Modeling anchoveta stocks with a hidden Markov model using spatio-temporal plankton measurements

James Rising, Kimberly Lai

1. Abstract:

Spatially distributed fishery data remains rare, and the geographical ranges of fish over the course of their lifespans remain unknown for many species. This paper combines port-disaggregated weekly anchoveta catch data with spatio-temporal phytoplankton and zooplankton measurements to estimate the range of Peru's anchoveta stock. A hidden Markov model is used to estimate weekly stocks between 1993 and 1999. Stocks are found to correlate with zooplankton heterogeneously over space and time, with the correlations in the northern region being greatest until 2 years before landing, followed by a uniform dependence across the coast between 40 and 100 weeks before landing.

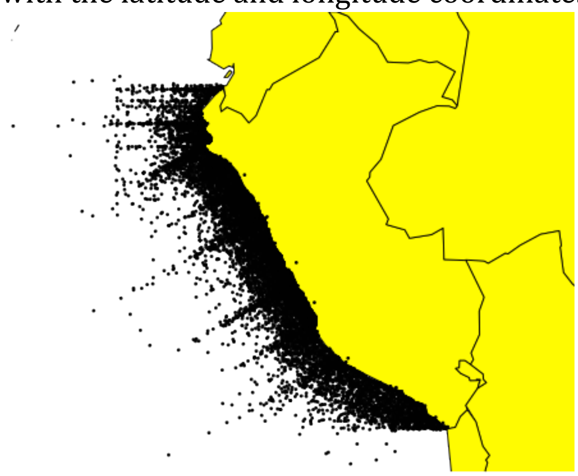
2. Literature review:

An extensive body of work on plankton and anchoveta dynamics was collected in the books *The Peruvian Anchoveta and Its Upwelling Ecosystem: Three Decades of Change* and *The Peruvian Upwelling Ecosystem: Dynamics and Interactions.* Multiple approaches were used for estimating anchoveta stock. The first approach involved collecting information on each step of catch-landing-processing to estimate under-reporting (Castillo et al., 1987). Other studies took a more standard stock-assessment approach with spawning, stock, and egg production data (Pauly et al., 1987), growth models (Palomares et al., 1987) and length-weight catch composition (Tsukayama et al., 1987). The papers collected in these volumes also studied plankton seasonal variation (Carrasco et al., 1989), the physical dynamics of the region (Brainard et al., 1987; Bakun, 1987), and bioeconomic modeling (Aguero, 1987).

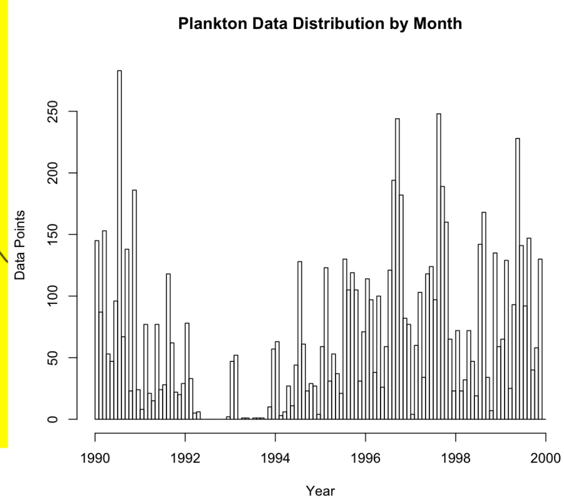
Scientists' understanding of plankton-fish dynamics has evolved over time. Originally, ecological modeling focused on modeling the physical system, including hydrodynamics such as temperature and nutrients (Denman, 1976; Weber et al., 1986). It was found that such hydrophysical features were relevant only on a macro scale (dozens of kilometers or more) and on a micro scale (100 meters or less); distribution on an intermediate scale was uncorrelated to hydrodynamics. To model these distributions, researchers looked for patterns based on predator-prey distribution and developed phytoplankton-zooplankton interaction models (Scheffer 1991, Malchow 1994). Several papers introduced local chaos based on seasonal oscillation of parameters (Doveri et al. 1993, Scheffer et al. 1997), multiple interacting plankton species (Ascioti et al., 1993), or diffusion of nutrient gradients (Pascual, 1993). Another branch of ecological modeling took off in the late 1990s: Instead of trying to model physical systems, papers applied statistical approaches (Punt and Hilborn, 1997, Chen et al., 2003). Other studies have incorporated logistic growth models (Pyo and Lee, 2003). More recently, focus has shifted to hierarchical Bayesian techniques, re-integrating physical assumptions to develop multilevel priors. For example, Hiruki-Raring used priors such as krill density and sea ice, krill density and dive depth intensity, and dive duration intensity in her paper on fur seal foraging and pup growth related to sea ice and prey abundance (Hiruki-Raring et al., 2009). When it is not possible to calculate the joint posterior distribution, Markov chain Monte Carlo simulations are used (Mantyniemi, 2001).

Our analysis builds on the statistical and Bayesian approaches. We apply a hidden Markov framework, informed by a logistic growth model. We separately apply a kriging analysis that steps through time to estimate plankton concentrations throughout our region.

3. Data:
This analysis draws on two key data sets. Plankton data comes from NOAA's World Ocean Plankton Database and was collected by Instituto del Mar del Peru (IMARPE). IMARPE was created in 1964 and started its plankton collection project the same year with the goal of better understanding the impacts of ENSO on the marine resources. IMARPE used a Hensen net (300 μm) to collect zooplankton and a 75 μm surface tow net for phytoplankton. The survey spanned the Peruvian coast and included all seasons between 1972 and 2005. It collected 16,099 data points for each of phytoplankton and zooplankton (ml/m^3), along with the latitude and longitude coordinates.



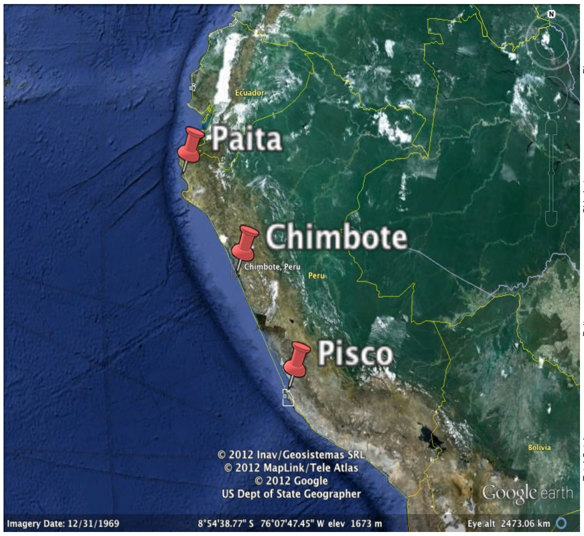
Plankton haul: IMARPE, 1972-2005.

