Metamodeling for Bias Estimation of Biological Reference Points.

Nick Grunloh

Introduction

- Hello. My name is Nick Grunloh.
- Thanks for coming. π day 2022!
- Collaboration with NOAA NMFS SWFSC, funded by NOAA Sea Grant.

Outline

- Metamodeling approach for assessing RP model misspecification in population dynamics models.
- Extensions to this work as I finish up my PhD.

Data and Basic Modeling Structure

- Context:
  - single species population dynamics models
  - as might be used in managing fisheries
- Simplest model that captures the essence of the managing objectives is the Surplus-production model
- Data comes to us in two parts:
  - Two panels shown: example data from the classic Namibian Hake data set.
  - **Left** A measure of population biomass thru time (called an index of abundance)
  - **Right** A Variety of other quantites, but at a minimum observe a matched time series of catches.
- We can't observe all the fish in the sea, but thru the index
- We assume we observe biomass upto a proportionality constant q
- q relates our index of abundance to actual biomass in the population.
- I_t forms a response variable with lognormal residuals.
- Most of the action in these models comes in through B process.
- Biomass is modeled as a (typically) nonlinear ODE.
  - the population biomass grows via a production function, $P(B)$
  - biomass is removed due to catch, $C(t)$.
  - Production maps the current biomass to some net growth of future biomass.
    - all Reproduction, maturation, and mortality processes other than fishing

Schaefer Model

- Classic choice: $P(B)$ =logistic growth curve => creates the Shaefer model.
- Parabola with parameters (r, K)
  - r controls the maximum rate of productivity increase
  - K is the so called "carrying capacity"
- Logistic models density dependence, and we can see the parameters on the parabola
  - low biomass: many resources & lack of competition => productivity increases quickly
  - biomass increases: more competition for resources => productivity decreases
- K is a stable equilibrium point
  - population above K decreases to return
  - population below K increases to K

Biological Reference points

- We want to mangage fisheries to promote future productivity.
- MSY: Peak of production function.
- introduce fishing so as to move the stable equilibrium nearer to MSY
- **Next**

- For management purposes we look at heuristic measures of the population called RP.
- Commonly RPs are ways of noticing MSY.
- Here I focus on two:
  - Fmsy: fishing rate to result in MSY
  - Bmsy: biomass of the populaiton at MSY (argmax)
  - Relative Bmsy
- for a parabola, Relative Bmsy is just half of carrying capacity

RP Constraints

- Conceptually F^* and $\frac{B^*}{B_0}$ coexist in an entire 2D space.
- (Mangel et.al., 2013) Canadian Journal of Fisheries and Aquatic Science
  - Two parameter BH model: RP space is limited to a 1D curve
  - **Right** Plot Relative Bmsy against Fmsy
    * black: posterior samples of the RPs for a 3 parameter Shepherd-like model. (cowcod)
    * red: posterior samples of the RPs for a 2 parameter BH model.
    * the red posterior is squashed into the curve $\frac{1}{x+2}$
  - Mangel et. al. suggests looking into 3-parameter curves
- **Next:**

- The Schaefer Model is a two parameter curve that suffers from a constrained RP-Space.
  - We already saw the constrain on the last slide
    * parobolic shape of logistic growth limits relative Bmsy to 1/2.
  - **Right**: This is the “Shaefer Line”
    * model misspecification in this context limits the space of RPs.
    * over constrains variance of estimates
    * and induces severe bias in estimated RPs.
  - For managment define a specific (highly relavant) sense of model misspecification
    * Think of each point in this RP-space as a different species.
    * if spp not on line the model is misspecified.

Outline

- In the next section I’ll outline a simulaiton method for exploring the inferential effects of this type of model misspecification.

Pella-Tomlinson Production Model

- Generate a species off of the Shaefer line, with a 3 parameter PT model.
- Same model structure as Shaefer, but production function is a 3 parameter curve.
- P even has a similar form.
  - PT has γ parameter specifically to move the peak of the curve left or right.
  - moves the relative Bmsy off of the shaefer line at 1/2.
- Simulate PT data and fit those data under the Schaefer Model.
  - Observe how RPs under Schaefer model are biased relative to the true PT RPs.

