Meta-Modeling for Bias Estimation of Biological Reference Points Under the Schaefer Model

Nicholas Grunloh

January 30, 2022

Abstract

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. (2013) 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. This study's findings suggest the useful limits of the misspecified Schaeffer model, and mechanisms of model failure which arise from mapping a three-dimensional parameter space into two dimensions.

1 Introduction

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

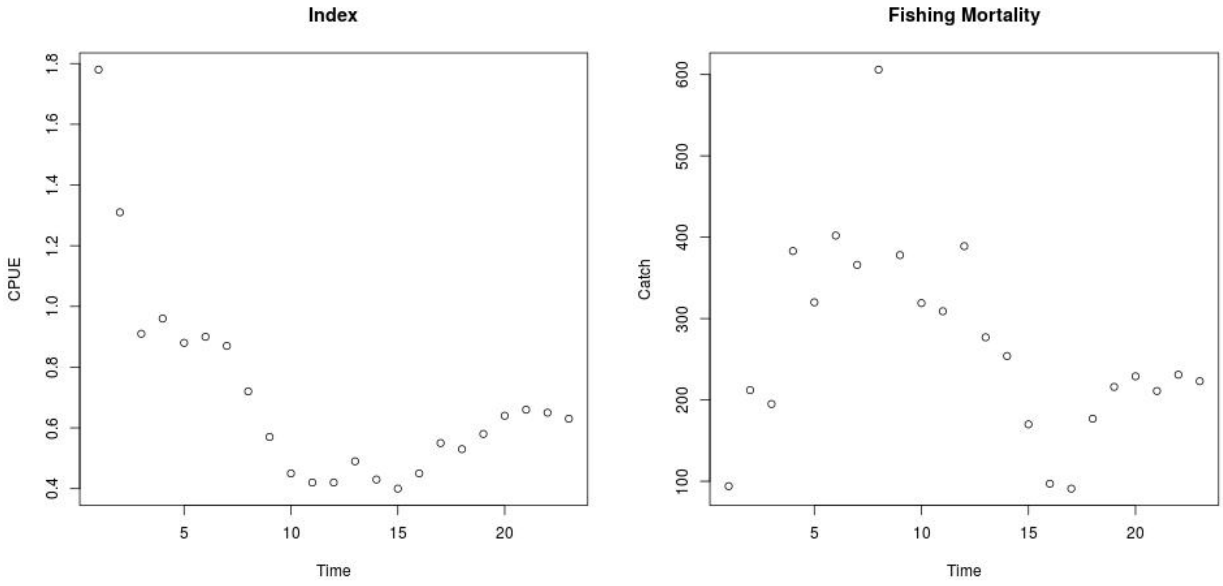

Figure 1: *left*: An observed series of index of abundance data for Namibian Hake from 1965 to 1987 (Hilborn & Mangel, 1997). *right*: The associated catch data for Namibian Hake over the same time period.

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 ordinary differential equation; in this case I focus on the following form,

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

Here biomass is assumed to change in time by two processes, net production of biomass into the population, and catches removing biomass from the population.

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

Secondly, the population decreases as biomass is removed due to catch, $C(t)$. While catches (aka yields) are observable quantities (Sen, 1984; Pearson & Erwin, 1997), the model assumes that catch is proportional to biomass with the proportionality constant representing the fishing rate, $F(t)$, so that $C(t) = F(t)B(t)$. From a management perspective a major goal of the model is to accurately infer a quantity known as *maximum sustainable yield* (MSY). One could maximize simple yield at a particular moment in time (and only for that moment) by fishing all available biomass in that moment. This strategy is penny-wise but pound-foolish (not to mention ecologically devastating) since it doesn't leave biomass in the population to reproduce for future time periods. We seek to fish in a way that allows (or even encourages) future productivity in the population. This is accomplished by maximizing the equilibrium level of catch over time. Equilibrium yield is considered by replacing the steady state biomass ($\bar{B}$) in the assumed form for catch, so that $\bar{C} = F\bar{B}(F)$, where $\bar{}$ indicates a value at steady state. Naturally the steady state biomass is a function of F; we will see a specific example of this in Section (2.2). MSY is found by optimizing $\bar{C}(F)$ with respect to F, and F^* is the fishing rate at MSY. Going forward let * decorate any value derived under the condition of MSY.

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

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

r is a parameter controlling the maximum reproductive rate of the population in the absence of competition for resources (i.e. the slope of SRR 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. $\bar{B}(0) = K$).

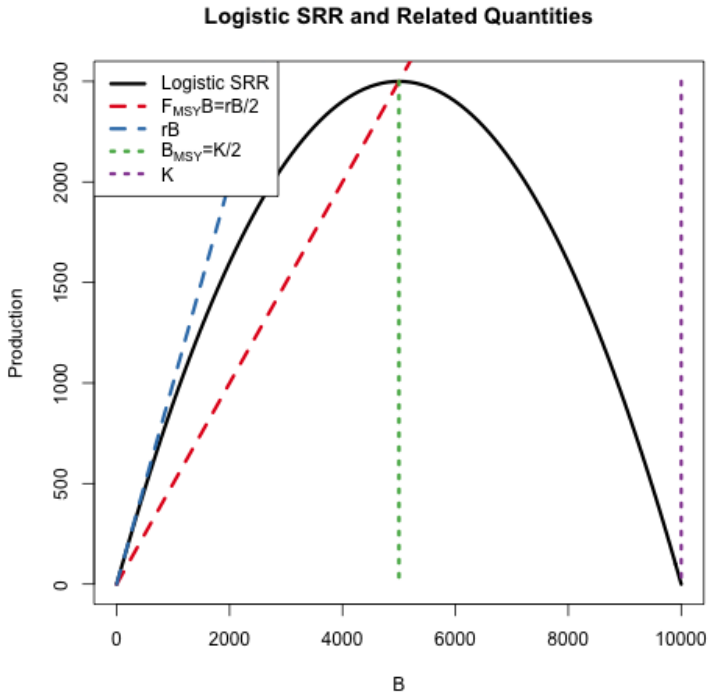

Figure 2:
The logistic production function in black plotted next to depictions of the key biological parameters and reference points. The slope at the origin (and thus r) is shown in blue, catch resulting in MSY in red, biomass at MSY in green, and K in purple at the right x-intercept. MSY is seen at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

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

maximum reproductive rate at the origin.