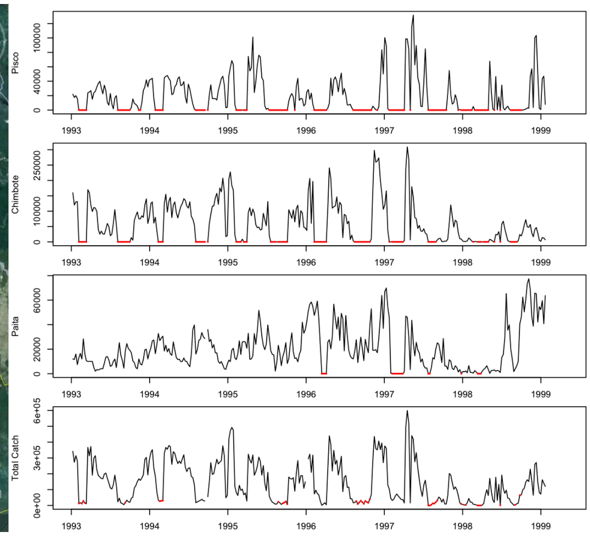


Distribution of Plankton casts by month, 1990-2000.

The second data set comes from the International Research Institute (IRI) Eastern Pacific Pelagic Fisheries project and was collected by the Fishmeal Exporters Organization (FEO). It contains weekly anchoveta landings in tons in three ports (Paita, Chimbote, and Pisco, Peru) as well as total anchoveta landings for Peru for the years 1993-99.

In the analysis, the entire range of plankton data is used to produce a kriging variogram. The 6,012 plankton casts taken before 1990 are used to construct a backdrop of plankton concentrations, while the 7,926 casts between 1990 and 2000 provide the variation for identifying plankton's effects on fish populations. See supplemental material A for a discussion of the selection of this data.



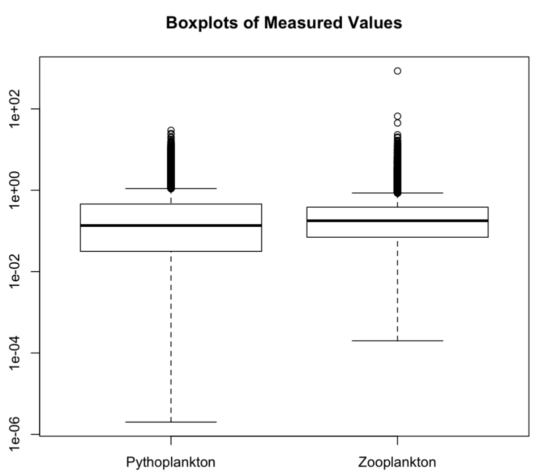


Anchoveta FEO ports: Google Earth. *Weekly catch data, for ports: Pisco, Chimbote, Paita, and total.*

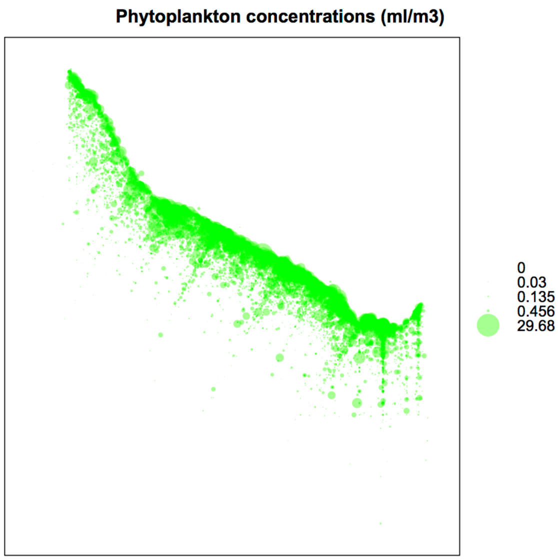
4. Methodology:

4.1 Cleaning the plankton data:

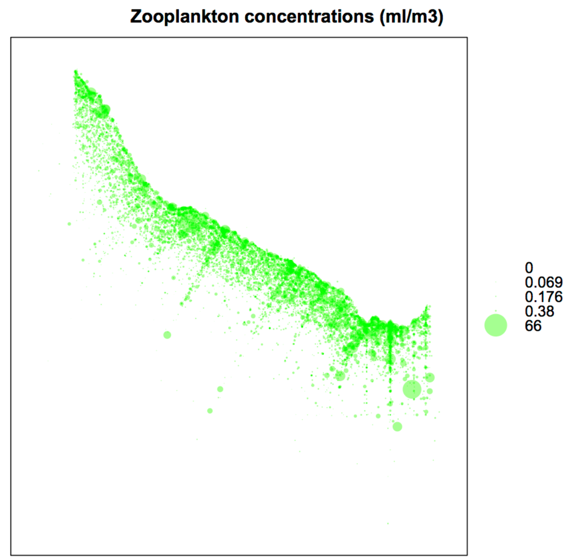
The phytoplankton and zooplankton data cover a wide range of values. The boxplot below shows only the values above 0, on a log scale (59 phytoplankton and 89 zooplankton zero values are excluded from the boxplot, but included in the analyzed data). We removed four data values recorded for Feb. 29 of non-leap years and the zooplankton outlier (a value of 855 ml/m^3, compared to the next highest value at 66 ml/m^3).



To handle the long tailed distribution of values, the plankton data is analyzed in ranks.



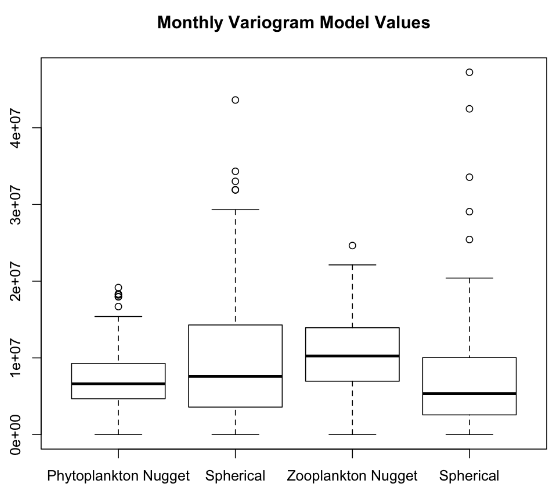
Phytoplankton bubble plot



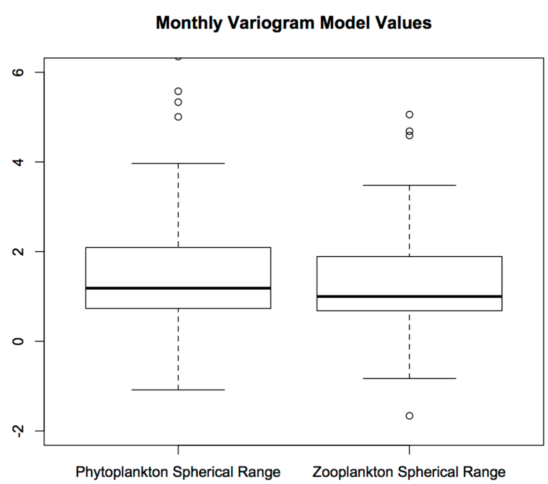
Zooplankton bubble plot

4.2 Gridding the plankton data:

Kriging was used to fill in the spatial data holes for phytoplankton and zooplankton. Separate variograms were estimated for each of the plankton types and each month in the data set. A spherical model fit was constructed for each month, and the median nugget and sill parameters were used to construct a single variogram model to be applied to all time.

