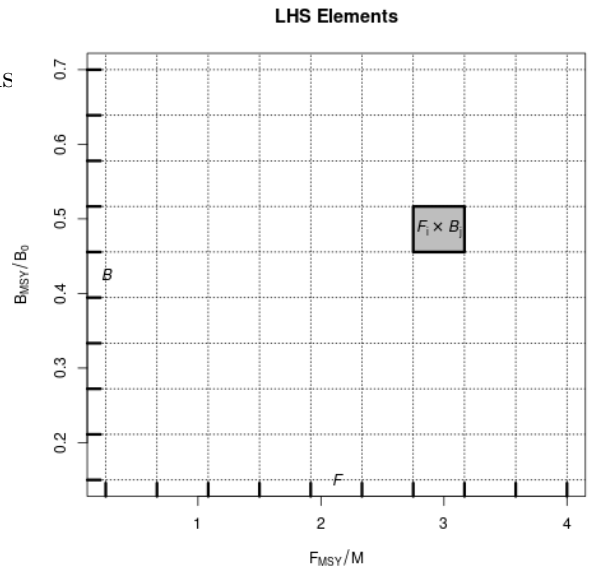

Figure 4: LHS grids. Intersecting $\mathcal{F}$ and $\mathcal{B}$ produces n^2 cells; a particular cell $\mathcal{F}_i \times \mathcal{B}_j$ is shown in grey. Maybe just show points.

n. By intersecting the grids of each dimension, cells are produced that evenly partition the design space. In two dimensions n^2 cells are produced, from which a total of n samples are taken. Crucially 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.

2 .3.1 PT Design

Letting $\mathcal{F}$ and $\mathcal{B}$ be regular grids, of size $n = 100$, on $F_{MSY} \in (0.1, 0.7)$ and $\frac{B_{MSY}}{B_0} \in (0.2, 0.6)$ 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. (12) is used to map the sampled RP design locations to PT productivity parameters.

2 .3.2 Schnute Design

Due to the lack of an analytical relationship mapping RPs $\mapsto \theta$, analogous to the PT model's Eq. (12), producing a LHS design over Schnute RPs requires a more tactful approach. The structured relationship between the RPs and productivity parameters, described in Section (2 .2.1), allows an approximate LHS to be obtained by a careful navigation of the system of equations seen in Eq. (21).

Under the Schnute model, let $\mathcal{F}$ and $\mathcal{B}$ represent regular grids on $\frac{F_{MSY}}{M} \in (0.25, 4)$ and $\frac{B_{MSY}}{B_0} \in (0.15, 0.7)$ respectively which can serve as the scaffolding for computing an approximate LHS

Since it is not practical 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 cumulative distribution function (CDF). The aim is to model γ as an easily sampled random

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: An outline of the sampling procedure for γ given B_0, M, and F_{MSY}.

variable 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,

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

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.

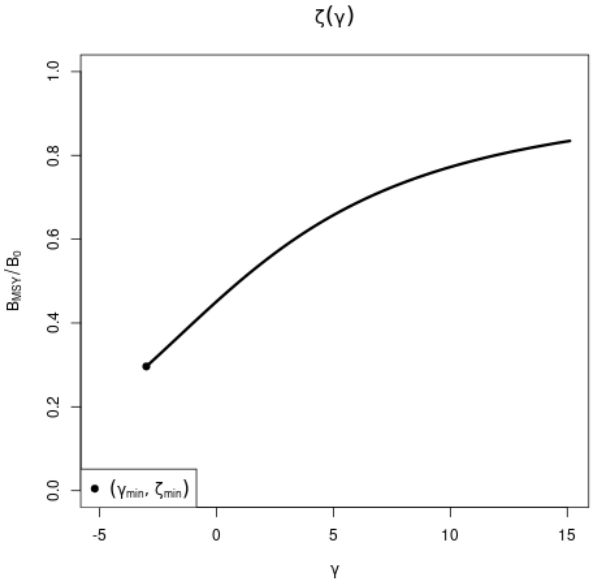

Figure 6: $\zeta(\gamma)$ Plotted for $F_{MSY} = 0.1$ and $M = 0.2$. The point $(\gamma_{min}, \zeta_{min})$ shows the lowest biologically meaningful value of γ; below which productivity is negative.

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 $\zeta(\gamma)$.