Fisheries are very often managed based upon reference points (RP) which serve as simplified heuristic measures of population behavior. The mathematical form of RPs depends upon the model assumptions primarily through the production function. Here the focus is primarily on the RPs F^* and $\frac{B^*}{B(0)}$.

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 logistic production these quantities take the following form,

$$F^* = \frac{r}{2} \qquad\qquad \frac{B^*}{B(0)} = \frac{1}{2} \tag{4}$$

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

In current practice, production functions are typically chosen to depend only on two parameters. The Schaefer model as presented depends only on the biological parameters r and K, but other common two parameter choices of the production function are the Beverton-Holt (BH) and Ricker curves. All of these two parameter production functions struggle similarly to model the full theoretical space of RPs (Mangel et al., 2013).

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

The bias-variance tradeoff (Ramasubramanian & Singh, 2017) makes it clear that the addition of a third parameter in the production function will neccessarily 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 managment quantities are not fully understood or described. Lee et al. (2012) provides some evidence that steepness estimation, and thus RPs via (Mangel et al., 2013), is dependent on model misspecification as well as data lacking contrast in biomass. However the particular behaviors of these biases are not fully described.

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

A key insight of this work is that bias is considered broadly across RP-space to uncover patterns of bias magnitude and direction. A flexible Gaussian Process (GP) meta-modeling approach is used to model bias patterns across RP-space. This technique not only informs the utility of reducing bias, but also suggest mechanisms for understanding what causes bias. Further, the effect of contrast on estimation is considered alongside the effect of model misspecification.

- ✓ Lead into this reaserch with other studies (Lit review)
- Enphasize a clear statement of the study goal
- a statement of novilty
  - ✓ model misspecification leads to bias

Sorta nonconstant bias of RP-space (a wholistic summary of model performance.)

Sorta methodology to account for bias in the presence of uncertainty

2 Methods

2.1 PT Model

The three parameter PT family has a convenient form that includes the logistic SRR as a special case to form the Schaefer model. The Pella-Tomlinson SRR is parameterized so that $\boldsymbol{\theta} = [r, K, \gamma]$ and the family takes the following form,

$$P(B; [r, K, \gamma]) = \frac{rB}{\gamma - 1}\left(1 - \frac{B}{K}\right)^{\gamma - 1}. \tag{5}$$

γ is a parameter which breaks PT out of the restrictive symmetry of the logistic curve. The parameters r and K maintain the same interpretation as they do in the logistic production function. In Figure (3) PT recruitment is shown for a range of parameter values so as to demonstrate the various recruitment shapes that can be achieved by PT recruitment.

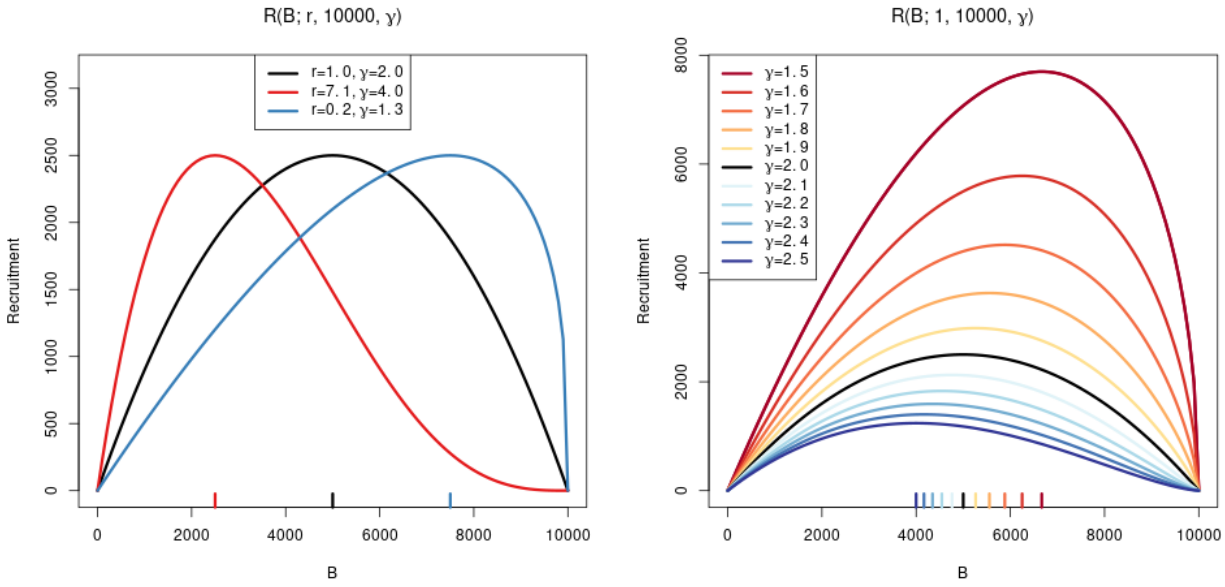

Figure 3: (*left*) PT production functions with parameters chosen so that MSY is consistent, but $\frac{B^*}{B(0)}$ is less than $\frac{1}{2}$ (in red), greater than $\frac{1}{2}$ (in blue), or equal to $\frac{1}{2}$ (in black; logistic production function). (*right*)

While the particular form of how γ appears in PT still produces some limitations to the form of the SRR, importantly the introduction of a third parameter allows enough

flexibility to fully describe the space of reference points used in management. To see this, the reference points are analytically derived for the PT model below.