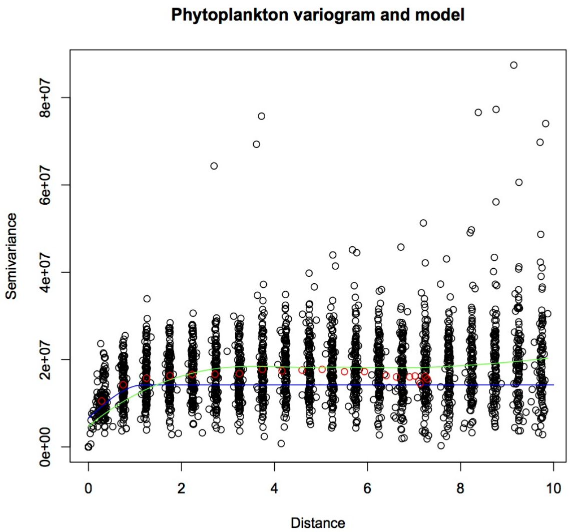


Sills for Variogram models

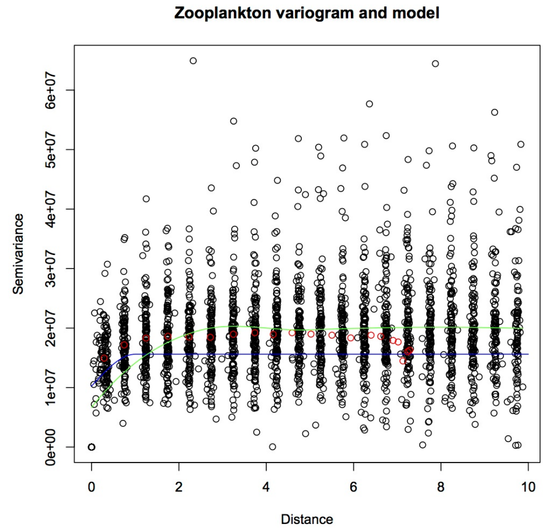


Ranges for variogram models

The below graphs show the variogram monthly semivariance data points, with our model in blue, the mean semivariances in red and a locfit line in green.

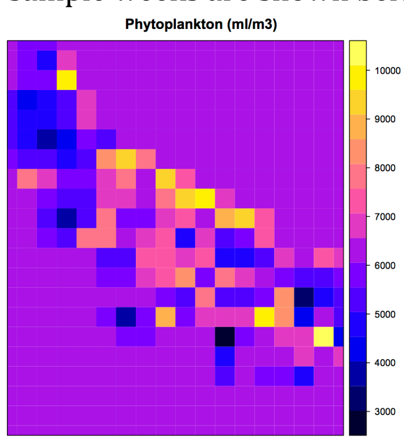


Phytoplankton variogram model verification

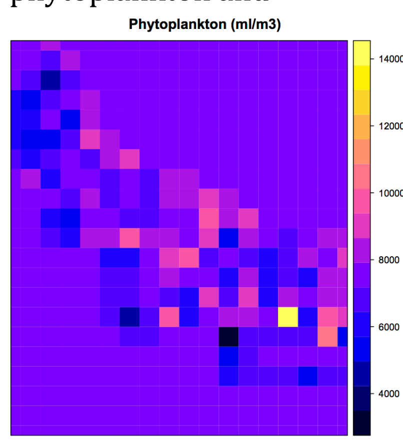


Zooplankton veriogram model verification

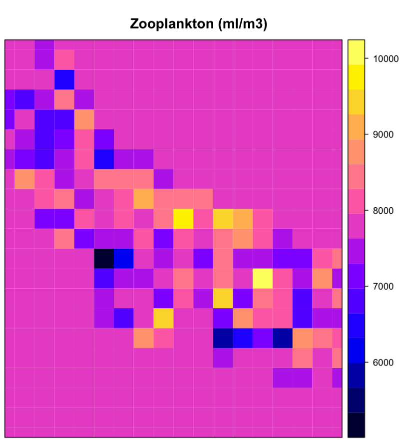
The resulting kriging graphs provide good spatial coverage and temperal variability. Two sample weeks are shown below for phytoplankton and zooplankton.



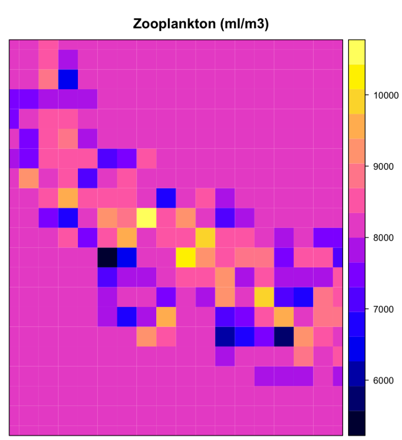
Week 1: Phytoplankton krige



Week 151: Phytoplankton krige



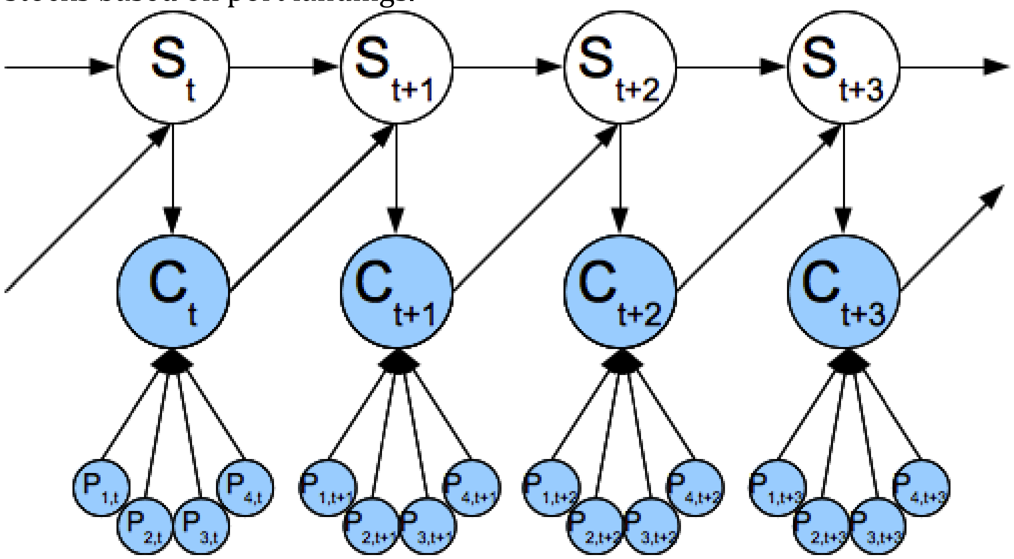
Week 1: Zooplankton krige



Week 151: Zooplankton krige

4.3 Modeling fish stocks with a hidden Markov model:

Next a hidden Markov model and expectation-maximization are used to estimate anchoveta stocks based on port landings.



In the diagram above, the observed catch (C_t) is informed both by a hidden stock variable (S_t) and by information denoting which of the ports are active ($P_{i,t}$). $P_{i,t}$ is a binary variable for each of the three ports for which data is available, and $P_{4,t}$ is a binary variable for the remaining undisaggregated ports (which are treated as one). $P_{i,t}$ is 1 if the port reported a catch greater than 0. For the undisaggregated ports, ($P_{4,t}$) is 1 if the country-wide total catch minus the sum of the available ports is greater than 0. The time periods when $P_{i,t}$ is 0 are marked in red in the catch figure above (labeled *Weekly catch data for ports Pisco, Chimbote, Paita, and total.)*

We assume that growth follows a logistic model,

$$\begin{aligned} S_{t+1} &= S_t + rS_t\left(1 - \frac{S_t}{K}\right) - C_t + \epsilon_t \\ &= \alpha S_t + \beta S_t^2 - C_t + \epsilon_t \end{aligned}$$

for some unknown parameters, S_0, α, and β.

