Canadian Journal of
Fisheries and
Aquatic Sciences

Article

Linking real world fisheries datasets for mapping of revenue from fishing grounds to dependent communities

Shay O'Farrell[a], **Larry Perruso**[b], **James N. Sanchirico**[c], and **Iliana Chollett**[a]

[a]Sea Cottage, Louisburgh, Co. Mayo, Ireland; [b]Southeast Fisheries Science Center, NOAA Fisheries, Miami, FL, USA; [c]Department of Environmental Science and Policy, University of California, Davis, CA, USA

Corresponding author: **Iliana Chollett** (email: iliana.chollett@gmail.com)

Abstract

Mapping the economic value of the ocean is pivotal to understand how marine ecosystems contribute to human well-being and to support fisheries management. We present a framework to analyse fisheries data and map fishing revenues by linking Vessel Monitoring System (VMS) information to logbooks and observer data. We provide a detailed step-by-step methodology and describe different approaches available to fulfill each step, with special notes for the processing of real-world messy data. The framework consists of six processing steps: (1) identifying the target fishery and subsetting VMS data, (2) extracting relevant variables, (3) linking observer and VMS data, (4) identifying fishing activity, (5) linking VMS and logbook data, and (6) extracting derived variables and mapping revenue back to the communities that extracted the resources. Building this framework opens a broad range of applications including marine spatial planning, rapid response analyses, high-resolution stock assessments, and spatially explicit-socioeconomic analyses. We demonstrate the framework in the reef fishery of the Gulf of Mexico, where spatial planning for aquaculture is currently underway.

Key words: Gulf of Mexico, fisheries management, logbooks, marine spatial planning, USA, vessel monitoring system

Introduction

Understanding the economic value of the ocean is critical for effective fisheries management and implementing marine spatial planning. Vessel monitoring systems (VMS) provide positional information for individual fishing vessels. Using this information, it is possible to determine where a vessel is fishing, and to estimate fishing effort. Positional data can also be linked to logbook catch and price information to map catch and revenue in space (Bastardie et al. 2010; Gerritsen and Lordan 2011).

When revenue, effort, and catch information are all available, revenue measures are generally preferred as a metric for depicting the relative importance of an area for fishing, because effort measures cannot gauge how effective vessels and gears are at capturing fish, and catch measures fail to capture the difference in value across species (Jin et al. 2013). In any case, the variable used should match closely the management objectives to ensure the ecological and/or social sustainability of the resulting plan (Chollett et al. 2022). For example, when planning for offshore wind areas, Bastardie et al. (2015) found fishing effort to be the most relevant variable to use to ensure social sustainability. In their study, the fishery was represented by vessels of very different size, and some geographic areas with low total revenue, not visited by large vessels, were very important for small-scale fishing activities and had low associated revenues but high fishing effort (Bastardie et al. 2015). Revenue information provides important insight into fisheries economics, but represents only one aspect of the activity. If available, a more comprehensive measure of economic value should be used, which considers both revenue earn from the activity and also its costs (e.g., fuel, time). This information is, however, rarely available and has been rarely used (see Bastardie et al. 2013 for an exception).

Integrating VMS data and logbooks has provided better input information for spatial stock assessments, marine spatial planning processes, and spatially explicit socioeconomic analyses (Booth 2000; Gilliland and Laffoley 2008). Linking these datasets has underpinned mapping of catch-per-unit-effort and estimating fish density (Afonso-Dias et al. 2002), mapping abundance (Ducharme-Barth et al. 2018), quantifying misreporting (Palmer and Wigley 2009), and trade-offs between landings value, habitat sensitivity, and fishing impacts to inform marine spatial planning (Jennings et al. 2012), among other uses.

