Methods

Emma Fuller

July 12, 2015

# 1. Question

How does participation diversity (and revenue variability?) vary with ecological, economic, and/or management characteristics?

# 2. Materials and Methods

This work makes use of landings data, habitat metrics, economic data, and management data. In the following sections we describe the datasets and processing to extract independent and response variables. Finally we describe the statistical models we use. All analyses are performed using R (R Core Team 2015).

## 2.1 Data

Data are consolidated from four databases which record fisheries landings, port, and vessel information; landscape and spatial management covariates; species diversity; and demographic data.

Fisheries landings data comes from the Pacific Fishery Information Network database (PacFIN 2014), and includes commercial landings from all vessels fishing in the US Extended Economic Zone (EEZ) off the US west coast between 2009 and 2013. Landings receipts report the amount of fish caught by market category. These market categories are approximately equivalent to species-level identification, although not exact. Rockfish are most likely to be approximate, as species are often difficult to distinguish between. However, the bigger the price differential between species, the more accurate the market categories.[[1]](#footnote-1) Thus in the following work we assume that vessels are targeting market categories, rather than species, as that is the economically relevant determination.

For each market category, landing tickets report price per pound, date, port of landing, and vessel identifying information. After adjusting for inflation using the 2009 Consumer Price Index, we calculate vessel revenues by multiplying the price per pound by the number of pounds landed. We include only vessels with average annual revenues above $10,000 (adjusted to 2009 values), to exclude vessels for which fishing was a part time activity. To focus on major species, we remove landings from species that show up in fewer than 100 trips across our five year dataset and whose median catch is less than 100 pounds.[[2]](#footnote-2) We categorize individual landings to fisheries by performing a metier analysis, where a fishery is defined as a target assemblage and gear type (see Appendix). We link landings data to vessel registration at both the state and federal level to obtain vessel length and horsepower, using the average if more than one record is found for a single year (i.e. registered in multiple states and/or state and federal registration). We categorize vessels as either small or large. Small boats are defined as 40ft based on distribution of vessel size for 2009-2013. The 40ft cutoff is also in line with the small boat category used by Kasperski and Holland (2013).

To capture differences in mangement across state lines, we categorize each vessel by the state in which the majority of landings occur by revenue for each year. We also categorize whether a vessel gets at least 25% of their annual revenue from a seasonal fishery. A fishery is considered seasonal if, in any quarter, a fishery is not landed at the coast-wide scale (a list of these fisheries can be found in the Appendix). We calculate market value for each fishery in which a vessel participates as the sum[[3]](#footnote-3) of coast-wide landings revenue (i.e. volume price) for each year. Finally we flag vessels participating in the groundfish trawl ITQ using DAHL sector rules developed by the Pacific Management Council.

We also use landings data to measure species diversity and taxonomic distinctness of commercially targeted species assemblages present around each port. We convert landings into a presence-absence records by port and calculate species richness and taxonomic distinctness. Taxonomic distinctness is measured as

where is the number of species present, and for the double summation, and range over these species. is the distinctness weight (Clarke and Warwick 1998).[[4]](#footnote-4)

To check that landings data provide reliable estimates of underlying trends in diversity, we subset landings to those species caught by commercial bottom trawls and compare to fisheries independent data using the West Coast Groundfish Bottom Trawl Survey conducted by the Northwest Fisheries Science Center, U.S. National Marine Fisheries Service (Keller et al. 2005). We use surveys[[5]](#footnote-5) for the same time duration (2009-2013) and assign survey trawls to a port if they occur within 100 km radius. We find these measures of species richness and taxonomic distinctness to be tightly correlated.[[6]](#footnote-6)

Landscape data was obtained from the Groundfish Essential Fish Habitat Synthesis (NMFS 2013) which provides geospatial layers detailing bathymetry, substrate, and spatial restrictions to fishing. Using port latitudes and longitudes from the Pacific Management Councils list of fishing communities (Norman et al. 2007), we collate bathymetry, substrate and spatial fishing restirctions within a 100 km radius of each port after first masking using high resolution coastline polygons from the Global Self-consistent, Hierarchical, High-resolution Geography Database (GSHHG) provided by NOAA (Wessel and Smith 2013). Using these substrate data we calculate habitat diversity as the Simpsons diversity index of substrate type

so for substrate types, is the proportion of cells within 100 km of port which has substrate . Substrate has three values: hard, mixed, and soft. Values of zero indicate no diversity, i.e. a single substrate type, and values close to one indicate high levels of diversity. Diversity indices are calculated using the vegan() package (Oksanen et al. 2015). We use the bathymetry data to calculate the minimum distance from shelf for each port, and use the spatial management data to calculated the percentage of cells with fishing restrictions.