2.2 PT Reference Points

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

$$\frac{dB}{dt} = \frac{rB}{\gamma - 1}\left(1 - \frac{B}{K}\right)^{\gamma - 1} - FB. \tag{6}$$

An expression for the equilibrium biomass is attained by setting Eq(6) equal to zero, and rearranging the resulting equation to solve for B. Thinking of the result as a function of F gives,

$$\bar{B}(F) = K\left(1 - \left(\frac{F(\gamma - 1)}{r}\right)^{\frac{1}{(\gamma - 1)}}\right). \tag{7}$$

At this point it is convenient to notice that $\bar{B}(0) = K$. The expression for B^* is given by evaluating Eq(7) at F^*.

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

$$F^* = \underset{F}{\operatorname{argmax}}\, F\bar{B}(F). \tag{8}$$

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

$$\frac{d\bar{Y}}{dF} = \bar{B}(F) + F\frac{d\bar{B}}{dF} \tag{9}$$

$$\frac{d\bar{B}}{dF} = -\frac{K}{F(\gamma - 1)}\left(\frac{F(\gamma - 1)}{r}\right)^{\frac{1}{\gamma - 1}}. \tag{10}$$

Setting Eq(9) equal to 0, substituting $\bar{B}(F)$ and $\frac{d\bar{B}}{dF}$ by Equations (7) and (10) respectively, and then solving for F produces the following expression for the fishing rate

required to produce MSY,

$$F^* = \frac{r}{\gamma - 1}\left(\frac{\gamma - 1}{\gamma}\right)^{\gamma - 1}. \tag{11}$$

Plugging the above expression for F^* back into Eq(7) gives the following expression for biomass at MSY,

$$B^* = \frac{K}{\gamma}. \tag{12}$$

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

$$F^* = \frac{r}{(\gamma - 1)}\left(\frac{\gamma - 1}{\gamma}\right)^{\gamma - 1} \qquad \frac{B^*}{\bar{B}(0)} = \frac{1}{\gamma} \tag{13}$$

2.3 Simulation Study

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

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

$$r = F^*\left(\frac{1 - \frac{B^*}{\bar{B}(0)}}{\frac{B^*}{\bar{B}(0)}}\right)\left(1 - \frac{B^*}{\bar{B}(0)}\right)^{\left(\frac{\frac{B^*}{\bar{B}(0)} - 1}{\frac{B^*}{\bar{B}(0)}}\right)} \qquad \gamma = \frac{1}{\frac{B^*}{\bar{B}(0)}}. \tag{14}$$

Indices are generated under the following conditions. Data are simulated at each point on the grid $\mathcal{F} \times \mathcal{B}$, with $F^* \in \mathcal{F}$ and $\frac{B^*}{B(0)} \in \mathcal{B}$, where $\mathcal{F} = \{0.1, 0.2, ..., 0.7\}$ and $\mathcal{B} = \{0.2, 0.3, ..., 0.6\}$ as seen in Figure (4). These ranges of values for F^* and $\frac{B^*}{B(0)}$ are selected to include the majority of most commonly assessed fish species cite. The red X's in Figure (4) show four simulation locations where the Schaefer model is misspecified to a large degree and will be considered in more detail in Section(3.1). For each $\left(F^*, \frac{B^*}{B(0)}\right)$, the associated pair (r, γ) are computed from Eq (14). Since K does not enter the RP calculation its value is fixed arbitrarily at 10000. The value of q is fixed at a typically small value of 0.0005. σ is fixed at the relatively small value of 0.01 to focus specifically on the behavior of population parameters. These parameters fully specify the PT model for the purposes of generating index data for each $\left(F^*, \frac{B^*}{B(0)}\right)$ pair.

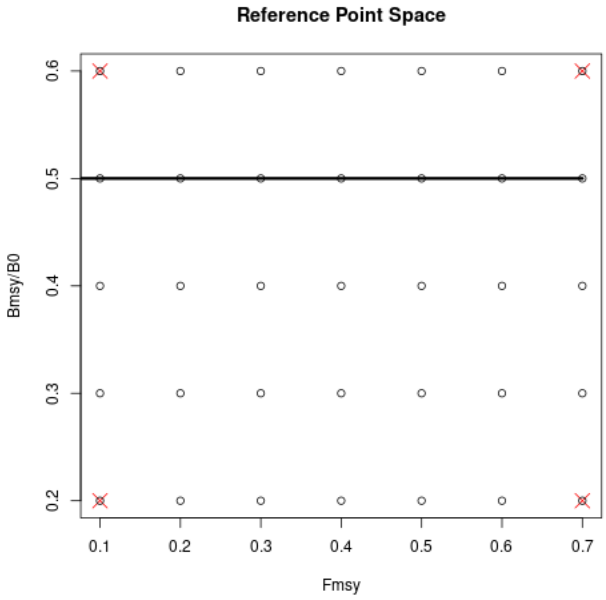

Figure 4: Open circles show the location of the simulation grid $\mathcal{F} \times \mathcal{B}$. The horizontal line shows the constrained space of RPs for the Schaefer model. The red X's indicated 4 simulation locations where the Schaefer model is a particularly misspecified.

2.4 Catch

It is known that the behavior of catch can effect inference on the biological parameters (Hilborn & Walters, 1992) cite walters). In particular it is understood that catch data must induce "contrast" in index data so as to well inform r. It is not well understood how contrast may factor into biases induced by model misspecification. To investigate the role that catch plays in understanding model misspecification errors a variety of catches are investigated.

Catch is parameterized so that $F(t)$ can be controlled with respect to F^*. Recall that catch is assumed to be proportional to biomass with the proportionality constant amounting to the fishing rate, so that $C(t) = F(t)B(t)$. To control $F(t)$ with respect to F^*, $C(t)$ is specified by defining the quantity $\frac{F(t)}{F^*}$ as the relative fishing rate. $B(t)$ is defined by the solution of the ODE, and F^* is defined by the biological parameters of the model, see Eq(11). Thus by defining $\frac{F(t)}{F^*}$, catch can then be written as $C(t) = F^* \left(\frac{F(t)}{F^*}\right) B(t)$.