A review of the literature on mapping fisheries for spatial planning found that almost half of published studies quantified fishing effort, whereas only 13% mapped revenue (Chollett et al. 2022). From those 13%, only half used VMS data to spatially locate fishing sites and map revenue. The paucity of studies mapping fisheries revenue using VMS data is in part related to the fact that linking VMS and logbook data is a complicated process. VMS data and logbooks are collected at different temporal scales and each dataset is subjected to different types of error (such as measurement error, unintentional, or intentional reporting error), which generates challenges when linking the information (Gerritsen and Lordan 2011). In particular, self-reported datasets such as logbooks or observer data are subject to human interpretation and manipulation when recorded, and are generally questioned for their accuracy (McCluskey and Lewison 2008).

Some researchers have linked VMS positions to logbook data to map the spatial distribution of catch or revenue. Several studies linked the datasets using fisheries statistical areas as reported in the logbook data by assigning catch and revenue to all VMS pings that fall within the reported statistical area (Pedersen et al. 2009; Jennings et al. 2012). This approach has shown to be inaccurate because the locations reported in the logbook are many times incorrect (Gerritsen and Lordan 2011). Additionally, while more than half of fishing operations cover more than one statistical area, 98% of fishers tend to report only one statistical area per trip, considerably underestimating the spatial footprint of fishing activity (Gerritsen and Lordan 2011). Other researchers have distributed catch or revenue evenly across VMS positions identified as fishing, using either trip-level landings (Afonso-Dias et al. 2002), day-level landings (Gerritsen and Lordan 2011), or a mixed approach that considers both statistical areas and trip-level landings (Bastardie et al. 2010; Hintzen et al. 2012). All of these studies have been carried out in Europe, partly because the European Union uses standardized data formats, which has expedited analyses and allowed the development of specialist software (Hintzen et al. 2012; Russo et al. 2014).

There are very few examples of linking VMS and logbook data to map catch or revenues in the USA, where management is atomized and data formats differ among regions and fisheries. Watson and Haynie (2016) identified and characterized fishing trips made by vessels targeting walleye pollock in the Eastern Bering Sea by linking VMS and fish tickets collected by fish processors. Ducharme-Barth et al. (2018) mapped species abundance matching VMS and trip-level logbook data from the vertical-line reef fisheries in the Gulf of Mexico. Berenshtein et al. (2019) matched trip data for the bottom longline and bandit-reel fisheries targeting reef fish in the Gulf of Mexico and calculated revenue forgone due to spatial closures. The first study provides a detailed methodology for the matching of VMS and statistical fishing areas recorded in the fish tickets (Watson and Haynie 2016), but the last two studies, as is commonly the case, do not provide details on the matching methodology.

Processing VMS and associated datasets is a complex task that requires a thorough understanding of the fishery and datasets in question, as well as the ability to solve analytical problems. Although methodological papers offering a general overview for the analysis of VMS data are available (Hintzen et al. 2012; Russo et al. 2014), most articles using VMS data to solve particular research questions fall short when describing the methodology used. Methodological details, are, however, useful to the scientific community to help deal with problems as they arise when analysing VMS data. There are persistent challenges associated to different fisheries that limit the transferability of methodologies. For instance, "real world" fisheries with particularly short fishing trips, those by multiple use vessels, and/or those without defined ports might require specific methods.

Here, we provide a framework to map revenue for any given fishery. The methodology allows linking imperfect VMS records to logbook and observer data and calculating maps of revenue, a useful spatial predictor of fisheries performance. It furthermore allows for mapping revenue back to ports, which can be useful for fisheries management and spatial decision-making (e.g., Russo et al. 2018). The methodology is explicit, considers alternative approaches to resolve each methodological step, and makes emphasis on techniques for the analysis of messy data that are frequently incomplete, with errors and typos, and inconsistent. We showcase the methodology in the reef fishery of the Gulf of Mexico using datasets curated by the National Oceanic and Atmospheric Administration (NOAA), and demonstrate the potential use of the maps for aquaculture planning, although the data produced could also be used in natural accounts and derived decision support tools. Building this methodological framework opens a broad range of applications including marine spatial planning for wind energy and aquaculture, rapid response analyses for oil spills, red tides or phosphate leaks, analysis of marine dead zones, high-resolution stock assessments, and spatially explicit socioeconomic analyses.