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

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 .3.3 Design Refinement

Since the behavior of RP inference, under misspecified models, will vary in yet-unknown 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" instantiations of the initial LHS design may necessitate adaptive design refinement, to accurately describe RP biases. Given the temperamental 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. New design points are collected based on areas of the RP design space which maximizes the minimum dis-

tance between all pairs of points in the current design, based on the following distance function

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

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 normalize 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 at low values of $\frac{B_{MSY}}{B_0}$, the calculation of x_a is further truncated based on a convex hull defined by the existing samples in the overall design.

Design refinement then proceeds as follows. An initial design is computed, $X_n = LHS_n(R_{full})$, 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 .4 Gaussian Process Metamodel

For assessing inference of productivity parameters over the simulated design a GP model is used as a flexible metamodel of how inference responds to various degrees of model misspecification of the restricted model. Design locations, $\boldsymbol{X}$, specify the degree of model misspecification relative to the restricted model. At each design location of the simulation fitting the restricted two parameter model results in a MLE of each of the productivity parameters (i.e. Schaefer:[$log(r)$, $log(K)$], BH:[$log(\alpha)$, $log(\beta)$]). Furthermore, since the maximum likelihood estimator is a random variable, MLE standard error estimates, on the variance scale (via the inverted Fisher information) are also outputs of the simulation. 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 at each $\mathbf{y}$. This simulation can be seen as the following mapping

$$\boldsymbol{X} \mapsto \mathbf{y} \pm \sqrt{\boldsymbol{\omega}}. \tag{26}$$

By constructing a metamodel of this mapping, it allows for a full characterization of inference under the misspecified restricted models.

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 of the fitted productivity parameter estimates are then modeled using independent

instances of the following GP metamodel.

$$\begin{aligned} \mathbf{y} &= \beta_0 + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v} + \boldsymbol{\epsilon} \\ \boldsymbol{v} &\sim N_n(\mathbf{0}, \tau^2 \boldsymbol{R}_{\boldsymbol{\ell}}) \\ \boldsymbol{\epsilon} &\sim N_n(\mathbf{0}, \boldsymbol{\omega}' \boldsymbol{I}) \end{aligned} \tag{27}$$

$\boldsymbol{X}$ is the n x 2 LHS design matrix of RPs, as derived above, for each respective three parameter data generating model. 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 heterogeneously.

The term, $\boldsymbol{v}$, contains spatially correlated GP effects. The correlation matrix, $\boldsymbol{R}_{\boldsymbol{\ell}}$ 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,

$$R(\boldsymbol{x}, \tilde{\boldsymbol{x}}) = \exp\left(\sum_{i=1}^{2} \frac{-(x_i - \tilde{x}_i)^2}{2\ell_j^2} \right). \tag{28}$$

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.

Fitting the metamodel allows for a full predictive description of inference under the misspecified restricted models. Predictive estimates are obtained via kriging (cite).

$$\hat{y}(\mathbf{x}) = \beta_0 + \mathbf{x}\boldsymbol{\beta} + \mathbf{r}(\mathbf{x})' \boldsymbol{R}_{\boldsymbol{\ell}}^{-1} \Big(\mathbf{y} - \big(\beta_0 + \boldsymbol{X}\boldsymbol{\beta} \big) \Big) \tag{29}$$

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

uses

2 .5 Catch

It is known that contrast in the observed index and catch time series 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 (7, *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. Thus catch is parameterized so as to allow for a spectrum of possible contrast simulation settings.

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. 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^*. When $\frac{F(t)}{F^*}$ is held constant in time 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 the steady state biomass equations 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 ultimatly brings fishing into equilibrium at F_{MSY}. 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.

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

The first term of Eq(30) 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_{MSY} 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 seen in Eqs (31, 32).

$$a = e^{\log(\chi_{max})-15b} \qquad b = \frac{1}{t-15}\log\left(\frac{\chi_{min}}{\chi_{max}}\right) \tag{31}$$

$$c = \frac{\chi_{max}-1}{15-1} \qquad d = 15c + \chi_{max} \tag{32}$$

By further specifying the following, the two parameters χ_{max}, and χ_{min} can be reduced to the single parameter χ.

$$\chi_{max} = 1.6^{\chi} \qquad \chi_{min} = 0.4^{\chi} \tag{33}$$

The tuning parameter χ singularly controls contrast that appears in time series data.

