https://doi.org/10.25923/7ag1-af45

NOAA Technical Memorandum NMFS

JANUARY 2020

MATCHING VESSEL MONITORING SYSTEM DATA TO TRAWL LOGBOOK AND FISH TICKET DATA FOR THE PACIFIC GROUNDFISH FISHERY

Aaron Mamula[1], Alice Thomas-Smyth[1,2], Cameron Speir[1], Rosemary Kosaka[1], and Don Pearson[1]

[1] NOAA Fisheries, SWFSC Fisheries Ecology Division
110 McAllister Way, Santa Cruz, CA 95060

[2] University of California Santa Cruz, Cooperative Institute for Marine Ecosystems and Climate (CIMEC)

NOAA-TM-NMFS-SWFSC-623

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Marine Fisheries Service
Southwest Fisheries Science Center

About the NOAA Technical Memorandum series

The National Oceanic and Atmospheric Administration (NOAA), organized in 1970, has evolved into an agency which establishes national policies and manages and conserves our oceanic, coastal, and atmospheric resources. An organizational element within NOAA, the Office of Fisheries is responsible for fisheries policy and the direction of the National Marine Fisheries Service (NMFS).

In addition to its formal publications, the NMFS uses the NOAA Technical Memorandum series to issue informal scientific and technical publications when complete formal review and editorial processing are not appropriate or feasible. Documents within this series, however, reflect sound professional work and may be referenced in the formal scientific and technical literature.

SWFSC Technical Memorandums are available online at the following websites:

SWFSC: https://swfsc.noaa.gov

NOAA Repository: https://repository.library.noaa.gov/

NTIS National Technical Reports Library: https://ntrl.ntis.gov/NTRL/

Accessibility information

NOAA Fisheries Southwest Fisheries Science Center (SWFSC) is committed to making our publications and supporting electronic documents accessible to individuals of all abilities. The complexity of some of SWFSC's publications, information, data, and products may make access difficult for some. If you encounter material in this document that you cannot access or use, please contact us so that we may assist you.
Phone: 858-546-7000

Recommended citation

Mamula, Aaron, Alice Thomas-Smyth, Cameron Speir, Rosemary Kosaka, and Don Pearson. 2020. Matching Vessel Monitoring System data to trawl logbook and fish ticket data for the Pacific groundfish fishery. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-SWFSC-623. 76 p.
https://doi.org/10.25923/7ag1-af45

Matching Vessel Monitoring System data to trawl logbook and fish ticket data for the Pacific groundfish fishery.

Aaron Mamula, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Santa Cruz, CA, USA.
Alice Thomas-Smyth[1], University of California, Santa Cruz, Cooperative Institute for Marine Ecosystems and Climate.
Cameron Speir, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Santa Cruz, CA, USA.
Rosemary Kosaka, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Santa Cruz, CA, USA.
Don Pearson, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Santa Cruz, CA, USA.

Abstract

High-resolution spatial data on fishing effort and catch is an increasingly important source of information for fisheries scientists and fisheries managers. In this report we detail how high-resolution spatial-temporal data from Vessel Monitoring Systems (VMS) can be matched with existing data on West Coast commercial fishing effort and catch in order to create rich data products. The primary purpose of the report is to provide descriptive summaries of West Coast VMS data and to relate these data to existing sources of West Coast commercial fishing data. A secondary objective of the report is to illustrate how VMS and complimentary commercial fishing data from groundfish trawl logbooks and fish tickets may be used to evaluate a range of research questions, including: (i) *how well do fishing locations reported on West Coast groundfish trawl logbooks agree with spatial records from VMS?, (ii) Do differences in spatial agreement between logbooks and VMS records vary systematically over time or across regions?, (iii) How are differences in spatial agreement between logbook and VMS fishing locations affected by modeling choices?,* and *(iv) How well can vessel speed discriminate between fishing and non-fishing VMS polls for West Coast groundfish vessels.* We found that the median distance between logbook fishing locations and VMS fishing locations ranged from 0.7 to 2 km depending on the method used to define logbook fishing locations. Differences in the method for constructing tow paths affected the degree to which logbook and VMS data agreed, with straight-line tow paths generating greater spatial agreement with VMS polls than bathymetry-influenced tow paths. The analysis found differences in agreement between regions with potentially different fishing strategies or bathymetric complexity. We found little difference in spatial agreement across years.

[1] Alice Thomas-Smyth contributed to this project while working at the Southwest Fisheries Science Center in Santa Cruz, CA. She currently works for the Environmental Defense Fund in San Francisco, CA.

Contents

Tables and Figures

1 Introduction

The increasing availability of high-resolution data from Vessel Monitoring Systems (VMS) has the potential to improve fisheries management. However, to realize the full potential of these data, fisheries managers and scientists generally must join VMS data with other sources of fishery dependent data. This report demonstrates how traditionally important West Coast commercial fishing data from groundfish trawl logbooks and fish tickets can be joined with VMS data to create rich data products. The primary objective of this report is to characterize key features of West Coast VMS data relative to existing West Coast commercial fishing data sources. A secondary objective is to illustrate a range of fisheries research topics that could be pursued using VMS data.

Relative to these two objectives, our analysis proceeds as follows. First we propose methods for quantifying how well spatial locations of fishing activity derived from VMS polls agree with fishing locations obtained from self-reported trawl logbooks. We then summarize this ‘spatial agreement’ along a number of different margins including over time and across regions. We also relate our measures of spatial agreement to observable behavioral characteristics such as fishery participation. We end our analysis by demonstrating how VMS data joined with logbook and fish ticket data can be used to refine understanding of the spatial distribution of fishing effort. This is done through predictive modeling that classifies VMS polls according to whether the poll is associated with fishing activity or not.

Fisheries management is increasingly being conducted at finer scales of spatial resolution. Understanding the distribution of fishing effort and catch is important for performing accurate stock assessments and in understanding the effects of spatial policies on fish stocks and fishermen. The effectiveness of such policies is contingent on the availability and quality of spatial data. Logbook data maintained by vessel captains during fishing operations have been a traditionally important source of information for fisheries scientists. For example, in our study area of the U.S. West Coast, spatially explicit data on fishing effort and retained catch from logbooks maintained by vessels in the limited entry groundfish trawl fishery have been extensively utilized by fisheries biologists and fisheries managers to: estimate spatially refined bycatch rates of Pacific halibut in the multi-species groundfish fishery (Pikitch et al. 1998), examine factors such as latitude gradient and depth affecting species mixing rates (Lee and Sampson 2000), and analyze the spatial distribution of trawls to assess the impacts of mobile fishing gear on benthic habitats (Bellman et al. 2005). In addition to these important biological and ecological uses, social scientists have used these data to develop a better understanding of how spatial distributions of fishing effort are affected by policy changes such as marine reserves (Valcic 2009; Mason et al. 2012), and how effort shifts can impact coastal economies (Speir et al. 2014). Economists have also used Pacific groundfish trawl logbooks to assess the economic performance of fishing firms (Collier et al. 2014).

The recent availability of high frequency position data collected by VMS has afforded researchers even greater opportunity to explore the spatial distribution of fishing effort and harvest. However, since VMS data typically do not include detailed information on catch or vessel characteristics, these data generally must be integrated with other sources of fishery information (e.g. logbooks or observer records) to address meaningful research questions.

Linking remotely-sensed VMS data with logbook data can be difficult for several reasons. VMS and logbook data are collected at different temporal scales, with VMS positions reported many times per day, regardless of whether fishing is occurring, and logbook entries completed each time gear is deployed or retrieved (Gerritsen and Lordan 2011). Also, each data set is subject to measurement error, due to data entry errors or malfunctioning equipment. As a result, assigning catch amounts or other information from logbooks or remotely-sensed positional point data can be challenging.

Agreement between positions recorded in trawl logbooks and VMS data can provide an indicator of the accuracy and precision of locational data used in analysis. Quantifying and summarizing this agreement is important for at least two reasons. First, the required precision of locational data depends on the scale of the analysis (Jennings and Lee 2012). For example, assessing the effect of bottom trawling on sensitive habitat may require a high degree of spatial precision (Demestre 2015, Bellman et al. 2005), while estimating spatial differences in catch per unit effort (CPUE) may require only that positions be recorded within the same larger statistical areas (Palmer and Wigley 2009). Since fisheries researchers are likely to use VMS and logbook data for a range of empirical applications, and each application will place unique demands on the data, it is important to provide information on the spatial properties of these data that researchers can use in determining how best to structure their VMS analysis.

Second, in many cases, logbooks provide a long historical record of the spatial distribution of fishing effort. The self-reported nature of these data, which introduces the possibility of intentional and unintentional reporting error, may raise some questions about the reliability of older logbook data. Quantifying agreement between recorded logbook positions and electronic monitoring data from more recent time periods may help develop a sense of the spatial accuracy and precision of historical records. Our analysis, therefore, matches data by fishing locations over the course of reported fishing events (e.g., trawl tows) and reports agreement in terms of absolute distance. This approach is consistent with previous studies that measured the distance between matched records (e.g. Skaar et al. 2011). Other previous studies examined positional agreement between logbooks and VMS at much coarser spatial scales or summed VMS point data to grid areas (Gerritsen et al. 2013; Palmer and Wigley 2009; Lee et al. 2010).

The first part of our analysis addresses the question of how closely spatial information in logbook and VMS data sets agree. We compare the degree of agreement of matched positions from the two data sets and describe the distribution of the distances between VMS polls and matched trawl lines[2]. We also systematically examine potential reasons why differences in position exist, including sensitivity to the method used to interpolate tow paths in the logbook data, differences in data characteristics between years or sub-regions, and sensitivity to choice of criteria for classifying VMS data points as fishing activity. To do this, we compare point-to-line agreement using logbook tow paths derived using bathymetric contours to tow paths constructed by drawing straight lines between recorded tow set and tow retrieval coordinates. The use of bathymetry to estimate detailed tow paths, to our knowledge, has not previously been applied in the literature. Our results show that, for this application, bathymetry tow paths provided an inferior fit

[2] The VMS program requires vessel captains to install data loggers on their vessels. This is an important regulatory detail that highlights a caveat of our analysis. Since our data include only vessels with active VMS loggers, we assume vessel captains know they are being monitored. Therefore, we cannot explore whether self-reported fishing locations are more accurate for monitored versus unmonitored vessels.

to VMS polls relative to straight line tow paths. We see trivial differences in agreement across years, but observe differences in agreement between sub-regions in our data set. These regional differences may be due to differences in fishing strategies.

The second part of our analysis explores how well fishing activity can be predicted from VMS data. Recently, researchers have developed an interest in identifying fishing activity from positional data (Watson and Haynie, 2016; Bez et al., 2011; Joo et al., 2011). Because VMS data loggers are relatively cheap and do not require input from vessel captains, VMS has the potential to monitor fishing effort in a way that is more cost effective and less burdensome than traditional data collection methods.

The remainder of our report is organized as follows. Section 2 describes the data sources used in the analysis. Section 3 details the methods used. This includes methods for joining VMS data with other fishery dependent data, methods for interpolating fishing paths given only starting and ending locations, methods used to evaluate distance between the two sources of spatial data collected at different temporal scales, and methods used to infer fishing versus non-fishing behavior from positional and other physical data. Section 4 summarizes the results and Section 5 provides a discussion of key results and their implications.

2 Data

2.1 Description of the data: groundfish trawl logbooks

Groundfish trawl logbook data (the logbooks) contain self-reported fishing information for the Pacific Coast commercial groundfish fishery. Logbook data used in this analysis is limited to vessels using departure ports in California. The unit of observation in the logbook data is a fishing event. For almost all fishing activity captured by the trawl logbooks, a fishing event[3] is a single tow of trawl gear. Each record contains a unique record identifier, vessel identification number, coordinates for the start and end points of each tow, times for the start and end of each tow, and basic catch composition as estimated by the vessel captain. The logbook data were provided by the Pacific States Marine Fisheries Commission's Pacific Fisheries Information Network (PacFIN), which curates trawl logbook data collected by state partners[4]. We use logbook data from 2008 to 2009 and from 2014 to 2015 in our analysis.

Before matching logbook observations to VMS polls, we imposed several ad-hoc filters on the logbook data. We excluded records that had an incomplete spatial record, i.e., do not contain both starting and ending tow coordinates and records that had start or end coordinates that occur on land. We retained tows

[3] To simplify the discussion, the remainder of our report will use the terminology "tow" to refer to logbook fishing activity. The scholarly literature on West Coast groundfish has used "trawl", "haul", and "tow" interchangeably to describe the operation of trawl fishing gear in commercial groundfish fisheries. We chose the term "tow" in order to establish consistency with field labels in the underlying data sources.

[4] Primary logbook data are collected by state agencies (California Department of Fish and Wildlife, Oregon Department of Fish and Wildlife, and Washington Department of Fish and Wildlife) and provided to PacFIN. PacFIN organize these data and provide access through their centralized database.

that contain identical coordinates for the start and end points. In these cases, we altered the latitude of the end point by 0.000001 decimal degrees.

As a practical matter, "logbook data" is a generic term that includes three specific data sources leveraged in our analysis. For our analysis it is sufficient for the reader to understand that logbook data contain three main types of features: i) characteristics specific to a fishing trip (such as departure and return port), ii) characteristics specific to a tow (such as set and retrieval locations of the gear), and iii) characteristics specific to a particular species or market grade. These three types of features are contained in three distinct database tables which can be joined together to produce a complete accounting of the pounds of each distinct species or market grade caught on each tow of each fishing trip.

As a matter of nomenclature, throughout the remainder of this report, we will refer to data originating from groundfish trawl logbooks as "logbook data." The reader should understand that this term refers specifically to information contained in PacFIN's Coastwide Trawl Logbook Subsystem. Metadata and a detailed description of this data source can be found here: https://pacfin.psmfc.org/data/trawl-logbooks/.

2.2 Description of the data: fish ticket data

Our analysis also makes use of PacFIN's Fish Ticket Reports data[5]. Fish tickets are generated when commercial landings occur. They track the weight, condition, and price paid for each fish landed. In addition to information about the specific landing (how many pounds of each market category[6] that were landed, what type of gear was used to harvest the fish, etc.), PacFIN's fish ticket data contain information on the vessel (length of the vessel, weight of the vessel, ownership information) and fishing trip (port of landing, date of landing) associated with each landing. In our analysis, fish tickets are used primarily to assign a gear type to each tow in the logbook data. While the logbooks contain a field for gear type, this field is not well documented and often is not precisely filled-in. Fish tickets can be linked with logbooks using the fish ticket identifier.

2.3 Description of the data: VMS

The VMS data are derived from positional data transmitted from units on each fishing vessel to enforcement agencies via satellite. The primary purpose of this system is to enforce closed area restrictions in the fishery. The unit of observations in these data are polls, which are reports showing the position, bearing, and speed of the vessels at particular points in time. Each poll includes a time-stamp

[5] These data are also sometimes referred to as landing receipts. For consistency, our report will use the term "fish tickets" or "fish ticket data" throughout. The reader should understand that this term references data contained in a specific database table maintained by PacFIN. This database table is extensively documented on PSFMC's PacFIN website here: https://pacfin.psmfc.org/data/documentation-2/.

[6] The Pacific Coast groundfish fishery includes over 90 distinct species. Some of these species are commonly landed together and treated by dockside buyers as homogenous aggregates. For this reason fish tickets organize landings by market category rather than by distinct species. Some market categories map to a single species while others denote a bundle of similar species.

indicating the day and time the positional record was made. VMS polling occurs at different intervals depending on the fishery. For our fishery of interest the VMS ping rate is set to record vessel positions each hour[7]. The VMS program began on a limited basis in 2007. All vessels in the groundfish fishery were required to operate VMS units beginning in 2008. We obtained these data from the National Oceanic and Atmospheric Administration (NOAA) Office of Law Enforcement.

Our analysis uses 4 years of VMS data including observations from 2008 – 2009 and observations from 2014 – 2015. Because the VMS data are quite large and difficult to process efficiently, we choose to work with a subset of the available years. The West Coast groundfish fishery underwent a significant regulatory change in 2011 with the introduction of ITQ management. Our analysis utilizes two years of VMS data preceding this change and two years of VMS data from the post policy regime.

3 Methods

In this section we first describe how VMS, logbook, and fish ticket data are joined. Then we discuss the methods used to evaluate spatial agreement and infer fishing versus non-fishing behavior from VMS data. Appendix Table A 1 provides a summary of the specific database tables that are used to link VMS, logbook, and fish ticket data.

3.1 Joining data Sets

3.1.1 Matching VMS polls to logbook tows

The primary task described here is assigning a VMS poll to a unique logbook fishing trip and, wherever possible, a unique logbook tow based on the time of the VMS poll. This operation is relatively straightforward but, because the two data sources have different temporal scales, deserves some discussion. Each tow reported on the logbooks can be linked to a unique vessel identifier, fishing trip identifier, and tow identifier. For the purposes of our analysis, a VMS poll may exist in one of three states:

1. It may be linked to a specific logbook tow
2. It may be linked to a specific logbook fishing trip but not to a specific tow
3. It may be linked to a fishing vessel from the logbooks but not to a particular logbook fishing trip.

Regarding possibility #1: consider a tow i carried out on fishing trip j by vessel k. Tow i has a starting time, defined by the time at which the gear was set (t_i^{set}), and an end time, defined by the time the gear

[7] See 50 CFR §660.14 (item 3).

was retrieved (t_i^{up})[8]. Finding the VMS polls corresponding to each logbook tow is a straightforward matter of filtering the VMS data for all polls from vessel k time-stamped between t_i^{set} and t_i^{up}.

Regarding possibility #2: logbook data provide a starting date and ending date for each fishing trip. VMS polls may occur between a fishing trip's starting and ending dates but may not occur between the starting and ending times for any particular tow. These polls may be associated with behaviors such as transiting (from port to fishing grounds or between fishing grounds) or sorting catch.

Regarding possibility #3: A VMS poll may be linked to a vessel present in the logbook data but occur at a time not matching any groundfish trips reported in the logbooks for that vessel. The average vessel in our logbook data sample has less than 60 fishing days per year. Since VMS data are polled every hour of every day, the vast majority of VMS polls are associated with dates and times during which the vessel was not only not actively fishing for groundfish but potentially not at sea at all[9].

In Appendix Figure A 1, we illustrate how VMS polls are matched to logbook tows using a common vessel identifier, starting and ending date-times of each logbook tow, and date-time stamps of VMS polls.

3.1.2 Matching VMS polls to fish ticket data

As discussed in Section 2, fish tickets contain important information on commercial fish landings. Individual fish tickets can be mapped to specific tows from the logbooks using a look-up table provided by PacFIN. This look-up table maps each fish ticket identifier to a logbook trip identifier and a logbook tow identifier.

Our analysis extracts information on fishing gear utilized from the fish ticket data and joins this with the VMS data. Although logbooks also contain self-reported information on fishing gear used, we choose to extract this information from the fish tickets because fish ticket gear information is easier to interpret and is generally more complete. It is also worth noting that fish tickets provide a full accounting of all market categories landed and prices paid for those species from a fishing trip. So, while our analysis relies on fish ticket data for gear information, fish ticket data could also be used to join market category specific landings and gross revenue with a collection of VMS polls defining a fishing trip.

[8] These times are reported in the logbooks as "SET_TIME" and "UP_TIME". To be consistent with this nomenclature, we use the terms "set" and "up."

[9] An important caveat here is that if a groundfish vessel was fishing in a non-groundfish fishery (such as Dungeness crab) our analysis would have no way of linking VMS polls to this activity since this activity would not be recorded in the logbooks. An important implication of this fact is that VMS polls matching a logbook vessel but not matching any logbook fishing trip may be associated either with times that this vessel was not fishing at all or times when the vessel was fishing in a fishery not monitored by the logbooks.

In the previous section we discussed how individual VMS polls are assigned to logbook trips and logbook tows. Adding gear information from fish tickets is a relatively straightforward matter of i) fortifying the logbook data with gear information from fish tickets and ii) joining the gear fortified logbook data back with the joined VMS-logbook data. First, fish ticket data (including type of gear used) are joined with trip identifiers and tow identifiers using a look up table. Vessel identifiers are then added using the table containing logbook tow characteristics. We then simplify the data with a final filtering step that removes any logbook fishing trips that use more than one type of gear. We also discard any fish tickets that could not be matched to a specific logbook fishing trip.

Finally, gear information can be added to each VMS poll by joining the gear-fortified logbook tow data to the merged VMS-logbook data using the common fields: vessel id, trip id, tow number. The process is illustrated in Appendix Figure A 2.

To summarize: gear type is assigned to each fish ticket identifier in the fish ticket data. The fish ticket look up table (LBK_FTID from Appendix Table A 1) assigns each fish ticket to a logbook trip identifier and tow number. Using these data sources each logbook trip and tow can be matched to fish ticket identifier and, with these fish ticket identifiers, gear type can be assigned to each logbook trip and tow.

3.2 Evaluating spatial agreement between VMS polls and logbook locations (tows)

The general strategy we use for evaluating the spatial agreement between VMS locations and logbook locations can be summarized as executing the following discrete tasks:

1. Constructing tow paths from set and up positions reported on the logbooks, and
2. Defining points along this path
3. Evaluating the distance in kilometers between VMS polls and logbook fishing points.

These tasks are described in detail below.

3.2.1 Constructing tow paths from logbook set and retrieval positions

For VMS polls that could be matched to a specific logbook tow we evaluate the distance between VMS polls and corresponding logbook fishing locations. The first step in this process is creating a line (tow path) that represents each logbook tow (defined in our data by a starting point and ending point). For a VMS poll at time *t* we generally do not observe a corresponding logbook vessel location precisely at time *t*, so we match the VMS poll to a time interval when gear is reported to have been in the water. Consider:

x_{it} – VMS poll assigned to logbook tow i which occurs at time t

t_i^{set} – the reported set time for logbook tow i

t_i^{up}- the reported retrieval time for logbook tow i

$l_i^{set} = (longitude_i^{set}, latitude_i^{set})$, the latitude/longitude coordinates for the set point for tow i.