Materials and methods

Three datasets are used to map and quantify revenues and their connection to dependent communities, namely VMS data, on-board observer data, and vessel logbooks or fishery tickets (Fig. 1). The analysis consists of six processing steps: (1) identifying the target fishery and subsetting VMS data, (2) extracting variables, (3) linking observer and VMS data, (4) identifying fishing activity, (5) linking VMS and logbook data, and (6) extracting derived variables and mapping revenue flows back to the communities that extracted the resource (Fig. 1). These steps are discussed in detail below.

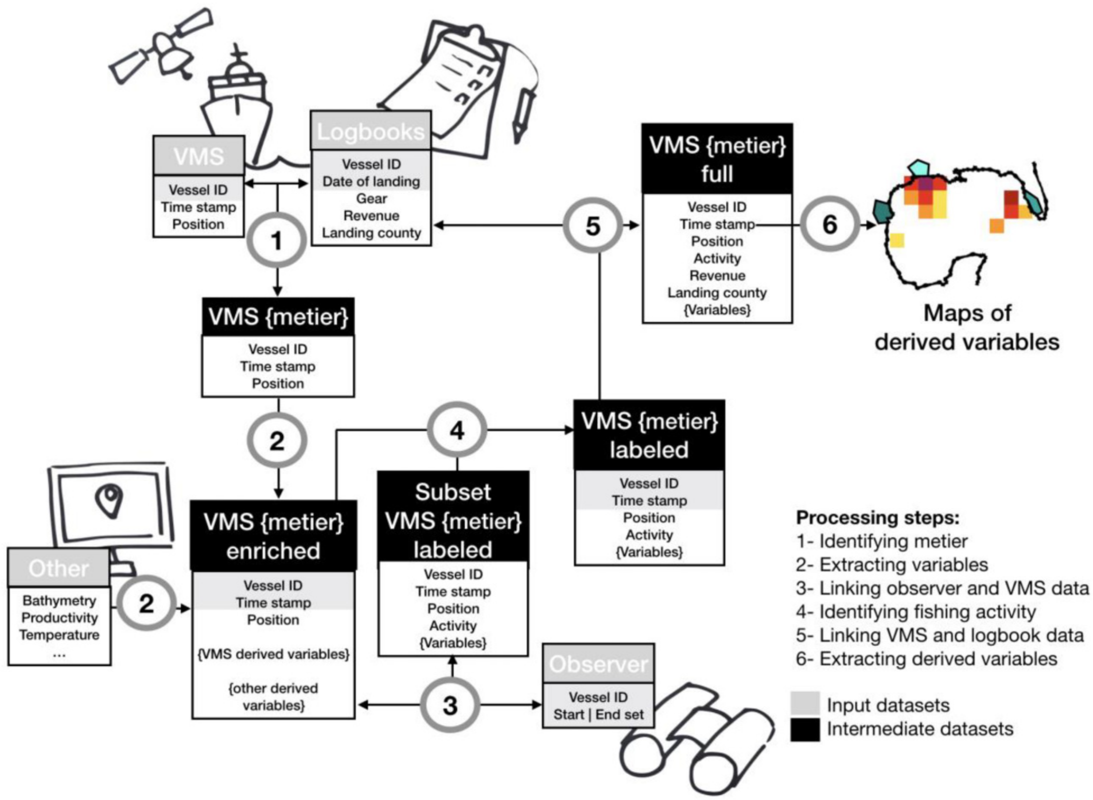
Identifying the target fishery: subsetting VMS data

Vessel behaviour is sensitive to the type of fishing carried out: for example, a vessel setting lobster pots moves differently than a vessel trawling. Therefore, analyses of fishing behaviour based on positional data need to be carried out within a metier (a group of fishing operations targeting a specific assemblage of species using a specific gear) or any relevant classification of fishing operations based on relevant management groups. VMS data do not contain information on gear used. However, this information can be obtained by doing a first link to the logbook data using the vessel identifier, a unique number that characterizes each vessel. To that end, vessel identifiers associated to the target gear of fish species can be extracted from logbook data and used to subset VMS data to include only relevant vessels (Fig. 1, step 1).