Intuitively $\frac{F(t)}{F^*}$ describes the fraction of F^* that $F(t)$ is specified to for the current $B(t)$. When $\frac{F(t)}{F^*} = 1$, $F(t)$ will be held at F^*, and the solution of the ODE brings $B(t)$ into equilibrium at B^*. For constant $\frac{F(t)}{F^*}$ the Schaefer model has an analytical solution, via integrating factors, so that $B(t)$ comes to equilibrium exponentially. The relative fishing rate is defined on $[0, \infty)$; when $\frac{F(t)}{F^*} < 1$, $F(t)$ is lower than F^* and $B(t)$ is pushed toward $\bar{B} > B^*$. Contrarily, when $\frac{F(t)}{F^*} > 1$, $F(t)$ is higher than F^* and $B(t)$ is pushed toward $\bar{B} < B^*$; the precise values of $\bar{B}$ can be calculated from Eq(7).

In practice, catch is determined by a series of observed, assumed known, catches. Catch observations are typically observed on a quarterly (or yearly) basis, so that the ODE may be discretized via Euler's method with integration step sizes to match the observation frequency of the modeled data. In this case, catch is sampled as would be done in practice however, the simulation can encounter a variate of issues working with the naively discretized ODE. As a result the ODE is integrated implicitly via the Livermore Solver (lsode), and catch is linearly interpolated between sampled epochs. More detail about issues of continuity and ODE integration are described in Appendix (4).

2.5 Model Fitting

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

$$I_t|q, \sigma^2, \boldsymbol{\theta} \sim LN(qB_t(\boldsymbol{\theta}), \sigma^2). \tag{15}$$

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

$$\frac{dB}{dt} = rB\left(1 - \frac{B}{K}\right) - FB. \tag{16}$$

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

$$\log \mathcal{L}(q, \sigma^2, \boldsymbol{\theta}; I) = -\frac{T}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_t \log\left(\frac{I_t}{qB_t(\boldsymbol{\theta})}\right)^2. \tag{17}$$

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

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

optimizer to refine optima within a local mode. The scheme functions by searching globally to iteratively improving hot starts for the local optimizer.

In Appendix (4) a profile likelihood method for estimating all of the parameters of the model is derived. The profile likelihood technique greatly improves the reliability of local optimizers when fitting the biological parameters alongside additional nuisance parameters. The catchability parameter q has the effect of rescaling biomass which can often function similarly the role of the carrying capacity parameter K. Thus, the structure of the likelihood may confound q and K, and for some data these parameters may only be weakly identifiable. Posing the model in a Bayesian context provides a convenient mechanism for managing these weak identifiability issues. In a tactful Bayesian formulation q and σ^2 may then be marginalized out of the joint posterior to yield fast and reliable inference (DeYoreo, n.d.).

2.6 Gaussian Process Metamodel

For assessing biological parameters over the simulated grid, as seen in Figure(4), a Gaussian Process (GP) model is used as a flexible, stochastic interpolator over RP space. As previously established, in Section (2.5), the biological parameters of interest are the Schaefer model's $log(r)$ and $log(K)$ parameters. Since the estimates of these parameters are random variables, with variances given by the inverse of the observed fisher information, interpolation of MLEs requires paying additional attention to propagating estimates of uncertainty into the metamodel.

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

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

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

$$\begin{aligned} \boldsymbol{x} &= \left(F^*, \frac{B^*}{\bar{B}(0)}\right) \\ \hat{\mu} &= \beta_0 + \boldsymbol{\beta'}\boldsymbol{x} + f(\boldsymbol{x}) + \boldsymbol{\epsilon} \\ f(\boldsymbol{x}) &\sim \text{GP}(0, \tau^2 R(\boldsymbol{x}, \boldsymbol{x'})) \\ \epsilon_i &\sim \text{N}(0, \hat{\omega}_i). \end{aligned} \tag{18}$$

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

Here R is the squared exponential correlation function.

$$R(\boldsymbol{x}, \boldsymbol{x'}) = \exp\left(\sum_{j=1}^{2} \frac{-(x_j - x'_j)^2}{2\ell_j^2}\right) \tag{19}$$

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

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

$$\check{\mu}(\check{s}) = \beta_0 + \boldsymbol{x}(\check{s})\boldsymbol{\beta} + R_{\boldsymbol{\ell}}(\check{s}, s)R_{\boldsymbol{\ell}}^{-1}(s, s)\Big(\hat{\mu}(s) - \big(\beta_0 + \boldsymbol{x}(s)\boldsymbol{\beta}\big)\Big) \tag{20}$$

3 Results

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

$$\check{B}^* = \frac{\check{K}}{2} \qquad \check{F}^* = \frac{\check{r}}{2}. \tag{21}$$

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

$$\text{Relative Bias} = \frac{\check{RP} - RP}{RP} \tag{22}$$

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

3.1 An MSY-Optimal Catch History

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

Figure(5) shows the biases in F^* and $\frac{B^*}{B(0)}$ over the space of simulated RPs. The (*top-right*) panel shows how data generated across the broader space of RPs are mapped onto the limited space of the Schaefer line. Below the Schaefer line, RP estimates are biased by over-estimating $\frac{B^*}{B(0)}$ and under-estimating F^*. Above the Schaefer line the vice-versa is true; $\frac{B^*}{B(0)}$ is under-estimated and F^* is over-estimated. In the (*left*) and (*bottom*) panels the bias in $\frac{B^*}{B(0)}$ and F^* are shown component-wise; each panel showing the same patterns, but focusing on only one component of the bias at a time. In these panels red coloring indicates over-estimation of the RP and blue indicates under-

estimation. Notice that the region of RPs near the Schaefer line enjoy relatively low bias since model misspecification is minor in this region.

Notice that under the Schaefer model, Eq(12) with $\gamma = 2$, B^* is necessarily half of K. Since $\frac{B^*}{B(0)}$ is always $\frac{1}{2}$ under the Schaefer model, the bias in $\frac{B^*}{B(0)}$ (as seen in Figure(5)) simply measures the distance from the data generating location vertically to the Schaefer line.