$l_i^{up} = (longitude_i^{up}, latitude_i^{up})$, the latitude/longitude coordinates for the retrieval point for tow i.

In general, we observe $t_i^{set} < t < t_i^{up}$. We do not observe the vessel's exact position at time t, we only know precisely where the logbook data place the vessel at t_i^{set} and t_i^{up}. Our analysis compares VMS locations to logbook locations by approximating the vessel's path between l_i^{set} and l_i^{up} and evaluating the distance from this path to the VMS poll x_{it}.

Tow paths between l_i^{set} and l_i^{up} are approximated using two methods: i) a simple straight line path between the two points and ii) a method which constructs a path by minimizing deviations from average bottom depth between the starting and ending points of the tow.

3.2.1.1 Bathymetry derived tow paths

The bathymetry tow paths[10] are constructed using a constrained minimization algorithm. Tow paths follow a least-cost path that is created by minimizing the distance the vessel travels subject to the constraint that it travels along a path with the least change in bathymetry between the origin and destination. We used the Minimum Cost Path class from the scikit-image package in Python to construct the least-cost path (Van der Walt et al. 2014) with bathymetry from the California Department of Fish and Wildlife's Marine Region GIS Unit[11].

3.2.1.2 Straight line tow paths

Straight line tow paths are constructed from tow starting and ending points using R's Simple Features Package (Pebesma, 2018). The function *st_cast()* is used to transform two objects (a starting point and ending point) of geometry type *POINT* into a single *LINESTRING* geometry.

[10] To simplify the tabular presentation of data in Section 4 we will refer to this bathymetry-based tow path method as the "bathy" method.
[11] The bathymetry data are available for download from: http://www.dfg.ca.gov/marine/gis/downloads.asp

3.2.2 Discretizing the tow paths

To assess the distance of the VMS poll to its corresponding tow path, we transform each tow path into a set of uniformly spaced points. Consider the following:

- Tow i is defined by starting and ending coordinates, l_i^{set} and l_i^{up}
- Let lbk_i be a line constructed by connecting the points l_i^{set} and l_i^{up}.
- Let $lbk_{pts_i} = [lbk_i^1, lbk_i^2, \dots, lbk_i^n]$ be a collection of points spaced evenly along the line lbk_i.
- x_i is a VMS poll assigned to tow i using the methodology in Section 3.1.1.

As a practical matter, we use the function *st_line_sample()* from Pebesma (2018) to define the points lbk_{pts_i}. The function accepts the input n which controls the number of points to define along the line. We discretize each tow path into 2,000 points. The distance calculation increases in accuracy as the parameter n increases; however, we found very minimal changes in the calculated distances for $n > 2{,}000$ in our study[12].

3.2.3 Calculating distance from VMS poll to logbook tow path

Our analysis approximates the VMS poll to logbook tow path distance by evaluating the point-to-point distances between each VMS poll and each point along the discretized tow path to which the VMS poll was assigned. The distance from the VMS poll x_i to the line $lbk_i$is defined to be the minimum of these point-to-point distances,

$$d(x_i) = \min(D^h\left(x_i, lbk_{pts_i}\right))$$

Where $d(x_i)$ is the distance from the VMS poll x_i to tow line lbk_i and D^h is the Haversine distance function defined for two points (two sets of latitude/longitude coordinates).

In Figure 1 we illustrate the process used to evaluate the distance between a VMS poll and its assigned logbook tow for the straight line tow paths. To review, the start and end points of the tow are connected to form a line. The line is converted to a set of uniformly spaced points. The distance from each point on the tow path to a particular VMS poll assigned to that tow is calculated. The smallest of these point-to-point distances is accepted as the distance from the VMS poll to the tow path. Figure 1 illustrates the distance calculation from the VMS poll to straight line tow. The process is identical for the bathymetry-derived tow paths.

[12] Appendix Table B 1 illustrates the impact of this parameter selection on the distance estimates.

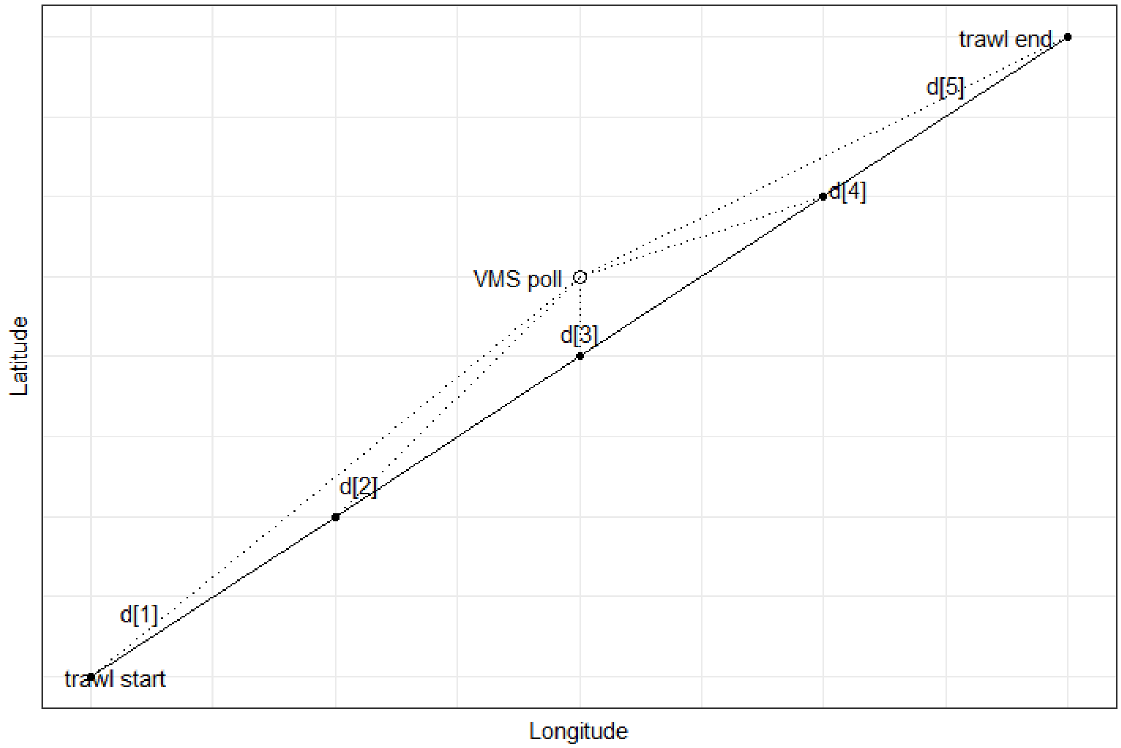


Figure 1. Illustration of relationships between tow paths and VMS polls for a hypothetical logbook tow. Straight line distances from the VMS poll to each point defined along the logbook trawl line are shown as dotted lines and annotated with the labels d[1] – d[5].

3.3 Evaluating feasibility of joined VMS-logbook data

In this section we propose a second approach to evaluate the overarching question of how well satellite tracked positions (from VMS) agree with self-reported positions (from logbooks). The approach here is as follows: we join logbook and VMS points belonging to the same tow and ask if the resulting fishing path is physically feasible given existing knowledge of fishing behavior.

1. Define a new approximate tow track as the points $track_i = [l_i^{set}, x_i^1, x_i^2, \ldots, l_i^{up}]$, where l_i^{set} and l_i^{up} are the logbook tow set and tow retrieval coordinates for tow i and $x_i^1, x_i^2, \ldots$ indicate VMS points assigned to tow i and ordered in time.
2. Create line segments connecting each pair of adjacent points.
3. Define the tow distance D_{track_i} as the summed length of the line segments. Let $s_{1,2}$ be the length in kilometers of the line segment connecting the first and second points in $track_i$. and $D_{track_i} = \sum_{k=2}^{n} s_{(k-1)k}$.

4. The average speed required to travel $track_i$ is $\frac{D_{track_i}}{t_i^{up} - t_i^{set}}$. To differentiate this speed from other calculated fishing speeds we will use, we call this s^{track_i}.

We compare this speed with the distribution of speeds calculated using only logbook data. The general approach here is to calculate fishing speed based on joined VMS-logbook data and compare these speeds to a reference distribution of fishing speeds based only on logbook data.

The intuition here is that if the track constructed by joining logbook and VMS data results in tow speeds far in excess of other observed tow speeds, then the fishing profile create by the joined data is unlikely to be an accurate depiction of tow i.

We propose here two distinct ways of constructing the reference distribution. First, we use all logbook observations and create a distribution of fishing speeds for each gear type. For clarity let:

- $s_{i,g}^{bathy}$ be fishing speed for tow *i* using gear *g* calculated using the bathymetry-defined logbook tow path
- $s_{i,g}^{straight}$ be fishing speed for tow *i* using gear *g* calculated using the straight line logbook tow path
- S_g^{bathy} be the distribution of all fishing speeds for tows using gear type *g* and calculated using the bathymetry-defined tow paths
- $S_g^{straight}$ be the distribution of all fishing speeds for tows using gear type *g* and calculated using the straight line tow paths

The method outlined above for creating a reference distribution of fishing speeds relies heavily on the 'gear type' field obtained from fish ticket data. In previous work, groundfish stock assessment scientists and fisheries managers have expressed concern about the accuracy and reliability of gear type information derived from California's historical fish ticket data (Pearson et al. 2008). To address this concern, we propose a second method for comparing fishing speeds from joined VMS-Logbook data to a reference distribution. This second method considers the speed of tow *i* from vessel *j* and compares it against a reference distribution constructed using the speeds of tow *i* and all other tows from vessel *j*. In this case, the reference distributions, defined at the vessel level, are denoted S_j^{bathy} and $S_j^{straight}$.

In both cases, the rules that we use to label a particular observation as infeasible are based on Tukey's rule for non-parametric outlier detection (Tukey, 1977). In the first case where the reference distributions are defined for all tows within a gear strata the rule is,

$$s^{track_i} > S(0.75)_g^{bathy} + 1.5 * [S(0.75)_g^{bathy} - S(0.25)_g^{bathy}]$$

$$s^{track_i} > S(0.75)_g^{straight} + 1.5 * [S(0.75)_g^{straight} - S(0.25)_g^{straight}]$$

In the second case where the reference distribution is defined at the vessel level, the rule is,

$$s^{track_i} > S(0.75)_j^{bathy} + 1.5 * \left[S(0.75)_j^{bathy} - S(0.25)_j^{bathy}\right]$$

$$s^{track_i} > S(0.75)_j^{straight} + 1.5 * \left[S(0.75)_j^{straight} - S(0.25)_j^{straight}\right]$$

For the inequalities above $S(0.75)_g$ and $S(0.25)_g$ indicate the speed values associated with the 75th and 25th percentile of the logbook only tow speed distributions stratified by gear type. And $S(0.75)_j$ and $S(0.25)_j$ are the speed values associated with the 75th and 25th percentiles of the logbook only tow speed distributions defined for each vessel.

3.4 Identification of fishing and non-fishing behaviors from VMS polls

One of the factors motivating interest in VMS data among fisheries scientists is the potential for VMS data to help refine our understanding of the spatial distribution of fishing effort. However, in order for VMS data to be useful in this context, individual VMS polls need to be labeled according to whether the vessel was actively fishing at the time of the poll or engaged in some other behavior (transiting between fishing grounds, sitting idle in port). Since raw VMS data contain only a time stamp and location (i.e., do not provide information on vessel behaviors), labeling individual polls as ‘fishing’ or ‘not fishing’ must either be done by joining auxiliary data sources to VMS data or through inference (presumably on the basis of observable characteristics of the poll). In our analysis, we examine some popular methods of inferring whether a particular VMS poll represents fishing versus non-fishing activity.

Classification algorithms based on vessel speed are simple, fast, and require no additional data since speed can be calculated directly from time stamped VMS polls. These properties have made speed-based inference very popular. Prior contributions to the fisheries literature by Murawski et al. (2005), Mills et al. (2006), Palmer and Wigley (2009), Gerritsen and Lordan (2011), Skaar et al. (2011), Murray et al. (2013), and Demestre et al. (2015) have established ranges of vessels speeds commonly associated with fishing activity for various vessel types. Most notable for our study is a paper by Bellman et al. (2005). The authors conducted interviews with Oregon commercial fishermen and established a range of groundfish bottom trawling speeds of 3.3 to 5.6 km/hr.

There is a large and growing literature on use of regression and machine learning models to infer fishing behavior from VMS data[13]. Joo et al. (2011) use an Artificial Neural Network to classify VMS polls in Peru as either fishing or not fishing. Muench et al. (2018) use a Generalized Linear (Logit) Model to classify individual VMS polls for commercial fishing vessels in the Northeastern United States. Watson and Haynie (2016) use a Generalized Additive Model to attempt to classify trips (collections of VMS polls) as either fishing trips or other trips (transiting between ports for example).

3.4.1 A fishing activity classification experiment

We test the classification accuracy of speed-based methods relative to the classification accuracy of some relatively simple regression models. We use the "hold-out" or cross validation paradigm common in applied statistics and machine learning to compare the classification accuracy of three models: a naïve-speed based classifier, a generalized linear model with a binominal link function (logistic regression) similar to Muench et al. (2018), and a generalized additive model with a binomial link function similar to Watson and Haynie (2016). The approach uses a simple 80/20 rule where 80% of the data are used for training the models and 20% of the data are set aside as 'testing' data. Classification accuracy is evaluated by how well each model classifies observations in the testing data. Details of the classification models are presented below.

Two important caveats accompany our model-based fishing/not fishing classification of VMS polls. First, we include only the three most utilized groundfish targeting gear types in our analysis: small footrope trawl gear, large footrope trawl gear, and selective flatfish gear. Our data include few observations for gear types GFL (otter trawl), MDT (midwater trawl), DNT (Danish seine), and BMT (beam trawl) and, in some cases, only a single vessel is included with the gear strata. For these reasons we have chosen to focus the estimations on the relatively data-rich gear strata. Second, our analysis does not include a full model selection exercise. Most notably we do not test all possible combinations of predictors and interactions. The emphasis of this manuscript is on characterizing the synthesized logbook, fish ticket, and VMS data. We provide the fishing classification analysis as an illustration of a potentially interesting use of these synthesized data. While we have chosen models and predictors to provide consistency with some notable previous VMS work, we don't claim to have found the "best" classification models for identifying fishing behavior from VMS polls.

Naïve speed based classifier

The speed classifier defines a 'fishing window' as speeds between a lower and upper quantile of the observed fishing speed distribution. VMS polls are then classified as fishing or not fishing based on whether the speed falls within the 'fishing window.' Fishing speed windows are calculated separately for each distinct gear type. For an unknown VMS poll, x, in the testing data,

[13] The use of hidden Markov and other predictive models to infer animal behavior from satellite tracking data preceded the literature on inferring behavior from VMS data by at least a decade. An exhaustive review of scholarly contributions in this area is beyond the scope of our analysis but a comprehensive review of methods can be found in Patterson et al. (2008).

$$y = \begin{cases} 1 \; if \; s_g^{lower} \leq speed(x) \leq s_g^{upper} \\ 0 \; otherwise \end{cases},$$

where y is an indicator of the fishing/non-fishing status of the VMS poll and s_g^{lower} and s_g^{upper} are the speeds values represented by quantiles of the training data for gear type g. In practice, we use the 25th and 75th percentiles of the speed distribution to define the parameters s_g^{lower} and s_g^{upper}.

A logit classifier

The logit classifier models the binary response variable (fishing v. not fishing) as a non-linear function. Specifically, the probability that a VMS poll x is fishing is conditional on a set of observed predictors (z),

$$P(y_i = 1|z_i) = \frac{e^{\beta z_i}}{1 + e^{\beta z_i}} = \widehat{p_i}$$

This probability $\hat{p}_i$ is transformed to a binary prediction using the decision rule,

$$\widehat{y_i} = \begin{cases} 1 \; if \; \widehat{p_i} > 0.5 \\ 0 \; otherwise \end{cases}$$

A separate logistic regression was estimated for each gear type. Predictors included in the regression were:

- Speed
- Bottom depth
- Vessel bearing (measured in radians)

Additionally, vessel bearing was binned in 30° increments and included in the logit model as a set of 11 dummy variables.

A GAM classifier

The generalized additive model for binary data replaces the predictor values from the logit equation with smooth functions of those predictors. We define our GAM using the same predictors as in the logit model above. We do not include the interaction between speed and depth in the GAM,

$$logit\{E(y)\} = \beta_1 f_1(speed, df) + \beta_2 f_2(bottom\ depth, df) + \beta_3(bearing, df) \quad y{\sim}binary.$$

In addition to the predictors, the GAM is defined by the functions f_1, f_2, f_3 and the smoothness or 'wiggliness' of these functions which is controlled through the degrees of freedom (df). Our GAM model specification is influenced heavily by Watson and Haynie (2016). We use the R package "mgcv" (Wood, 2017) to estimate the GAM using default thin plate splines with 4 degrees of freedom.

Like the logit model, the GAM produces a predicted probability which we transform to a binary prediction using the same 50% rule as with the logit regression above.

3.4.1.1 Additional data

The predictive models we propose here include two quantities not previously discussed: bearing and bottom depth. Bearing (in degrees) was calculated for each VMS poll using the current and proceeding poll. Bottom depth at each VMS poll was approximated using spatial interpolation. Data for the bottom depth interpolation comes from NOAA's Coastal Relief Model[14]. Bottom depths are measured in meters over a fine grid with a grid step of 0.01666°. Depths for individual latitude/longitude coordinates are approximated using inverse distance weighted interpolation.

4 Results

4.1 Tow path interpolation

Our analysis relies on two methods of interpolation to construct tow paths from a set of reported starting and ending coordinates for each tow. Straight line tow paths are constructed using a straight line to connect the starting and ending coordinates of the tow. It is reasonable to suspect that this method would provide an underestimate of area fished and trawling time, as it is the shortest possible path between the set and up locations and is calculated without consideration of any physical features that may impede a vessel's progress along the line. Bathymetry-derived tow paths are constructed by choosing a path between tow set and up locations that minimizes the change in ocean bottom depth. This method of interpolation has the advantage of conforming to conventional wisdom regarding the nature of trawl fishing[15]. A notable disadvantage of this method is that the rigid adherence to optimization conditions can

[14] The data for California and the Pacific Northwest were obtained from the ERDAPP website. Specific data sets downloaded include dataset IDs USGS CeCrm 7 (https://coastwatch.pfeg.noaa.gov/erddap/griddap/usgsCeCrm8.html) and USGS CeCrm8 (https://coastwatch.pfeg.noaa.gov/erddap/griddap/usgsCeCrm7.html).

[15] It is generally accepted that trawl fishing tends to follow bathymetric contours.

produce tow paths that are visibly nonsensical. Figure 2[16] provides an example of a logbook tow for which the bathymetry-based tow path interpolation method creates an unrealistic tow.

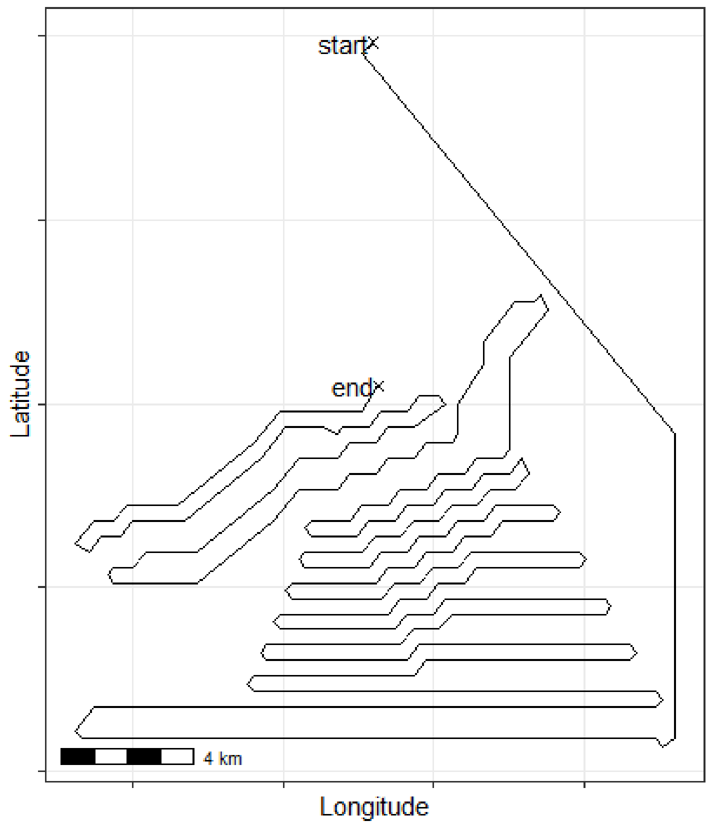


Figure 2. Example of an unrealistic fishing profile created by the bathymetry-based tow path interpolation method.

In this section we provide a simple comparison between the straight line method and bathymetry method used to infer fishing paths from logbook reported tow set and tow retrieval coordinates. Table 1 provides a summary of the difference between tow lengths constructed using the bathymetry-based method versus the straight line method.

[16] Figure 2 obscures the precise latitude and longitude coordinates of the vessel locations in order to protect confidentiality. This convention (leaving vertical and horizontal axes unlabeled) will be applied throughout the manuscript when plotting vessel locations.

Table 1. Quantile values for the empirical distribution of differences in tow lengths between bathymetry constructed and straight line tow paths (in km). Column headers indicate proportions of the data less than each cell value.

0.5	0.75	0.9	0.95	0.99
0.67	7.15	18.43	25.41	38.16

Table 1 is meant to be descriptive and is not meant to provide support for a determination about whether straight lines or bathymetry lines are a better representation of actual fishing paths. Bathymetry tow paths are always longer than straight line tow paths. For 90% of the tows in our logbook sample, the total distances fished for the bathymetry tows paths are less than 18.43 km longer than the total distance fished for the straight line tow paths.

