Using data on where vessels went, when they fished, and what they caught, we developed a set of algorithms to 1) identify fisheries from landings data, 2) identify fishing trips based on spatial trajectories and match them to fisheries defined in 1) and 3) identify fishing events. In the following, we describe the datasets, and each step of identifying fishing behavior.

**Description of data sources**

*Catch data*

We collected vessel landings tickets for all commercial landings on the US west-coast between 2009-2013 from the Pacific Fisheries Information Network (PacFIN) database ([www.psmfc.org](http://www.psmfc.org/)). Landing tickets identify vessels, species and amount caught for each landing of commercially caught fish.

*Locational data*

We used vessel monitoring system (VMS) data collected by the Office of Law Enforcement. VMS are GPS systems required for all vessels commercially fishing federally managed groundfish in the past 5 years. Vessels locations are recorded approximately hourly, with an accuracy of approximately 500 meters.

**Matching algorithms**

*Identifying fisheries from landings data*

Fisheries are defined as harvest assemblages caught with a specific gear (van Putten et al. 2012; Boonstra and Hentati Sundberg 2014). The Pacific Fisheries Management Council (PFMC) has developed a set of sector-based definitions similar to this approach for the federally managed groundfish landings (www.pcouncil.org), but no equivalent exists for non-groundfish fisheries (Northwest Fisheries Science Center 2015). In order to treat the landings dataset uniformly, we applied a métier analysis to this landing data (Deporte et al. 2012) to build a set of realized fisheries. A métier analysis identifies realized fisheries by clustering the species composition of landings. This methodology requires choices in the way similarity among trips are measured, a clustering algorithm for grouping similar trips together, and a constraint that the methods can scale across hundreds of thousands of landings.

For our distance metric we used the Hellinger distance *D* (P. Legendre and Legendre 2012) to calculate the similarity in revenue profiles between trips and generated a pairwise distance matrix. This distance metric has the benefit that it is asymmetric, where the presence of a species in both trips is considered more informative than the absence of a species. The Hellinger distance between the species composition of two fishing trips *A* and *B* is defined as

|  |  |
| --- | --- |
|  | (1) |

where *ai* is the fraction of revenue derived from species *i* on trip *A*, *bi* is the fraction of revenue derived from species *i* on trip *B*, and *S* indicates the total number of species collected in both trips. With this metric, trips *A* and *B* become increasingly similar (and the Hellinger distance declines) as the proportion of revenue attributable to each of the *S* species becomes increasingly matched.

We identified realized fisheries as groups of trips with similar target assemblages using the infoMap community detection algorithm (Rosvall and Bergstrom 2008). This algorithm examines networks for subgraphs more interconnected to one another than the network in which it is embedded. To generate the required network we transformed the distance matrix into a similarity matrix by subtracting the distance metric’s upper limit (i.e. ) from each pairwise distance. The result is a weighted, undirected network where trips are connected by edges proportional to their similarity. However, because our dataset contained 340,466 unique trips, we were not able to perform clustering using a single matrix containing all pairwise similarities. To obtain manageable matrix sizes we used one year of landings (2010) which we split by gear. Pairwise distances among trips and community detection were calculated within each gear partition, which grouped trips into target assemblage categories. To classify the 2009, 2012 and 2013 trips to fisheries, we assigned each unclassified trip to the same realized fishery as the 2010 trip to which it was closest in multi-dimensional space using a k-nearest neighbors algorithm.

*Identifying fishing trips and matching them to fisheries*

VMS data lacks any markers for when the vessels depart for commercial fishing trips. To define a fishing trip, we first filtered the data to remove any successive on-land points by overlaying the relocations with NOAA’s GSHHG high-resolution geography dataset.[[1]](#footnote-1) Using this coastline, we calculated minimum pairwise-distances between each relocation and the coastline, and labeled any segments that were at least 3 hours in duration and more than 1.5 km from the coastline as a potential trip. These potential trips were matched with the landings tickets using the recorded landing date on the landing ticket.

1. Wessel, P., and W. H. F. Smith, A Global Self-consistent, Hierarchical, High-resolution Shoreline Database, *J. Geophys. Res., 101*, 8741-8743, 1996 [[PDF](http://www.soest.hawaii.edu/pwessel/gshhg/Wessel+Smith_1996_JGR.pdf)]. [↑](#footnote-ref-1)