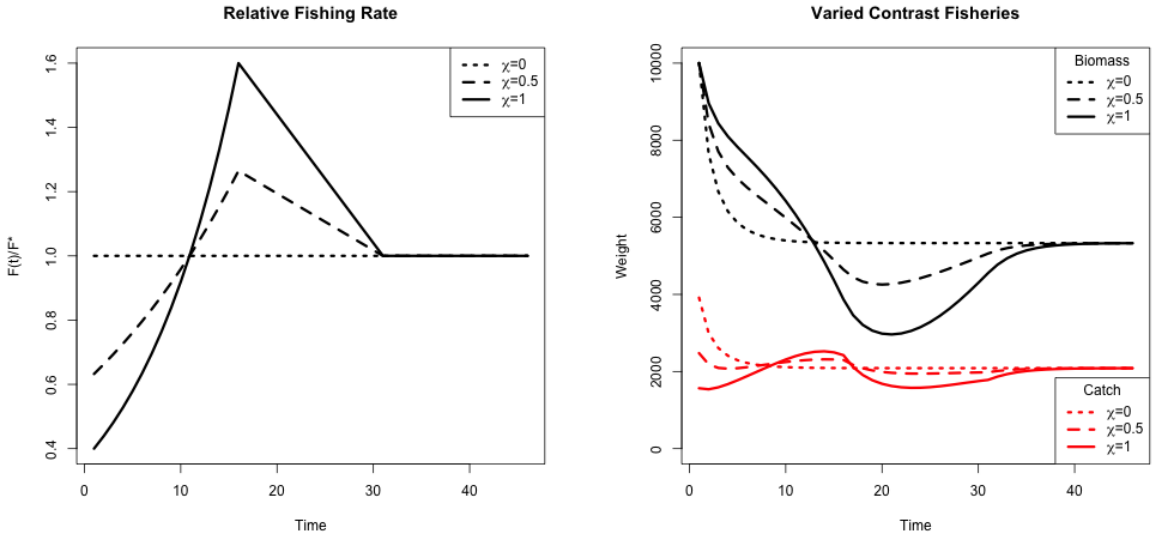

Figure 7: (*left*) Relative fishing with low, medium, and high contrast. (*right*) Population biomass and catch at each associated level of contrast.

When $\chi = 0$, the relative fishing rate is a constant at 1 to create a low contrast simulation environment. As χ increases Eq (30) induces more and more contrast in the observed index and catch time series until $\chi = 1$ which produces a high contrast simualtion environment. Figure (7) demonstrates a spectrum of contrast simulation environments as well as the time series data they induce in the solution of the production model ODE.

2 .6 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(2)). 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(2) 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 discretized 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 .6.1 Integration and Stiffness

As previously mentioned, the overwhelming majority of implementations of population dynamics models discretized 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 since it runs relatively quickly and has a relatively smaller footprint in system memory.

3 Results

3 .1 PT/Schaefer

3 .1.1 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 (Hilborn & Walters, 1992).

Figure (8) 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 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 (8) 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 (8) 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

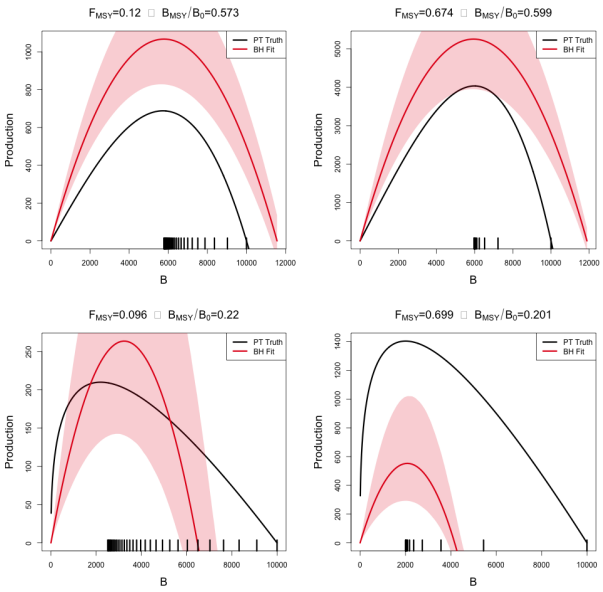

Figure 8: A comparison of the true PT production function (in black) and the estimated logistic curve (in red) with 95% CI shown. The examples shown represent the four corners of maximum model misspecification in the simulated RP-space. Observed biomasses are plotted in the rug plots below the curves.

logistic parabola estimate tends to under estimate F^*.

3 .1.2 Metamodeled Trends

Each point in the space of the RPs F_{MSY} 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_{MSY}. So the line through RP space defined by $\frac{B^*}{B(0)} = 0.5 \ \ \forall \ \ F_{MSY}$, defines the subset of RP space where $\gamma = 2$ and where the PT model is equivalent to the Schaefer model. For brevit this subset of RP were $\frac{B^*}{B(0)} = 0.5$ will be refered 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 (8) 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. (11), 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 inferencial RP bias, induced by model misspecification of the Schaefer model, can be seen to be seen.