Extracting variables

VMS data ought to be preprocessed to remove empty data records, records without vessel identifier, position or time

Fig. 1. Methodological framework to identify fishing locations and derived fisheries performance metrics from VMS, logbook, and observer data for a given target fishery.



stamp, and outliers or duplicated data. Character data should be transformed into the relevant data type (e.g., numeric or date). To decrease the size of the files and improve processing speed, it is also advisable to delete unused variables and/or delete records outside the target region of study.

After data preprocessing, variables that could aid in the identification of fishing activity can be extracted (see step 4, "Identifying fishing activity"). Variables can help understand different attributes of the locations, related to vessel movement patterns, fishing practice, gear or management restrictions, and resource distribution or abundance (Table 1). Common metrics include variables derived from the VMS dataset. Complementary variables can also be extracted from external databases, such as depth, or variables specific to the targeted species (Fig. 1, step 2). For example, Scales et al. (2017) found zonal wind speed, sea surface temperature, sea surface height anomalies, and isothermal layer depth good descriptors of swordfish catches. Although depth has been incorporated into classifiers using VMS data to identify fishing activity (e.g., Muench et al. 2018; O'Farrell et al. 2017; Watson et al. 2018), including other external variables has not been routinely done when using VMS to identify fishing locations.

After extracting relevant variables, obvious outliers in the data can be deleted to decrease the size of the dataset and improve computational speed, such as vessels traveling at impossible speeds or fishing locations that appear to be on land.

Linking observer and VMS data

VMS data linked to observer data and labelled as "fishing" and "not fishing" can be used to create a training and validation dataset used as input to any supervised classification methodology to identify fishing activity (step 4), or merely to assess the accuracy of an unsupervised classification. Observer data ought to be preprocessed to remove empty data records, records without vessel identifier, position or time stamp, outliers or duplicated data. Character data should be transformed into the relevant data type (e.g., numeric or date) and format to match the VMS data.

An important issue when handling spatiotemporal data in large regions is the presence of multiple time zones. Generally, the time zone of the observer's timing device (watch, phone, and GPS or ship's clock) is not recorded in the dataset, adding uncertainty to the accuracy of the reported time and the matching process. For example, although a GPS unit will automatically update the local time to reflect changes in the current time zone, a wristwatch will not, and a mobile phone may or may not depending on whether it is within signal range of a cell network. In these situations, an initial assumption on time zone needs to be made to operationalize matching. For example, assuming the observer's time zone is that of the trip origin, and then transform time to UTC to match VMS data format. This problem could be resolved systemically within observer programs by requiring times to be recorded in UTC, as VMS systems do.

Table 1. Description of variables that have been used in the literature to predict fishing activity for each VMS record.

Variable	Description	Attribute measured	Reference
Leg distance	Distance from the preceding GPS position	Movement pattern	Muench et al. (2018), O'Farrell et al. (2017), and Watson et al. (2018)
Compass heading	Direction toward true north in degrees	Movement pattern	Joo et al. (2013) and O'Farrell et al. (2017)
Travel speed	Vessel speed between the current and previous position	Movement pattern	Bastardie et al. (2010), Joo et al. (2013), Muench et al. (2018), O'Farrell et al. (2017), and Watson et al. (2018)
Turning angles	Magnitude of change in angle	Movement pattern	Bastardie et al. (2010), Joo et al. (2013), Muench et al. (2018), and O'Farrell et al. (2017)
Coordinates	Latitude and longitude	Resource distribution	Watson et al. (2018)
Position weight	Relative position within a trip, assumes fishing happens in the middle	Fishing practice	Muench et al. (2018)
Distance	Distance to the nearest coast	Fishing practice	Watson et al. (2018)
Time	Time of the day	Fishing practice, resource abundance	Muench et al. (2018), O'Farrell et al. (2017), and Watson et al. (2018)
Year	Year in record	Management restrictions	Muench et al. (2018) and Watson et al. (2018)
Month	Month of the year	Resource abundance, management restrictions	Muench et al. (2018), O'Farrell et al. 2017, and Watson et al. (2018)
Lunar phase	State of the moon	Resource abundance	Muench et al. (2018)
Depth	Distance to the seabottom	Resource distribution, gear restrictions	Muench et al. (2018), O'Farrell et al. (2017), and Watson et al. (2018)

Note: All variables but depth can be obtained from VMS data. The table is meant to be illustrative and not fully comprehensive.