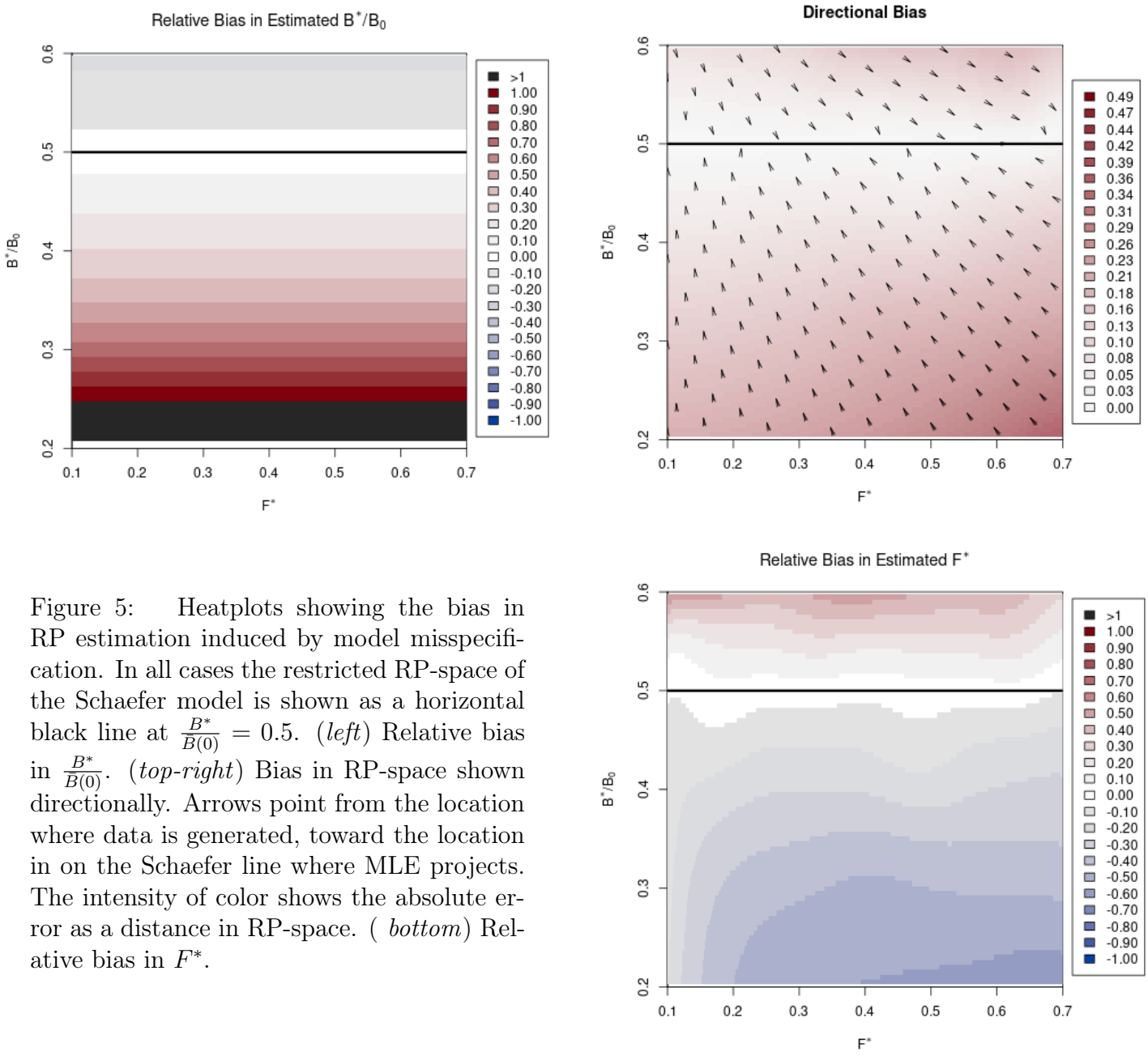

Figure 5: Heatplots showing the bias in RP estimation induced by model misspecification. In all cases the restricted RP-space of the Schaefer model is shown as a horizontal black line at $\frac{B^*}{B(0)} = 0.5$. (*left*) Relative bias in $\frac{B^*}{B(0)}$. (*top-right*) Bias in RP-space shown directionally. Arrows point from the location where data is generated, toward the location in on the Schaefer line where MLE projects. The intensity of color shows the absolute error as a distance in RP-space. (*bottom*) Relative bias in F^*.