We also assume that the catch in each time period is drawn from a probability distribution which scales linearly with the effective stock. The effective stock is

$$\bar{S}_t = S_t \sum_i \gamma_i P_{i,t}$$

where γ_i is the factional influence of each port. Let $\sum_i \gamma_i = 1$. Therefore, we assume that there exists a probability distribution function f such that

$$P(C_t|\bar{S}_t) \sim f(\frac{C_t}{\bar{S}_t})$$

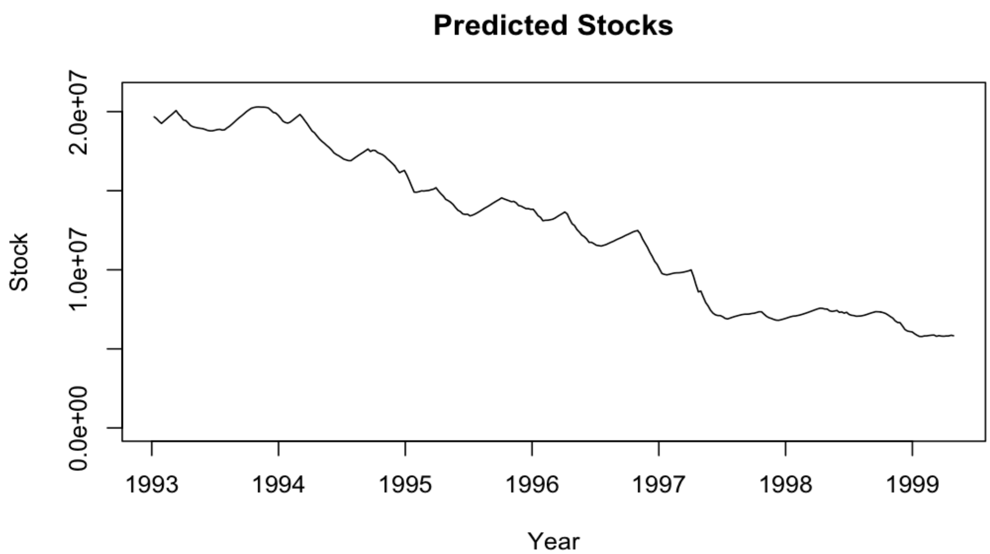
This assumption is valid if the effort did not significantly change over the time period and if there are no density-dependent effects in anchoveta schooling behavior.

The aim of the estimation-maximization procedure is to simultaneously estimate this function f and determine the model parameter estimates which produce a maximum likelihood under it:

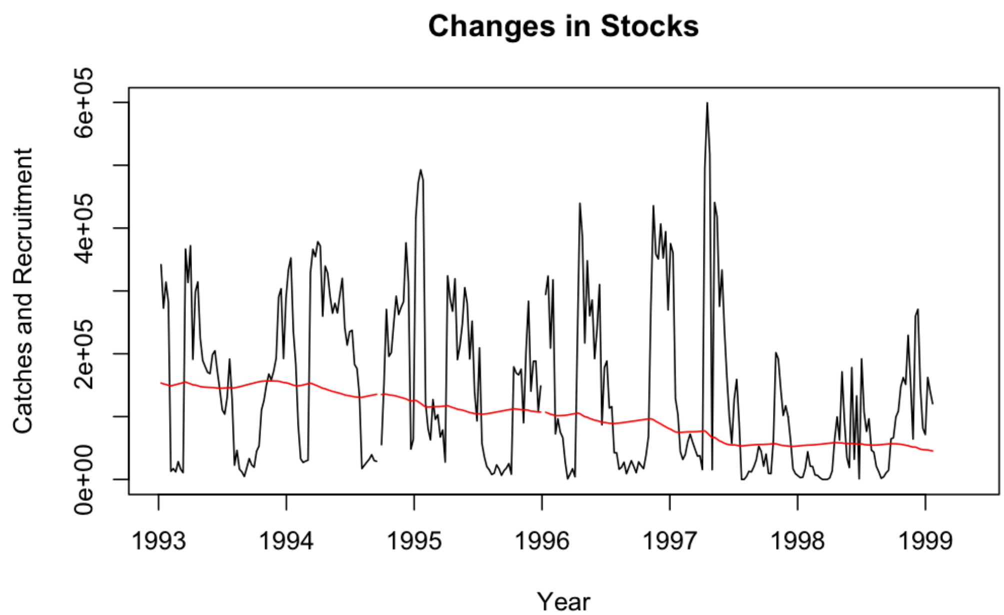
$$\max_{S_0,\alpha,\beta,\{\gamma_i\}} \sum_t \log f(\frac{C_t}{S_t(S_0,\alpha,\beta)\left(\sum_i \gamma_i P_{i,t}\right)})$$

The specific steps are enumerated in supplemental material section B.

The predicted stock is shown below.

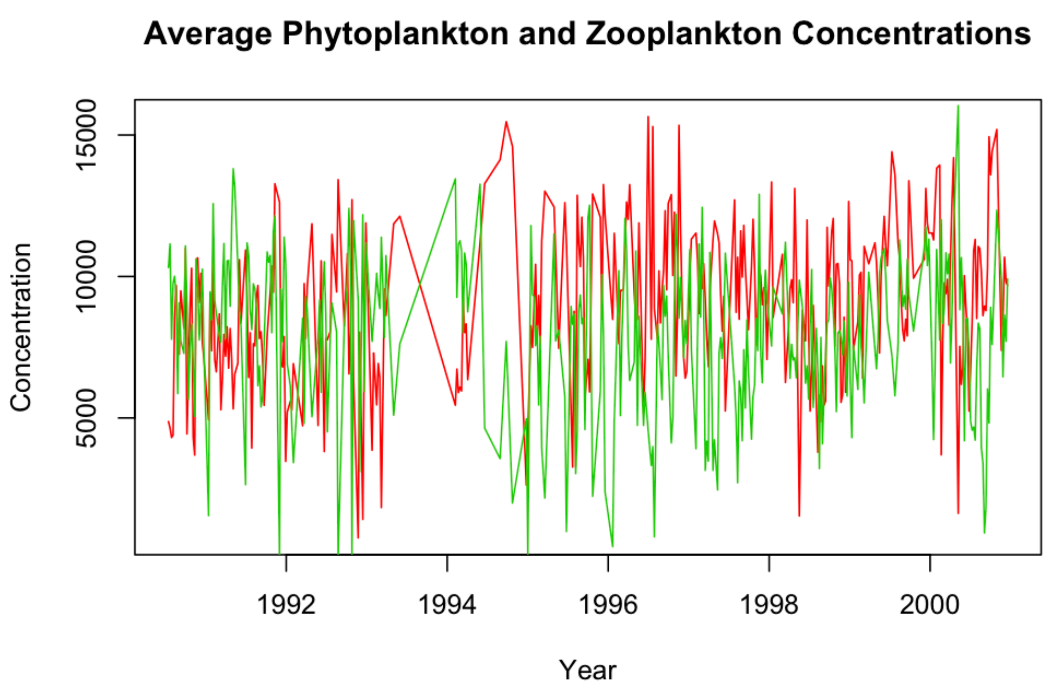


From the total stock, we can also estimate the stock growth ($G_t = S_{t+1} - S_t + C_t$), which we will correlate with plankton across