To match the observer data to the VMS data, we first extract VMS data within the start and end dates of each observer trip. An observer dataset generally contains a separate entry for each set, each containing a trip number, vessel ID, and start and end date of the trip. A common problem in observer data is the recording of varying start and end dates within the same trip because of typos. This issue can be solved by using the mode of all records of a specific variable within a trip, assuming the most commonly recorded value for a given trip is correct.

The most common approach to labelling VMS pings using observer data is a point-labelling method, in which each VMS record, or *ping,* is labelled either as "fishing" or "not fishing" if an observer has recorded that gears were or were not deployed at the time. This method has shown to be inaccurate for short-set fisheries when fishing sets fall between pings and/or vessels are engaged in fishing-related behaviour (e.g., gear sorting) before and after the gear is in the water (O'Farrell et al. 2017, their Fig. 1). In those cases, a window-labelling method is more appropriate, and VMS pings can be labelled using a window of half the interval between pings around the observer start and end time of the set (O'Farrell et al. 2017).

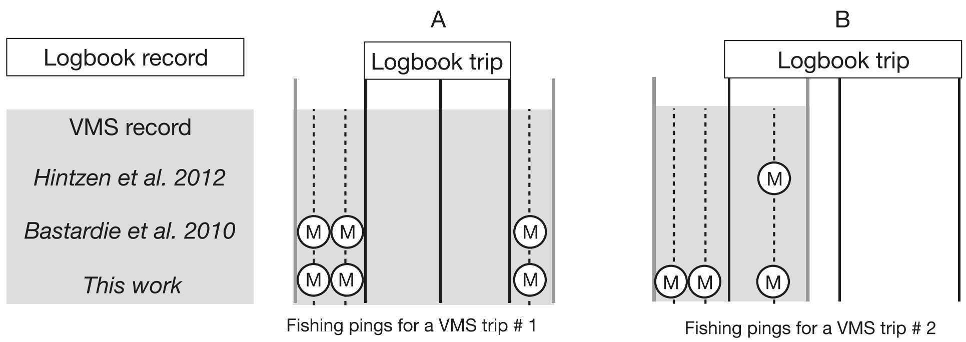
Identifying fishing activity

Vessels that are actively fishing have characteristic movement patterns that can be used to separate fishing tracks into fishing, resting, and/or steaming (Bastardie et al. 2010; Fig. 1, step 4). Many different methods have been used to separate fishing activity from other movement patterns. The simplest way of doing so is using speed filters, assuming, for example, that fishing occurs within a particular speed range (e.g., Hintzen et al. 2012; Gerritsen et al. 2013). Other unsupervised methods include classification using a data mining approach (e.g., de Souza et al. 2016). Another type of classification methods involve supervised algorithms that use on-board observer data as ground truthing for a variety of training methods, such as artificial neural networks (e.g., Russo et al. 2011), hidden Markov models (e.g., Joo et al. 2013), random forests (e.g., O'Farrell et al. 2017), generalized additive models (e.g., Watson et al. 2018), and generalized linear models (e.g., Muench et al. 2018). In general, supervised methods have shown considerable improvement over unsupervised methods (e.g., Bertrand et al. 2008; Muench et al. 2018), but the best method is dependent on the dataset in question.