In Figure 3 and Figure 4 we provide two illustrations to add context to the statistics reported in Table 1. Figure 3 shows a tow where the difference in total distance fished between the bathymetry-based tow path and the straight line tow path is 18 kilometers. The VMS polls assigned to this tow suggest that the vessel was not fishing in a straight line but rather appears to have been following the general shape of the bathymetry-based path. Figure 4 shows another tow where the difference in total distance fished between bathymetry-based tow path and the straight line tow path is 18 kilometers. For this tow, the matched VMS polls suggest that the straight line tow path is a better representation of actual area fished than the bathymetry path.

Our analysis of VMS and logbook data utilizes fishing paths inferred from starting and ending coordinates for self-reported tows. Figure 3 and Figure 4 were provided in order to illustrate an interesting feature of our logbook and VMS data that will be discussed further throughout Section 4: bathymetry-based and straight line tow paths can paint notably different picture of fishing activity. Sometimes bathymetry-based tow paths appear to match the general shape of VMS fishing polls and sometimes VMS polls show fishing activity to be carried out a relative straight line. Figure 2 was included in order to emphasize that, in a small number of cases, very large differences between bathymetry-based tow lengths and straight line tow lengths can arise. This is often due to the greedy nature of the bathymetry-based interpolation algorithm forcing the vessel to make unrealistic movements in search of a path that will minimize bathymetric change.

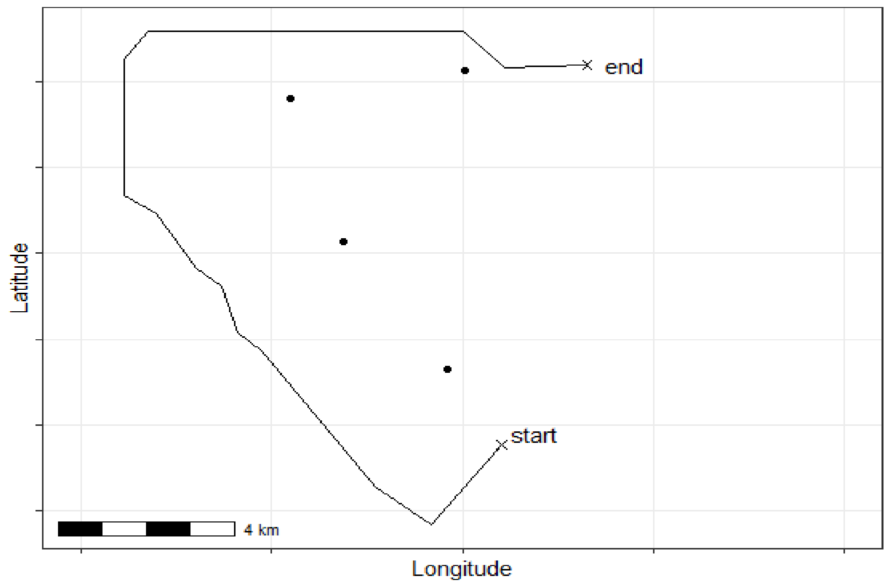


Figure 3. Bathymetry-based tow paths for an observation where VMS polls do not form a straight line. Starting and ending points for the tow are annotated and filled circles indicated VMS polls matched to the tow.

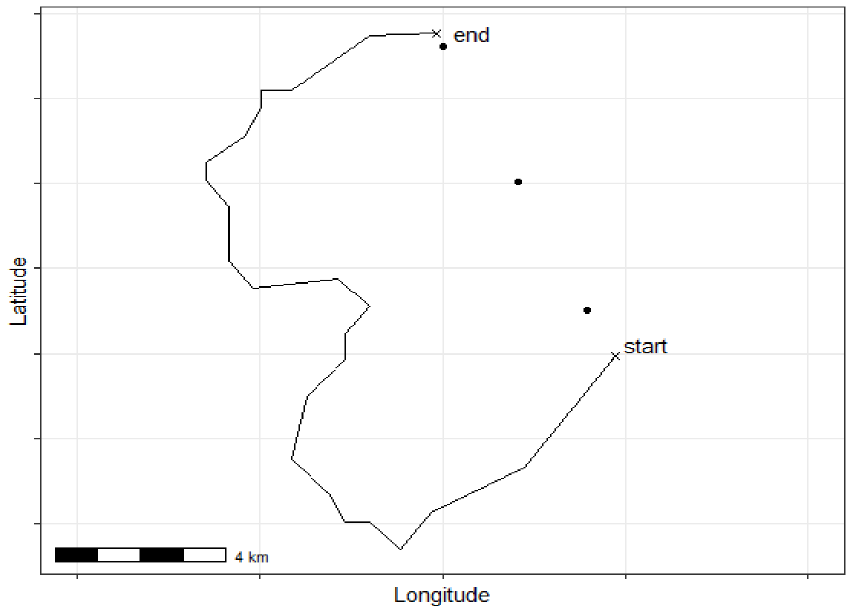


Figure 4. Bathymetry-based tow paths for an observation where VMS polls do form a straight line Starting and ending points for the tow are annotated and filled circles indicated VMS polls matched to the tow.

4.2 Matching

4.2.1 VMS and logbook data

We match individual VMS polls to fishing trips and tows from the logbooks according to the procedure from Section 3.1.1. The analysis begins with 252,655 VMS polls from the years 2008-2009 and 2014-2015 that could be matched to fishing trips reported on the logbooks. Of these polls, a little over 80,000 could be matched to specific logbook tows and roughly 163,000 could be matched to logbook fishing trips but not to specific tows. In the matched VMS-logbook data, over half of the logbook tows were matched to two or fewer VMS polls. Figure 5 shows the distribution of VMS polls per logbook tow for tows in our logbook data sample. Table 2 summarizes the matching of VMS and logbook data.

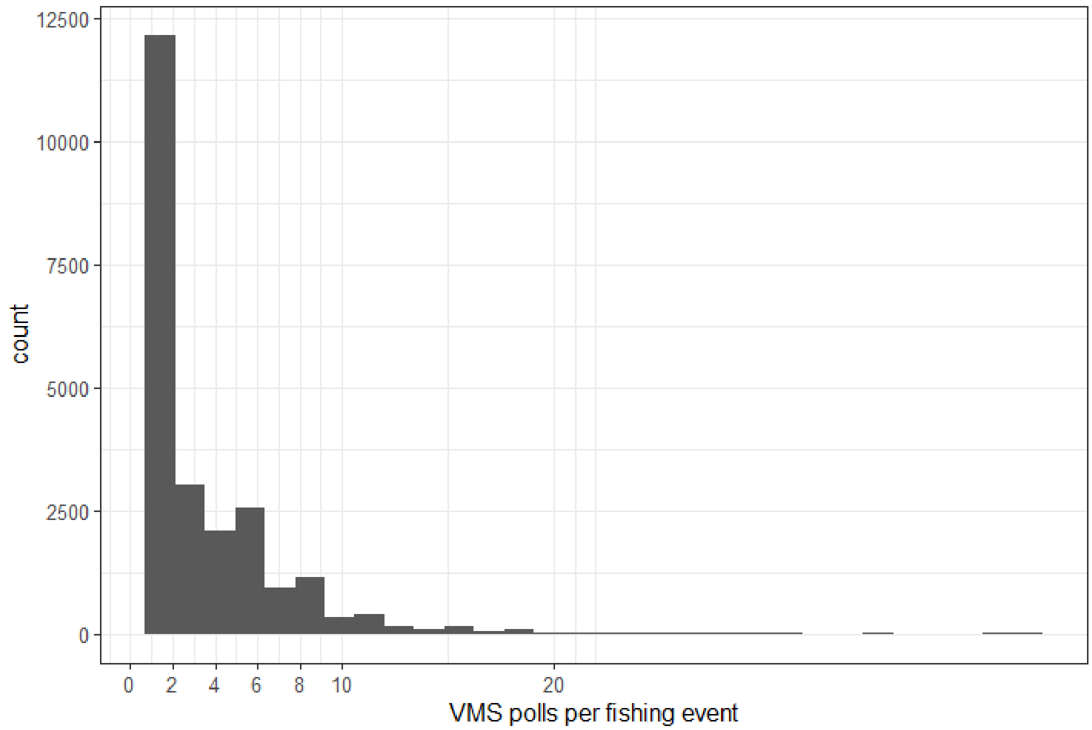


Figure 5. Distribution of VMS polls per logbook tow.

Table 2. Summary of VMS polls matched to logbook tows.

Year	Vessels	Fishing Trips	Logbook Tows	VMS Polls Matched to Trip Only	VMS Polls Matched to Logbook Tows	Total VMS Polls
2008	55	1,814	7,194	33,577	19,163	52,740
2009	52	1,937	7,163	50,411	26,101	76,512
2014	43	1,622	7,430	41,793	18,911	60,704
2015	39	1,806	8,248	44,213	18,486	62,699
Total	79	7,178[17]	30,035	163,108	80,671	252,655

Appendix C provides a detailed accounting of logbook observations that could not be matched to any VMS polls.

[17] The column total for unique logbook trips by year is greater than the total number of unique logbook trips because there was one fishing trip in our logbook data sample that spanned multiple years. In Table 2 this trip counted as a unique trip in both 2008 and 2009.

When we encountered instances where a vessel reported fishing on the logbooks but no VMS polls could be found corresponding to the reported trip, it was generally the case that the vessel had full months of VMS data missing. This observation is illustrated in Figure 6, which shows daily VMS polling for a single vessel in the year 2009. Note that a large chunk of VMS polling is missing between July and August when a non-trivial amount of fishing activity happened. One possible explanation for prolonged gaps in VMS coverage could be seasonal participation in fisheries where continuous VMS monitoring is not required[18]. While this may plausibly explain some VMS coverage gaps, we note there are important cases where we observe significant vessel activity in the limited entry groundfish fishery with no corresponding VMS data.

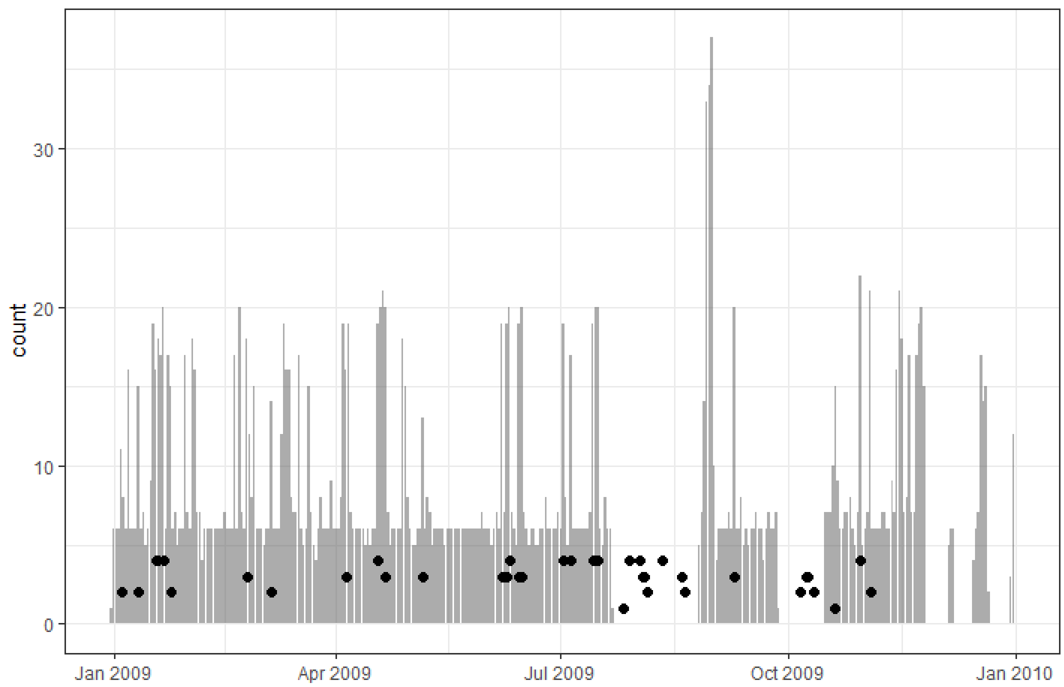


Figure 6. VMS polls per day and limited entry groundfish trawl tows per day for a particular vessel in 2009. Dots indicate number of limited entry trawl tows and bars show the number of VMS polls for each day.

[18] For example, the trawl logbooks contain information on hundreds of trips that target the state managed California Halibut stock. It is our understanding that vessels trawling for California Halibut are not required to have VMS. For vessels participating in both the limited entry groundfish fisheries and directed California Halibut fishery one might expect to see temporal clusters of VMS polls during participation in the limited entry fishery, punctuated by gaps in VMS coverage during times the vessel was targeting California Halibut.

4.2.2 Matching VMS and fish ticket data

Our raw logbook data contain information on 30,061 tows executed on 7,182 unique fishing trips. Joining these data to gear information from the fish tickets data using fish ticket identifiers, we are able to match 6,682 fishing trips to gear types. About 270 of these trips reported using more than 1 gear type during the trip. Filtering these multiple gear type trips out, we are left with a data set containing 27,027 tows executed over 6,409 unique trips.

4.3 Distance agreement between logbook tow paths and VMS polls

In this section we examine the distance between logbook reported locations and VMS polls over time and latitude strata. We begin by assessing whether the bathymetry derived tow paths provide a better fit to the VMS data than straight line tow paths.

4.3.1 VMS and logbook spatial agreement by tow path interpolation method

Table 3 shows that the straight line tow paths result in smaller distances from VMS polls to logbook tows than the bathymetry tow paths. In the next section we examine whether bathymetry tow paths provide a closer spatial match to VMS polls for particular types of tows.

Table 3. Quantile values of the empirical distribution of distances (in km) from VMS fishing polls to corresponding logbook tow lines.

Tow Path Method	0.1	0.25	0.5	0.75	0.9	0.95	0.99
Bathy	0.22	0.72	2.12	4.81	8.14	10.70	30.54
Straight	0.06	0.22	0.76	2.10	4.90	8.06	30.70

As discussed in Section 3 there are 2,000 discretized points on each logbook tow line. Appendix C, Table C1, illustrates that increasing the number of vertices beyond 2,000 did not produce notable changes in the calculated distances between VMS polls and logbook tows.

4.3.1.1 Do bathymetry tow paths match VMS data better than straight line tow paths in particular areas?

One might expect straight line tows path to adequately represent true fishing activity if trawling occurs over areas with relatively constant bottom depths. Conversely, in areas with more complex bathymetry, one might expect the bathymetry derived tow paths to better represent the true location of fishing activity. To the extent that bathymetry varies by latitude gradient, one might expect the bathymetry-derived tow paths to provide a better spatial match to VMS polls in some areas but not others. To address this issue empirically, we compare the distances between VMS polls and straight-line tow paths and between VMS polls and bathymetry tow paths across different latitude strata.

We define regions based on latitude strata used in prior groundfish research by Holland and Jannot (2012). The latitude strata used are defined in Table 4[19].

Table 4. Latitude strata and associated coordinates.

Latitude Strata	Latitude Range
1	South of $36°N.$
2	$36°N.-38°N.$
3	$38°N.-40°10\ N.$
4	$40°10\ N.-42°30\ N.$
5	$42°30\ N.-44°\ N.$

Table 5 shows that for all latitude strata, representing potentially different seafloor complexity and/or fishing strategies, straight line tow paths produce a greater degree of spatial agreement ($\hat{\delta} > 0$) than bathymetry-derived tow paths.

[19] Our analysis utilizes data from logbooks observations with departure ports in California. However, vessels departing from ports along the northern part of the California coast often move north to fish in Oregon waters. As a result, our analysis includes tow paths and VMS polls from fishing activity taking place north of $42°\ N.$

Table 5. Differences in VMS-logbook distances between bathymetry-derived logbook tow paths and straight-line logbook tow paths.

Latitude Strata	Number of VMS Fishing Polls	Mean Difference (km) in Distance ($\hat{\delta} = d_{bathy} - d_{straight}$)	$\Pr(\hat{\delta} \neq 0)$
1	19,232	0.476	0.000
2	15,543	0.792	0.000
3	18,442	2.278	0.000
4	29,446	1.845	0.000
5	12	0.106	0.430

4.3.1.2 Do bathymetry tow paths provide a better spatial fit to VMS data for longer tows?

On shorter tows one might expect a trawl vessel to encounter fewer natural changes in bottom depth. In this case, bottom depth could be kept constant with a straight line. Over longer distances however, more departures from the straight line may be required in order to maintain a constant bottom depth. In this case, one might expect the bathymetry tow paths to provide a better fit to the VMS data.

We can test whether the bathymetry tow paths result in smaller calculated distances between VMS and logbook data for longer tows by regressing the VMS-to-logbook bathymetry distance on the tow duration. We do this separately for each latitude strata to control for any differences in this potential relationship across latitude strata. Regression results are reported in Table 6.

Table 6. Regression of VMS-to-logbook distances (bathymetry method) on tow duration.

Latitude Strata	Number of VMS Fishing Polls	Adjusted R-Squared	Estimated Coefficient for Tow Duration (p-values)
1	19,232	0.028	0.741 (0.000)
2	15,543	0.057	0.627 (0.000)
3	18,442	0.086	0.424 (0.000)
4	29,446	0.092	0.545 (0.000)
5	12	0.084	0.864 (0.187)

There are no latitude strata in which the VMS-to-logbook bathymetry tow path distance declines with tow duration. Without exception, this distance increases as tow length increases, as indicated by the strongly positive regression coefficients for the tow duration variable.

4.3.1.3 Summary of bathymetry-derived versus straight line tow path approximations

By examining individual logbook tow paths relative to VMS data it is straightforward to conclude that vessels generally do not tow in straight lines. However, our analysis shows that, for our logbook data, straight lines tend to match VMS fishing polls better than an algorithmic approach which minimizes change in the average bathymetry. We would like to emphasize however that this conclusion should not be interpreted as an endorsement of either method in the estimation of the spatial distribution of fishing effort. Our analysis is meant to provide informational summaries of the VMS data relative to existing commercial fishing data sources. The primary purpose of this information is educating future data users about the strengths and limitation of joined VMS-logbook data.

In Appendix E we provide illustrations of some common spatial inconsistencies found between VMS polls and logbook tows. These inconsistencies can be organized as follows:

1. Case 1: Bathymetry drawn tow-paths provide a better visual fit to the shape of the VMS fishing path and bathymetry drawn tow-paths have greater spatial agreement with VMS data than straight line tow paths.
2. Case 2: Bathymetry drawn tow paths provide a good visual fit to the shape of the VMS points (in that both VMS fishing polls and the bathymetry tow path exhibit similar curvature) but the straight line logbook path results in smaller distances between VMS polls and the tow path.
3. Case 3: The VMS fishing path appears to be relatively straight and the straight line tow path results in the smallest distances between VMS polls and the logbook tow.
4. Case 4: Both bathy tow paths and straight tow paths offer a poor fit to the VMS fishing path due to obvious errors in either the logbook or VMS data.

4.3.2 VMS and logbook spatial agreement by year

In Table 7 we summarize the distributions of distances between logbook tows and VMS fishing polls for each year. In general, we observe differences in the empirical quantiles of distances over time that are small in absolute terms (fractions of a kilometer) but can be large in relative terms (note the difference in median distance from 2008 to 2015 is almost 30%).

Table 7. Quantile values of the empirical distributions of distances (in km) from VMS poll to logbook tow lines reported by year.

Year	Tow Path Method	0.1	0.25	0.5	0.75	0.9	0.95	0.99	VMS Polls
2008	Straight	0.078	0.25	0.81	2.08	4.45	7.29	27.21	19,163
2008	Bathy	0.275	0.88	2.47	4.97	7.99	9.94	26.87	19,163
2009	Straight	0.08	0.28	0.88	2.36	5.46	9.56	32.34	26,101
2009	Bathy	0.27	0.92	2.52	5.08	8.23	11.27	31.73	26,101
2014	Straight	0.05	0.19	0.68	1.98	4.78	7.98	29.65	18,911
2014	Bathy	0.20	0.64	1.85	4.78	8.32	11.30	29.66	18,911
2015	Straight	0.04	0.14	0.54	1.71	4.48	7.23	32.15	18,486
2015	Bathy	0.14	0.46	1.38	3.77	7.62	10.20	32.16	18,486

In Table 8 we provide a test of the significance of interannual differences in VMS-logbook distances. We first log-transform the distances to addressed the skewed nature of the distributions. We then conducted a simple linear regression relating the log-transformed distances to year. The p-values for the year covariate are functionally 0 for both distance calculation methods (straight-line and bathymetry-based), indicating the average agreement between VMS locations and logbook locations improved with the passage of time. However, the practical implications of these results will likely depend on the research question under evaluation. As an example, in social science applications, researchers are likely to be interested in whether spatial agreement between VMS and logbook data increased significantly over time. A trend toward greater spatial agreement over time could indicate more consistent compliance in later years, perhaps driven by more effective enforcement or confusion about regulatory requirements in the initial implementation of the program. If VMS data from the early stages of program implementation were spatially inconsistent with existing data sources due to fishers' confusion about the regulatory regime or lax enforcement, researcher seeking to use VMS data for analytical purposes would be wise to focus analysis on more recent time periods. While our results from Table 8 do indicate increasing spatial agreement over time, the magnitudes of the interannual differences do not, in our opinion, provide any reason to be skeptical of the data quality for the early years of the VMS program.

Table 8. Linear regression of log transformed VMS-to-logbook distances on year.

Tow Path Method	Degrees of Freedom	Year	p-value
Straight	82,673	-0.059	0.000
Bathy	82,673	-0.058	0.000

4.3.3 VMS and logbook spatial agreement by region

We previously compared the fit of bathymetry-based and straight-line tow paths with the VMS data relative to tow duration across years. Here we summarize the distributions of VMS-to-logbook distances within each latitude strata. Table 9 summarizes the quantiles of the VMS-logbook distance by tow interpolation method (bathymetry-based and straight line) and latitude strata.