Catch

- Assume synthetic catch series, To complete the model specification.
- Data informs parameters differently than we normally expect.
- It is known: information content is about how biomass and catch series change.
  - not so much sample size
- I show Catch in red and the population in black.
- Two cases. All cases, Population initial condition starting at Carrying Capacity.
- Low Contrast:
  - Catch held at to come to equilibrium at MSY; Optimially managed species
  - low contrast, relatively low information setting
  - Exponential decay from K to Bmsy
- High Contrast:
  - fishing increases accelerates as technology and fishing techniques improve rapidly until management practices are applied to bring the stock into equilibrium at MSY.
  - high contrast, relatively high information setting
  - wiggles about until coming to equilibrium

Simulation

- Again, use PT to generate data broadly in RP-space.
- Shaefer model estimated RPs will neccessarily land somewhere on that $\frac{1}{2}$ Schaefer line.
- Here I show my grid of RP locations for simulating data.
- **Next** 4 red X's in the corners are examples of large model misspecification of Schaefer
  - I show an example biomass series of in the high contrast setting in each corner.
  - $Bmsy/B0$: describes where the biomass comes to equilibrium
  - $Fmsy$: describes how quickly the stock responds to fishing and how fast it rebuilds.
- **Next** At every simulation point, the parameters of the PT can be uniquely identified.
  - Analytical for PT
  - Generally inverting the RP-parameter relationship can be difficult.

Metamodel

- Particular model fits are only as helpful as their standard errors allow, but when you observe trends in RP bias on repeated sampling broadly across RP space you can start to gain confidence in patterns of inferential bias.
- Here a Squared Exponential GP Metamodel is used as a flexible approximator for mapping the true PT RPs to the estimated RPs under limited setting of the Schaefer model.
-
- Since the MLE estimator is a random variable it is important to propogate that uncertainty into the metamodel.
- The GP residual variation provides an ideal mechanism for propagating that uncertainty, to better understand biases.
- While the constrained RP space limits the extent of RP standard errors, accounting for estimate uncertainty has a smoothing effect, similar to a knugget, that improves the interpretability of RP biases.
-
- This metamodeling approach explicitly highlights the inferencial trade-offs imposed by productivity model misspecification in terms of specific bottom-line metrics that important to managing fisheries.
- While previous studies have considered the factors neccisary to estimate model parameters, the limiting constraints of RP model misspecification have not been explicitly considered, with a direct focus on RP estimation.

Outline

- Split results into two parts
- First low contrast, low information, setting.

Directionally

- Metamodel allows for a high level look at biases.
- What does inference do over the entire space of PT data
- Arrows indicate direction of mapping; color indicates the magnitude of bias.
- Give example.
  - Map vertically to the horizontal Shaefer line
  - below line : underestimate $Fmsy$
  - above line: overestimate $Fmsy$

Components of Bias

- Split out thosse arrows into the components of bias.
- The color scale is in percent error.
- Red indicates over estimation, blue indicates underestimation by the Schaefer model.
- **Left** $Fmsy$ story
  - Below the line we underestimate $Fmsy$
  - Above the line we overestimate $Fmsy$
  - Larger Model misspecification => larger bias.
- **Right** Relative $Bmsy$ story
  - Law of Nature under Schaefer
  - Only looking at bias vertically
  - estimate must land on the line

$Fmsy$ Curves

- How to understand $Fmsy$ bias?
- production functions in the 4 corners (examples of large model misspecification)
  - red lines represent the Schaefer fit
  - black represent the true PT production function
  - rug plot are the data
- Recall: low contrast, biomass is exponential decay from K -> $Bmsy$.

- Tell top story
  - The model only observes the right portion of the true SRR
  - Due to the leaning of the true PT curves, and the symmetry of the logistic parabola, the logistic curve is learning about its slope at the origin entirely from data where Biomass>Bmsy, and above the schaefer line the PT is steeper that on the right half than it is on the left, and so we over estimate F^* for data generated above the line.

  - The vice versa phenomena occurs below the schaffer line.

  - $Fmsy$ is underestimated because of the shallow slopes on the right half of PT.

Ratio

- Recall: realtive $Bmsy$ pattern fixed in place

- Use Metamodel to split $Bmsy$ from carrying capacity (Constrained to divide back up).

  - $Bmsy$ shows large swaths of relatively little bias

  - Carrying capacity has substantial bias

    * Share one degree of freedom between the two qantities
    * K estimated to serve $Bmsy$
    * repeatedly observed pattern across all catch histories tried

Contrast

- I show $Fmsy$ and $Bmsy$
  - Constraint: carrying capacity bias is similar to previous slide
- When contrast is present, Bias is generally less over a large sets of largely misspecified settings.
  - As expected $Fmsy$ is largely estimated well in the presence of contrast.