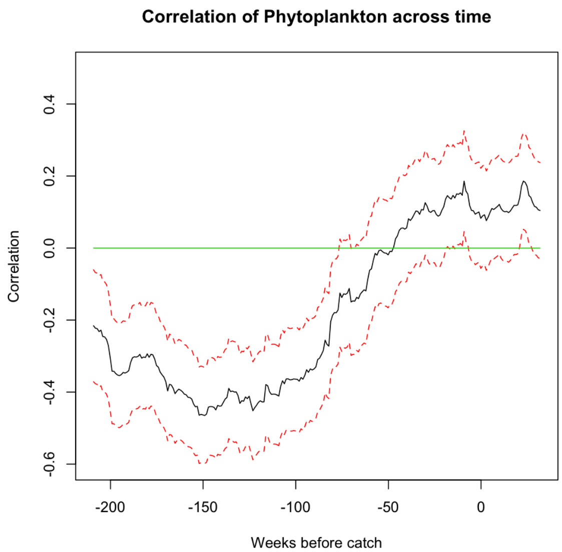


space:
4.4 Average correlations

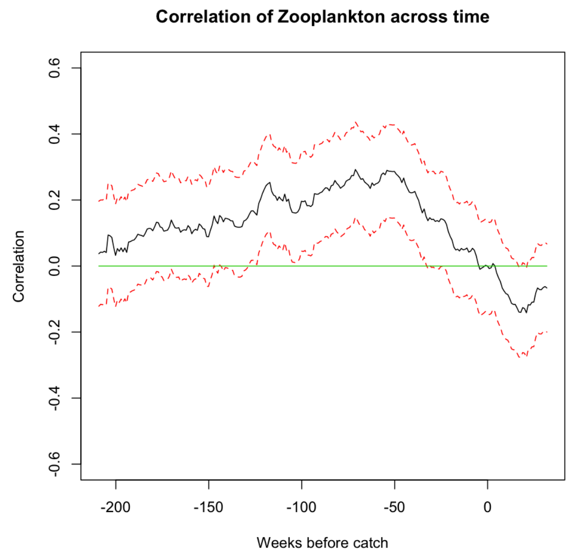
Anchoveta eat zooplankton, so we expect to see a correlation between anchoveta growth and average levels of plankton. That correlation could be delayed, based on the maturation and spawning time of the anchoveta. The average plankton concentrations are shown below to give an indication of their variability.



We correlate the predicted anchoveta stock growth against measured average plankton concentrations, delayed from 0 to 200 weeks. We also include some leading weeks, as a falsification test.



Correlation between growth and phytoplankton

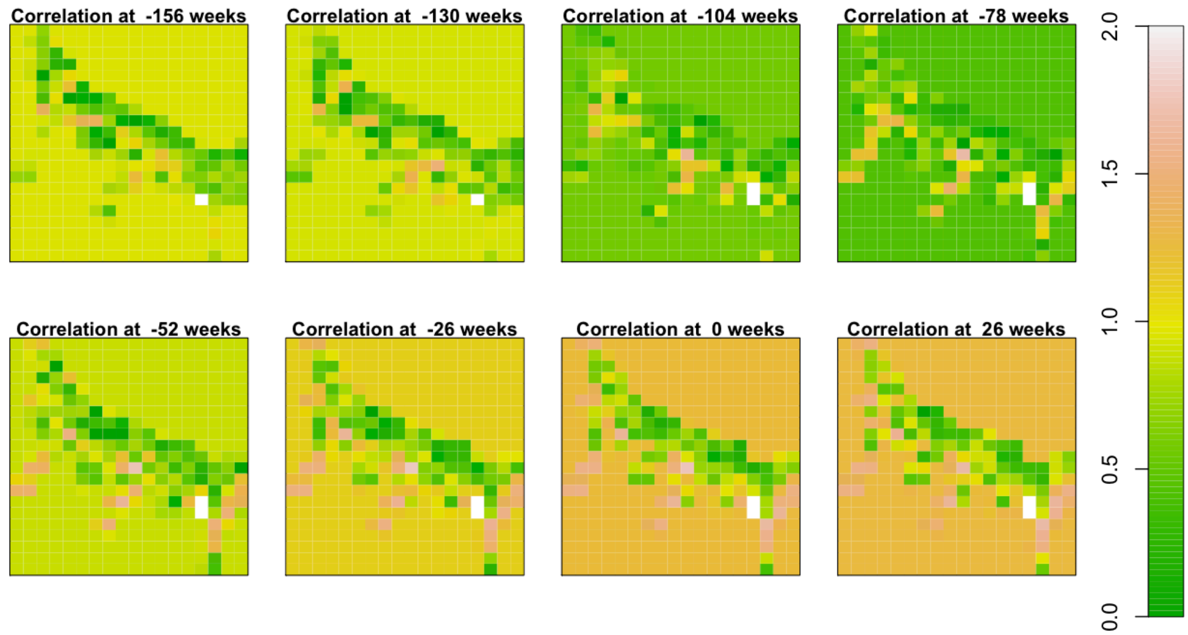


Correlation between growth and zooplankton

The results are promising. Phytoplankton abundance in the preceding two to four years correlates negatively with anchoveta stock growth. It is possible that this reflects the growth of competitors. Zooplankton abundance correlates positively with anchoveta growth, with 95% statistical significance in the preceding 33 to 128 weeks.

4.5 Spatial correlation

Finally, we perform the delayed correlation for each grid cell. Where the 95% limits include 0, the correlation is set to 0. Elsewhere, we sum the absolute value of the correlation for both phytoplankton and zooplankton. This gives an indication of what regions anchoveta, and their food chain, draw upon for various states stages of their lives.



Qualitatively, the following results are suggested:

In the period 2-3 years before spawning, very few regions show importance.

Around 1.5-2 years before spawning, the northern region is important, and regions far from the shore.

Around 0.5-1 year before spawning, the region closer to the shore becomes important, and the stock more integrated across the entire region.

Conclusions:

It was found that stock growth correlates positively with the abundance of zooplankton in the period from 33 to 128 weeks prior (95% statistical significance). The correlation varies considerably by region.

These interesting results suggest a need for further research to better understand the dynamics of the plankton-Anchoveta relationship. By improving our kriging model, we could improve plankton estimates, which would be useful in confirming or modifying the correlations found here. This could be accomplished by adding predictors to the variograms in our kriging model, such as latitude, distance from shore, and meridional Atmospheric Circulation Index, which has been associated elsewhere with increases in Anchoveta population (FAO, 2001). Closer study of the correlations between phytoplankton and zooplankton could also help to shed light on the reasons for the time-lagged negative correlation between phytoplankton and anchoveta. Finally, using a hidden Markov-Monte Carlo model might allow the results here to be compared more directly with findings from Bayesian models in the existing literature.