Figure (9) 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 (9) 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 $\frac{B_{MSY}}{B_0}$ must be 0.5 under the Schaefer model, biases in the $\frac{B_{MSY}}{B_0}$ direction must simply map vertically onto the Schaefer set. Due to this simplified RP geometry under the

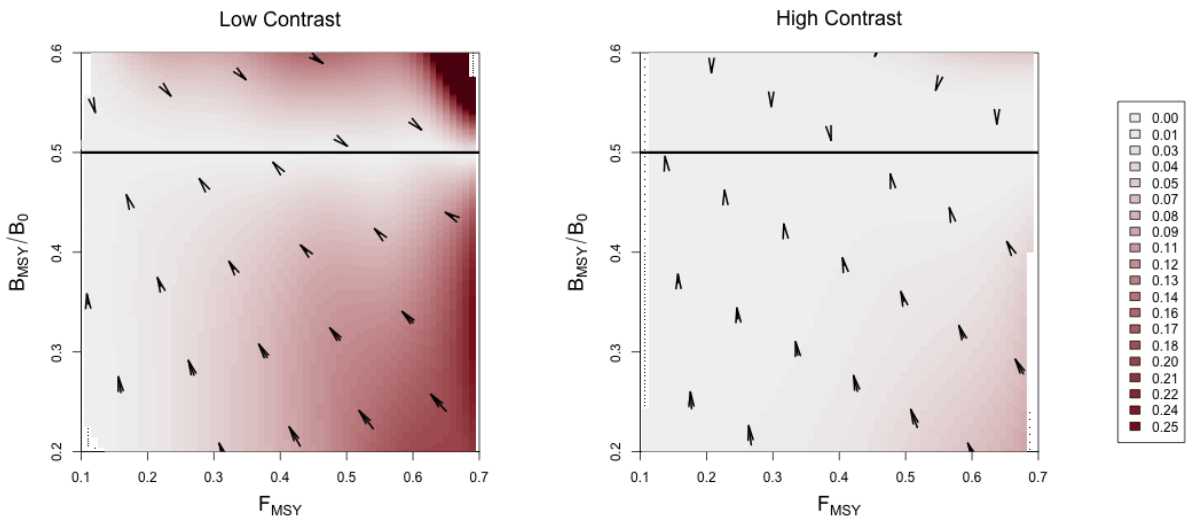

Figure 9: Joint bias direction for (F_{MSY}, $\frac{B_{MSY}}{B_0}$) estimates under the misspecified Schaefer Model. The intensity of color represents the excess bias relative to the shortest possible mapping. Results in the low contrast setting are shown $left$, and the high contrast setting is shown $right$.

Schaefer model, the degree of bias in $\frac{B_{MSY}}{B_0}$ estimation is entirely defined solely by the degree of model misspecification irrespective of F_{MSY}. Furthermore, the closest possible point along the Schaefer set that Schaefer model inference could map RPs would be the perfectly vertical mapping. This pattern only contains the strictly neccessary bias present in $\frac{B_{MSY}}{B_0}$, and zero bias in F_{MSY}. Any deviation from this minimal bias pattern necessarily to be due to added bias in F_{MSY}.

The two simulation settings shown in Figure (9) are identical except for the amount of contrast present in the simulated index. The left panel of Figure (9) shows RP biases in the low contrast setting, while the right panel shows the high contrast setting. Notice that in the low contrast setting the RP bias pattern is far from the minimum distance mapping, however when contrast is added the mapping becomes much closer to a minimal bias mapping. In the low contrast setting the observed bias is consistent with the pattern and mechanism described in Figure (8), where F_{MSY} is underestimated for data generated below the Schaefer line and overestimated above the Schaefer set. In the high contrast simulation the mapping is nearly minimal distance with the exception of PT data generated with simulatiously low $\frac{B_{MSY}}{B_0}$ and high F_{MSY}.

