Fishermen vary in how social they are when they fish. We think that social foraging might be more efficient at fishing (lower CPUE) due to reduced uncertainty in fish location, but no one has been able to measure this empirically before. Therefore I’m quantifying how social the fishermen and whether it relates to their efficiency at fishing.

Introduction

*Motivation for studying fishing behavior*

* To better predict how changes in management will play out (Little et al. 2004 - Sente)
* Empirically studied to look at impacts on habitat (Bellman et al. 2005 – Sente)
* To determine if CPUE actually is a good proxy for fish abundance (Lorenzen et al. 2006, McCluskey & Lewison 2008, Sente)
  + Fishing effort, even if not a direct measure of abundance may still tell us informative things about the distribution of fish populations (Stewart et al. 2010, Sente)

*Fishing behavior has been treated in a variety of ways*

* people have tried to build neural networks to see if realistic fishing behavior has emerged (Dreyfus-Leon 1999, sente)
* Built ABMs to see if realistic fishing behavior and social norms have emerged (Wilson et al. 2007, Sente)

In terms of fishing skill: *Differences in fishing strategy have been studied, often focusing on differences in skill*

* Random effect chosen over fixed effect model for Pacific Coast Trawl Fishery

Interference has been studied: *Fishing in close proximity to other boats could either indicate interference or prosocial fishing.*

* Found that interference exists and can affect CPUE estimates (Gills & Peterman 1996, Sente)

Fishing norms: *The distribution of pink shrimp cooperation could be affected by social norms rather than ecological/economic conditions alone.*

* Found that fishing norms strongly affected fishing strategies (Gezelius 2007, Sente)

Fishing as foraging: *Fishing behavior is often studied using foraging theory as a theoretical framework*

* Pacific Hake vessels respond to changes in catch rate by shifting fishing patches (Dorn 1997, Sente)
* Integrated entry-exit-foraging model of Hake (Dorn 2001, Sente)
* As ideal free distribution (Abernethy et al. 2007, Sente)

Fishing location choice: *One of the main ways fishing behavior has been studied is trying to predict where fishermen will fish*

* Predict change spatial distribution of fishing after change in management (Vermard et al. 2008, Sente)
* Understanding the location choice decision (Holland 2008, Sente)

Fishing for information: *Fishing behavior has also been framed as a search for information*

* Trade offs of costs and benefits of sharing information (Wilson 1990, Sente)
* Theoretically allowing vessels to communicate in ABM (Little et al. 2004 –Sente)

Fishing behavior and CPUE:

* Looked at correlations between spatial distribution and fishing patterns (Bertrand et al. 2008, Sente)
* Tropical tuna purse seine fishery is managed using CPUE, wanted a more accurate estimation of effort (Bez et al. 2011, sente).

West coast fisheries generally: *The US West coast fisheries system, and fleet dynamics in particular, have been previously described.*

* Hake: Dorn 1997, Sente
* Hake: Bellman 2005, Sente (calls it limited entry groundfish)

*The Pacific Coast shrimp fishery fleet dynamics have also been well studied*

* Whether skipper skill is managerial or just based on finding fish (Squires & Kirkley 1999)

*However there is no stock assessment for the species*

* <http://www.fishchoice.com/buying-guide/pink-shrimp>

Methods

**Behavioral Segmentation**

Defining fishing patches

* Spatial clustering to define fishing opportunities and grounds (Branch et al. 2005 Sente)