References:

Aguero, V. (1987) *A Bioeconomic Model of the Peruvian Pelagic Fishery.* In: Pauly, D., Tsukayama, I. (ed.) *The Peruvian Anchoveta and its Upwelling Ecosystem: Three Decades of Change*. ICLARM Studies and Reviews No. 15. Callao, Peru, IMARPE. p. 307-324.

Ascioti, F., Beltrami, E., Carroll, T., Wirick, C., 1993. *Is there chaos in plankton dynamics.* Journal of Plankton Research. 15, 603–617.

Bakun, A. 1987. *Monthly variability in the ocean habitat off Peru as deduced from maritime observations, 1953-84.* In: Pauly, D., Tsukayama, I. (ed.) *The Peruvian Anchoveta and its Upwelling Ecosystem: Three Decades of Change*. ICLARM Studies and Reviews No. 15. Callao, Peru, IMARPE. *p.* 46-74*.*

Brainard, R.E. and D.R. McLain. 1987. *Seasonal and interannual subsurface temperature variability off Peru, 1952 to 1984.* In: Pauly, D., Tsukayama, I. (ed.) *The Peruvian Anchoveta and its Upwelling Ecosystem: Three Decades of Change*. ICLARM Studies and Reviews No. 15. Callao, Peru, IMARPE. *p*. 14-45*.*

Carrasco, S. and O. Lozano*. 1987 Seasonal and long-term variations of zooplankton volumes in the Peruvian Sea, 1964–1987.* In: Pauly, D., Tsukayama, I. (ed.) *The Peruvian Upwelling Ecosystem: Dynamics and Interations*. ICLARM Studies and Reviews No. 15. Callao, Peru, IMARPE. *p. 82-85.*

Castillo, S. and J. Mendo 1987. *Estimation of unregistered Peruvian anchoveta (Engraulisringens) in official catch statistics, 1951 to 1982.* In: Pauly, D., Tsukayama, I. (ed.) *The Peruvian Anchoveta and its Upwelling Ecosystem: Three Decades of Change*. ICLARM Studies and Reviews No. 15. Callao, Peru, IMARPE. *p. 109-116.*

Chen, Y., Jiao, Y., and Chen, L. 2003. *Developing robust frequentist and Bayesian fish stock assessment methods*. Fish and Fisheries. 4:105-120.

Climate change and long-term fluctuations of commercial catches: the possibility of forecasting. FAO Fisheries Technical Paper. No. 410. Rome, FAO. 2001. 86p.

Denman, K., 1976. *Covariability of chlorophyll and temperature in the sea*. Deep Sea Research. 23, 539–550.

Doveri, F., Scheffer, M., Rinaldi, S., Muratori, S., Kuznetsov, Y., 1993. *Seasonality and chaos in a plankton–fish model*. Theoretical Population Biology. 43, 159–183.

Hiruki-Raring et al. 2012. *A Bayesian hierarchical model of Antarctic fur seal foraging and pup growth related to sea ice and prey abundance.* Ecological Applications. 22:668-684.

Malchow, H., 1994. *Nonequilibrium structures in plankton dynamics.* Ecological Modelling. 75/76, 123–134.

Mäntyniemi, S. 2001. *A hierarchical Bayes model for assessing salmon (Salmo salar L.) parr and smolt populations.* Fisheries Management and Ecology. 5, 209-223.

Palomares, M.L. and P. Muck. 1987. *Growth of the Peruvian Anchoveta (Engraulisringens), 1953 to 1982.* In: Pauly, D., Tsukayama, I. (ed.) *The Peruvian Anchoveta and its Upwelling Ecosystem: Three Decades of Change*. ICLARM Studies and Reviews No. 15. Callao, Peru, IMARPE. *p. 117-142.*

Pascual, M., 1993. *Diffusion-induced chaos in a spatial predator–prey system.* Proceedings of the Royal Society B. 251, 1–7.

Pauly, D. and M. Soriano. 1987. *Monthly spawning stock and egg production of Peruvian anchoveta* (Engraulis ringens), 1953 to 1982. In: Pauly, D., Tsukayama, I. (ed.) *The Peruvian Anchoveta and its Upwelling Ecosystem: Three Decades of Change*. ICLARM Studies and Reviews No. 15. Callao, Peru, IMARPE. p.167-178.

Punt, A.E. & R. Hilborn. 1997. *Fisheries stock assessment and decision analysis*: *A review of the Bayesian approach*. Reviews in Fish Biology and Fisheries. 7: 35-63.

Pyo, H.D. and K.N. Lee. 2003. *A Comparative Analysis of Surplus Production Model and Maximum Entropy Model for Estimating the Anchovy's Stock*. EAFE Conference, Ifremer, Brest, France. Session 12.

Scheffer, M., 1991b. *Should we expect strange attractors behind plankton dynamics*—*and if so, should we bother*. Journal of Plankton Research. 13, 1291–1305.

Scheffer, M., Rinaldi, S., Kuznetsov, Y., van Nes, E., 1997. *Seasonal dynamics of daphnia and algae explained as a periodically forced predator–prey system*. OIKOS 80, 519–532.

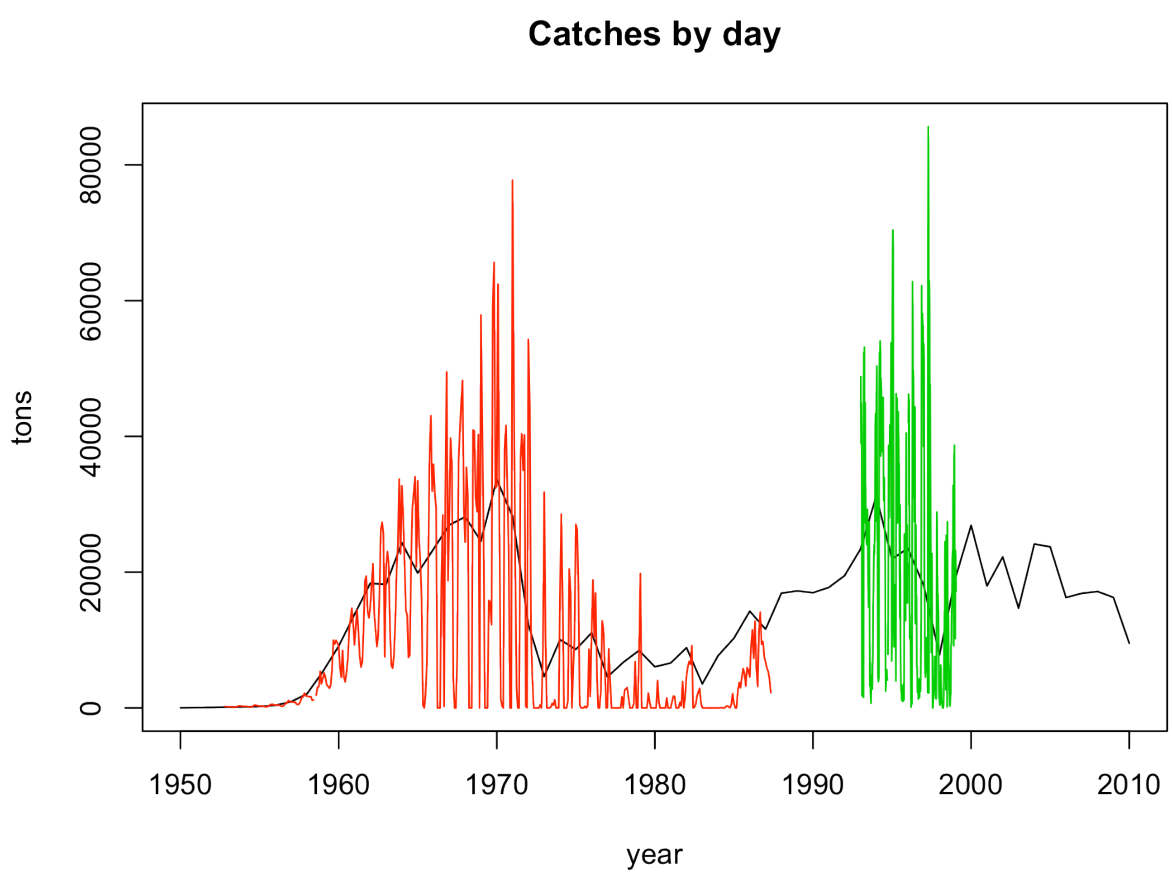
Tsukayama, I. and M.L. Palomares. 1987. *Monthly catch and catch composition of Peruvian anchoveta* (*Engraulis ringens*) (Northern-Central Stock, 4-140S), 1953 to **1982.** In: Pauly, D., Tsukayama, I. (ed.) *The Peruvian Anchoveta and its Upwelling Ecosystem: Three Decades of Change*. ICLARM Studies and Reviews No. 15. Callao, Peru, IMARPE. p. 89-109.