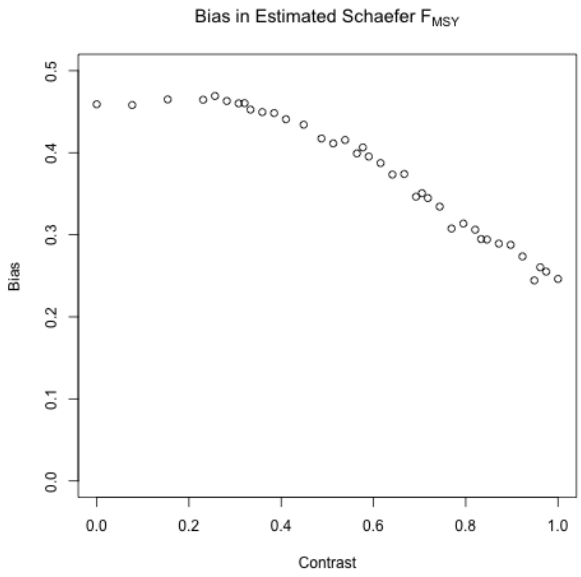

Figure 10: Bias in F_{MSY} under the Schaefer model when PT data are generated with increasing contrast so that F_{MSY} and $\frac{B_{MSY}}{B_0}$ are fixed at 0.699 and 0.201 respectively.

Figure (3 .1.2) demonstrates how bias in F_{MSY} estimation decreases as contrast is added to PT data as generated in the low $\frac{B_{MSY}}{B_0}$ and high F_{MSY} regime. By including additional contrast F_{MSY} bias is decreased, however parameterizing contrast so as to fully exstinguish F_{MSY} bias may require a more complex model of fishing.

3 .2 Schnute/BH

3 .2.1 Design

Algorithm (1) enforces uniform marginals in $\frac{F_{MSY}}{M}$ directly, as well as the adherence of the overall design to latin squares. Figure (11) shows a uniform Q-Q plot for sampled ζ, using Algorithm (1), against theoretical uniform quantiles. As evidence by the excellent coherence to the theoretical uniform quantiles, the approximation in Section (2 .3.2) for sampling γ (and therefore $\zeta(\gamma)$), is very effective. Furthermore since numerical inversion of $\zeta(\gamma)$ is costly and unreliable, the relative speed and accuracy that this approx-

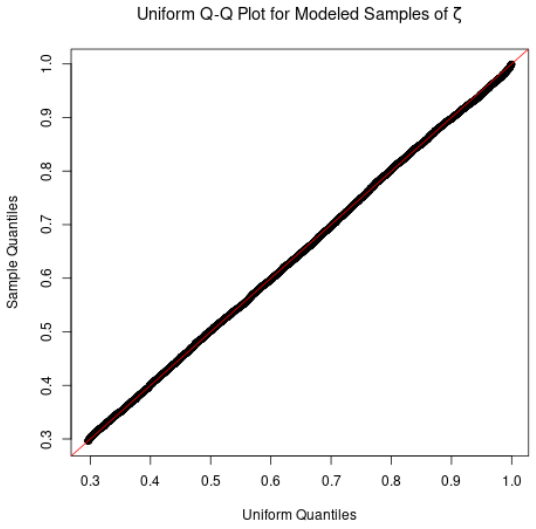

Figure 11: Uniform Q-Q plot for ζ plotted for $F_{MSY} = 0.1$ and $M = 0.2$.

imate LHS sampling method provides is pivotal for the rest of the work presented here.

Similarly to the PT model, the three parameter Schnute model is uniquely identified by each point in the space of $\frac{F_{MSY}}{M}$ and $\frac{B_{MSY}}{B_0}$ RPs. As seen in Figure (12), Schnute production has different behaviors in different ranges of RPs space, which are entirely defined by the value of γ (shown in Figure (3)). When $\gamma \geq 1$ the Schnute model produces a family of Logistic-like curves that are increasingly right leaning as γ increases. For $1 > \gamma \geq 0$, Schnute production takes a family of left leaning Ricker-like curves that all, at least, approach the x-axis. For $0 > \gamma > -1$ there are a family of BH-like curves that do not approach the x-axis but still have decreasing productivity for large biomass stocks. When γ is exactly -1 Schnute reduces to BH production which has asymptoting production for large biomass. Finally when $-1 > \gamma$, Schnute produces a family of increasing curves that do no asymptote, and produce Cushing-like production as γ becomes large.