Next

- The second row is the same high contrast setting, but appended an additionl period of catch near MSY.
  - Now, $Fmsy$ is more biased for a wide range of model misspecification
  - but, $Bmsy$ is better estimated now
- While contrast generally improves RP estimation,
  - the over constrained Shaefer model lives in a zero sum world

Summary

- "All model are wrong, but some models are useful"
- How useful is the Schaefer model?, and if useful, how useful? and when?
- This is a rich simulation based method understand that those questions at a deeper level.
  - The Shaefer model presents a simple setting to build the basic concepts at play, but
  - I plan an extensions into a more broad class of production functions, as well as
  - extensions across models that include dynamics of individul growth and maturity.
- In this severly overconstrained settings we pay for our modeling mistakes primarily in estimate bias.
- In practice, when Schaefer model is unlikely to be correctly specified, one should at best expect to only reasonably estimate either $Bmsy$ or $Fmsy$ correctly depending on the particular degree of model misspecification.
- The observed contrast serves to distribute the available information among B^* and F^*
  - So good models of catch are important to contextulize the interpretation of RP estimation.

Outline: Start with the extension of growth and productivity

Productivity Extension

- PT-Schaefer simulation setting explores a limited range of model misspecification matchups.
- Extend simulation: replacing 3parm-PT with the 3parm-Dersio production function
  - Deriso has a number of 2 parameter special cases (including the Logistic)
    * Importantly Beverton-Holt and Ricker

Beverton-Holt

- Restricted inference under BH is of primary interest
  - Due to overwhelming popularity in Stock Assessment
- See 2 parmater form α & β.
  - α function similarly to r in the Shaefer model
  - $\frac{1}{\beta}$ function similarly to the carrying capacity.
- **Right:**
  - Again I show the constrained space=$\frac{1}{x+2}$
  - While equally constrained due to 2-parmater form,
  - Cuts through RP-Space where we think many commonly assessed fish species exist in RP space.

Growth Extension

- Models that include individual growth and maturity can get more complex fast.
- Deriso-Schnute model:
  - a relatively compact (fast) delay differential model to include growth effects, in the same simulation setting outlined here.
  - Models recruitment into numbers alongside a coupled biomass equation.
  - Tracking numbers and biomass separately allows the model to account for the averge individual size and growth.
- Still RPs are largly determined by the recruitment function.
  - fills role of the production function
  - very similar 2 v. 3 parameter parametric forms
  - Species Properties: $a0$, κ, w_∞, M estimated externally to the model.
- I will explore if the presence of these growth rate parameters effect RP bias

Details

- Individuals simultaneously sexually mature, and start dieing and being fished at age a_0
- Individuals grow via von bertilanfy growth (Derived directly from).
- Biomass dynamics, can then be seen to describe biomass by an account
  - B(new recruits), growth(existing biomass), mortality

Outline

- Given the importance of catch for interpreting RPs
- I have a catch model to better contextualize the dynamics model.

Common Discretization

- The motivating biomass dynamics models are commonly discretized
- discretized for ease of implementation and data handling
- discrete version: catch observations assumed known and directly plugged in
- **Next**
  - The top requires instantaneous catch
  - The bottom only requires the aggregate catch over the year.
  - bottom matches the stucture of data as typiclly collected.
  - but the integral implies a useful latent stucture for modeling.

Catch Interpolation

- **Right** I show the Namibian Hake catch again.
  - Each point represents the aggragate catch over a year.
- **Left**
  - Model to get at the continuous latent structure of catch (consistent with statement of dynamics)
  - Continuous catch model is a linear spline.
  - Represents a sensible linear interpolation assumption on continuous catch.
- and so what of uncertainty in catch?
- Catch is commonly assumed known without uncertainty
- and there are biological hypotheses that could make use of jitters
- A Statisticians eye will notice what looks like uncertainty in catch.
- so, I investigate what becomes of the uncertainty assumption?

Next

- Working the linear spline assumption forward (integrating over time) to get a spline in terms of observables
- I show a model that allows uncertainty
- you can also turn uncertainty off (typically assumed)
- the plot below (**Left**) shows this model fit w/ and w/o assumpted uncertainty
  - Connect the dots vs. smooth

- **Next** (**Right**)
  - Implied latent instantaneous catchs
  - catch uncertaitny smooths jitters and implies a sensible instantaneous catch.
  - assuming catch known implies wild oscillations on instantneous catch.
- **Next** Back to the motivating models again,
  - this exposes worrying assumptions
  - If we think that our discretizations bare any resemblance to the continuous motivating dynamics, then the assumption of catch known without uncertainty is also assuming that instantaneous catch can be negative (not an admissible assumption).
- I'll investigate:
  - Models of various smoothness on instantaneous catch
  - how accounting for catch uncertainty affects inference in the biomass dynamics model

Timeline