To measure market availability we calculate the distance to nearest large cities (defined as population > 500,000 as of the 2010 census) using the ggmap() package (Kahle and Wickham 2013). Finally we measure the yearly average number of first receivers present in each port as the number of unique processorid codes by year available in the landings data.

We measure revenue diversity (or portfolio diversification) with the Simpson’s diversity index (Kasperski and Holland 2013; Sethi et al. 2014). Simpson’s diversity index is calculated per vessel; i.e. have an index of where is the number of vessels. Thus is defined as previously in eq. 2. As above, values of 0 indicate no diversity (a single fishery), where values close to 1 indicate high levels of diversity. We calculate the diversity indices each year a vessel is active and averaged across years. Vessels had to be active for at least two years in order to be included in this analysis.

Revenue variability is calculated as the coefficient of variation in annual revenues. For a single vessel , the coefficient of variation () is calculated as

where is a vector of the annual revenues for vessel .

untitled:CNH:Analysis:Metiers:writing:variable_table.pdf

## 2.2 Statistical Models

In the following we examine two sets of models: revenue variability as predicted by participation diversity and participation diversity predicted by proxies for market, management and ecology conditions.

### 2.2.1 Revenue variability predicted by participation diversity

Before we examine the effect of the ecology, management and market variables on participation diversity, we first ensure that these variables act only via the effect they have on participation diversity. To do this, we use GLMs to examine the relationship between revenue variability and diversity alone and in the presence of a selection of the above covariates. These variables are: whether a vessel ever landed in an ITQ fishery, the market value of a fishery, the state in which a majority of landings occur (across all years) and whether the vessel participates in a seasonal fishery. In these models the replicates are vessels and predictor covariates are adjusted by taking the mean if previous designations were by landing or by year. **[needs work]** To interpret the main effects in the presence of interactions, we mean-centered continuous covariates.

### 2.2.2 Participation diversity predicted by ecology, management and market variables

After first checking to ensure correlation among predictor variables are low, we examine the relationship between participation diversity and ecology, management and market variables. Here vessel-years are replicates and we use beta GLMs with a logit link (as Simpsons diversity is constrained between zero and one). We use the vessel ID as a random effect and standardize categorical covariates (Schielzeth 2010 - Methods in Ecology and Evolution). We present effect size plots for models and examine interdependencies of different predictor variables.

# 3. Appendix

## 3.1 Metier analysis

Presently there exists no universal way to define a fishery. In US West coast commercial landings, the Pacific Management Council has developed a set of sector based definitions for groundfish landings, but no equivalent exists for non-groundfish fisheries. In order to treat the dataset uniformly, we apply a metier analysis to this landing data. To assign each trip to a fishery we perform a metier analysis for the US west coast commercial fisheries (Deporte et al. 2012). Previous metier analyses occur largely in Europe, although metier-like analyses have been performed in the Northeast US, classifying fishing data to define “operational fisheries” of New England (Lucey and Fogarty 2013). While promising as a way to classify fisheries for use in ecosystem-based management, these methods introduced spatial and temporal structure prior to defining fisheries. In our analysis such structure emerges from the data, and we are able to recover the commonly recognized major fisheries and their seasonality, along with more spatially and temporally restricted fisheries. These methods have the additional benefit of only requiring the catch composition of trips, making it possible to integrate data from both state and federal management databases which lack consistent permitting data.

A metier is defined as a gear-species target combination (Deporte et al. 2012). We first define species targets and then assign these targets to gear to make the final metier designation. To find species target data we classify target species assemblages by first subsetting to all 2010 trips and searching for characteristic catch compositions. To find these assemblages we first split trips by gear type (using PacFIN grgroups designation) and calculate a pairwise dis-similiarity index for each trip within a gear/year subset using the Bray-Curtis dissimilarity index. This metric has the advantage of avoiding the double-0 problem common in species count data. The Bray-Curtis dissimilarity index is defined as

where is the biomass of the lesser value for only those two species in common between both sites. and are the total number of individuals counted in both trips. This index ranges between zero and one, with zero meaning the sites have the same composition and one meaning they share no species.