Table 9. Quantile values of the empirical distributions of distances (in km) from VMS poll to logbook tow lines reported by latitude strata.

Latitude Strata	Tow Path Method	0.1	0.25	0.5	0.75	0.9	0.95	0.99	VMS Polls
1	Straight	0.03	0.08	0.30	0.99	4.25	13.12	54.16	19,232
1	Bathy	0.10	0.31	0.83	1.99	5.02	13.13	53.02	19,232
2	Straight	0.09	0.32	0.95	2.28	4.31	6.21	16.22	15,543
2	Bathy	0.18	0.57	1.63	3.81	6.31	7.84	16.40	15,543
3	Straight	0.07	0.23	0.69	1.71	3.80	6.04	13.79	18,442
3	Bathy	0.37	1.28	3.20	5.53	8.10	9.36	16.67	18,442
4	Straight	0.11	0.37	1.09	2.77	5.93	9.01	28.14	29,446
4	Bathy	0.42	1.26	3.01	5.95	9.35	12.25	28.19	29,446
5	Straight	0.61	1.30	5.84	10.78	13.82	14.11	14.11	12
5	Bathy	0.88	1.88	5.84	10.78	13.82	14.11	14.11	12

From Table 9, two observations stand out:

1. Median distance between VMS polls and logbook tow lines is notably lower for the southern-most latitude zone relative to other latitude strata.
2. Extreme distance values are notably larger for tows in the southern-most latitude zone.

Differences in VMS-to-logbook distance across latitude strata are likely influenced by a number of factors including differences in the nature of fishing activity and differences in the physical characteristics of the fishing grounds over space. While we leave a more rigorous examination of these factors to future research, we present here two observable sources of heterogeneity correlated with fishing location: tow duration and species targeting.

Logbook reported tows in the southern-most latitude zone are considerably shorter in duration than tows in other areas (Figure 7). A significant amount of fishing activity in this area targets ridgeback prawns and California halibut, while fishing activity in other latitude strata mainly targets federally managed groundfish stocks (Figure 8).

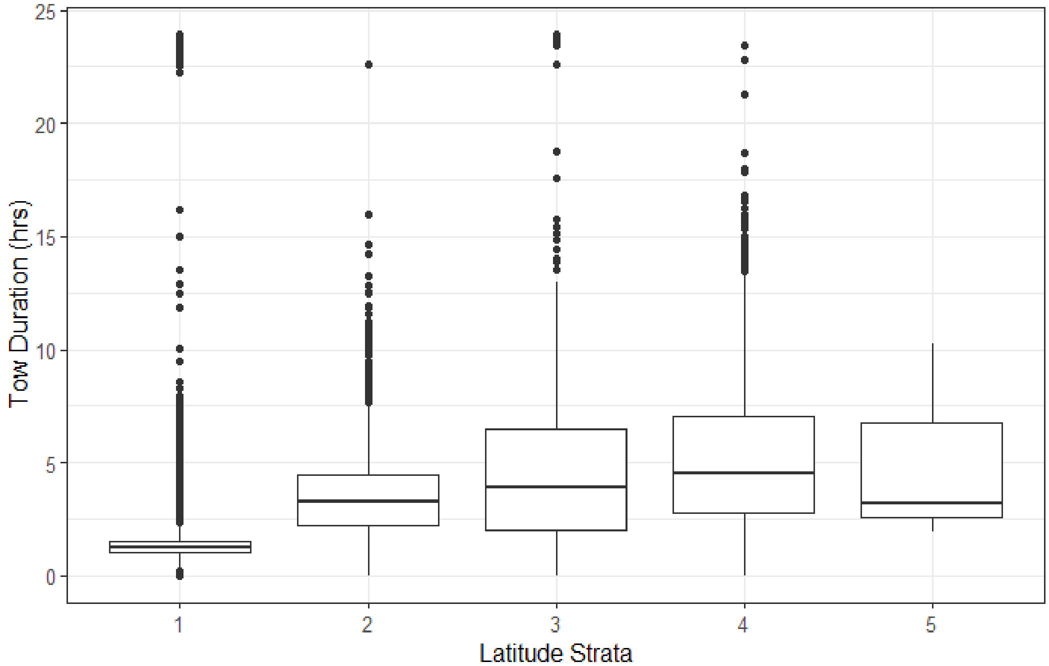


Figure 7. Duration of logbook tows by latitude strata.

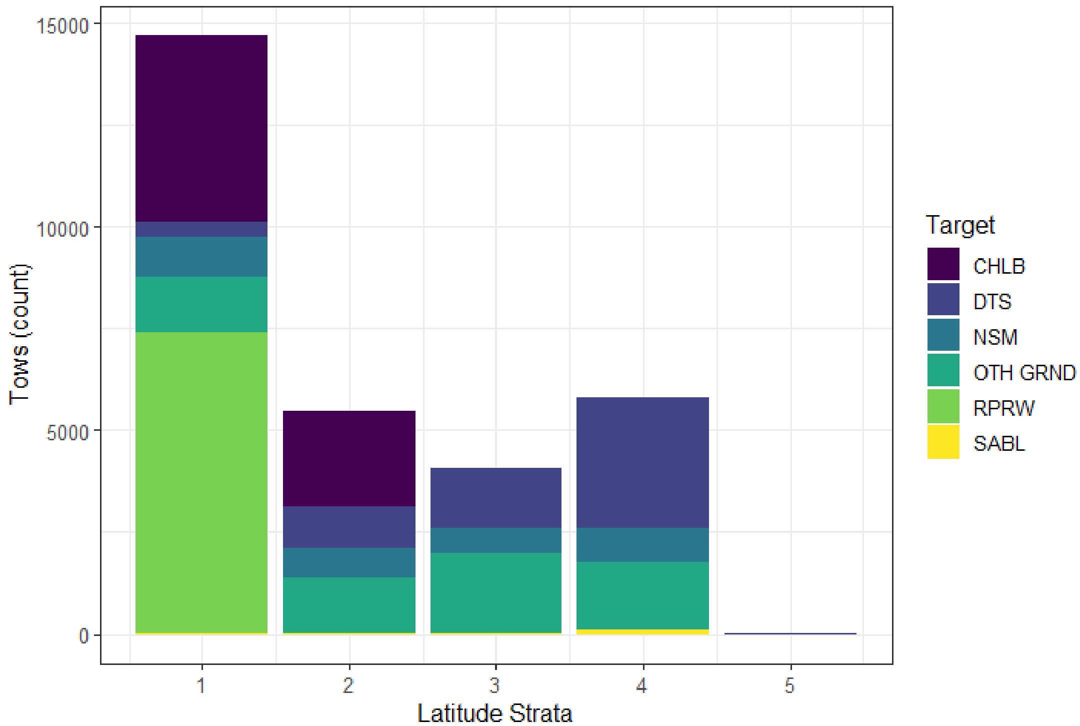


Figure 8. Species and species complexes targeted on logbook tows by latitude strata. Targets for each tow are self-reported by vessel captains. Target abbreviations are as follows: CHLB is California Halibut; DTS indicates targeting deep water groundfish dominated by dover sole, short and longspine thorny heads, and sablefish; NSM indicates targeting a nearshore mix (often comprised of assorted flatfish such as sanddabs); OTH GRND indicates targeting other groundfish species; RPRW indicates ridgeback prawn; SABL indicates tows targeted[20].

4.3.4 VMS and logbook spatial agreement by Dahl Groundfish Code

In this section we evaluate differences in the distance between VMS polls and logbook tows for fishing activity subject to VMS requirements relative to fishing activity not subject to VMS requirements. This data summary is meant to partially address the question of whether VMS data could be used by fisheries researchers to fill data gaps. VMS monitoring is required for groundfish activity in the limited entry trawl and IFQ sectors. These sectors are also monitored through other programs such as the West Coast Groundfish Observer Program. However, as an empirical matter, some logbook reported tows from sectors and fisheries not required to carry VMS units match to polling from the VMS data[21]. These sectors and fisheries are subject to less rigorous monitoring requirements than the limited entry trawl and

[20] Species targets appear in the logbook data under the field PACFIN_TAR. Fishermen reported 31 distinct targets on logbooks. Here we retain the most frequently used target codes (CHLB, DTS, NSM, RPRW, and SABL) and aggregate the remaining entries into the "OTH GRND" code.
[21] An intuitive explanation for why this might be the case is the following: if a fisher participates in multiple groundfish sectors where some activity is required to be monitored by VMS and other activity is not, the fisher may find it operationally convenient to simply keep the VMS unit functioning continuously.

IFQ groundfish sectors, so less is known about them. The fact that VMS data offer some coverage of fisheries not explicitly subject to VMS requirements suggests the possibility that VMS could provide an important source of information on fisheries for which spatial data is currently limited.

Dahl Groundfish Codes[22] can serve as a proxy for fishing activity required by law to be covered by VMS observation. Logbook data include fishing activity in the limited entry trawl and IFQ sectors of the groundfish fishery (Dahl Groundfish Codes 1-4 and 20). These logbook fishing trips should have corresponding VMS data as a matter of regulation[23]. Logbook data also include a significant amount of fishing activity assigned to groundfish sectors and fisheries not required by law to be monitored by VMS[24].

Table 10 summarizes the distance between VMS polls and logbook tows for fishing activity required, and fishing activity not required, to carry VMS units. These results do not suggest large scale differences in VMS-logbook distance for the two groups. There are however, two additional points to consider when comparing tows for which VMS is not required with tows subject to a VMS requirement:

1. As shown in Appendix C, VMS coverage of tows outside the limited entry trawl and IFQ groundfish sectors is relatively sparse, and
2. It is likely that VMS data on tows not required to have VMS is generated primarily from fishers who also participate in fisheries where VMS is required. These fishers may exhibit different behavioral characteristics from fishers that exclusively participate in fisheries not subject to VMS requirements.

To summarize, our data provide no evidence that VMS-logbook spatial agreement is systematically worse for fisheries not subject to a VMS requirement. However, VMS data coverage of fisheries and sectors outside the limited entry trawl and groundfish IFQ sectors may not constitute a representative sample of fishing activity in these fisheries/sectors.

[22] Dahl Groundfish Codes are designations used by PacFIN to assign fishing activity to management sectors. These codes are described in detail here: https://pacfin.psmfc.org/data/documentation-2/.

[23] It should be noted that Dahl Sectors 1-3 refer to the various sectors of the directed Pacific whiting fishery. It is our understanding that these sectors have a VMS requirement. For the analysis of VMS and logbook spatial agreement by Dahl Sector, logbook observations assigned to Dahl Sectors 1-3 are included. However, since Pacific whiting generally are not targeted in California waters, less than 0.5% of the fishing trips in our logbook data sample were assigned to Dahl Sectors 1-3.

[24] These sectors and fisheries are detailed in Appendix Table C 1.

Table 10. Quantile values of the distributions of distances between VMS polls and logbook tows for tows required and not required to carry VMS units.

VMS Mandate	Tow Path Method	Observations	0.1	0.25	0.5	0.75	0.9	0.95	0.99
Required	Straight	61,019	0.07	0.26	0.84	2.22	4.85	7.45	25.70
Not Required	Straight	21,656	0.04	0.13	0.47	1.55	4.86	11.47	38.43
Required	Bathy	61,019	0.31	1.02	2.72	5.35	8.39	10.51	25.70
Not Required	Bathy	21,656	0.11	0.35	0.91	2.22	5.64	11.47	38.32

4.4 Data agreement based on imputed speed

In this section we discuss the results of a proposed method for evaluating VMS and logbook spatial agreement by creating fishing tracks using joined VMS and logbook data. As discussed in Section 3.3., VMS polls between the starting and ending times for a logbook reported tow are joined with the logbook reported starting and ending coordinates for the tow. Next, line segments are drawn connecting each consecutive pair of coordinates. The total distance traveled is approximated by summing the lengths of the individual line segments. The speed required to fish this path is evaluated using the logbook reported starting and ending time for the tow and the total distance covered along the path. These speeds are then compared to the reference distributions of fishing speeds calculated using only the logbook data.

The first method of comparison constructs a reference distributions for each gear type. Table 11 summarizes the frequency of gear types observed in our logbook data sample and presents the empirical quantiles of the fishing speed distributions for each gear[25]. These are the fishing speeds calculated exclusively from logbook data by dividing the length of the tow by the tow duration (the difference between reported tow set time and tow retrieval time).

[25] Table 11 does not include non-trawl gear types. Logbook observations assigned to non-trawl gear types (such as hook and line, drift gillnet, and dredge gear) and observations that could not be matched to a gear type were approximately 15% of our logbook sample.

Table 11. Quantile values for the distributions of fishing speeds for logbook observations by assigned gear type. Bathymetry-based speeds appear first in each cell with straight line speeds in ().

PacFIN Gear Code	Description	Number of Tows	0.1	0.25	0.5	0.75	0.9
GFS	Small footrope trawl	9,335	0.69 (0.67)	2.79 (2.51)	4.40 (3.73)	6.49 (4.42)	9.89 (5.05)
GFL	Large footrope trawl	7,115	2.89 (2.51)	3.89 (3.16)	5.24 (3.70)	7.03 (4.14)	9.25 (4.72)
SST	Single rigged shrimp trawl	6,938	2.81 (2.46)	3.85 (3.41)	5.00 (4.35)	6.72 (5.00)	10.31 (5.62)
FTS	Selective flatfish trawl	1,280	2.89 (2.70)	4.16 (3.56)	5.96 (4.11)	8.30 (4.42)	10.07 (4.72)
GFT	Otter trawl	810	0.1 (0.90)	3.11 (2.63)	5.53 (4.41)	7.93 (5.17)	11.00 (5.77)
MDT	Midwater trawl	95	1.72 (1.71)	4.71 (4.33)	6.08 (5.35)	9.76 (6.04)	13.13 (8.09)
DNT	Danish seine	22	0.58 (0.49)	0.84 (0.77)	1.79 (1.79)	4.90 (4.24)	15.10 (8.17)
BMT	Beam trawl	11	3.89 (2.84)	3.89 (3.99)	5.79 (5.20)	8.38 (5.44)	18.91 (14.68)

Table 12 summarizes the number of tows with calculated speeds considered as outliers when evaluated against the reference distributions and using Tukey's outlier detection rule. Comparisons using the bathymetry-based reference distributions resulted in more outliers. This is because bathymetry-derived tows paths are longer than straight line tow paths, a fact that results in faster speeds when using logbook times and distances to construct the fishing speed distributions.

To provide some context for these percentages we performed a bootstrapping exercise to determine how many outliers one would expect to detect with this method if the speeds were drawn from the same underlying distribution. In this exercise we created a bootstrapped distribution of average tow speeds by randomly sampling the logbook data. For each random draw we compared the bathymetry-based tow speed and straight-line tow speed with the upper bound (calculated for each gear type using Tukey's rule) of the relevant gear-specific distribution of tow speeds. We performed 100 bootstrapping replicates with 1,000 draws in each replicate. The results were as follows: for the bathymetry-based speeds, outliers as a percent of total observations ranged from 3% to 7.5% with a median of 5%. For the straight-line speeds, outliers as a percent of total observations ranged from 1% to 4% with a median of 2.6%.The purpose of the bootstrapping exercise described here is to provide additional context for the percentages reported in Table 12.

The analysis undertaken in this section compares a sample of fishing speeds from joined VMS-logbook data to a sample of fishing speeds calculated using only logbook data. We report the percent of tows in the joined VMS-logbook data with average speeds exceeding a threshold value defined using the quantiles of the distribution of fishing speeds observed in the logbook data. These tows are labeled as outliers. The bootstrapping exercise provides context by calculating how many outliers one would expect to detect if the two fishing speed samples were drawn from the same underlying distribution.

Table 12. Speed outliers in joined VMS-logbook data.

Tow Path Method	Observation Count	Reference Distribution	Outliers
Bathy	19,975	Gear	881 (4%)
Bathy	19,975	Vessel	1,364 (6.8%)
Straight	19,975	Gear	1,769 (8.8%)
Straight	19,975	Vessel	2,275 (11%)

In Figure 9 and Figure 10 we provide visual illustrations of the method used here to evaluate consistency between VMS and logbook spatial data. Figure 9 shows a tow not determined to be an outlier. Using the logbook reported times for this tow and the distance implied by connecting the VMS poll locations to logbook reported starting and end coordinates, we calculate an average speed of 3.66 km/hr. The gear type for this tow (taken from the fish ticket data) is gear code "GFS", or small footrope trawl. Since 3.66 km/hr. is within the interquartile range of the fishing speed distribution for small footrope trawl gear (Table 11), the tow is not labeled as an outlier.

Figure 10 illustrates a case where the speed required to fish the path created by joining logbook and VMS data exceeds the speed threshold for defining an outlier. In this case the path constructed by connecting logbook start and end points and imposing the condition that the path run through the VMS polls matched to that tow results in a tow covering a total of 180 kilometers in 4 hours. The average speed required to fish this path is approximately 45 km/hr. In this case, joining logbook and VMS data results in a tow that is unlikely to have happened as reported. In this case, we consider the VMS and logbook data to be inconsistent for this tow.

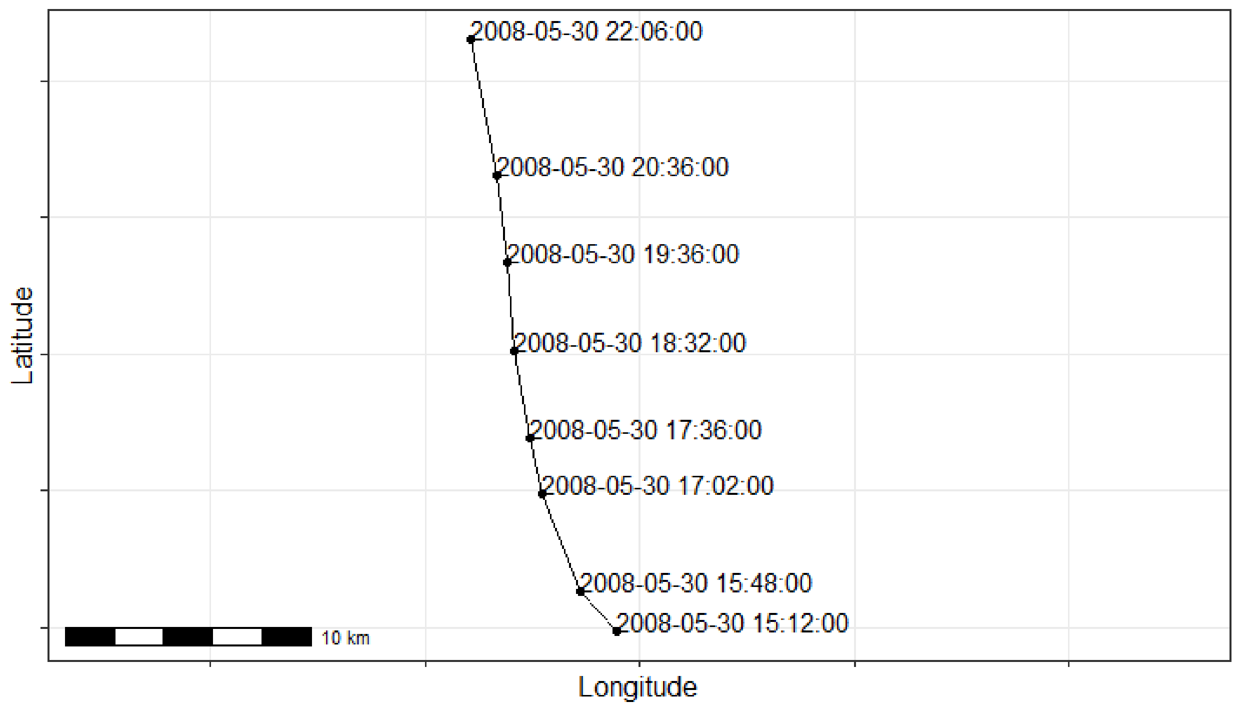


Figure 9. An imputed tow path from logbook start to logbook end through matching VMS poll locations for a feasible tow.

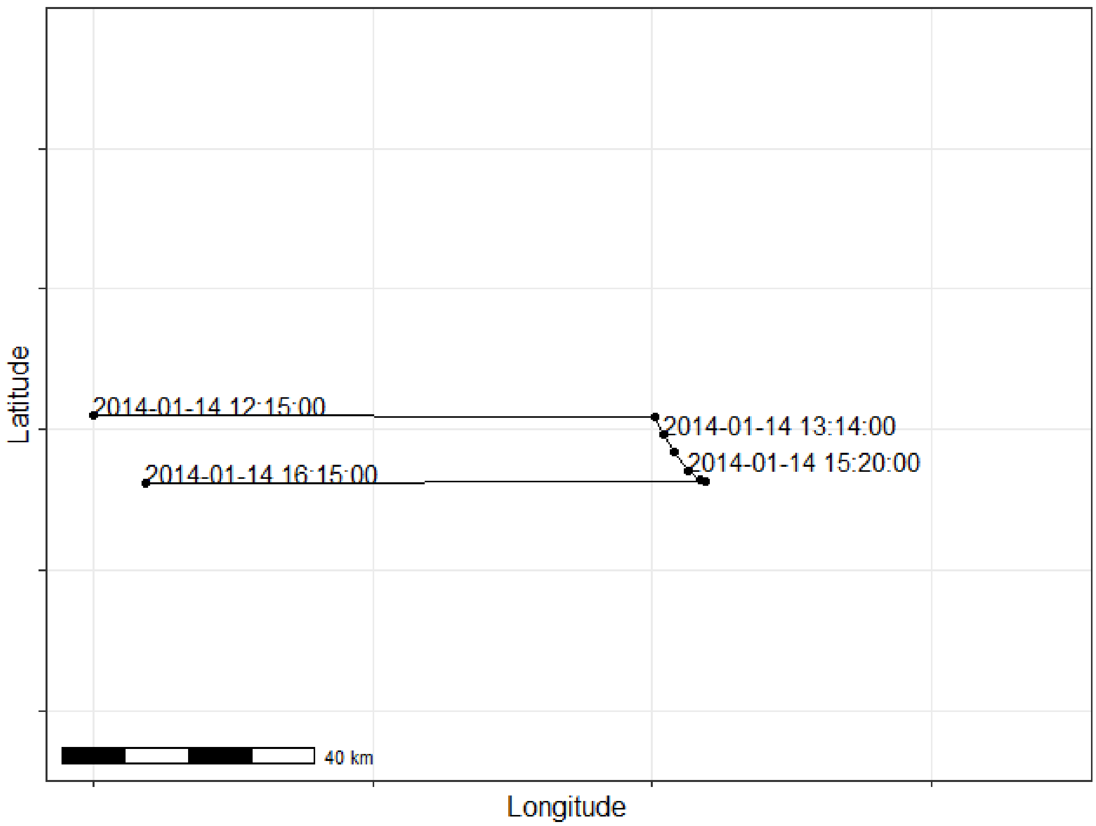


Figure 10. An imputed tow path from logbook start to logbook end through matching VMS polls for an infeasible tow.