Linking VMS and logbook data

Logbook data ought to be preprocessed to remove empty data records, exact duplicates or records with missing vessel identifier, landings data, etc. Logbook data generally collect trip-level information for every fishing trip made. In Europe, logbooks provide information on the catch and value of the species caught, date, statistical area fished and metier, among other variables (Bastardie et al. 2010). In the US, different fisheries and management areas collect information differently, but mandatory reporting requirements generally include gear, ex-vessel landings (catch, generally in pounds), value (revenue in USD), discards, and port or county of origin. Revenue is usually calculated as the price per pound at first purchase of the commercial landings multiplied by the total pounds landed. If logbook data are not available, other fishery records such as seafood dealer's or data from regional

Fig. 2. Comparison of methods for matching logbooks and VMS data. Logbook-trip time-windows are indicated with white boxes and VMS-trip time-windows with grey boxes, with chronological sequence along the *x*-axis. VMS pings are marked with vertical dotted lines. Two hypothetical scenarios are shown, representing common data quality issues: when there are (A) no pings or (B) few pings falling within the reported logbook date range. Relevant times for matching are marked with vertical solid lines, namely: (left to right) beginning of VMS trip, beginning of logbook trip, middle of logbook trip, end of logbook trip, and end of VMS trip. VMS pings successfully matched by the three different methods are marked with an "M." (A) Hintzen's method of identifying pings within the beginning and end of the logbook trip fails in matching any VMS ping; both Bastardie's midpoint method and the present method successfully match all three available pings. (B) Hinzen's methods matches 33% of VMS pings, Bastardie's midpoint method fails in matching any VMS ping, and the present method again matches all three available pings.



observer programs could also be used to complement VMS effort information.

Several methods have been used to match VMS pings to logbook data (Afonso-Dias et al. 2002; Bastardie et al. 2010; Gerritsen and Lordan 2011; Hintzen et al. 2012; Watson and Haynie 2016). Although this step is crucial, it is rarely well documented in the literature. Generally, the process of matching the datasets involves two to three steps. First, VMS data are segmented into trips. Then, these VMS trips are matched to logbook trips. Finally, in a minority of methods, unmatched logbook trips are handled.

To identify individual trips in the VMS data, it is useful to pinpoint when a vessel leaves port. This is a relatively easy task when fishing fleets depart from well-defined fishing ports (Bastardie et al. 2010), but it is more complicated with fishing fleets that operate small boats launched from trailers at private jetties or even up rivers. If ports are well defined, VMS trips can be demarcated by identifying VMS positions as "at port" and marking a trip between departures and arrivals at that location (Bastardie et al. 2010). When ports are not well defined, we present an alternative method. Here, VMS trips can be identified using information about the status of the VMS position relative to land. A trip starts whenever a nearshore VMS position is followed by an offshore position, and a trip ends whenever the opposite occurs. The exact definition of what is considered nearshore versus offshore will vary with the geography of the area. Technical issues and gaps in VMS records can result in spuriously long VMS trips. These can be split using the maximum trip duration for that particular vessel observed in the logbook data. After trip segmentation, each trip in the VMS dataset is assigned a unique identifier.

To match logbook trips to VMS trips, Hintzen et al. (2012) used a simple approach: they selected all VMS positions that occurred between the start and end dates of each logbook trip and assigned them to that logbook entry. This approach, however, misses many VMS points that lay just around the logbook trip window, which will be the case if there is error during data entry in the logbook records, a common issue in logbook data (Fig. 2A). Bastardie et al. (2010) match the VMS trip by searching for the nearest logbook trip based on its temporal midpoint. This approach produces a match even when the start and end time of VMS and logbook trips do not match exactly (Bastardie et al. 2010; Fig. 2A). This method, however, fails to match trips that end or start before or after the midpoint, as can happen in short fishing trips (Fig. 2B). To handle this issue, we match VMS trips if any of three conditions is met: if the start of any VMS trip lies within the logbook trip window, if the end of any VMS trip lies within the logbook trip window, or if the end of the logbook trip lies within the VMS trip window. This method is robust-to-short trip durations as shown in Fig. 2B.

