**CHAPTER THREE**

**Sociality in US west coast shrimp fishery: correlated to success?**

*Abstract*

*Introduction*

Here I seek to determine a) whether there’s a gradient of sociality among shrimp vessels on the US west coast, b) whether this sociality is different than what would be expected by chance and c) whether it is correlated to fishing success.

*Materials and Methods*

Here I determine how often shrimp vessels fish near one another, and whether it differs from proximity patterns due to random chance. To answer these questions I need to know when and where a vessels are fishing. Further, fishing in the US west coast shrimp fishery is patchy, with fishing occurring on a number of distinct fishing grounds. Based on informal interviews, which patch to visit is the first decision vessels make. Thus to determine whether vessels respond to the presence or absence of other vessels in making their fishing decisions, I needed to delineate patches. To fisheries data nominally captures all aspects of a fishing trip in order to do this calculation, however substantial data aggregation, filtering and processing had to occur to pull out the relevant information. In the following I describe the raw data sources, processing and aggregation.

**Description of Data sources**

We used three data sets to examine fishing strategies in the US west coast pink shrimp trawl fishery (hereafter shrimp trawl fishery): landing tickets, vessel monitoring system (VMS), and observer data. All data sets covered the period 2009-2013. Each dataset has different coverage of vessels which I describe below.

To identify trips to a particular fishery I used landings tickets that record all commercial landings on the US west coast from the Pacific Fisheries Information Network (PacFIN) database ([www.psmfc.org](http://www.psmfc.org/)). Landings tickets provide species composition, price, date, and vessel identification information for all commercial landings on the US west coast during this time. Using a métier analysis (Fuller et al. 2016) I subset landings to those of the pink shrimp trawl fishery (n = **XXX**).

To determine where vessels fished, I used VMS data from the Office of Law Enforcement (OLE). VMS are GPS systems which transmit location, speed, and bearing approximately every hour with a resolution of approximately **XXX**meters. As of 2009, all vessels which participate in federal groundfish fisheries are required to carry a VMS. I filtered this data for coordinates in the US west coast exclusive economic zone (EEZ), and filtered to vessels for which I have pink shrimp landings data (n vessel = **XXX**).

To determine when vessels were fishing, I used federal observer data came from the Northwest Fisheries Science Center Observer Program. In the pink shrimp fishery, observers are required to be on **XX**% of shrimp trips, during which time they record haul locations, durations and species composition including bycatch.

**DETECTING FISHING BEHAVIOR**

Using these three datasets I matched landings, movement and fishing behavior to fully capture a pink shrimp fishing trip. Landing tickets and VMS data were matched by vessel ID and date of landing (see Appendix for details). Using these matched trips, I filter VMS data for trips linked to pink shrimp trawl fishery landings. Using this set of vessels for which we have landing tickets and observer data, I subset to GPS data corresponding to these recorded trips. Using this set of validated data for which I have landing tickets, observer data and VMS data I built a random forest (RF) classification model. Random forest is a non-parametric machine learning algorithm which makes uses of training sets to develop discrimitive rules for classifying multivariate data (see Appendix for details).

**DEFINING FISHING GROUNDS**

Fishing effort is patchy in the US west coast shrimp fishery. These patches vary in their size and distance apart, with some patches of intense fishing activity flowing into neighboring patches. To segment this fishing activity into patches, I used VMS data classified as “fishing” to spatially bin vessel locations along the US west coast in a 2x2 kilometer raster. Using a kernel smoothed intensity function as calculated by the *density* function in the *R* package *spatstat*, I calculated a polygon of the **XXX** percentile of points. This results in 11 polygons, which are approximate mediods of fishing activity in each patch. Using the polygons as mediods, I calculated pair-wise distances between all fishing points and assigned them to the closest patch.

**Defining social fishing**

I define social fishing as vessels which, when given the choice, opt for patches with vessels already present. Thus for each patch I counted the number of vessels present on each patch in each hour. Using these time series, I calculated for each trip, what percentage of the time the vessel was predicted to be fishing was spent on a patch occupied by another vessel.

To correct for different total numbers of vessels fishing in different patches, I permuted my data in the following way. First for each trip I calculated the maximum distance traveled. Using this as a radius, I calculated a circle and determined all patches a vessel traveling the same distance might have visited. I then subset each of the patch occupation time series for the relevant patches to the time period of the trip. I then reshuffled the occupation time series randomly across all relevant patches and let the trip occur again, re-calculating the percentage of time the vessel spent fishing in occupied patches as expected by chance. For each trip I repeated this process 100 times, generating a distribution. To determine if the trip was more or less social than random, I calculated a t-test that the differences between the observed and distribution of random values was different than zero.

**MEASURING FISHING SUCCESS**

To determine how successful a fishing trip was I calculated the total revenue earned by multiplying the pounds of fish caught by the price-per-pound given on each fish ticket and adjusted to 2009 dollars to remove inflation. To control for variable effort I also calculated the revenue per hour for each trip by taking the total revenue divided by the total duration of the trip in hours.

*Results*

**DETECTING FISHING BEHAVIOR**

I found significant errors in the observer data used as validation data. Often observed trawls were at a time that when plotted alongside VMS data, indicated that a vessel started fishing as it exited port, or placed the vessel trawling several nautical miles away from its location according to VMS. To correct for these inaccuracies, I filtered observer and VMS data to only ‘high quality’ fishing locations: those observed hauls which occurred when the VMS data put the vessel within 5 kilometers of either the set or recovery location of a haul AND when these locations were recorded as being within the reported time interval given in the observer data. Using only high quality fishing data I was able to classify fishing behavior with **XX** percent accuracy and non-fishing with **YY** percent accuracy.

**DEFINING SOCIAL FISHING**

I found substantial variability in the average percent of time vessels spent fishing in an occupied patch. However,

*Appendix*

**Matching landing tickets and VMS**

Matching of tickets and VMS was completed using a combination of spatial and temporal filters. VMS data was first filtered and any segment of > 2 continuous locations > 1.5 kilometers offshore. These segments were labelled as “fishing trips”. Using landing tickets for a given vessel I start from the last date and search back for VMS trips which occurred within a 24 hour window of this date. Any VMS trips that are present in this time window are linked to the landing ticket. If more than one landing ticket or fishing trip occurred in the prior 24 hour window I group all landing tickets and VMS fishing trips.

**FISHING BEHAVIOR SEGMENTATION**

To detect fishing behavior during shrimp trips I built a random forest model which predicts whether a vessel is “fishing” or “transiting” based on statistics derived from the vessels movements. Movement statistics are calculated in the following way.

For each vessel trajectory, I derived a series of candidate movement statistics to use as covariates (Table A1). Once all metrics were created for all individuals, I normalized covariates.

|  |  |
| --- | --- |
| **Movement statistic** | **Description** |
| Displacement | Euclidean distance between the first and last relocation in a given time period. |
| Turning angle | Absolute and relative turning angles as computed by ltraj package in R. The absolute turning angle is measured as the angle between each move and the x axis. Relative angle is calculated as the angle between successive moves [1]. |
| R2n | The squared net displacement between the current relocation and the first relocation of the trajectory [1]. |
| Step length | Distance between relocations |
| Speed | Kilometers per hour |