We transform the dissimilarity index to be a measure of similarity

and build an undirected, weighted network in which nodes are trips, and edge widths are the similarity in species composition between trips. This allows a vessel to be represented in multiple nodes if it makes trips that vary substantially in catch composition. With this network we use the infoMap algorithm to find communities (clusters or subgraphs) within the network (Martin Rosvall and Bergstrom 2008; M Rosvall, Axelsson, and Bergstrom 2009).

infoMap is an information theoretic approach, which uses the probability flow of a random walker on a network as a proxy for the information flows in a real system. The objective of infoMap is to compress the description of the probability flow, and in doing so partitions the network into modules. infoMap works by computing the fraction of time a node is visited by a random walker using a deterministically greedy search algorithm. Merges between modules that give the largest decrease in description length are made until further merging leads to increases of description length. Results are refined with a simulated annealing approach, starting at several different temperatures, with the run selected as the one that gives the shortest description of the network. We found that other commonly used clustering algorithms (i.e. k-means, hierarchical clustering) did poorly with this data. Many clustering algorithms do best when clusters are spherical in n-dimensional space, and/or require the number of clusters decided a priori. In this data we have fisheries participation may vary by several orders of magnitude (100s of trips to 100,000s of trips), and we wanted to avoid having to decide subjectively on the number of clusters.

After dropping any modules that have fewer than five trips, we use a k-nearest-neighbor (knn) classifier to assign all other trips of each gear subset to those possible metiers. The nearest neighbor to each trip was found using the Bray-Curtis dissimilarity index (transformed into a similarity).

## 3.2 Seasonal Fisheries

**FILL IN**

# References

Clarke, K. R., and R. M. Warwick. 1998. A taxonomic distinctnes insdex and its statistical properties. Journal of Applied Ecology 35:523–531.

Deporte, N., C. Ulrich, S. Mahevas, S. Demaneche, and F. Bastardie. 2012. Regional metier definition: a comparative investigation of statistical methods using a workflow applied to international otter trawl fisheries in the North Sea. ICES Journal of Marine Science 69:331–342.

Kahle, D., and H. Wickham. 2013. Ggmap: Spatial visualization with ggplot2. The R Journal 5:144–161.

Kasperski, S., and D. S. Holland. 2013. Income diversification and risk for fishermen. Proceedings of the National Academy of Sciences.

Keller, A. A., T. L. Wick, E. L. Fruh, K. L. Bosley, D. J. Kamikawa, W. J. J, and B. H. Horness. 2005. NOAA Technical Memorandum NMFS-NWFSC-70, The 2000 U.S. West Coast Upper Continental Slope Trawl Survey of Groundfish Resources off Washington, Oregon, and California: Estimates of Distribution, Abundance, and Length Composition 1–179.

Lucey, S. M., and M. J. Fogarty. 2013. Operational fisheries in New England: Linking current fishing patterns to proposed ecological production units. Fisheries Research.

NMFS. 2013. *Groundfish essential fish habitat synthesis: A report to the pacific fishery management council*. NOAA NMFS Northwest Fisheries Science Center, Seattle, WA.

Norman, K., J. Sepez, H. Lazarus, N. Milne, C. Package, S. Russell, K. Grant, et al. 2007. NOAA Technical Memorandum NMFS-NWFSC-85. Community Profiles for West Coast and North Pacific Fisheries, Washington, Oregon, California, and other U.S. States 1–617.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O’Hara, G. L. Simpson, et al. 2015. Vegan: Community ecology package.

PacFIN. 2014. Pacific Fisheries Information Network retrieval dated 2014-07-09. Pacific States Marine Fisheries Commission, Portland, Oregon (www.psmfc.org).

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sethi, S. A., M. Reimer, and G. Knapp. 2014. Alaskan fishing community revenues and the stabilizing role of fishing portfolios. Marine Policy.

Wessel, P., and W. H. Smith. 2013. A global, self-consistent, hierarchical, high-resolution shoreline database. Journal of Geophysical Research 101:8741–8743.

1. Personal communication from Brad Stenberg [↑](#footnote-ref-1)
2. Am going to try the clustering without this, I think this was based on trying to do k-means and is no longer necessary. Also it's an awkward thing to have to defend. And then when we're looking at species diversity in landings data, we want to include those species again. So it seems overly unclear. [↑](#footnote-ref-2)
3. or average - check if correlated [↑](#footnote-ref-3)
4. From Oksanen et al. (2015): *These can be any distance structure among species, but usually it is found from established hierarchic taxonomy. Typical coding is that differences among species in the same genus is 1, among the same family it is 2 etc. However, the taxonomic differences are scaled to maximum 100 for easier comparison between different data sets and taxonomies. Alternatively, it is possible to scale steps between taxonomic level proportional to the reduction in the number of categories (Clarke and Warwick, 1999): if almost all genera have only one species, it does not make a great difference if two individuals belong to a different species or to a different genus* [↑](#footnote-ref-4)
5. West coast annual? Triennial? Fall + Spring both? [↑](#footnote-ref-5)
6. Not completed yet, fingers crossed! [↑](#footnote-ref-6)