The second method of comparison outlined in Section 3.3 uses vessel-specific tow speed distributions to evaluate possible speed outliers. From Table 12, one can see that this method results in more fishing profiles being classified as outliers when compared to the method using gear strata to construct the reference distribution.

This section has taken the approach of looking for fishing profiles in the joined logbook-VMS data that, based on prior available information, appear infeasible. The method of discovery used was to compare fishing speeds in the joined logbook-VMS data with fishing speeds derived from only logbook data. Outliers, or infeasible fishing profiles, were defined to be those with fishing speeds in excess of threshold values based on empirical quantiles of predetermined reference distributions. It is important to stress that both logbook and VMS data sources have imperfections that can result in a logbook-VMS fishing profile being deemed an outlier according to the methodology employed here. Examples of such imperfections include:

1. Logbook entries can be subject to data entry errors. Consider the example of an outlier fishing profile from Figure 10: from a visual inspection of the VMS polls relative to the logbook reported coordinates it is reasonable to suspect that the logbook set and up points were reported

incorrectly[26]. Moreover, if the longitude coordinates for the logbook set and up points were adjusted by exactly *1°*, the resulting fishing profile would have an average speed of approximately 3.4 km/hr and a conventional north-south orientation.

2. VMS polls can suffer data transmission anomalies. We have observed cases in the VMS data where a single vessel has multiple VMS polls separated by a kilometer or more but timestamped within minutes of one another.

It is important to again emphasize that the objective of our analysis is not to make a binary conclusion about whether VMS and logbook data 'agree' or 'disagree.' Neither is it our intention to use VMS data to validate or invalidate self-reported logbook data. The primary purpose of this analysis is to provide a detailed characterization of VMS data relative to logbook and other fishery dependent data sources. We contend that this analysis is important as it provides future data users with enhanced understanding of the benefits and potential limitation of VMS data for fisheries science and management. In this section we have shown that between 5 and 10% of logbook reported tows have corresponding VMS data points that appear spatially inconsistent with logbook reported locations. Individual researchers will need to determine, likely on a case-by-case basis, whether/how detrimental this fact is to their chosen application.

4.5 Classification of fishing versus not fishing using VMS polls

4.5.1 Analysis of observed fishing speeds

Our speed analysis is based on data from logbook recorded tows in 2008, 2009, 2014, and 2015 but includes only trips for which a single gear type was used for the entire trip[27]. Table 11 showed the gear types with abbreviations used by PacFIN and selected empirical quantiles of the fishing speed distributions for each gear type. Table 11 shows a wide range of fishing speeds calculated from logbook observations. For the two most utilized gear types, large footrope bottom trawl (GFL) and small footrope bottom trawl (GFS), fishing speeds were between approximately 0.7 and 10 km/hr. with median speeds of 4.4 km/hr. for GFS and 5.2 km/hr. for GFL gear types. For the lesser utilized bottom trawl gears (GFT and BMT) speeds were calculated between 0.1 km/hr. and 19 km/hr. with median fishing speeds of 5.5 for GFT and 5.8 for BMT gear. For selective flatfish trawl gear (FTS), speeds were generally between 2.9 and 10 km/hr., with a median speed of roughly 6 km/hr. For Danish Seine gear the speed range was 0.6 – 15 km/hr. with a median fishing speed of 6.1 km/hr. Finally, for single rigged shrimp trawl gear (SST) fishing speeds were generally between 2.8 and 10.3 km/hr., with a median speed of 5 km/hr.

---

[26] The coordinates may have been recorded incorrectly (either accidentally or intentionally) or the coordinates may have been correctly recorded from gear that was not functioning properly.

[27] We impose this filter because of difficulties associated with matching gear types to individual tows. Gear types are assigned to each fish ticket. An individual fish ticket can be mapped back to a trip identifier and tow number from the logbook data but this match is not 1 to1. Some fish tickets contain catch from more than 1 tow. Fish tickets also often include multiple landings on the same ticket and so a fish ticket may have more than one gear type associated with it (the maximum number of distinct gear types for a single fish ticket observed from the fish ticket data 2008-2009 was 3). To avoid the complication of mapping a single tow to more than 1 gear type we select only trips where a single gear type is used for the trip.

4.5.1.1 In sample prediction using speed ranges

Using the merged VMS-logbook data, we conduct an analysis of how well vessel speed identifies fishing activity. This exercise classifies a VMS poll as 'fishing' if the observed speed is between the speeds defining the 25th and 75th percentiles of the speed distributions shown in Table 11. For simplicity we have referred to this classification method as the "naïve" method. Table 13 summarizes the results of this classification exercise.

Table 13. In-sample prediction of VMS polls as "fishing" or "not fishing" using a fishing speed window defined by the 25th and 75th percentile of the speed distribution for each gear type.

Gear	Tow Path Method	Total Observations	Total Fishing Polls	Fishing Polls Correctly Predicted (% of fishing polls)	Fishing Polls Incorrectly Predicted (% of fishing polls)	Total Non-Fishing Polls	Non-Fishing Polls Correctly Predicted (% of non-fishing polls)	Non-Fishing Polls Incorrectly Predicted (% of non-fishing polls)
GFL	Bathy	88,395	37,852	15,346 (0.41)	22,506 (0.59)	50,543	46,769 (0.93)	3,774 (0.07)
GFS	Bathy	80,377	24,348	18,312 (0.75)	6,036 (0.25)	56,029	47,685 (0.85)	8,344 (0.15)
SST	Bathy	28,677	6,625	2,539 (0.38)	4,086 (0.62)	22,052	19,775 (0.90)	2,277 (0.10)
FTS	Bathy	11,823	3,182	1,482 (0.47)	1,700 (0.53)	8,641	7,858 (0.90)	783 (0.10)
GFT	Bathy	3,907	990	592 (0.60)	398 (0.4)	2,917	2,399 (0.82)	518 (0.18)
MDT	Bathy	2,908	177	106 (0.60)	71 (0.40)	2,731	2,388 (0.87)	343 (0.13)
DNT	Bathy	314	24	19 (0.79)	5 (0.21)	290	171 (0.59)	119 (0.41)
BMT	Bathy	44	10	1 (0.1)	9 (0.90)	34	31 (0.91)	3 (0.09)
GFL	Straight	88,395	37,852	21,818 (0.58)	16,034 (0.42)	50,543	47,425 (0.94)	3,118 (0.06)
GFS	Straight	80,377	24,348	13,091 (0.54)	11,257 (0.46)	56,029	49,910 (0.89)	6,119 (0.11)
SST	Straight	28,677	6,625	2,600 (0.39)	4,025 (0.61)	22,052	19,937 (0.90)	2,115 (0.10)
FTS	Straight	11,823	3,182	1,449 (0.46)	1,733 (0.54)	8,641	8,084 (0.96)	557 (0.04)
GFT	Straight	3,907	990	489 (0.49)	501 (0.51)	2,917	2,474 (0.85)	443 (0.15)
MDT	Straight	2,908	177	78 (0.44)	99 (0.56)	2,731	2,590 (0.95)	141 (0.05)
DNT	Straight	314	24	17 (0.71)	7 (0.29)	290	176 (0.61)	114 (0.39)
BMT	Straight	44	10	1 (0.10)	9 (0.90)	34	30 (0.88)	4 (0.12)

For large footrope trawl gear (GFL), the naïve method correctly classifies about 40% of the fishing polls when using speeds based on the bathymetry tows paths and almost 60% when using the straight line speeds. In the case of the large footrope trawl gear, the fishing speed distribution calculated from logbooks using bathymetry lines and the distribution of VMS calculated fishing speeds are quite different. Figure 11 shows the distributions of speeds for logbook tows (using bathymetry tow paths) relative to speeds calculated from VMS polls assigned to tows using large footrope trawl gear. In the next section we will discuss the implications for fishing classification accuracy of using larger speed windows for identifying fishing and non-fishing VMS polls.

For small footrope trawl gear (GFS), the fishing classifier based on logbook speeds drawn from the bathymetry tow paths correctly predicted 75% of fishing polls and 84% of non-fishing polls. Classification predictions for small footrope trawl gear made using straight line tow speeds correctly classified 54% of VMS fishing polls and 89% of non-fishing polls.

For selective flatfish gear the speed-based criterion is effective at discriminating non-fishing behavior but has a difficult time correctly identifying fishing polls. The speed cutoffs calculated from bathymetry-derived tow paths had a lower bound of 4.15 km/hr. and an upper bound of 8.27 km/hr. This window captures more of the VMS fishing polls at higher speeds. The speed cutoffs calculated from straight line tow speeds were 3.5 km/hr. to 4.41 km.hr. This window correctly classifies more fishing polls with speeds less than 4.15 km/hr. but incorrectly classifies fishing polls at higher speeds. VMS calculated fishing speeds and logbook calculated fishing speeds for selective flatfish trawl gear are shown in Figure 12.

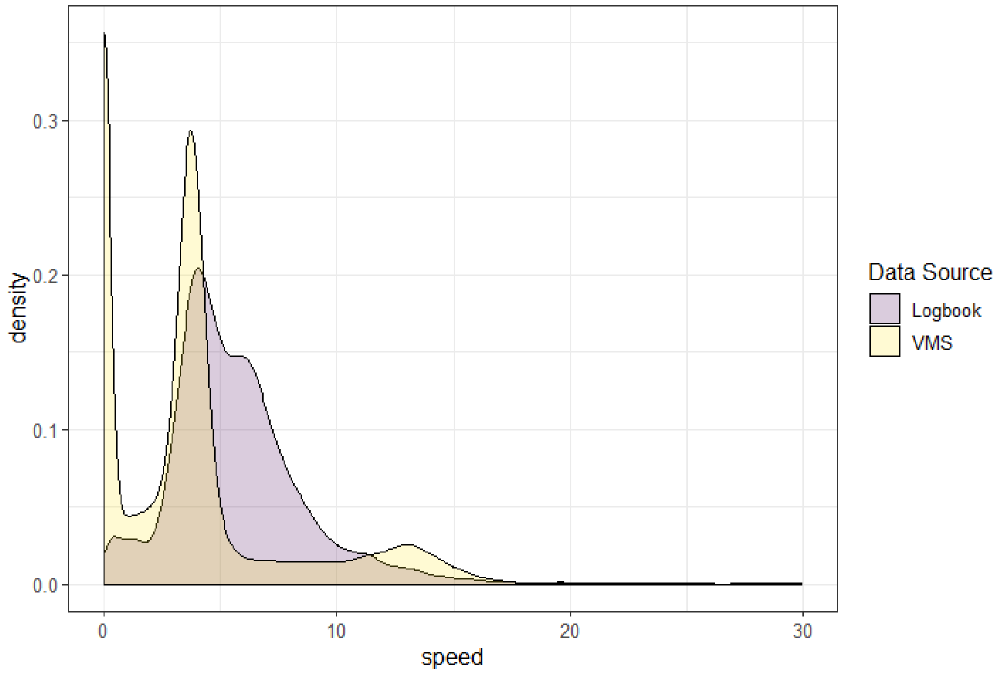


Figure 11. Distribution of logbook fishing speeds (km/hr.) calculated from bathymetry tow lines and VMS speeds for fishing observations using gear type GFL.

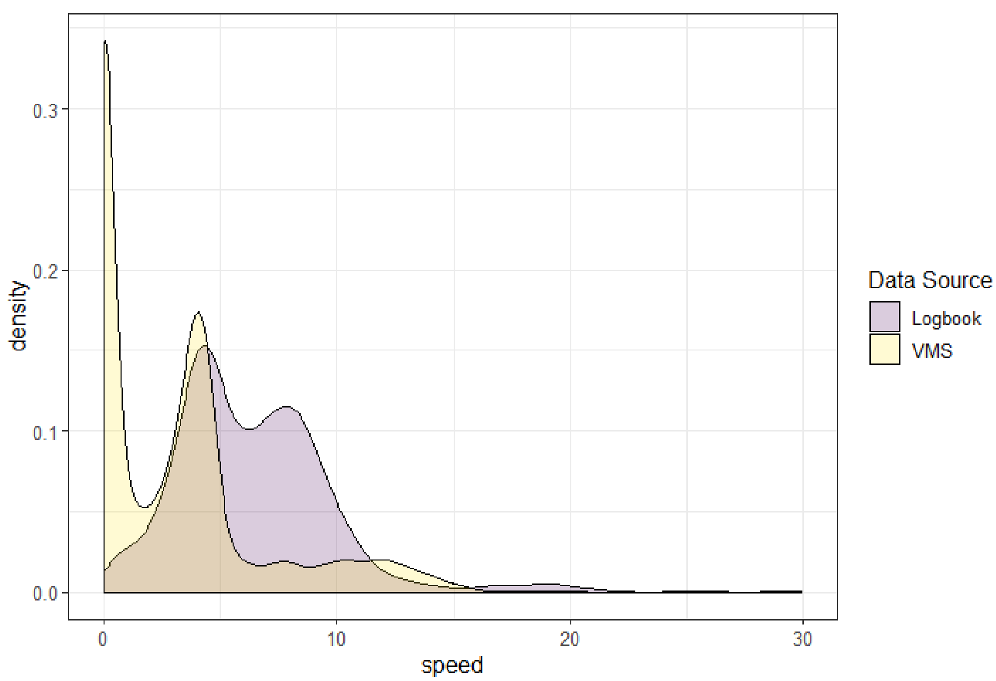


Figure 12. Distribution of logbook fishing speeds (km/hr.) calculated from bathymetry tow paths and VMS speeds for fishing observations using gear type FTS.

4.5.1.2 Sensitivity analysis of the naïve classification method

Since the decision to use the interquartile range of the logbook speed distributions to classify VMS polls is inherently ad-hoc, we provide a brief sensitivity analysis here. With speed-based classification methods the correct identification of fishing polls can generally be increased by increasing the size of the “fishing speed window”. A notable cost of expanding the fishing speed window is that it produces more false positives (non-fishing polls classified as fishing). An empirical question likely to be of interest to fisheries scientists is, “what is the magnitude of this tradeoff for the different ad-hoc speed-based classification rules?” To provide a partial answer to the question we summarize results of an additional speed-based classification rules in Table 14.

Using a naïve classifier based on the 10^{th} and 90^{th} percentiles of the logbook bathymetry-based tow speed distributions increases correct identification of fishing polls for GFL gear from 41% to 88% and decreases correct identification of non-fishing polls from 93% to 82%. Using the logbook straight line tow speed distribution for GFL gear, correct identification of VMS fishing polls increased from 58% to 84% while correct identification of non-fishing polls decreased from 94% to 88%.

For small footrope trawl gear (GFS), using values for the 10^{th} and 90^{th} percentile of the logbook speed distributions increases correct identification of fishing polls by 18% and 25% for bathymetry and straight line-based speeds respectively. For this gear type correct identification of non-fishing polls was decreased by 16% (for bathymetry-based speeds) and 11% (for straight line-based speeds).

For other, less utilized gear types, the increase in correct identification of VMS fishing polls produced by the 10/90 classifier relative to the 25/75 classifier can be observed directly from Table 14. In the next section we will examine the additional improvements in classification accuracy that can be obtained by introducing additional data.

Table 14. In-sample prediction of VMS polls as "fishing" or "not fishing" using a fishing speed window defined by the 10th and 90th percentile of the speed distribution for each gear type.

Gear	Tow Path Method	Total Observations	Total Fishing Polls	Fishing Polls Correctly Predicted (% of fishing polls)	Fishing Polls Incorrectly Predicted (% of fishing polls)	Total Non-Fishing Polls	Non-Fishing Polls Correctly Predicted (% of non-fishing polls)	Non-Fishing Polls Incorrectly Predicted (% of non-fishing polls)
GFL	Bathy	88,395	37,852	33,983 (0.88)	3,869 (0.12)	50,543	41,273 (0.82)	9,270 (0.18)
GFS	Bathy	80,377	24,348	22,698 (0.93)	1,650 (0.07)	56,029	38,595 (0.69)	17,434 (0.31)
SST	Bathy	28,677	6,625	4,134 (0.62)	2,491 (0.38)	22,052	17,217 (0.78)	4,835 (0.22)
FTS	Bathy	11,823	3,182	2,749 (0.86)	433 (0.14)	8,641	6,668 (0.77)	1,973 (0.23)
GFT	Bathy	3,907	990	890 (0.90)	100 (0.1)	2,917	2,016 (0.69)	901 (0.31)
MDT	Bathy	2,908	177	157 (0.89)	20 (0.11)	2,731	1,844 (0.68)	887 (0.32)
DNT	Bathy	314	24	24 (1)	0 (0)	290	86 (0.30)	204 (0.70)
BMT	Bathy	44	10	2 (0.2)	8 (0.8)	34	21 (0.62)	13 (0.38)
GFL	Straight	88,395	37,852	31,769 (0.84)	6,083 (0.22)	50,543	44,352 (0.88)	6,191 (0.22)
GFS	Straight	80,377	24,348	19,259 (0.79)	5,089 (0.21)	56,029	43,551 (0.78)	12,478 (0.22)
SST	Straight	28,677	6,625	3,993 (0.60)	2,632 (0.40)	22,052	18,128 (0.82)	3,924 (0.18)
FTS	Straight	11,823	3,182	2,344 (0.74)	838 (0.26)	8,641	7,400 (0.86)	1,241 (0.14)
GFT	Straight	3,907	990	808 (0.82)	182 (0.18)	2,917	2,202 (0.75)	715 (0.25)
MDT	Straight	2,908	177	148 (0.84)	29 (0.16)	2,731	2,256 (0.83)	475 (0.17)
DNT	Straight	314	24	21 (0.88)	3 (0.12)	290	121 (0.42)	169 (0.58)
BMT	Straight	44	10	2 (0.2)	8 (0.80)	34	23 (0.68)	11 (0.32)

4.5.2 A supervised learning experiment

As outlined in Section 3.3, we compare the prediction of a speed-rule classifier to two regression-based models. In addition to vessel speed, the regression approaches use information on ocean bottom depth and vessel bearing in order to classify VMS polls. The inclusion of these additional covariates is supported by prior VMS research (Watson and Haynie, 2016; Muench et al., 2018).

In Section 4.5.1 we used the naïve classifier to generate in-sample predictions. In this section we use the popular testing data/training data paradigm. Specifically, we estimate the models using a randomly selected 80% of the data and use the estimated models to generate predictions for the remaining 20% of observations. Additionally, all models were run separately for each gear type. Then the classification experiment was repeated using a single data sample with all trawl gear types pooled.

4.5.2.1 Classification by gear strata

To avoid perfect separation in the case of the logit model, we have filtered out a small number of extreme value data points. Specifically, VMS polls associated with bottom depths exceeding 800 fathoms (94 observations) and VMS polls with speeds exceeding 30 km/hr. (152 observations) have been removed. While this data truncation is admittedly ad-hoc, the mechanical characteristics of bottom trawling make it extremely unlikely for fishing activity to occur at depths greater than 800 fathoms or speeds greater than 30 km/hr.

Figure 13 shows that VMS polls matched to logbook tows suggest a north-south or south-north orientation to groundfish trawling. Figure 14 shows that fishing tends to happen over deeper bottom depths than non-fishing activity.

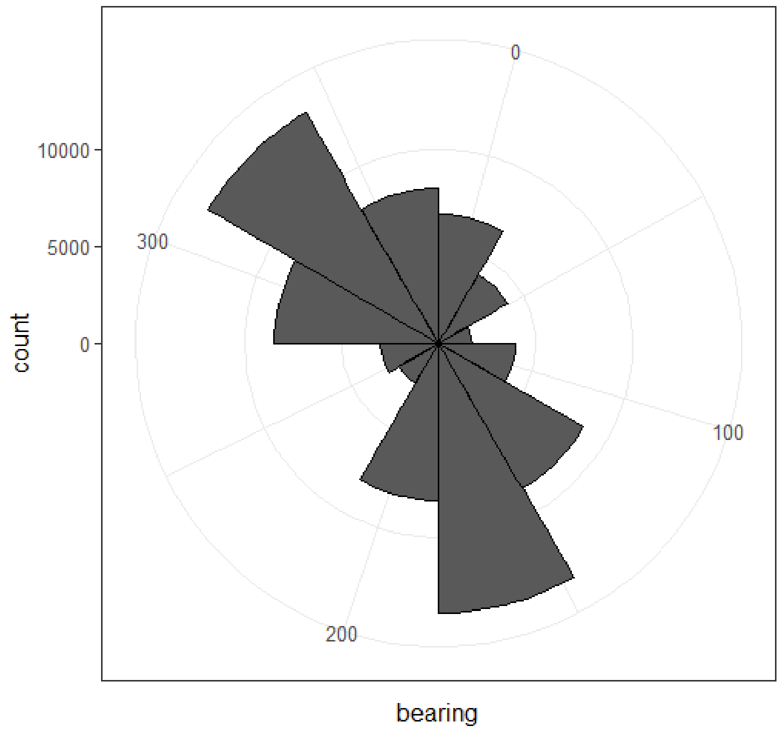


Figure 13. Histogram of bearing (measured in degrees) between consecutive fishing polls in the VMS data. Bearing is calculated at individual VMS polls using the current poll and previous poll.