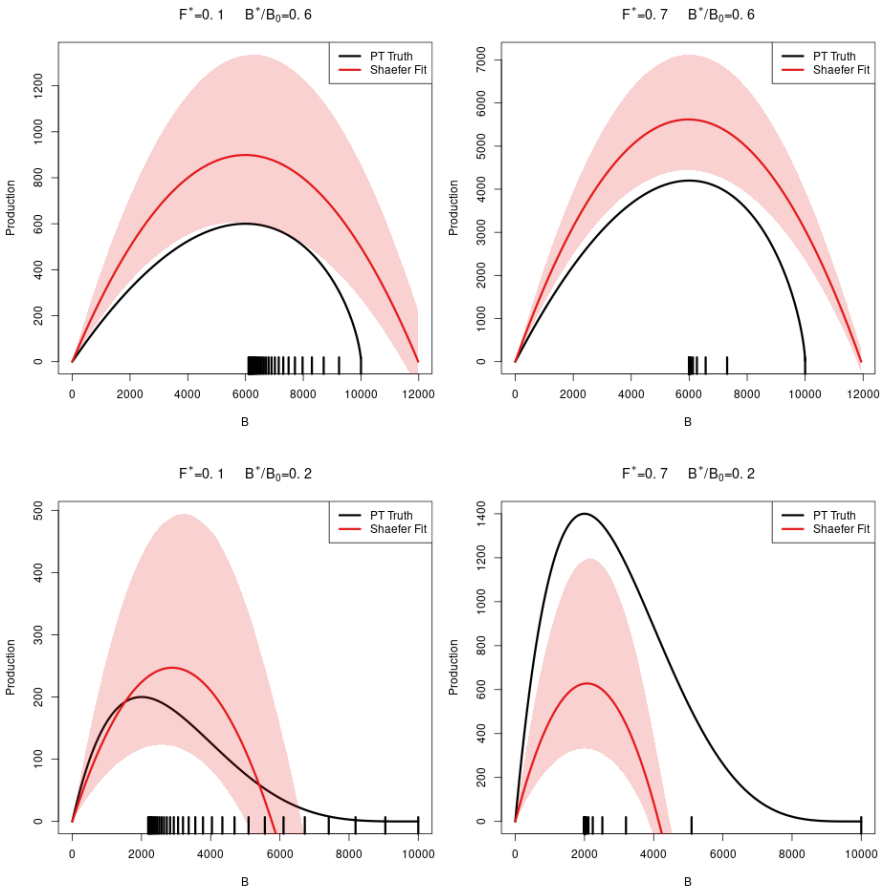

Figure 6: A comparison of the true PT production function (in black) and the estimated logistic curve (in red) with 95% CI shown. The examples shown represent the four corners of maximum model misspecification in the simulated RP-space. Observed biomasses are plotted in the rug plots below the curves.

Figure(6) shows four of the most misspecified example production function fits as compared to the true data generating PT production functions. In the rug plots below each set of curves the observed biomasses demonstrate the exponential decay from K to B^* in each case. In particular, notice how only biomasses greater than the PT B^* are observed. Due to the leaning of the true PT curves, and the symmetry of the logistic parabola, the logistic curve only observes information about its slope at the origin from data observed on the right portion of the PT curves. Above the Schaefer line PT is steeper on the right of B^* than it is on the left, and so the the logistic curve over-estimates r, and thus F^*, for data generated above the Schaefer line. Below the Schaefer line the vice versa phenomena occurs. Below the Schaefer line PT is shallower to the right of B^* than it is on the left and so the logistic parabola estimate tends to under estimate F^*.

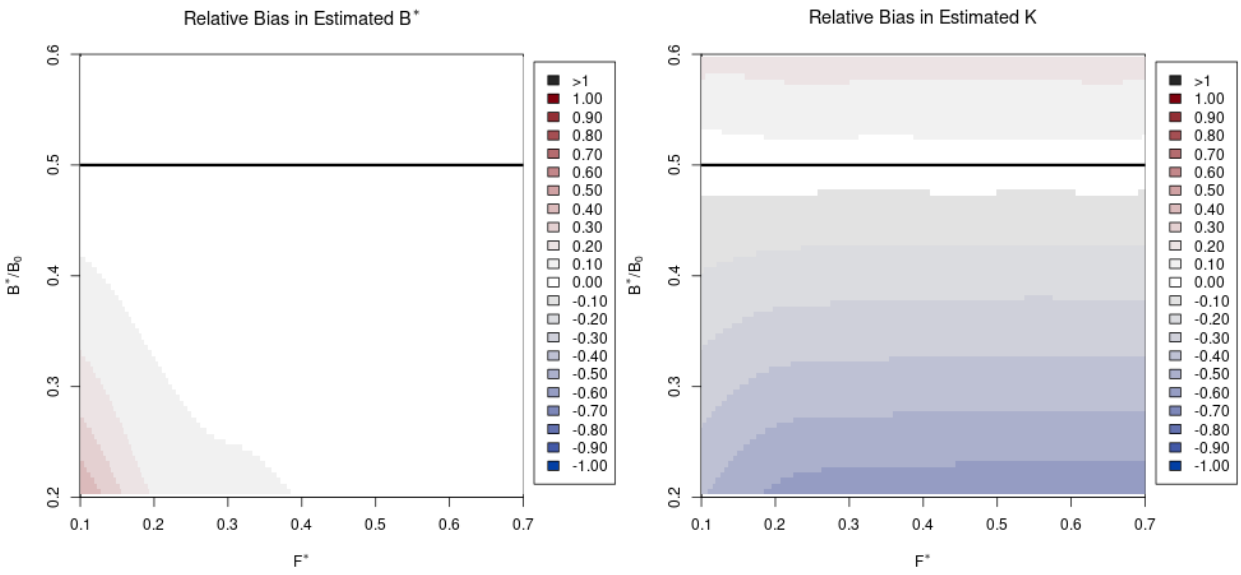

Figure 7: MLE Bias surfaces for B^* (*left*) and K (*right*) individually.

Figure(6) also gives some examples of the relative behavior of B^* and K. In Figure(5) it is clear that the bias behavior of $\frac{B^*}{\bar{B}(0)}$ is locked in a fixed pattern under the Schaefer model. Figure(6) indicates that the individual biases of B^* and K may behave quite differently. B^* appears to be estimated fairly accurately while K does not.

Figure(7) teases apart $\frac{B^*}{\bar{B}(0)}$ into individual bias surfaces for B^* and K respectively. Interestingly B^* enjoys a large region of RP-space with relatively low bias. Given that B^* has relatively consistently low bias K maintains the expected inverse relationship with $\frac{B^*}{\bar{B}(0)}$ bias. Since the parabolic structure of the logistic function ties the ratio of B^* and $\bar{B}(0)$ to $\frac{1}{2}$, their is only one degree of freedom shared between B^* and $\bar{B}(0)$ so that their ratio is maintained at $\frac{1}{2}$. In this setting it appears that B^* estimation is largely conserved at the cost of K.