Segmenting fishing behavior

* Using Artificial Neural Networks on Peruvian Anchovetta purse seiners, correctly classified 83% of fishing sets and outperformed linear model. Found that fishing location was a good proxy for fish spatial distribution. Speed tends to overestimate fishing (+182%) due to similarities in speeds between fishing, searching and drifting
  + VMS reports approximately ever hour and set are on average 2 hours long. Fishing trips last between a few hours and a 2 days. And between one and 5 fishing sets occur on a trip. Sample was about 1350 vessels starting in 2000-2002. Accuracy was ~100m.
  + Pre-processed to methods described in Bertrand et al. (2005, 2007). Removed VMS < 2 nautical miles from shore. Also because trips were between 2 and 120 hours, discarded trips that were shorter or longer. And because fishing requires speed of < 3 knots, discarded trips where the speed was always higher.
  + Used Turchin (1998) method of describing moves where trips are split into elementary moves, defined as the sum of the consecutive positions of a vessel in the same direction. Because instantenous direction at sea is liable to be highly variable, used a threshold of 10° and considered that in the same direction. Move lengths were calculated as great-circle distances
  + Derived higher order variables from lat/lon/time stamps. These included speed, absolute turning angle, speed variation (difference in speed between two time segments). Of raw and derived variables selected absolute turn angle, speed, hour, two measurements of speed variation (current and next position and current and previous). Standarized and linearized these features (linearizing the circular variables). Used mean squares error function (MSE[[1]](#footnote-1)) between simulated output and observed output. To prevent overfitting, use early stopping method. To do this they split data into three segments: training, test and validation. ANN is trained using training data and validates to validation set until the error starts to rise on validation set. Then stop and predict to test dataset. This prevents over-fitting to training model but ensures an accurate estimate of error. With the optimal number of neurons they got about 83% correct classification of fishing events[[2]](#footnote-2)
* Vermard et al. (2010, Sente) segmented VMS data using a Bayesian Hierarchical Model (HBM) using a Hidden Markov Process (HMP) on pelagic trawlers in the Bay of Biscay. Define 3 states of fishing, steaming and stopping at port/sea. Parameters of speed and turning angles are defined conditionally on behavioral states. First test on simulated data with different parameter structures. Find that if relocations occur at the instantaneous point of behavioral change that the methods are unbiased. But if observations are not gathered with sufficiently high frequency the estimation method could be drastically impacted. The model then is applied to real data. However it appears that they don’t have validation data.
  + Time steps are approximately 2 hours, gaps are not uncommon. Trawling duration was between 5-8 hours.
  + Real data had < 2% fishing events at higher frequency than data.
  + Simulations accuracy hard to decipher defined as the relative discrepancy between the estimated and simulated mean and defined as . But doesn’t define what is. But generally score between -0.2 and 0. Depending on behavioral state.
* Bez et al. (2011, Sente) used a Bayesian state-space model and classified tropical tuna (yellowfin, skipjack, bigeye) purse seine fishing into 3 states: fishing, tracking and cruising.
  + Had data between 2006-2008, 131 fishing trips performed by 18 french purse seiners.
  + Used observer data to validate, had 10% of trips validated. 11 trips, 301 days at sea, 265 sets were available
  + Dropped any still phases during the night, since fishing only occurs during daylight.
  + Resampled to get Turchin (1998) moves, so requires identically sampled relocations.
  + Correctly estimated 97% of the fishing sets. But doesn’t say anything about false positives[[3]](#footnote-3). Although previous paper (Walker et al. 2010) has MSE 10.3% of hourly steps were misclassified.
* Joo et al. 2013 compares HMMs and “discriminative models” (classification models) performance on Peruvian anchovetta. VMS data is approximately hourly and look for three states: fishing, searching and cruising. Have observer data for validation.
  + Models to compare include HMMs, HSMMs, RF, ANN and SVM
  + HSMMs show the highest accuracy (80%) significantly outperforming HMMs and discriminative models.
  + With higher resolution HSMMS reach even higher accuracy (approaching 100%).
  + Results demonstrate the sequential nature of bheaivor is critical for accurately inferring behavioral modes.
  + Resolution is ~100m, with some irregularities in time between relocations. But because no optimal way to infer these gaps work with data as they are.
  + For each VMS track compue speed, heading, changes of speed and turning angles (delta speed and delta theta) between previous and next step.
  + Have observers that classify all behavioral states, so have continuous record.
  + Use 2008 data which corresponds to 242 fishing trips for observation out of ~360000 total that year. [[4]](#footnote-4)
  + Groundtruthed dataset is split into two sub-samples. First partition is used for training. And second is used for partition. And are built by repeated random sub-sampling (20 repetitions)[[5]](#footnote-5).
  + Classify model performance at 3 scales[[6]](#footnote-6)
    - Accuracy: Use percentage of individual steps that were classified correctly.
    - Recall: predicted & true/# true
    - Precision: inferred & true/# inferred
    - F1 = 2xprecision x recall/precision + recal;
    - Duration = sum(F – G)^2/n (F = real fishing, G = predicted fishing)
  + Found all models had accuracy > 75%. But did this for each behavioral mode. Cruising was the easiest to get.
  + HSMM was best with F1 scores of fishing: 77%, searching: 67% and cruising 89%. HMM was next best, and ANN was the best classification model.
  + Found that classification models which didn’t take into account the order of sequences tended to over-estimate the duration of modes due to over-segmentation.[[7]](#footnote-7)

**Measure of sociability**

Previous quantitative measures of sociability are often disease, animal behavior, or cross-species contacts, but all tend to measure the proportion of time spent near conspecifics relative to total time observed. In fisheries, proximity is mostly thought of as related to information sharing.

Disease

* Cowie et al. {\*Cowie:2015hs} look at interaction rates between cows, red deer, and wild boar. Had GPS proximity loggers that were triggered to log if an animal came within 1.5 meters. Measured direct contacts as the time within 1.5 meters over total time.

Animal behavior

* Cote et al. {\*Cote:2012da} looked at sociability of fish and related it to shoal characteristics. Measured sociability as proportion of time a fish spent near a group of conspecifics.

Fisheries

* Little attention has been paid to the role of information sharing in RUMs with the assumption that fishermen generally all know the value of sites and indeed there is evidence in the anthropological literature that it is quite the opposite {Curtis:2004ec}. Curtis & McConnell {\*Curtis:2004ec}found that after parameterizing a RUM for the Hawaii longline fishery with three different ways vessels could update their ideas about patch quality, that the model with only information at the start of the trip did best.

1. where is the true classification value for observation *I* [[Wikipedia](https://en.wikipedia.org/wiki/Mean_squared_error)]. [↑](#footnote-ref-1)
2. I think this refers to the number of correct fishing events/total fishing events. So says nothing about whether they’re overclassifying non-fishing events. Although if their MSE is indeed around 0.03, it’s likely extremely small. [↑](#footnote-ref-2)
3. This is particularly frustrating because my first run, crappy RF gets 91% accuracy on fishing sets, but dramatically over-estimates non-fishing events. [↑](#footnote-ref-3)
4. For perspective, using burst\_ids we have 1081 trips on which to train. [↑](#footnote-ref-4)
5. Not clear if the models are bootstrapped 20 times from randomly resampling training partition. Or if the partition of data into training and validation is repeated 20 times. [↑](#footnote-ref-5)
6. For reference my out of the box RF was precision = 0.6769, recall = 0.9108 and F1 = 0.7766. This was for fishing. For not fishing was much worse, recall = 0.097, precision 0.344, f1 = 0.151 [↑](#footnote-ref-6)
7. I think this means flipping back and forth between behavioral states and chopping up one behavioral state into a bunch of little ones. [↑](#footnote-ref-7)