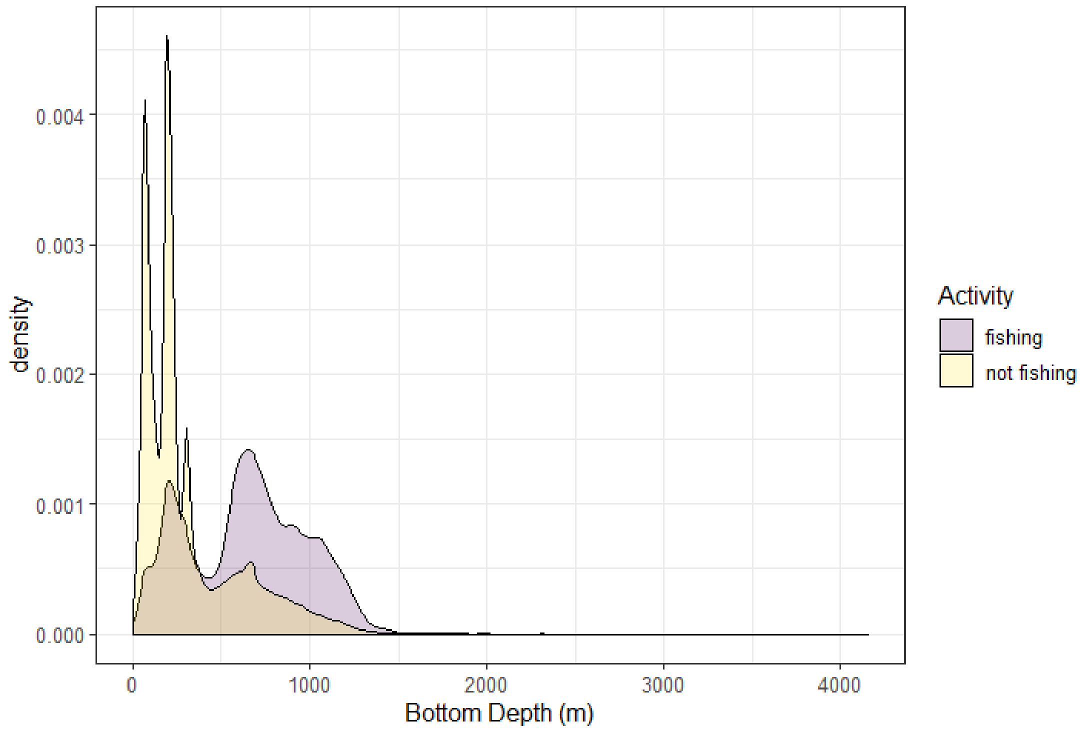


Figure 14. Bottom depth (measured in meters) at VMS polls for fishing and non-fishing polls.

All individual models included speed, bottom depth, and bearing as predictors. Classification accuracy for each model and gear type is summarized in Table 15. The simple models presented here offered mixed results with respect to whether the linear (GLM) or non-linear (GAM) predictors could classify fishing/not fishing activity of VMS polls more effectively than the naïve speed-based approach. For large footrope trawl gear, the GAM outperformed both the speed-based classifier and the logit classifier in correctly identifying fishing activity and non-fishing activity. For both the small footrope trawl gear and selective flatfish gear types, the GAM correctly classified more fishing polls than either the logit or naïve models. For these gear types, the logit model produced better classification accuracy for non-fishing polls.

Speed-based algorithms for classifying VMS polls are simple to implement and require minimal data. Table 15, illustrates that adding two additional, easily observable, covariates (vessel bearing and ocean bottom depth) can substantially improve fishing/not fishing classification of VMS polls. As mentioned in Section 3.4, the classification models employed here are quite simple. Future research in this area will benefit from exploring more advanced feature selection and feature engineering techniques.

Table 15. Comparison of models predicting activity status of individual VMS polls.

Model	Gear	Total Observations	Total Fishing Polls	Fishing Polls Correctly Predicted (% of fishing polls)	Fishing Polls Incorrectly predicted (% of fishing polls)	Total Non-Fishing Polls	Non-Fishing Polls Correctly Predicted (% of non-fishing polls)	Non-Fishing Polls Incorrectly Predicted (% of non-fishing polls)
Naïve	GFL	17,394	7,570	6,146 (0.81)	1,424 (0.19)	9,824	7,187 (0.73)	2,637 (0.27)
Logit	GFL	17,394	7,570	5,702 (0.75)	1,868 (0.25)	9,824	7,669 (0.78)	2,155 (0.22)
GAM	GFL	17,394	7,570	7,024 (0.93)	546 (0.08)	9,824	8,154 (0.83)	1,670 (0.17)
Naïve	GFS	15,608	4,893	3,421 (0.70)	1,472 (0.30)	10,715	2,428 (0.23)	8,287 (0.78)
Logit	GFS	15,608	4,893	2,285 (0.47)	2,608 (0.53)	10,715	9,408 (0.88)	1,307 (0.12)
GAM	GFS	15,608	4,893	3,698 (0.76)	1,195 (0.24)	10,715	9,157 (0.85)	1,558 (0.15)
Naïve	FTS	2,320	667	436 (0.65)	231 (0.35)	1,653	961 (0.58)	692 (0.42)
Logit	FTS	2,320	667	147 (0.22)	520 (0.88)	1,653	1,519 (0.92)	134 (0.09)
GAM	FTS	2,320	667	517 (0.78)	150 (0.22)	1,653	1,452 (0.88)	201 (0.12)

4.5.2.2 Classification with pooled gear types

The supervised classification experiment detailed in Section 4.4.2 is subject to the constraint that gear type is observable. In some situations, gear type and fishing activity may be observed simultaneously. While this is not true in all fisheries, for the groundfish trawl fishery, a predictive model that does not depend on gear type may be more useful, as gear type is often only observed during fishing activity. This section repeats the experiment from Section 4.4.2 and removes the covariate gear type from the analysis.

Although specific types of trawl gear are only reported on logbooks during tows, there are two gear types that can be removed from our analysis here. The first is the Danish Seine gear type: there is only 1 vessel in our data sample fishing groundfish with this gear type. The second is single rigged shrimp trawl. Since vessels must declare into a fishery before leaving port, declarations data should be capable of discriminating between vessels using this gear and vessels potentially using other groundfish trawl gear[28]. The models summarized in Table 16 include all remaining gear types ('GFL', 'GFS', 'GFT', 'FTS', 'BMT'). The same filtering of extreme values described in Section 4.5.2 is employed here (observations with speed in excess of 30 km/hr. or bottom depth greater than 800 fathoms have been removed).

Table 16. Comparison of models predicting activity status of individual VMS polls without gear type segmentation.

Model	Total Polls	Fishing Polls	Fishing Polls Correctly Predicted (% of fishing polls)	Fishing Polls Incorrectly Predicted (% of fishing polls)	Non-Fishing Polls	Non-Fishing Polls Correctly Predicted (% of non-fishing polls)	Non-fishing Polls Incorrectly Predicted (% of non-fishing polls)
Naïve	36,649	13,399	10,245 (0.76)	3,154 (0.24)	23,250	15,029 (0.65)	8,221 (0.35)
Logit	36,649	13,399	8,207 (0.61)	5,192 (0.39)	23,250	19,548 (0.84)	3,702 (0.16)
GAM	36,649	13,399	11,350 (0.85)	2,049 (0.15)	23,250	19,553 (0.85)	3,697 (0.15)

Results of the classification experiment run on a single data sample including all trawl gear types are consistent with those from Section 4.5.2.1. Specifically, the general observation that adding vessel bearing and bottom depth improves classification accuracy relative to the speed only classifier holds up even when the models are not stratified by gear type.

A notable caveat here is that the analysis included bottom trawl gears as well as midwater trawl gear. Midwater trawl fishing is associated with different speed and bottom depths profiles than bottom trawl fishing. Future research in this area should explore the potential value of models capable of jointly predicting gear type and fishing behavior.

Additionally, a caveat from Section 4.5.2.1 should be repeated here: our classification exercise does not include a rigorous feature selection component. The predictive performance of these model classes could likely be enhanced by adding interaction terms or 2nd order terms. We contend that this type of analysis will be best pursued in a separate, more focused research effort.

[28] Vessels using single rigged shrimp trawl gear would declare into one of the state managed shrimp fisheries before departing on the trip. Vessels declaring into one of the limited entry or open access groundfish trawl fisheries would be using one of the groundfish trawl gear types.

5 Discussion

The primary purpose of this report is to characterize the properties of West Coast VMS data relative to other important sources of data on commercial groundfish fishing. A secondary objective was to demonstrate how VMS data can be paired with other fisheries dependent data sources in order to evaluate important research questions.

We first assessed the question: *how well do self-reported fishing locations from logbooks agree with satellite positions recorded in VMS?* This question is of interest to fisheries managers and scientists on the West Coast because VMS is a relatively new source of data. Logbooks can provide a much richer historical picture of groundfish fishing off the Pacific Coast as they span a much longer time period. Comparing logbook recorded location to remotely sensed location can provide a measure of the precision and accuracy of logbook data.

The median distance between logbook reported locations and VMS tracked locations is less than 1km when comparing VMS polls to straight-line derived logbook tow paths. We find that straight-line tow paths are closer than bathymetry-derived tow paths to VMS polls, a finding that is consistent across regions and tow duration. Importantly, our analysis takes a simplistic view of spatial agreement in that we consider only raw VMS poll to logbook interpolated tow path distances. In many empirical fisheries applications involving VMS or logbook data there will be important context for researchers to consider, such as whether bathymetry-derived tow paths provide a better representation of area swept/fishing effort. The focus of our distance analysis is on measuring and summarizing important data relationships. Our intent is not to provide methodological recommendations but rather to provide important background information on the data.

VMS-logbook spatial agreement improved somewhat from earlier to later time periods. Further investigation of changes in spatial agreement over time may be warranted and may yield insights about changing technologies, reporting practices, fishing strategies, and observer coverage. Spatial agreement also differs somewhat between regions and some of these differences may be explained by differences in tow duration and target species/gear type.

Additionally, we assessed spatial agreement using an approach that asked whether tow tracks constructed by joining logbook and VMS data were feasible given existing information on fishing speeds. Our analysis found that 6 – 11% of tows were evaluated to be outliers according to our methodology.

It is important to emphasize that our assessment of spatial agreement does not specify (or assume) the source of any disagreement. Logbook data are subject to measurement error associated with potentially inaccurate recording of locations and times of tows. However, VMS data are also subject to measurement errors: positions are not perfectly recorded and timestamps can be inaccurate due to transmission speeds. In addition to measurement error, our methods are subject to process error related mainly to the fact that we do not observe the true path traveled by any vessel during a tow. Our analysis infers fishing paths using two different approaches, straight lines between set and up points and paths that follow bathymetry contours. The accuracy of our computation of fishing speeds and distances from VMS polls to their assigned tow lines depends on the extent to which true unobservable fishing paths deviate from the inferred paths.

A second research question evaluated here was, *how well can fishing activity be predicted from VMS polls?* Vessel speed has been an accepted criteria for inferring fishing activity from unlabeled data (see Murawski et al., 2005; Palmer and Wigley, 2009; Gerritson and Lordan, 2011). Our analysis focused on assessing whether or not speed-based methods for classification of unknown polls could be improved by incorporating information on bottom depth, bearing, and gear type. We found that regression-based classification methods performed notably better than the naive speed-based criteria for designating VMS polls as fishing or non-fishing (Table 12).

We found that including bottom depth and bearing covariates improved classification accuracy when the analysis is conditioned on gear types. This was true for the three groundfish targeting trawl gears used on over 95% of tows in our sample (large footrope trawl, small footrope trawl, and selective flatfish gear).

Classifying unknown VMS polls without conditioning on gear type was most accurately performed using a generalized additive model. The simple predictive models were estimated primarily to illustrate interesting lines of future research that may be pursued using VMS data along with logbook and fish ticket data. A number of potentially interesting extensions to this simple modeling exercise are worth highlighting. First, more rigorous feature selection and engineering may be employed in order to optimize predictive performance of fishing/not fishing classifiers with West Coast groundfish data. Additionally, as highlighted in Krigsman et al. (2012), the choice of cutoff values in binomial predictive models can significantly affect classification accuracy.

Finally, our analysis does not address the important question of whether VMS units affect the accuracy of self-reported fishing locations. We were unable to test for this potential effect using the available data. Future research in this area may benefit from exploring whether additional data sources could be leveraged to test for differences in reporting accuracy between monitored and unmonitored vessels.

6 Acknowledgements

The authors wish to acknowledge E.J. Dick for helpful comments on an earlier draft of this manuscript. We also acknowledge financial support from the NOAA Office of Science and Technology, Economics and Human Dimensions Program.

7 References

Baker, M.S., M.B. Sciance, and J.N. Halls. 2016. Potential for a simple GPS-based binary logit model to predict fishing effort tin a vertical hook-and-line reef fish fishery. *Marine and Coastal Fisheries*, v.8(1): 118-131.

Bellman, M.A., S. Heppell, and C. Goldfinger. 2005. Evaluation of a US west coast groundfish habitat conservation regulation via analysis of spatial and temporal patterns of trawl fishing effort. *Canadian Journal of Fisheries and Aquatic Sciences* 62:2886-2900.

Bertrand, S., A. Bertrand, R. Guevara-Carrasco, and F. Gerlotto. 2007. Scale-invariant movements of fishermen: the same foraging strategy as natural predators. *Ecological Applications* 17(2):331-337.

Bez, N., E. Walker, D. Gaertner, J. Rivoirard, and P. Gaspar. 2011. Fishing activity of tuna purse seiners estimated from vessel monitoring system (VMS) data. *Canadian Journal of Fisheries and Aquatic Sciences*, v.68: 1998-2010.

Collier, T.C., A. Mamula, and J. Ruggiero. 2014. Estimation of multi-output production functions in commercial fisheries. *Omega* 42(1):157-165.

Demestre, M., A. Muntadas, S. de Juan, C. Mitilineou, P. Sartor, J. Mas, S. Kavadas, and J. Martín. 2015. The need for fine-scale assessment of trawl fishing effort to inform on an ecosystem approach to fisheries: Exploring three data sources in Mediterranean trawling grounds. *Marine Policy* 62:134-143.

Gerritsen, H., and C. Lordan. 2011. Integrating vessel monitoring systems (VMS) data with daily catch data from logbooks to explore the spatial distribution of catch and effort at high resolution. *ICES Journal of Marine Science* 68(1):245-252.

Gerritsen, H., C. Minto, and C. Lordan. 2013. How much of the seabed is impacted by mobile fishing gear? Absolute estimates from Vessel Monitoring System (VMS) point data. *ICES Journal of Marine Science* 70(3):523-531.

Holland, D.S. and J.E. Jannot. 2012. Bycatch risk pools for the US West Coast Groundfish Fishery. *Ecological Economics*, v.78: 132-147.

Jennings, S. and J. Lee. 2012. Defining fishing grounds with vessel monitoring system data. *ICES Journal of Marine Science* 69(1):51-63.

Joo, R. S. Bertrand, A. Chaigneau, and M. Niquen. 2011. Optimization of an artificial neural network for identifying fishing set positions from VMS data: An example from the Peruvian anchovy purse seine fishery. *Ecological Modelling*, v.222(4): 1048-1059.

Krigsman, L.M., M.M. Yoklavich, E.J. Dick, G.R. Cochrane. 2012. Models and maps: predicting the distribution of corals and other benthic macro-invertebrates in shelf habitats. *Ecosphere*, v.3(1).

Lee, J., A.B. South, and S. Jennings. 2010. Developing reliable, repeatable, and accessible methods to provide high-resolution estimates of fishing-effort distributions from vessel monitoring system data. *ICES Journal of Marine Science* 67:1260-1271.

Lee, Y.W., and D.B. Sampson. 2000. Spatial and temporal stability of commercial groundfish assemblages off Oregon and Washington as inferred from Oregon trawl logbooks. *Canadian Journal of Fish and Aquatic Sciences* 57:2443-2454.

Mason, J, R. Kosaka, A. Mamula, and C. Speir. 2012. Effort changes around a marine reserve: The case of the California Rockfish Conservation Area. *Marine Policy* 36(5):1054-1063.

Mills, C.M., S.E. Townsend, S. Jennings, P.D. Eastwood, and C.A. Houghton. 2006. Estimating high resolution trawl fishing effort from satellite-based monitoring system data. *ICES Journal of Marine Science,* v.64(2): 248-255.

Muench, A., G.S. Depiper, and C. Demerest. 2018. On the precision of predicted fishing location using data from the vessel monitoring system (VMS). *Canadian Journal of Fisheries and Aquatic Sciences*, v.75(7): 1036-1047.

Murawski, S.A., S.E. Wigley, M.J. Fogarty, P.J. Rago, and D.G. Mountain. 2005. Effort distribution and catch patterns adjacent to temperate MPAs. *ICES Journal of Marine Science* 62(6):1150-1167.

Murray, L.G., H. Hinz, N. Hold, and M.J. Kaiser. 2013. The effectiveness of using CPUE data derived from vessel monitoring systems and fisheries logbooks to estimate scallop biomass. *ICES Journal of Marine Science* 70(7):1330-1340.

National Marine Fisheries Service (NMFS). 2015. Midwater trawl restrictions and prohibited species retention for the shorebased trawl individual fishing quota program: Draft environmental assessment. Retrieved from: https://archive.fisheries.noaa.gov/wcr/publications/nepa/groundfish/misc_ea/midwatertrawlcleanupea.pdf, 10/2/2019.

Palmer, M.C., and S.E. Wigley. 2009. Using positional data from vessel monitoring systems to validate the logbook-reported areas fished and the stock allocation of commercial fisheries landings. *North American Journal of Fisheries Management* 29(4):928-942.

Patterson, T.A., L. Thomas, C. Wilcox, O. Ovaskainen, J. Matthiopoulos. 2008. State-space models of individual animal movement. *Trends in Ecology and Evolution*, v.23(2): 87-94.

Pearson, D.E., B. Erwin, and M. Key. 2008. Reliability of California's groundfish landing estimates from 1969-2006. NOAA Technical Memorandum, NOAA-TM-NMFS-SWFSC-431.

Pebesma, E., 2018. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, https://journal.r-project.org/archive/2018/RJ-2018-009/

Pikitch, E.K., J.R. Wallace, E.A. Babcock, D.L. Erickson, M. Saelens, and G. Oddsson. 1998. Pacific halibut bycatch in the Washington, Oregon and California groundfish and shrimp trawl fisheries. *North American Journal of Fisheries Management* 18(3):569-586.

Skaar, K.L., T. Jørgensen, B.K.H. Ulvestad, and A. Engås. 2011. Accuracy of VMS data from Norwegian demersal stern trawlers for estimating trawled areas in the Barents Sea. *ICES Journal of Marine Science* 68(8):1615-1620.

Speir, C.S., J.G. Sutinen, and C. Pomeroy. 2014. Port-level fishing dynamics: Assessing changes in the distribution of fishing activity over time. *Marine Policy* 46:171-191.

Valcic, B. 2009. Spatial policy and the behavior of fishermen. *Marine Policy* 33(2):215-222.

Van der Walt, S., J.L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J.D. Warner et al. (2014) scikit-image: Image processing in Python. *PeerJ Bioinformatics and Genomics Section*. 2:e336v2

Walker, E., D. Gaertner, P. Gaspar, and N. Benz. 2010. Fishing activity of tuna purse seiners estimated from VMS data and validated by observers' data. Collective Volume of Scientific Papers ICCAT 65(6):2376-2391.

Watson, J.T., and A.C. Haynie. 2016. Using Vessel Monitoring System data to identify and characterize fishing trips made by fishing vessels in the United States North Pacific. *PLOS One*, DOI: https://doi.org/10.1371/journal.pone.0165173

Wood S. 2017. *Generalized Additive Models: An introduction with R, 2nd edition*. Chapman and Hall/CRC, Boca Raton.

Appendix A. Data tables and descriptions

Our analysis joins precise spatial-temporal data on vessel locations from NOAA Office of Law Enforcement's (OLE) VMS program with existing fishery dependent data from the Pacific Groundfish Fishery. In the interest of clarity and reproducibility, this Appendix provides detail on the data sources used and the procedures used to join information from different database tables.

VMS data was acquired directly from NOAA OLE and database tables were copied in their entirety to a local database. For this analysis we utilized only two tables from the VMS database: a table containing the individual VMS polls and a vessel look-up table mapping the vessel identifiers used by OLE to vessel identifiers used by other agencies. Appendix Table A 1 provides some detail on the naming convention and important data fields from these database tables.

Fishery dependent data used for this analysis comes from the Pacific Fishery Information Network (PacFIN) Database. More specifically, we access data from two PacFIN subsystems: the PacFIN Trawl Logbook Database[29] and the PacFIN Comprehensive Fish Tickets Table[30]. PacFIN uses an Oracle System to manage their databases. In Appendix Table A 1 below we reference each database table used in the analysis by the PacFIN subsystem in which it resides as well as the specific database schema where the table is located.

Appendix Figure A 1and Appendix Figure A 2 illustrate the relational algebra used to join different data tables and create the datasets used in our analysis. Figures should be read from left to right then top to bottom. Appendix Figure A 1 illustrates how the VMS positional data is joined with various data tables from the Trawl Logbook Database to in order to assign fishing trip and fishing tow level characteristics to each VMS poll. Appendix Figure A 2 illustrates how data tables from the Trawl Logbook Database are joined with fish ticket data in order to assign a gear type field to each logbook fishing trip[31].