Weber, L., El-Sayed, S., Hampton, I., 1986. *The variance spectra of phytoplankton, krill and water temperature in the Antarctic ocean south of Africa*. Deep Sea Research. 33,1327–1343.

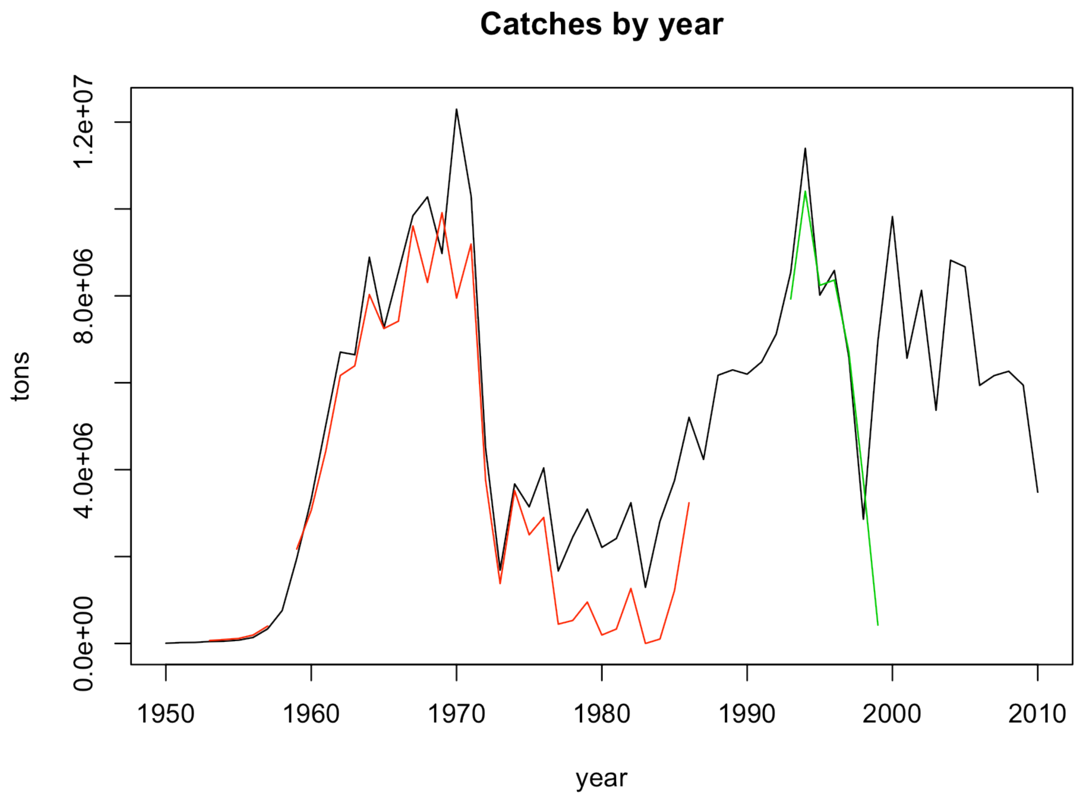
Spatial fisheries supplemental material

A. Data Selection

We considered a variety of anchoveta landing series for Peru to analyze in conjunction with the plankton data. FAO provides yearly production data since 1950, while IRI's EFFP contains both monthly catches from 1953 to 1987 and weekly catches from 1993 to 1999. All sources provide the catch in tonnes. The graph below displays these three sources, scaled to daily values.



Between these, the weekly data set seems superior for three reasons. (1) It provides the most data points within the time period for which plankton data is available (330 as compared to 34 (yearly) and 322 (monthly)). (2) It corresponds very well to the yearly totals, while the aggregate monthly data diverges considerably after 1976 (see figure below). (3) The weekly data is available separately at three ports, Pisco, Chimbote, and Paita. These ports account for 65% of the Peru anchoveta catch during this time period.



The monthly data is available disaggregated into length classes (monthly tonnes by length, 4 - 20 cm). Incorporating this length data as a proxy for age distributions could be used to improve our lifecycle results.

B. Expectation Maximization Procedure

As usual, the EM algorithm alternates between estimating a probability distribution and identifying the maximum likelihood parameters under it. However, the complicated functional forms involved make identifying the maximum likelihood parameters difficult. Instead, a randomized algorithm is used.

The estimation maximization procedure used the following steps:

1. Assume initial values for S_0, α, β, and $\{\gamma_i\}$. The procedure is very sensitive to this initial set of values. The values used are listed in the table below.
2. Model the stock growth under these parameters, producing a sequence $\{S_t\}$.

<u>Estimation</u>

3. Estimate the probability distribution for catch given effective stock, $f(\frac{C_t}{S_t \sum_i \gamma_i P_{i,t}})$, as a kernel-estimated empirical distribution.
4. Determine the summed log-likelihood of the sequence of observed catch and estimated stocks, under the estimated distribution.

