Teaching Objectives for Peruvian Stock Assessment Training

To be held Oct 2016

Dan Ovando

10/28/2016

# Objectives

The goal of this workshop is to empower scientists at IMARPE and TNC Peru to perform basic evaluations of stock status, based on the work resulting from the Data Limited Fisheries SNAP working group and TNC. The workshop will be a three day affair, with day 1 dedicated to basic fisheries science training led by Ricky Amoroso and Maite Pons. Day 2, covered here, will be dedicated to developing R based skills for stock assessment. Day 3 will be used to review the assessment outputs produced by Dawn. This document covers day 2 and 3.

# Day 2 - R training

Our last workshop demonstrated substantial interest, and baseline capacity, for learning R based tools for fishery science. This day will focus on capacity building in this regard.

Dawn's report on Chita (*Anisotremus scapularis*, or Peruvian grunt) covered data on life history, landings, CPUE, and length frequency for Chita up and down the coast. She also covered assessments using LBSPR, equilibrium SPR and YPR, catch curve, and CPUE standardization. It remains to be seen which of these we will focus on, but the goal will be to work through 1-2 of these methods in details

The goal of the R training agenda will be to have participants leaving feeling confident in basic reproducible data science and fisheries science tools. I lay out the key topics for potential areas of exploration below

### R Fundamentals

The hope is that folks have basic R skills (installing, opening, loading packages, reading and writing files, projects) ready to go and this will be an extension.

I can send around a short training exercise for participants to run prior to the workshop to make sure that they are up and running on these, and if folks need extra help we could meet Sunday night or Monday.

## Data Fundamentals

* Reading in data
  + Keep raw data raw, process in R
  + Common pitfalls in data storage
  + Data types (beware of factors)
* ABP - Always Be Plotting
  + Using dplyr for summary statistics
  + Using ggplot for summary plotting (make histograms of length, cpue, etc.)
* Break to discuss results before jumping into assessment
* Saving data

## CPUE Standardization

If we go down this route...

* Reminder of rationale for CPUE (Simfish example)
  + Goal is to control for other factors that might influence CPUE through *q*, instead of *B*
  + Regression based approach, goal is to extract "year effects", controlling for other factors
* Regression reminder
  + Why use regression for CPUE standardization?
  + Assumptions of regression
    - i.i.d
    - normality
    - endogeneity
  + Assumptions are always wrong, need to check for when they're really wrong
  + Key is to recognize places where assumption violations result in incorrect results
* Regression in R
  + using lm
    - broom package
    - writing formulas
    - extracting summary statistics
    - extracting coefficients
    - getting predictions
  + Regression diagnostics
    - Histogram of residuals
    - resids vs fitted
    - qqplot
* Application to Chita
  + Go back to our raw CPUE data
  + Brainstorm things that could be influencing this that we could control for
  + Select control variables for CPUE
    - Make plots of control variables to check for outliers, correlation
    - Particularly check for correlation with year terms
  + Fit model to raw CPUE data
  + Look at QQ plot, diagnose non-normality
  + Log transform, run the thing again
  + zero problem
  + Check QQplot, good!
  + Look at other model diagnostics
  + Apply bias correction to year terms, get new index of abundance
* Using index of abundance
  + Illustration in fitting to surplus production model
  + Revisit surplus production example from Ricky and Maite
  + Refit using standardized CPUE, compare and contrast
* Discussion!
  + how does this apply to other potential fisheries in Peru?

## Catch-Curve

* Theory of catch curve
  + Walk through a cohort
  + Show how a the slope equals the mortality rate
* Assumptions of catch curve
  + Equilibrium
  + Constant mortality
  + Mean constant recruitment
  + Logistic selectivity
  + Explore consequences of these using SimFish
* Apply Catch Curve
  + Plot length frequency
  + Translate length to age using VBK
  + Identify mode and max of distribution to fit to
    - Mode and first zero as rules of thumb
  + Fit regression and extract slope + uncertainty
* Interpret catch curve
  + Methods for extracting F
    - Empirical equations for M
    - MPAs
* Application of catch curve
* F/M ratio (function of selectivity) \* selectivity weighted F/M
* Discussion

## LBSPR

* Theory of LBSPR
  + Discussion of fish maturity ogives
  + Walk through a cohort under different selectivities:
  + Discussion of proxy reference points
* Assumptions of LBSPR
  + Equilibrium
  + Constant mortality
  + Mean constant recruitment
  + Logistic selectivity
  + Proxy reference point
* Fundamentals of LBSPR (from scratch)
  + Review of required life history parameters
    - Life history invariants
  + Project unfished population
    - calculate YPR, SPR
  + Project fished population
    - calculate YPR, SPR
  + Compare overlay of lengths under alternative mortalities and selectivities
* Fitting LBSPR
  + Introduction to optimization
    - how it works
    - how to write functions for it
    - Do example finding the minimum/max of a quadratic function
    - Troubleshooting optimization
  + Write function to calculate deviations from observed vs expected length frequency as a function of F
  + Use function to find F that best explains observed length frequency, given passed selectivity setc.
    - Calculate YPR and SPR that comes from that F
* Discuss!
* An alternative LBSPR - using Adrian's package
  + Does it still require ADMB?
* Advantage: Automates processes, provides estimates of selectivity
* Disadvantage: Automates processes, black box

# Day 3 - Stock Assessment Outcome Results

Given tools from Day 1-2, we now turn to the stock assessment report prepared by Dawn. Goal is to have participants interpret the document, identify strengths, weaknesses, and action items based on report. Jono moderates.

* Summary of Report (Dan)
  + Data included
  + Methods run
  + Overview of results
* Data
  + Review of input data: What issues do participants see with data? Are there other data that could be leveraged?
  + What are the relative strengths and weaknesses of the different data sources?
* Review of length based assessments
  + Given when they've learned this week, what questions/comments/concerns do they have?
  + How do they interpret the results of the length based assessments?
  + How do these results conform with local perceptions of Chita status?
* Review of CPUE based assessments
  + Given when they've learned this week, what questions/comments/concerns do they have?
  + How do they interpret the results of the length based assessments?
  + How do these results conform with local perceptions of Chita status?
* Synthesis of assessments
  + What picture do the suite of assessment outputs paint?
  + Are they in agreement? If so, why, if not why
  + Which assessments do participants have the most confidence in. Why?
* Management Actions
  + Given the results of the assessments, and building off of the decision tree developed at the last workshop, what possible decisions could be made based on these?
  + What are the pros/cons/risks of different potential actions?
* Next Steps
  + What would be needed to act on these results?
    - New data?
    - New Analyses?
  + What is needed to monitor the outcomes of any actions decided as a result of this process?
    - Review of adaptive management
  + How would they envision rolling this process out for other species?
* Closing discussion
  + What remains unclear?
  + What concerns do participants have?
  + Resources for future problems

Note to dan: Bias = cesgado