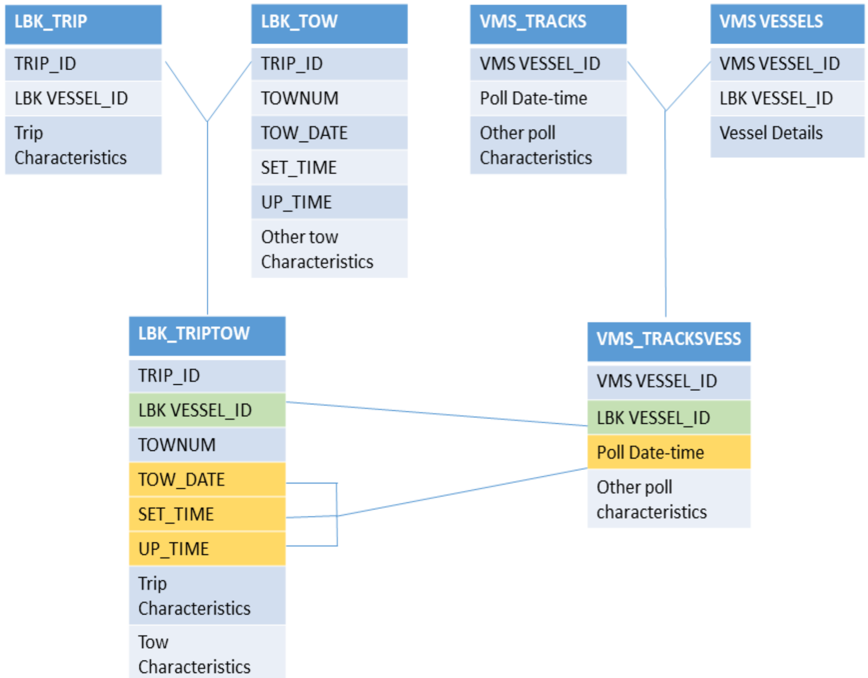
[29]Documentation for this database is available at: https://pacfin.psmfc.org/data/trawl-logbooks/

[30] Further documentation on this database table is available at: https://pacfin.psmfc.org//wp-content/uploads/2016/06/PacFIN_Comprehensive_Fish_Tickets.pdf

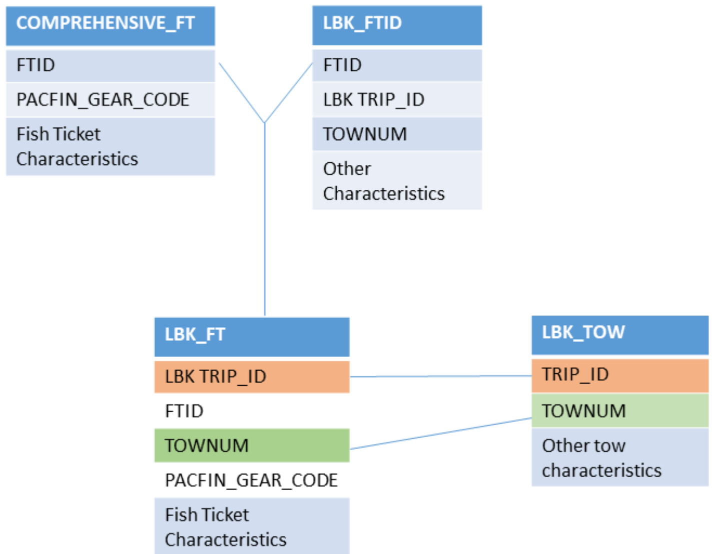
[31] Here we would like to emphasize that the gear type field could also be pulled from the LBK_TOW table in in the Trawl Logbook Database. However, our informal observations suggest that gear information contained in the fish tickets is more complete as well as easier to interpret.

Appendix Table A 1. Data sources and descriptions.

Table Name	Source	Important Fields	Description
LBK_TRIP	PacFIN Coastwide Trawl Logbook Database, “pacfin” schema	Vessel identifier (DRVID), trip identifier (TRIP_ID), trip departure and return dates (DDATE, RDATE)	This table contains important summary information on groundfish fishing trips such as departure and return ports, and departure and return date/ times.
LBK_TOW	PacFIN Coastwide Trawl Logbook Database, “pacfin” schema	Trip identifier (TRIP_ID), haul identifier (TOWNUM), haul start and end dates and times (SET_DATE, UP_DATE, SET_TIME, UP_TIME)	This table contains information on individual tows such as dates and time of each event and a trip identifier so that each tow can be linked back to a fishing trip
COMPREHENSIVE_FT	PacFIN Coastwide Trawl Logbook Database, “pacfin_marts” schema	Vessel identifier (vessel_num), fish ticket identifier (FTID), vessel length (VESSEL_LENGTH), gear identifier (PACFIN_GEAR_COD E)	COMPREHENSIVE_FT is a flattened version of the fish ticket data that contain the pounds and value of all commercial fish landed. The table also contains information on the gear types used to harvest all fish and characteristics of each fishing vessel.
LBK_FTID	PacFIN Coastwide Trawl Logbook Database, “pacfin” schema	Logbook trip identifier (TRIP_ID), logbook tow identifier (TOWNUM), fish ticket identifier (FTID)	This is a look up table that allows a particular fish ticket from the COMPREHENSIVE_FT table to be linked to a unique trip identifier and tow number from the groundfish trawl logbooks
VMS_TRACKS	NMFS office of law enforcement	Vessel identifier (FMC_LOGVESS_ID)	The primary VMS table of interest. The table contains the individual VMS polls including: date/time, vessel, latitude, and longitude
VMS_VESSEL_INFO	NMFS OLE	VMS vessel identifier, (FMC_LOGVESS_ID), coastguard vessel identifier or state vessel identifier (RADIO)	A look-up table that allows vessels in the VMS data to be matched to vessels in the groundfish trawl logbooks.



Appendix Figure A 1. Relational algebra relating VMS polls to logbook fishing vessels, fishing trips, and tows. Boxes along the top row of the figure depict primary database tables, lines indicates common fields used to construct derived data tables which are shown below the primary tables.



Appendix Figure A 2. Relational algebra to add gear types to logbook data. Boxes along the top row of the figure depict primary database tables, lines indicates common fields used to construct derived data tables which are shown below the primary tables.

Appendix B. Matching VMS polls to logbook tows

Our method for calculating distance from a VMS poll to its corresponding logbook tow path involves defining points along the tow path and computing the point-to-point distance between a VMS poll and each point along the tow path. The precision of this approach depends on the number of points defined along the tow path. In Section 3.2.2 we discussed the computational methods used to define points along each tow path. The methods rely on a parameter, n, controlling the total number of points to divide each line into. Appendix Table B 1 illustrates the impact this parameter has on our calculation of distances from VMS polls to logbook two paths. The table shows the gain in accuracy resulting from changes in the parameter n. Gain is calculated by differencing the distances calculated between a VMS poll and its corresponding logbook tow path at different values of the parameter n. For example, define $d_i(20)$ to be the distance between VMS poll i and its assigned logbook tow path using 20 discrete points to define the logbook tow path. The gain in accuracy for VMS poll i, realized by increasing the number of discrete tow path points from 20 to 200, is $d_i(20) - d_i(200)$. The average gain reported in Appendix Table B 1 is computed by calculating the gain at each individual VMS poll and taking the average gain over all VMS polls ($\bar{d}_n = \frac{1}{V}\sum_{i=1}^{V} d_i(n)$). It is important to note that for our application the relationship between the VMS poll to logbook distance and the number of points used to define each tow path was strictly non-increasing ($d_i(n) > d_i(n+k)$ for all i and $k > 0$).

Appendix Table B 1. Average estimated gain in accuracy of distance calculation from increasing the number of sampling point for each tow line.

Tow Path Method	$(n; m)$	Average Gain $(\overline{d_m} - \overline{d_n})$
Bathy	(20;200)	0.145
Bathy	(200;2,000)	0.104
Bathy	(2,000;4,000)	0.0006
Straight	(20;200)	0.071
Straight	(200;2,000)	0.003
Straight	(2,000;4,000)	0.0001

Appendix C. Analysis of logbook activity not matched to any VMS polls

C.1. Summary of logbook data by sectors

Our data included a sample of just over 30,000 tows from groundfish trawl logbooks with trip departure ports in California. These data include fishing trips from a variety of distinct fisheries and sectors that land groundfish either as the targeted species or as incidental catch. The fisheries and sectors represented in our logbook data can be broadly categorized as:

1. Limited entry or IFQ groundfish trawl – this is activity that directly targets groundfish species with trawl gear through participation in the limited entry and IFQ sectors.
2. Open access groundfish trawl – this activity uses trawl gear to harvest an array of groundfish species in the open access fishery
3. California Halibut trawl – this activity targets California Halibut in the state managed California Halibut fishery and lands other incidentally caught groundfish species.
4. Other trawl fisheries landing groundfish – this mainly includes participation in various state-managed shrimp trawl fisheries that land incidentally caught groundfish.

Data collected from the Pacific Fisheries Information Network assigns commercial groundfish landings to sectors defined by Dahl Sector Codes[32]. These codes are assigned to landings based on species harvested, gear used, and permits held by the fisherman. We summarize our logbook data by Dahl Sector Code using the Dahl Code assigned to the plurality of landings from each trip[33]. This data summary is provided in Appendix Table C 1.

[32] Business rules used to assign landings to a Dahl Sector are provided by PacFIN here: https://pacfin.psmfc.org//wp-content/uploads/2015/10/PacFIN_groundfish_sector_codes.pdf

[33] Dahl Codes are assigned to each landing on a fish ticket. Each logbook fishing trip may have several fish tickets. In most cases, all fish tickets from the same trip are assigned to the same Dahl Groundfish Sector. However, it is possible for a fishing trip to contain fish tickets assigned to more than one Dahl Groundfish Code. We aggregate landings on each trip by Dahl Groundfish Code and assign the fishing trip to the code assigned to the plurality of landings.

Appendix Table C 1. Distribution of logbook fishing trips by Dahl Groundfish Code.

Dahl Code	Share of Trips	Description
4	0.45	Shoreside Non-whiting Trawl
13	0.24	Exempted Trawl
15	0.18	Commercial Nongroundfish
NA	0.06	Logbook trip not matching any fish ticket
14	0.05	EFP and Miscellaneous
3	0.005	Shoreside whiting
XX	0.004	Other landing not accounted for
11	0	Non-fixed gear directed open access (shrimp trawl or net)
20	0	Shoreside IFQ non-trawl
7	0	Non-nearshore limited entry (fixed gear)
10	0	Non-nearshore non-sablefish open access (fixed gear)
6	0	Nearshore open access
12	0	Incidental open access
8	0	Non-nearshore open access

C.2. Summary of logbook fishing activity not matching any VMS polls

Our data include approximately 1,300 fishing trips reported in logbook data to which no VMS polls could be matched. These fishing trips matched to approximately 1,500 unique fish tickets. The following tables summarize these unmatched logbook fishing trips.

Appendix Table C 2. Logbook trips not matched to any VMS polls by return port and year.

Port	Port Code	2008	2009	2014	2015
Avila	AVL	21	16	0	26
Fort Bragg	BRG	4	0	0	0
Crescent City	CRS	9	1	1	0
Eureka	ERK	54	10	1	0
Monterey	MNT	2	0	0	0
Moss Landing	MOS	6	0	0	0
Morro Bay	MRO	12	0	0	1
Oxnard	OXN	18	2	4	18
Princeton	PRN	32	16	21	4
Santa Barbara	SB	144	35	15	138
San Francisco	SF	76	54	0	0
San Simeon	SIM	0	3	1	0
Terminal Island	TRM	10	8	0	0
Ventura	VEN	275	130	38	64
Multiple WA Ports	WLB	0	0	0	0
Westport	WPT	4	0	0	0

Appendix Table C 3. Share of logbook trips not matching any VMS polls by Dahl Groundfish Code. Trips were assigned to Dahl Groundfish Code matching the plurality of landed weight from the trip.

Dahl Code	Share of Unmatched Trips	Fishery Description
3	0.01	Shoreside Whiting Sector
4	0.12	Shoreside Nonwhiting Trawl Sector
6	0	Nearshore Sector (Open Access)
13	0.34	Exempted Trawl Sector
14	0.05	EFP and Miscellaneous Sector
15	0.42	Commercial Nongroundfish Sector
NA	0.06	Logbook trips not matching VMS polls and not matching any fish ticket

There are two Dahl Groundfish Codes associated with the shoreside nonwhiting limited entry groundfish fishery: 4 and 20. Dahl Groundfish Code 4 is assigned to shoreside limited entry non-whiting groundfish trawl trips and Dahl Groundfish Code 20 is assigned to shoreside groundfish IFQ trips using fixed gear[34].

[34] These landings are sometimes informally referred to as the 'gear switching' sector.

Appendix Table C 4. Logbook fishing trips assigned to Dahl Groundfish Sectors 4 or 20 not matching any VMS polls.

Year	Matched	Unmatched
2008	752	112
2009	858	28
2014	396	16
2015	306	0

Appendix Table C 5. Top species landed by Dahl Groundfish Code for logbook observations not matching any VMS polls

Dahl Groundfish Code	PacFIN Species Code	Species Lbs	Species Description
3	PWHT	2,331,049	Pacific whiting
	WDOW	10,740	widow rockfish
	COHO	2,122	coho salmon
	CLPR	1,677	chilipepper rockfish
	CHNK	243	Chinook salmon
4	DOVR	1,118,338	Dover sole
	PTRL	263,470	petrale sole
	SABL	234,972	sablefish
	LSPN	215,148	longspine thornyhead
	SSPN	97,686	shortspine thornyhead
6	SCR1	390	California scorpionfish
	CHLB	8	California halibut
	BLUR	4	blue rockfish
	VRML	3	vermillion rockfish
	BLCK	1	black rockfish
13	RPRW	290,478	ridgeback prawn
	CHLB	46,626	California halibut
	UFLT	29,819	unidentified flatfish
	WCRK	24,374	wreckfish
	MSC2	5,570	miscellaneous
14	CHLB	14,676	California halibut
	DOVR	12,719	Dover sole
	SABL	8,183	sablefish
	USCU	5,831	sea cucumber
	LCOD	4,025	lingcod
15	PWHT	136,354	Pacific whiting
	RPRW	120,369	ridgeback prawn
	ALBC	107,856	albacore
	CHLB	802,316	California halibut
	DOVR	36,696	Dover sole

Based on the tables above we make the following observations regarding logbook fishing activity not matching VMS polls:

1. Unmatched logbook tows were more prevalent in 2008 than in later years.
2. Unmatched logbook tows were concentrated in the Santa Barbara and Ventura areas

3. Landings classified as "exempted trawl" and "commercial non groundfish" accounted for a significant number of unmatched logbook observations.
4. Within the "exempted trawl" and "commercial non groundfish" sectors, the most landed species among logbook records not matched to any VMS polls were California Halibut, Ridgeback Prawn, and Albacore.

On the West Coast, VMS is required aboard all vessels holding a limited entry groundfish permit and fishing in state or federal waters[35]. We observe that a large fraction of fishing activity reported in the logbooks that could not be matched to any VMS polls appears to have come from non-limited entry groundfish fisheries (e.g. state-managed California Halibut and ridgeback prawn fisheries). This unmatched activity could be characterized as 'expected.' There is however a non-trivial amount of unmatched logbook fishing activity for which we cannot offer an explanation. Appendix Table C 4 shows that over 150 logbook fishing trips from our logbook data sample categorized as Dahl Sectors 4 or 20 (limited entry groundfish fisheries) could not be matched to any VMS polls.

[35] https://www.fisheries.noaa.gov/national/enforcement/regional-vessel-monitoring-information

Appendix D: Distance outliers

Our analysis produced a small number of relatively large distances between VMS polls and corresponding logbook fishing locations. Here we summarize the extreme distance values along several margins and we provide some one-off examples illustrating cases where large spatial inconsistencies between VMS polls and logbooks were found.

We start with an analysis of VMS polls that were more than 8 km from their corresponding logbook tow path. This includes the 95th percentile of the distribution of VMS polls-to-logbook-lines for the straight line tow paths. A summary of these points is provided in Appendix Table D 1.

Appendix Table D 1. Summary of distance outliers by year for the VMS-poll-to-logbook-straight-line distances.

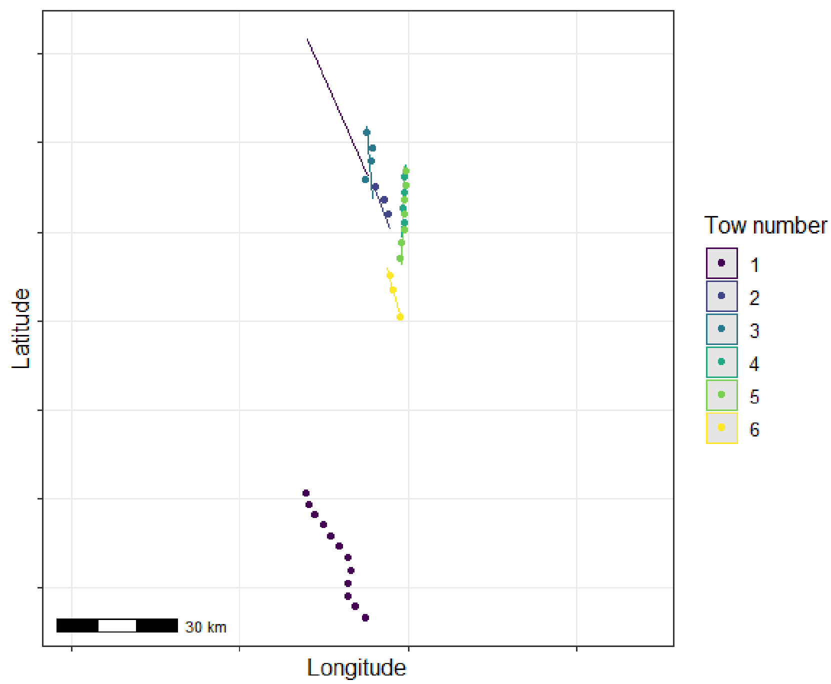
Year	Outliers	Total VMS Polls
2008	819	52,740
2009	1,582	76,512
2014	939	60,704
2015	792	62,699

Appendix Table D 2. Spatial distribution of extreme distance values between VMS and logbook locations.

Latitude Strata	VMS Polls	Outliers	Share of Total VMS Polls	Share of Total Outliers
1	88,944	1,294	0.35	0.31
2	46,572	471	0.18	0.11
3	41,007	572	0.16	0.14
4	75,383	1,793	0.30	0.44
5	581	4	0.00	0.00

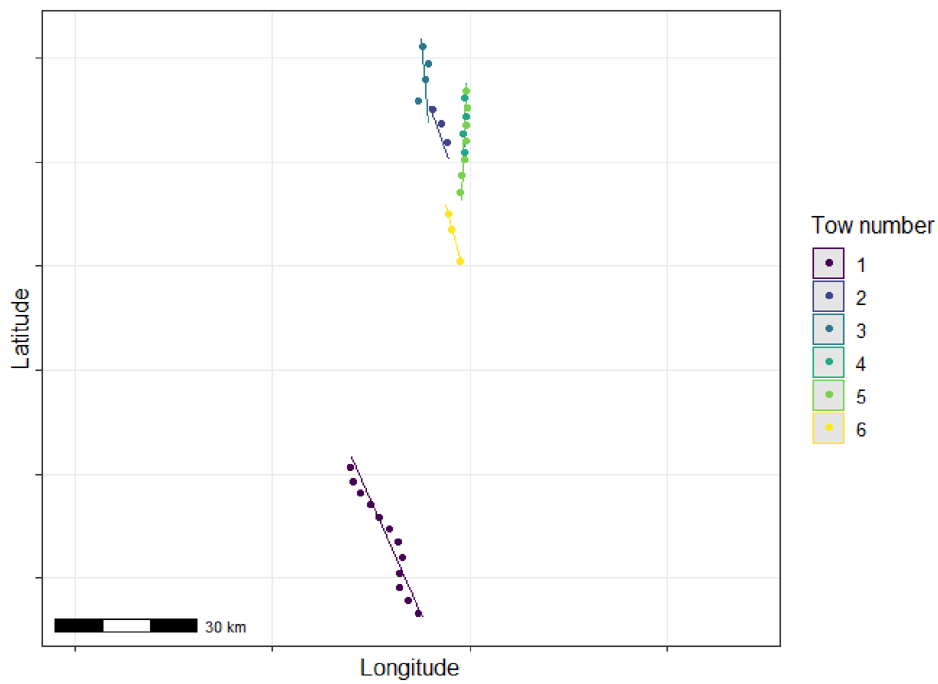
In the remainder of this section we illustrate three cases where large spatial inconsistencies between VMS polls and logbook tows were found. These maps help illustrate some of the sources of error in matching VMS polls to logbook tows. We have organized the illustrations below according to our subjective beliefs about the most likely source of the spatial inconsistency between logbook and VMS data. Appendix Figure D 1, Appendix Figure D 2, Appendix Figure D 3, and Appendix Figure D 4 illustrate cases where we believe the most likely source of spatial inconsistency to be limitations of the logbook data. Appendix Figure D 5 illustrates a case where we believe the most likely cause of large spatial disagreement between VMS and logbook data involves limitations of the VMS data.

In Appendix Figure D 1 we show a fishing trip where VMS-logbook distances are relatively small with the exception of a single tow on the trip. For tow number 1 on the trip illustrated below, VMS-logbook distances range from 83 to 111 km. For tows 2 – 6 distances are between 0.06 and 2.5 km.



Appendix Figure D 1. Example of a logbook tow location possibly incorrectly recorded. Colored lines mark the straight line tow path from logbook coordinates. Dots mark the VMS polls assigned to each logbook tow based on time stamps.

Tow number 1 (long tow line located towards the northwest corner of the plot window) does not run through any VMS polls. In contrast, each of tows 2 – 6 for this trip pass almost directly through multiple VMS polls. Additionally, there is a cluster of VMS polls towards the southwest corner of the plot window that are associated with a vessel speed commonly believed to be indicative of groundfish trawling. In this case, a likely explanation for the large spatial disagreement between logbook and VMS for tow number 1 is human error in reporting the location of the tow. In fact, if we manually alter the logbook coordinates by moving only the location of tow number 1 south by exactly 1 degree we get the map in Appendix Figure D 2.