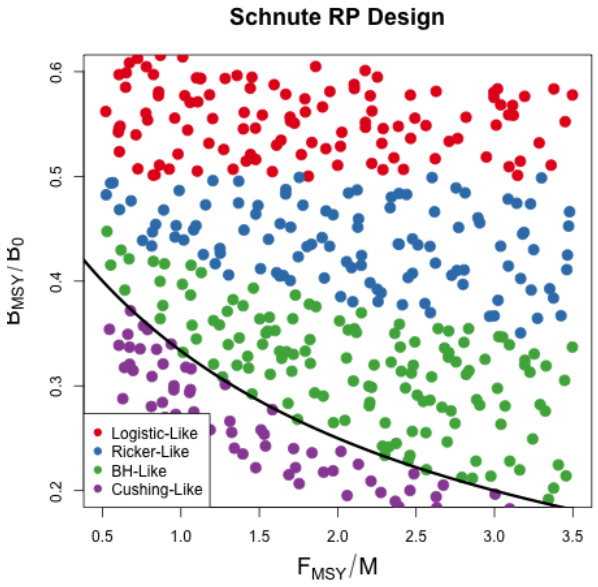

Figure 12: A Schnute RP design. Colors indicate different regimes of Schnute production. The black curve shows the BH set.

Modeling index data that are simulated broadly over the theoretical space of RPs with misspecified BH production greatly limits the range of possible RPs that can be inferred. Under BH production the full theoretical space of RPs are limited to the curve $\frac{B_{MSY}}{B_0} = \frac{1}{F_{MSY}/M+2}$. Define the "BH set" as the set of RPs defined by this limited space, i.e. the curve $\left\{ \left(\frac{B_{MSY}}{B_0}, \frac{F_{MSY}}{M} \right) \middle| \frac{B_{MSY}}{B_0} = \frac{1}{F_{MSY}/M+2} \right\}$. as seen in the black curve in Figure (12). The farther away from this set that Schnute data are simulated, the worse the BH model is misspecified for those data.

3 .2.2 Metamodeled Trends

Unlike the Schaefer model, the BH set is not a constant in $\frac{B_{MSY}}{B_0}$. Under the BH model, bias in $\frac{B_{MSY}}{B_0}$ is no longer entirely defined by the degree of model mispecification, but rather the structure of BH RPs allows bias in both $\frac{B_{MSY}}{B_0}$ and $\frac{F_{MSY}}{M}$ to interact as a function of contrast in the data.