3.2 More Typical Catch Histories

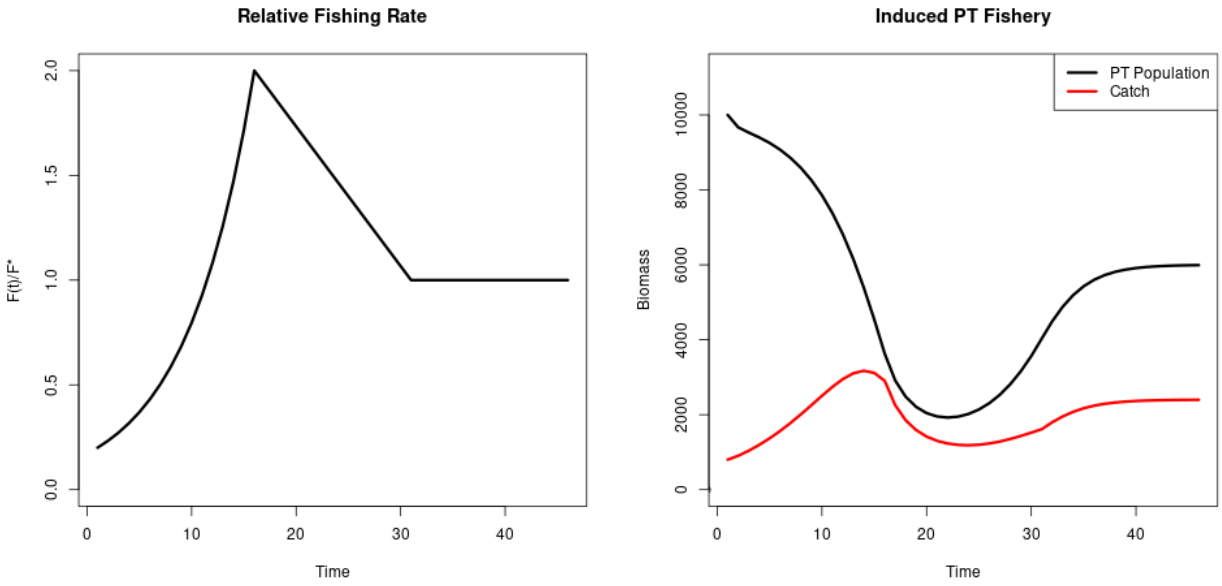

Figure 8: (*left*) Relative fishing specified so as to induce contrast. (*right*) Population biomass and catch demonstrating contrast in a PT population with $F^* = 0.4$ and $\frac{B^*}{B(0)} = 0.6$.

The setting of constant relative fishing rate is a useful simplification for building understanding of the dynamics that induce bias, but in practice constant fishing rate is a somewhat oversimplified setting. It is far more common to observe fishing increase rapidly as technology and fishing techniques improved around the turn of the century, followed by a rapid drop in fishing as management proceedures are applied (cite). Figure(8) demonstrates a more realistic, while still simplified, fishing behavior that exposes the population to a variety of fishing rates, which induce contrast in the generated indicies and allow the fitting model to observe the population decrease in size and subsequently rebuild. This represents a "high contrast" setting that is known to better inform growth rate parameters, such as r.

Figure(**??**) shows the relative bias surfaces for B^* and F^* under the high contrast setting. On the one hand, notice the relative lack of bias in F^* over a large swath of RPs far from the Schaefer line. On the other hand, notice that bias in B^* increases here relative to the low contrast setting. The pattern of bias in B^* maintians a similar pattern, and overall scale, as the low contrast setting seen in Figure(7), however a smaller region

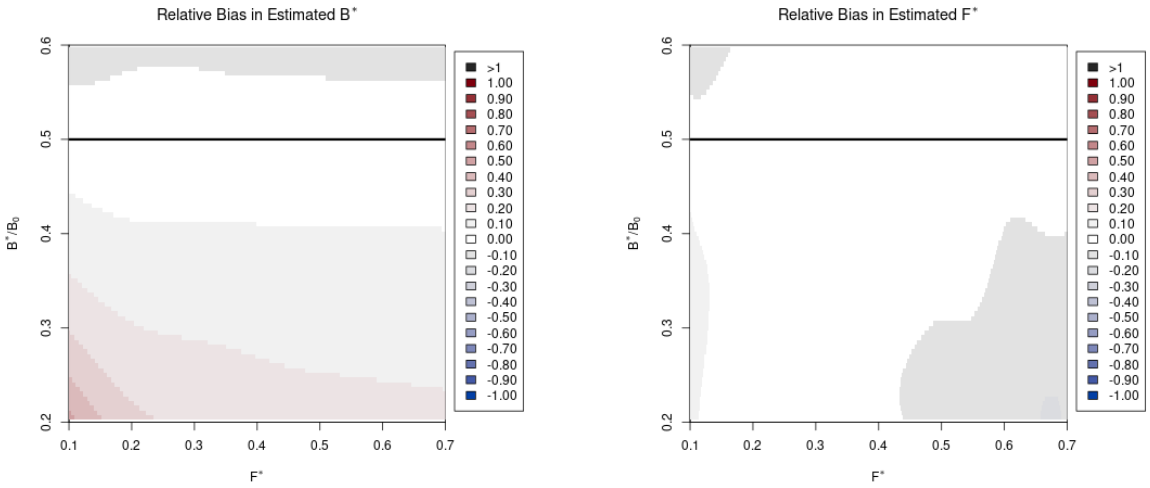

Figure 9: MLE Bias surfaces for B^* (*left*) and F^* (*right*) with relative fishing rate as specified in Figure(8)

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

If the data is augmented so that the fishing rate is held at F^* for an additional 15 time epocs, so that slower growing stocks may observe more data near B^*, Figure(10) shows the updated bias surfaces. The scale of bias in B^* is reduced, but the scope and general pattern of bias remains similar. While the bias behavior of B^* estimates are somewhat diminished, F^* biases are generally magnified.