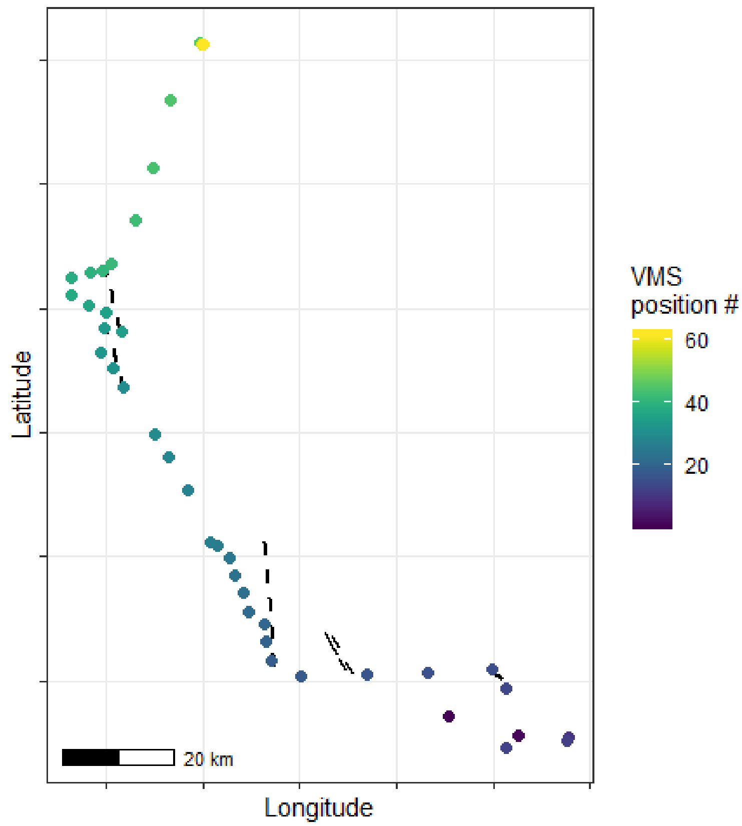


Appendix Figure D 2. A potentially erroneous logbook entry with manually altered tow coordinates for tow #1.

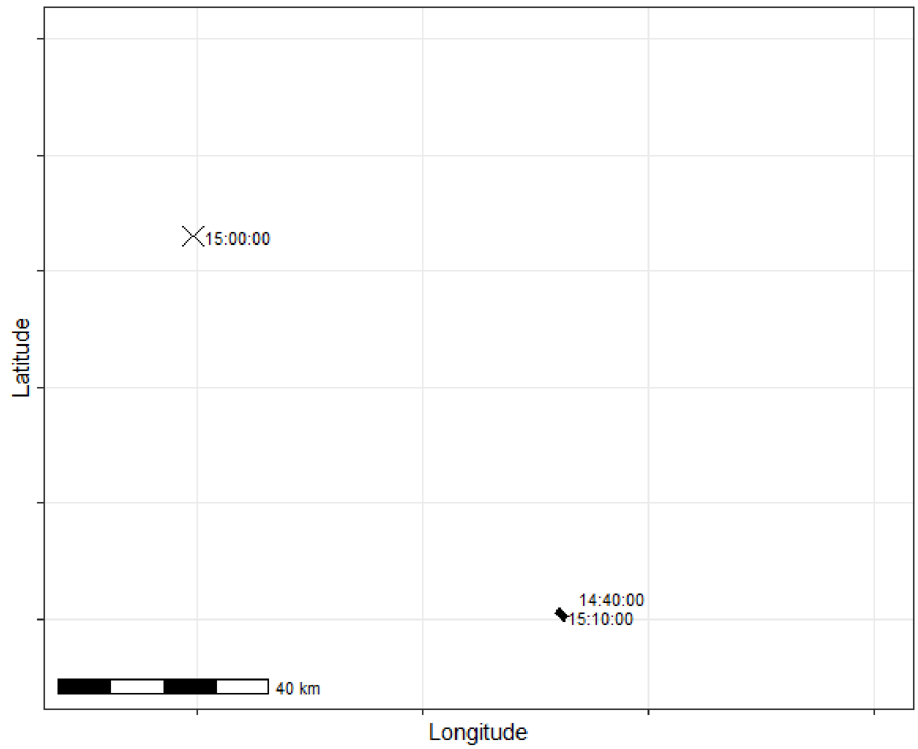
In Appendix Figure D 3 and Appendix Figure D 4 we provide an illustration of a logbook reported fishing trip where VMS and logbook data have relatively large spatial disagreements. Appendix Figure D 3 shows all reported tows for this trip and all VMS polls between the trip start and trip end. From a visual inspection of these data it appears that VMS polls place the vessel very close to the self-reported logbook locations during this fishing trip. However, the average calculated distance from VMS polls to matched logbook tow lines (using the straight line tow paths) is 58 kilometers. This fishing trip includes 6 tows that matched to 62 VMS polls using the matching algorithm outlined in Section 3. Each tow on this trip has between 1 and 5 corresponding VMS polls with a minimum VMS-to-logbook-straight line distance of 26.7 kilometers and a maximum distance of 100.1 kilometers.

In Appendix Figure D 4 we show a single tow from this fishing trip with its matched VMS data. This was a relatively short tow of 30 minutes in duration. The closest VMS poll, matched based on the VMS timestamp and the logbook reported starting and ending times, is over 100 kilometers away. Interestingly, if distance were calculated from VMS polls in Appendix Figure D 3 to the closest tow lines irrespective of reported tow set and tow end times, the VMS-to-logbook distance in Appendix Figure D 4 would be less than 2 kilometers. Moreover, if each tow in Appendix Figure D 3 were matched to the closest VMS poll for the same vessel and day, the average VMS-to-logbook distance for the fishing trip would be less than 3 kilometers.

To summarize, Appendix Figure D 3 and Appendix Figure D 4 illustrate an edge case where VMS data and logbook data appear to provide visually consistent spatial information about the location of fishing activity. Despite this apparent consistency, implementing the VMS poll-to-logbook tow matching algorithm from Section 3 results in relatively large calculated distance values between logbook tow paths and matched VMS polls. While there are many possible explanations for the phenomenon illustrated with Appendix Figure D 3 and Appendix Figure D 4, a likely candidate is misreported logbook tow times. As evidence of this claim, we reiterate that the average VMS-to-logbook distance for this fishing trip is less than 3 kilometers when VMS-to-logbook distance is calculated without relying on logbook reported fishing times.



Appendix Figure D 3. Example of a fishing trip with a potentially misreported tow. Points mark VMS polls and dashed lines are straight line tracks for tows on the fishing trip.

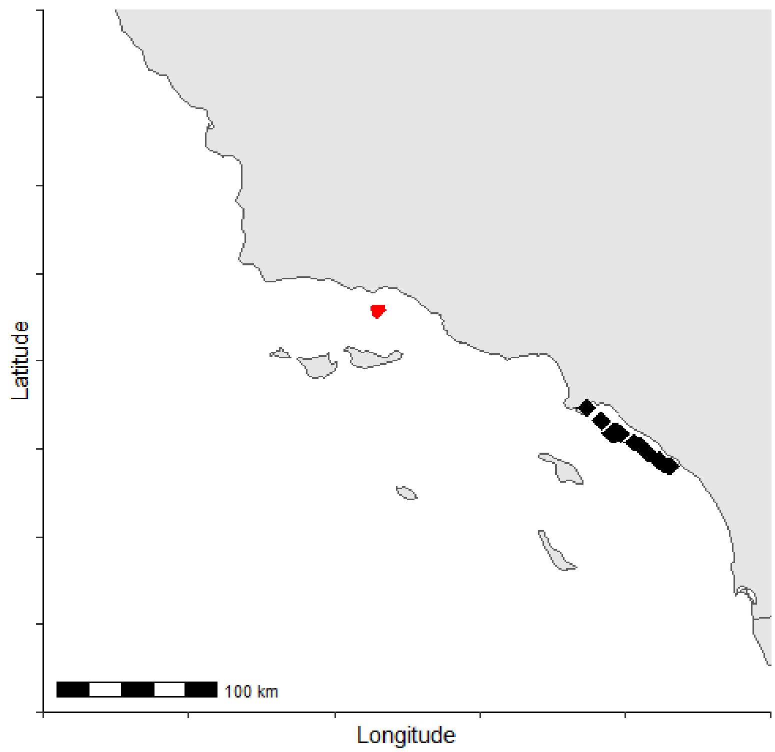


Appendix Figure D 4. Example of a tow with an extreme distance value possible caused by inaccurate logbook data. The tow is shown as a very short line (lower right portion of the plot window) with annotated starting and ending times. The VMS poll matched to this tow is shown as an "X" with annotated timestamp. The date for both the VMS poll and logbook set and up coordinates is the same but has been removed from the figure to protect confidentiality.

Some instances of large spatial disagreements between logbook coordinates and VMS positions appear attributable to limitations of the VMS data. In Appendix Figure D 5 we illustrate a logbook fishing trip that reported tows on a particular day all in the general vicinity of Santa Barbara, CA. VMS polling for this vessel on this day place the vessel in an area considerably further south. Logbook data indicate the vessel was fishing approximately 150 km north of the VMS logged locations.

To investigate whether this large disagreement might have resulted from a vessel captain inadvertently recording the wrong day on the logbook, we examined VMS polls for this vessel within a window beginning 13 hours before the first recorded tow of the trip and ending 2.5 days after the last recorded tow. During this time window, there were no VMS polls within 100km of the logbook reported fishing activity.

The simplest, and in our assessment most likely, explanation for the large spatial disagreement between VMS and logbook data in this case is that the fisher's VMS unit was somewhere other than the fishing vessel. This particular fishing trip appeared to be targeting ridgeback prawn in state waters. This fishing activity generally would not be required to maintain VMS coverage. We present this case in order to illustrate that some care must be taken when matching VMS polls to logbook reported fishing activity. Even in cases (such as non-limited entry groundfish fishing) where one might not expect logbook data to match to VMS data, they sometimes do. However, the matched data can present a confusing picture of vessel activity if the full context of the data is not preserved.



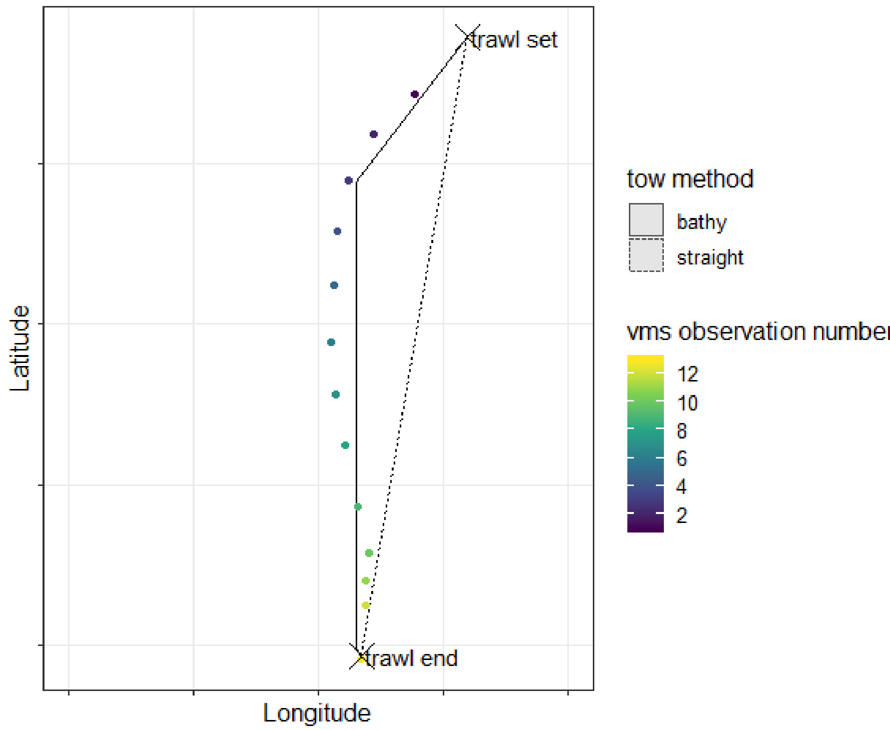
Appendix Figure D 5. VMS (black) and logbook locations (red) for a fishing trip with large calculated distances between logbook reported fishing locations and matched VMS polls.

Appendix E. Illustration of bathymetry-derived versus straight-line tow paths

In Section 4.2 we summarized the differences in logbook-VMS spatial agreement resulting from using a bathymetry-based method for approximating fishing paths relative to a straight line approximation. Overall, we found that straight line tow paths tend to produce better spatial agreement between logbook and VMS fishing locations. However, when evaluated on a tow-by-tow basis, there were interesting variations in how well bathymetry-derived tow paths and straight line tow paths fit to VMS fishing polls. In this Appendix we illustrate some of these interesting cases. Here we present illustrations of four observed cases:

1. Bathymetry lines fit the general shape of VMS polls better than straight lines
2. VMS polls indicate some curvature but the straight line tow path provides a better fit
3. VMS fishing polls are relatively straight
4. VMS fishing polls are curves but both bathymetry-based and straight line interpolated tow paths provide a visually inconsistent fit to VMS polls.

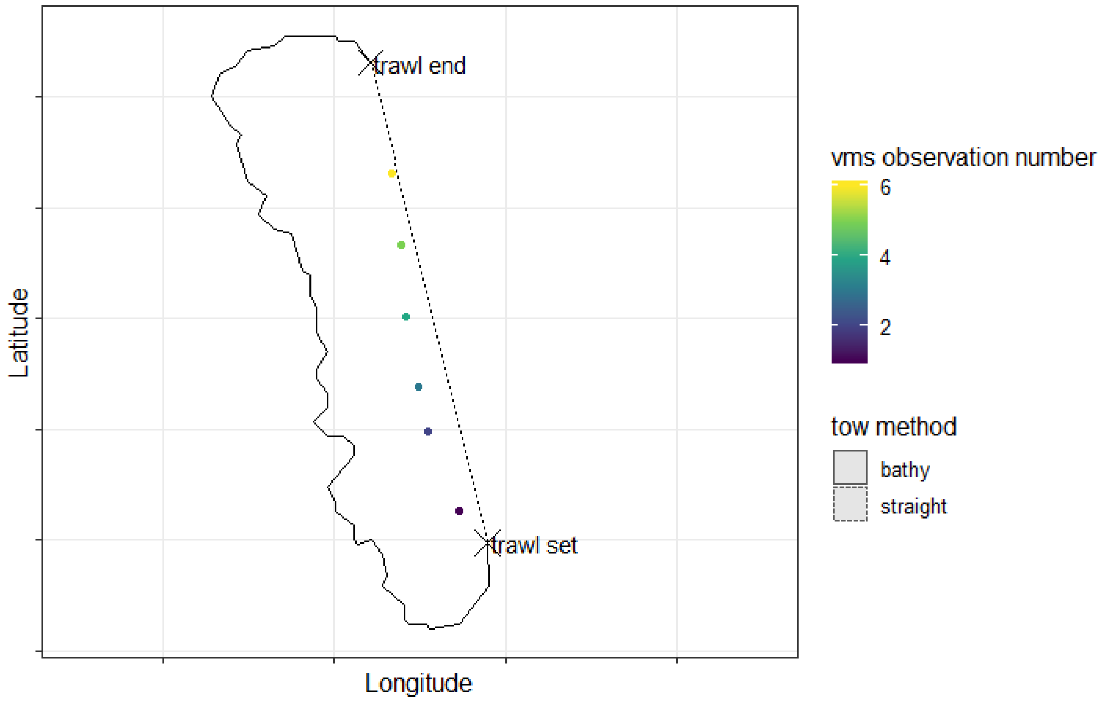
First, we examine a case where the bathymetry-based tow path interpolation appears to fit the shape of VMS fishing polls better than the straight line tow paths.



Appendix Figure E 1. Example of VMS fishing polls with bathymetry tow path roughly matching the shape of the VMS polls.

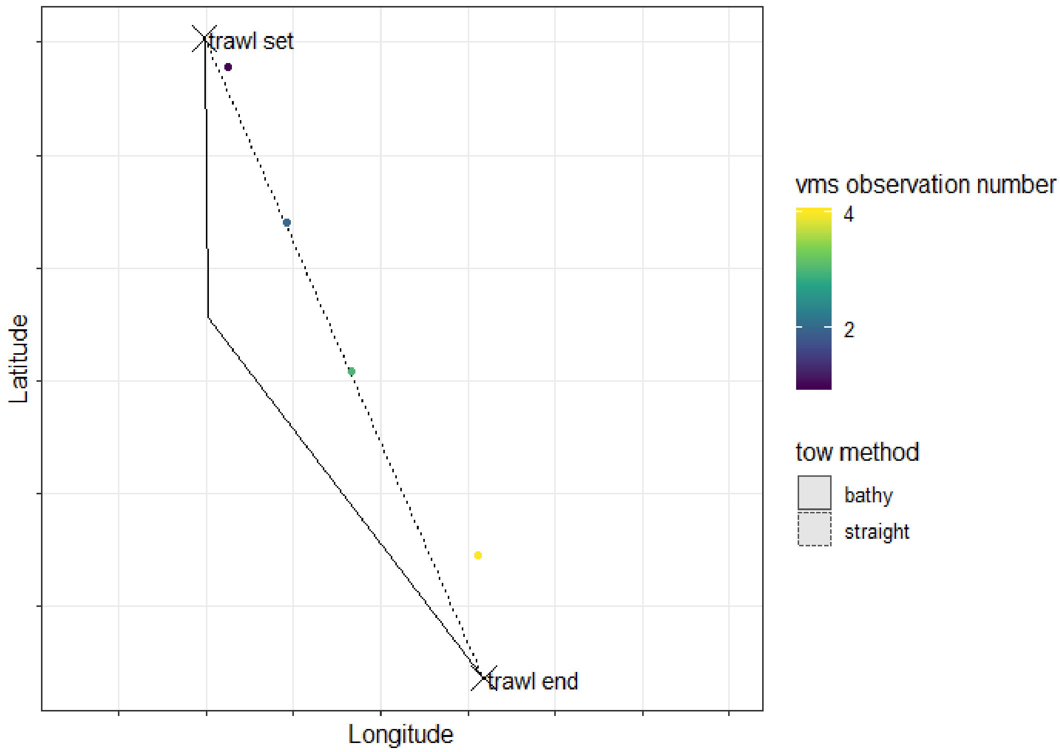
Additionally, Appendix Figure D 3 also illustrates a case where the bathymetry derived tow path matches the pattern of VMS fishing points more closely than the straight line tow path.

Next we illustrate a case where VMS fishing polls do not appear to be located on a straight line but the straight line tow path provides a better visual fit to the VMS polls than the bathymetry path.



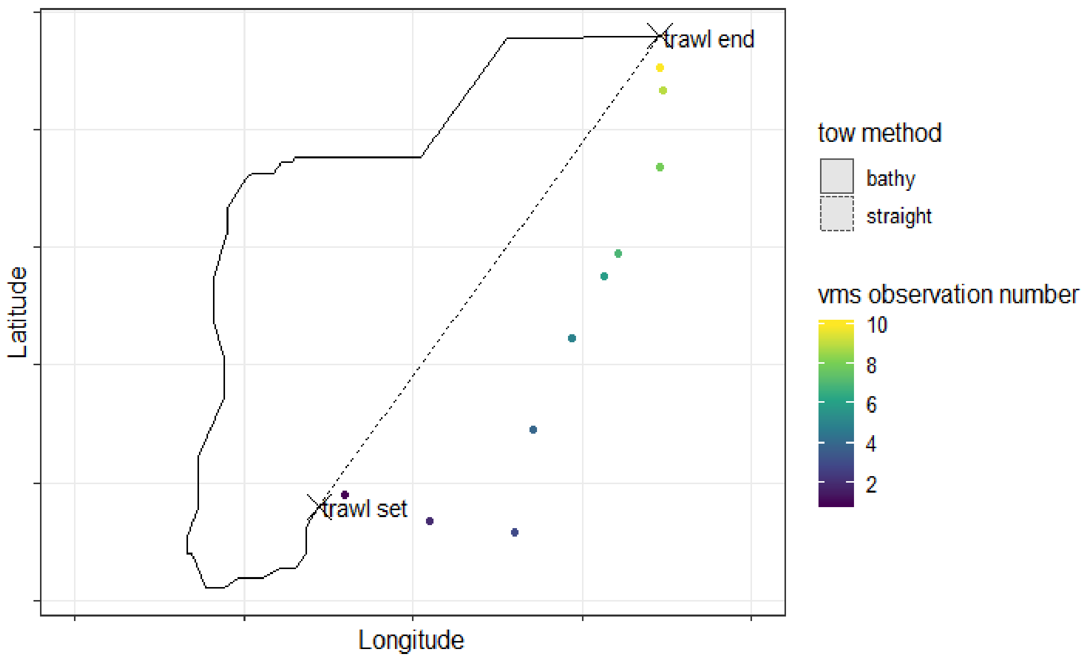
Appendix Figure E 2. Example of VMS fishing polls exhibiting some curvature but for which the straight line tow path provides a better fit than bathymetry-derived path.

In Appendix Figure E 3 we illustrate a case where VMS fishing polls occur on an approximately straight line. In this case the straight line tow path provides a better visual fit to the VMS fishing polls than the bathymetry-derived tow path.



Appendix Figure E 3. Example of VMS fishing polls located on an approximately straight line.

Finally, there are some cases in our data where both the bathymetry and straight line interpolation methods result in a tow path providing a poor visual fit to the matched VMS data. One such case is illustrated in Appendix Figure E 4.



Appendix Figure E 4. Example of VMS fishing polls poorly fit by both bathymetry and straight line tow paths.