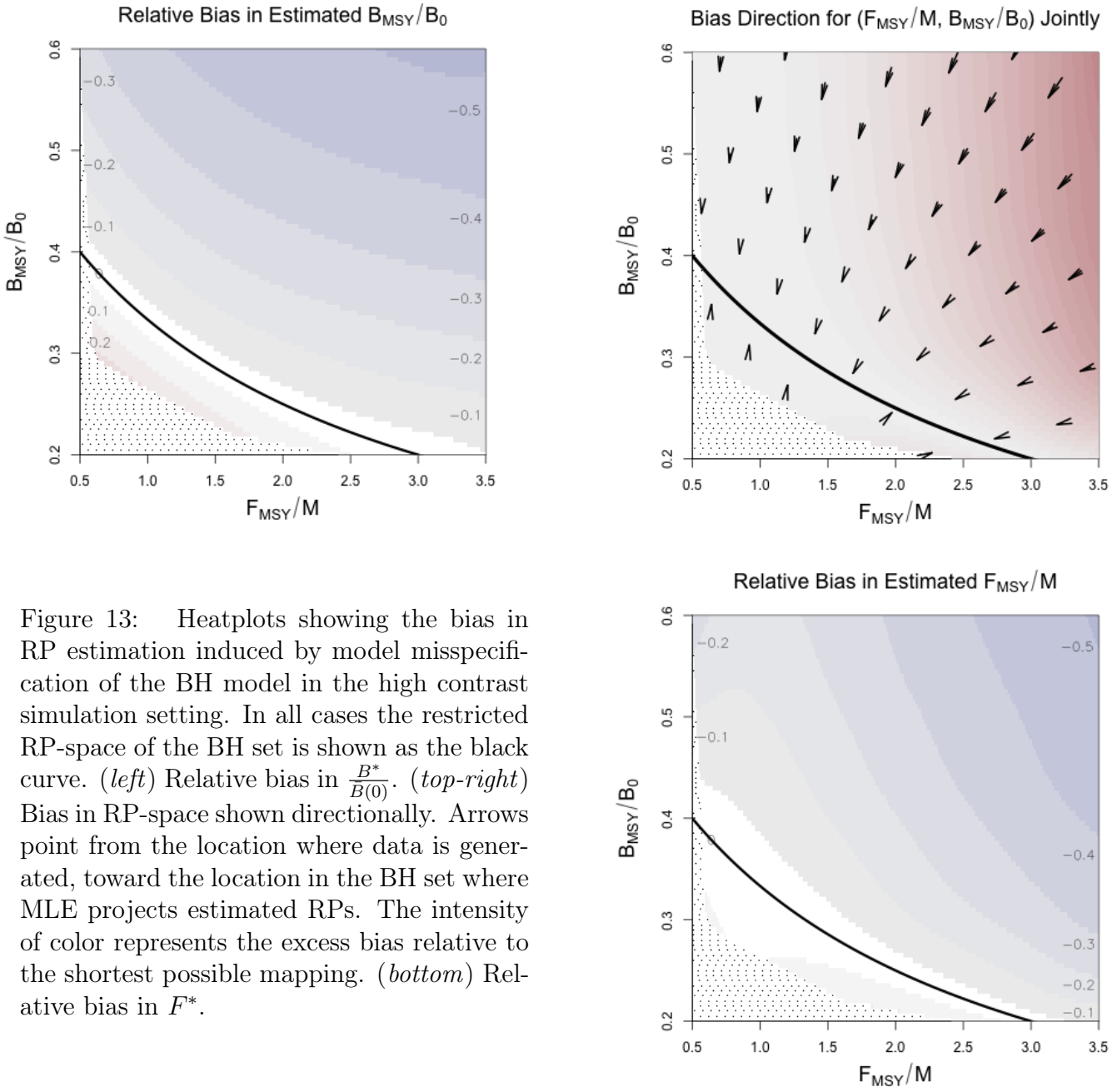

Figure 13: Heatplots showing the bias in RP estimation induced by model misspecification of the BH model in the high contrast simulation setting. In all cases the restricted RP-space of the BH set is shown as the black curve. (*left*) Relative bias in $\frac{B^*}{B(0)}$. (*top-right*) Bias in RP-space shown directionally. Arrows point from the location where data is generated, toward the location in the BH set where MLE projects estimated RPs. The intensity of color represents the excess bias relative to the shortest possible mapping. (*bottom*) Relative bias in F^*.

Figure (13) shows metamodeled RP bias surfaces for inference under the BH model in the high contrast setting. The (*left*) and (*bottom*) panels focus only on the $\frac{B^*}{B(0)}$ and $\frac{F_{MSY}}{M}$ components of bias respectively. In these panels bias is shown as relative bias, $\frac{\hat{RP}-RP}{RP}$, similar to a percent error calculation. Where RP represents the true value of the three parameter RP, and $\hat{RP}$ refers to the metamodel estimate.

Figure (13, *top-right*) combines the components of bias to show the overall mapping of RPs under BH inference in the high contrast simulation setting. Unlike RP inference under the Schaefer model in the high contrast setting, the BH model does shows significant bias in both RPs here. Despite the bias in $\frac{B^*}{B(0)}$ and $\frac{F_{MSY}}{M}$ these results are similar to that of the Schaefer model in that the overall mapping of RPs is very near to a minimal distance mapping onto the constrained set of RPs.

Figure (15) shows the mapping of RPs in the low contrast simulation setting. Figures (15) and (13, *top-right*) share a common scale for the intensity of color to facilitate comparison. Notice that the mildly misspecified area around the BH set produces mappings onto the BH set which resemble the minimal distance mapping seen in Figure (13, *top-right*) under the high contrast simulation setting. The primary difference in the low constrast setting, seen here, is the break point around $\frac{B^*}{B(0)} = 0.4$ above which $\frac{F_{MSY}}{M}$ is sharply underestimated. By comparison of Figure (15), with the location of the BH-Like/Ricker-Like regimes seen in Figure (12), this break point coincides with a regime change in the Schnute model. The Ricker model lies along this regime interface, and represents the first model to approach the x-axis for large biomasses as γ increases. However, even in the low contrast setting the BH model manages to recover the minimal distance mapping for data generated under the

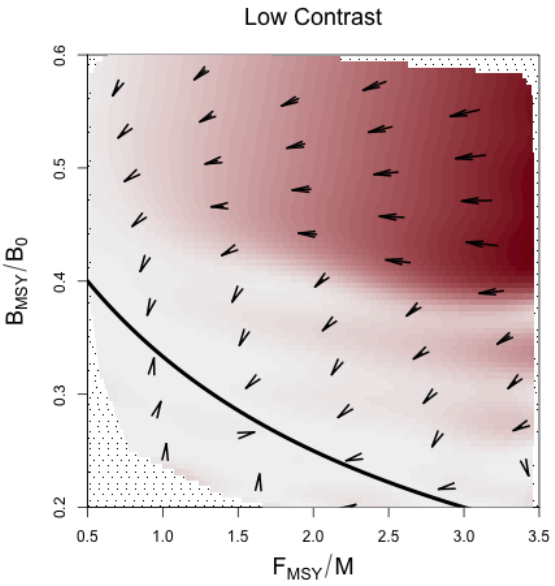

Figure 14: Joint bias direction of RP inference in the low contrast simulation setting. The intensity of color represents the excess bias relative to the shortest possible mapping.