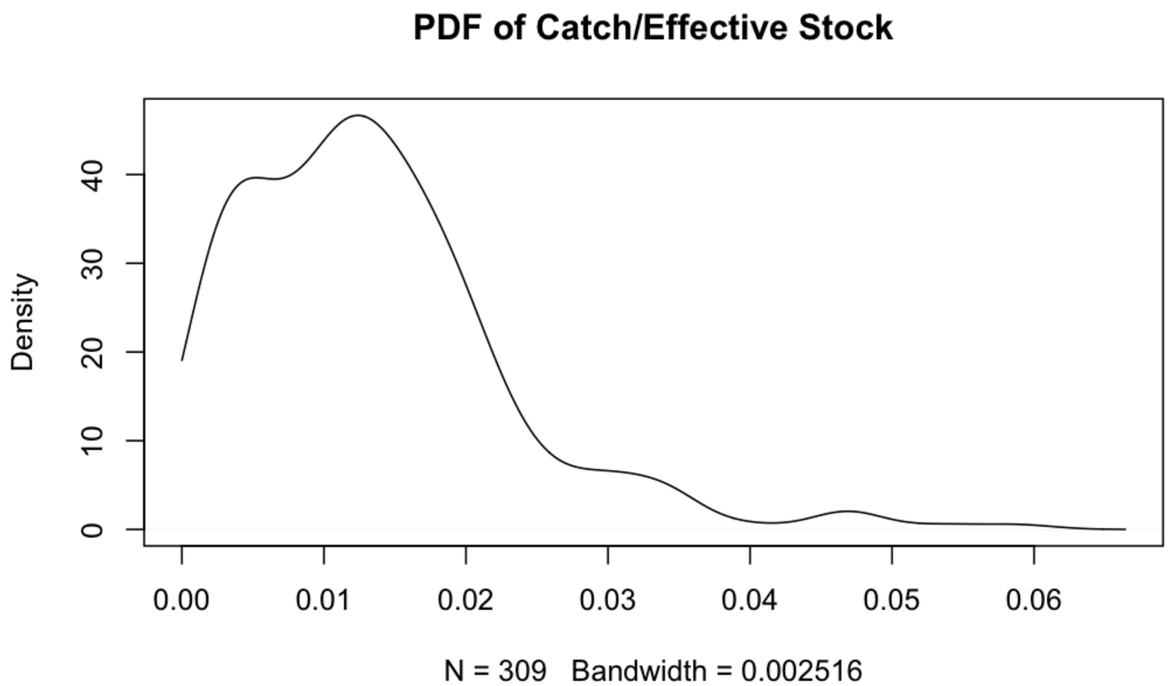
Maximization

5. Construct a randomly adjusted new set of parameters, slightly changed from the most optimal set identified.
6. Model the growth under these parameters.
7. Calculate the summed log-likelihood under the existing empirical distribution.
8. If the summed log-likelihood for the new parameters is greater than the previous most-optimal parameter set, use these new parameters as the new most-optimal set and return to step 3.
9. Otherwise, return to step 5.

Below are the initial parameters used, a rationale for selecting them, and the final parameters determined by the EM algorithm. Rather than reporting α and β, we show the underlying logistic parameters, R (the growth rate) and K (the carrying capacity).

	Initial	**Rationale**	**Final**
S_0	16 790 000	28 x maximum catch	22 790 000
$R = \alpha - 1$	.01	arbitrary	0.007745
$K = -(\alpha - 1)/\beta$	1 679 000 000	100 x S_0	10 400 000 000
γ_1 (Pisco)	0.4168	$\max(P_{1,t})$/sum maxes	0.2127
γ_2 (Chimbote)	0.1045	$\max(P_{2,t})$/sum maxes	0.1395
γ_3 (Paita)	0.3013	$\max(P_{3,t})$/sum maxes	0.2692
γ_4 (remaining)	0.1775	$\max(P_{4,t})$/sum maxes	0.3785

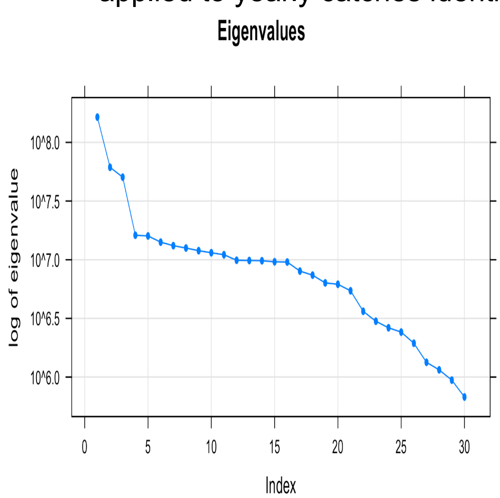
The estimated distribution for catch/effective stock is below.

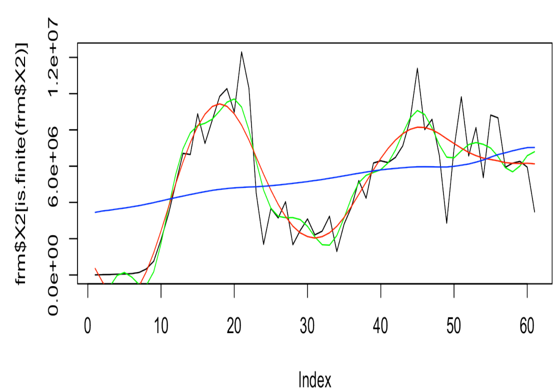


C. Some missing patterns

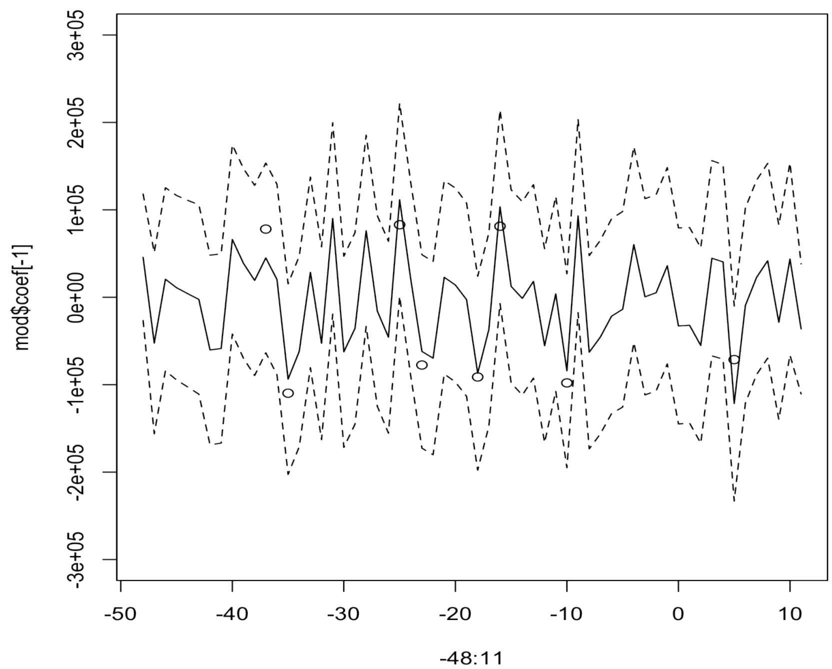
We failed to find some patterns that might be expected.

- The correlation analysis of plankton with catch shows no pattern and no extended regions of statistically significant correlation. This suggests that fishers do not use plankton as an effective indicator for regions to fish. Based on our paper's results, the low correlation of plankton with current fish stocks suggests that this is reasonable.
- Catches are not well-predicted by NINO3 or any combinations of it delayed. SSA applied to yearly catches identified no cycles of the relevant length:





An AIC model selection from of two years of delayed monthly NINO3 values, linearly predicting monthly catches, identified eight delays which combined produce a R^2 of only .056. The graph below shows the model parameter estimates for all delays, and identifies (dots) the delays selected by the AIC



criteria.