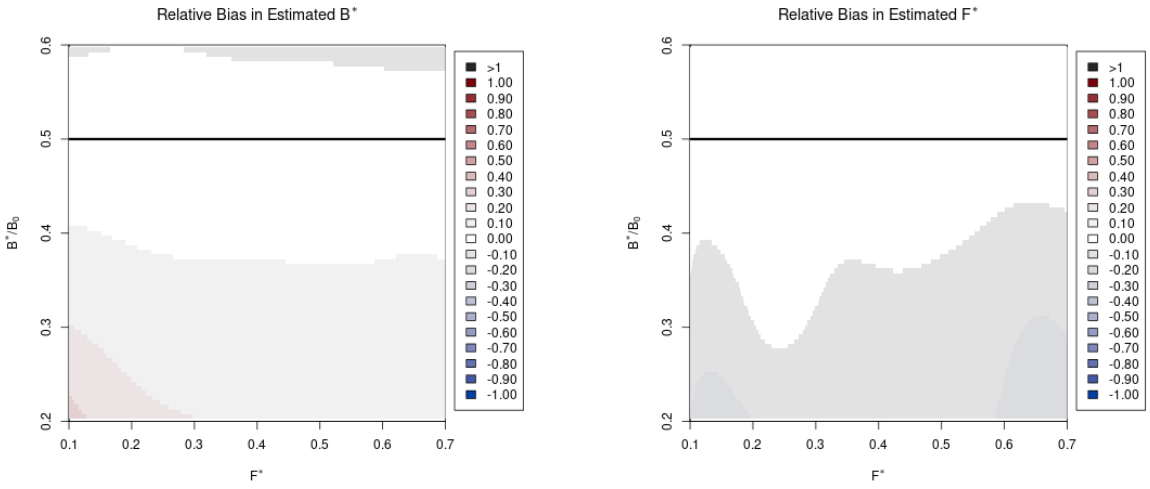

Figure 10: MLE Bias surfaces for B^* (*left*) and F^* (*right*) with relative fishing rate augmented with additional observations near equilibrium.

4 Discussion

• The statistician George Box famously said "All model are wrong, but some models are useful" • My question is how useful are our models?, and if our models are useful. How useful are they? and when are they useful?

In the high contrast setting, F^* estimation is largely conserved, especially when compared to the dramatic F^* biases observed in Figure(5), however this

Although the shape of the SRR is important, it is often ambiguous what the appropriate SSR is given the available stock assess- ment information. For example, Dorn (2002) concluded in a meta-analysis of rockfish (Sebastes spp.) stock and recruitment data that neither the R-SRR nor the BH-SRR can be distinguished as a preferred model on the basis of statistical goodness of fit . Brodziak (2002) reached a similar conclusion in an analysis of USA west coast groundfish stock–recruitment data. Alternatively, in a meta-analysis of 128 diverse fish stocks, Punt et al. (2005) con- cluded that as a general model , the BH-SRR is strongly preferred over the R-SRR, but that "there are also indications that other (more complicated) forms may provide better representations of the existing data" (p. 76). This conclusion is not unexpected because models with more parameters can approximate variability in the observations used to estimate the SRR more accurately than models with fewer parameters.

References

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

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

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

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

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

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

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

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

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

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

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

Punt, A. E., & 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

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

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

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

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

Sen, A. R. (1984). Sampling commercial rockfish landings in California.

Appendix A: Profile Likelihood MLE

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

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

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

$$\log \mathcal{L}(q, \sigma^2, \boldsymbol{\theta}; I) = -\frac{T}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_t \log\left(\frac{I_t}{qB_t(\boldsymbol{\theta})}\right)^2. \tag{23}$$

First Eq(23) is maximized with respect to q by partial differentiation of Eq(23) with respect to q,

$$\frac{\partial \log \mathcal{L}}{\partial q} = -\frac{1}{q\sigma^2}\left(\sum_t \log\left(\frac{I_t}{B_t(\boldsymbol{\theta})}\right) - T\log(q)\right) \tag{24}$$

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

$$q(\boldsymbol{\theta}) = e^{\frac{1}{T}\sum_t \log\left(\frac{I_t}{B_t(\boldsymbol{\theta})}\right)} = \left(\prod_t \frac{I_t}{B_t(\boldsymbol{\theta})}\right)^{\frac{1}{T}}. \tag{25}$$

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

Now to maximize in the σ^2 direction Eq(23) is differentiated with respect to σ^2,

$$\frac{\partial \log \mathcal{L}}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_t \log \left(\frac{I_t}{qB_t(\boldsymbol{\theta})} \right)^2. \tag{26}$$

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

$$\sigma^2(\boldsymbol{\theta}) = \frac{1}{T} \sum_t \log \left(\frac{I_t}{q(\boldsymbol{\theta})B_t(\boldsymbol{\theta})} \right)^2 \tag{27}$$

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

By combining Eq(25) and Eq(27) the MLEs of q and σ^2 can be written entirely in terms of $\boldsymbol{\theta}$. Furthermore, this realization allows the joint maximization of the likelihood to be reduced to the following profile log-likelihood,

$$\log \mathcal{L}(\boldsymbol{\theta}; I) = -\frac{T}{2} \log\big(\sigma^2(\boldsymbol{\theta})\big) - \frac{1}{2\sigma^2(\boldsymbol{\theta})} \sum_t \log \left(\frac{I_t}{q(\boldsymbol{\theta})B_t(\boldsymbol{\theta})} \right)^2. \tag{28}$$

This profile log-likelihood is maximized numerically over $\boldsymbol{\theta}$, and the estimates for q and σ^2 are given by evaluating Equations (25) and (27) at $\hat{\boldsymbol{\theta}}$.

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \log \mathcal{L}(\boldsymbol{\theta}; I) \tag{29}$$

$$\hat{\sigma}^2 = \sigma^2(\hat{\boldsymbol{\theta}}) \tag{30}$$

$$\hat{q} = q(\hat{\boldsymbol{\theta}}) \tag{31}$$

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

Appendix B: Integrating ODEs, Stiffness, Interpolation, and Chaos.

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

4.0.1 Uniqueness, Continuity, and Identifiability

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 population production, $P(B)$, and a term for removals via catch, FB. 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 $P(B)$ and FB separately. Typically any choice of $P(B)$ 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 $P(B)$ does not typically introduce continuity concerns, unlike some potential assumptions for catch.

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

4.0.2 Integration and Stiffness

4.0.3 Integration, Stiffness, and Chaos

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

The concept of stiffness in ODEs is hard to precisely characterize (cite). Hairer and Wanner [5, p. 2] describe stiffness in the following pragmatic sense, "Stiff equations are problems for which explicit methods don't work". It is hard to make this definition more mathematically precise, but this

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 explicate methods result in wildly varying solutions to the ODE, and in still regions of parameter space explicate methods completely fail to represent the model as stated in the stiff regions of parameter space. Results shown here are computed using the lsode integration method since it runs relatively quickly and has a relatively smaller footprint in system memory.