Cushing-like or BH-like regimes.

- "safe" regiem similar to the high contrast minimal distance mapping for mild model misspcificaiton.
- Ricker-like break point. Complete model failure beyond.
- what is going on with SRR when model completely fails.
- Mechanism plot on SRR LHS Grid plot
- mapping distance as a function of contrast at (3.5, 0.5)
- for LHS grid locations show $\frac{B_{MSY}}{B_0}$ and F_{MSY} biases for grids in $M \in (0, 0.5)$ For sure in High Contrast, maybe also in Low??.

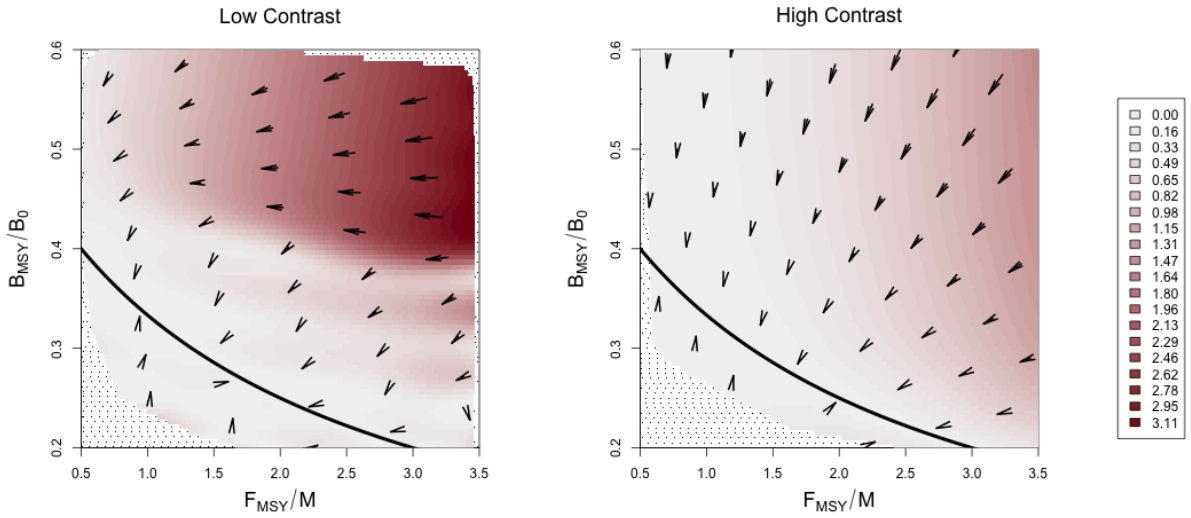

Figure 15: Joint bias direction of RP infernce in the low contrast simulation setting. The intensity of color represents the excess bias relative to the shortest possible mapping.

- summary of σ over RP space comparing between models (PT, Schnute, Schnute DD) to show areas of model breakdown.
  - miss-identifying signal for noise.
  - It happens more as the dynamics get more complex.
  - point to the full age structed models.

4 Discussion

- model mispecification story as in advancement
- inference is oriented to the shape of the constrained space
- conjecture that with perfectly informed models mapping would be orthoganal
- Best case data result in a minimum distance mapping (discussion?)
- Due to the shape of the BH RP set this ensures bias in both $\frac{B_{MSY}}{B_0}$ and F_{MSY}
- Unlike the Schaefer models geometry the BH set has a geometry that will mix bias directions in the best case.

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

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

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

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.

$$\begin{aligned}
\gamma\log(\zeta) &= W\left(\zeta\log(\zeta)\right) \\
\gamma &= \frac{W\left(\zeta\log(\zeta)\right)}{\log(\zeta)}
\end{aligned} \tag{34}$$

The Lambert product logarithm is a multivalued 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,

$$\gamma = \begin{cases} \frac{W_0(\zeta\log(\zeta))}{\log(\zeta)} & \zeta \in \left(0, \frac{1}{e}\right) \\ \frac{W_{-1}(\zeta\log(\zeta))}{\log(\zeta)} & \zeta \in \left(\frac{1}{e}, 1\right) \end{cases}. \tag{35}$$

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

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

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*.

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 as-

sessment 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

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

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. 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