Whenever logbook trips remain unmatched, they can be linked to the nearest unmatched VMS trip (Bastardie et al. 2010), uniformly to pings for that vessel ID, or distributed uniformly among all fishing pings within a year (Hintzen et al. 2012), constrained by situation-appropriate decision rules.

Extracting derived variables

To spatially summarize the data, fishing positions can be rasterized using grids of different spatial resolution (Fig. 1, step 6). The spatial resolution should match the objectives of the fisheries management question or the spatial management unit, and the ecology of the species involved (Mills et al. 2007; Piet and Quirijns 2009).

Within each cell, the number of VMS fishing records can be summed to estimate fishing effort. By linking VMS and logbook data, maps of total or species-specific catch and revenue can also be produced. To distribute logbook information on trip-level catch and revenue among VMS positions, most researchers have assumed a uniform distribution (Bastardie et al. 2010; Ducharme-Barth et al. 2018; Berenshtein et al. 2019; Mamula et al. 2020). This ignores spatial heterogeneity in catch rates that cannot be documented with the data available. However, the assumption of uniform distribution has been shown to provide relatively accurate depictions of catches in space (Gerritsen and Lordan 2011).

Finally, the linked VMS-logbook datasets can be used to relate revenues from fishing grounds back to the fishing communities where the landings occur by linking the fishing grounds to the port information also available in the logbooks. Logbooks might include port information as the specific location of the harbour, or territorial divisions such as counties or towns. If port information is not explicitly included in the logbook dataset, it can be estimated from the VMS data if vessels have high fidelity in home ports (e.g., Bastardie et al. 2010). This information is useful in marine spatial planning to allow for planning outcomes that are equitable and socially acceptable (Saunders et al. 2019). To do so, it is possible to partition catches or revenues either to landing ports or fishing communities if the information is available in the logbook data (e.g., St. Martin and Olson 2017; Berenshtein et al. 2019), or use the VMS data to identify an associated port for each trip and use that information as input.

Case study

Here we applied the framework for mapping revenues to the bandit-reel and bottom-longline reef fisheries in the Gulf of Mexico (GoM) during the year 2021. These fisheries have the most complete datasets in the GoM (Perruso et al. in prep) and have been analyzed by our group previously (O’Farrell et al. 2017; Berenshtein et al. 2019). Three datasets were used to quantify revenues, namely VMS data, on-board observer programme data, and vessel logbooks (Fig. 1).

VMS dataset

VMS transponders sending hourly or better reports have been mandated on all commercial vessels targeting reef fish in the GoM since 2006 (eCFR 2016). VMS datasets contain information on vessel ID, geographic position, and time stamp.

Observer programme

To satisfy requirements of the Magnuson Stevens Fishery Conservation and Management Act, the Marine Mammal Protection Act, and the Endangered Species Act, among others, the US National Observer programme monitors nearly 50 different fisheries in the US (NMFS 2016). Observers accompany a sample of commercial fishing trips, recording the trip number, vessel ID, gears used, the start and end date of the trip, as well as the start and end time of each set. Time zone of observer’s records was unavailable; therefore, we assumed the observer’s time zone is that of the trip origin. i.e., aboard a trip originating in Florida, the observer used EST.

Logbook dataset

All commercial fishing vessels with a reef fish permit in the GoM are required to maintain a logbook recording the type of gear used on each trip, together with information on vessel ID, catch, revenue earned, date of landing, and county of landing. Vessels submit logbooks to the Southeast Fisheries Science Center (SEFSC) in Florida.

Processing.

VMS data were subset to retain only target gears used in this analysis: bandit, handline, and bottom longline. VMS data were preprocessed as described in the section *“2. Extracting variables”* above.

To classify fishing activity, we used 13 variables: depth, vessel length, bearing, prior leg distance, prior velocity, post velocity, mean velocity, velocity change, absolute velocity change, turn angle, change in turn angle, decimal hour of the day, and lunar phase. Depth was calculated from the ETOPO 1 database (Amante and Eakings 2009). The direction of travel and distance between consecutive positions were calculated using rhumb lines or tracks of constant true course. Velocity for each position was characterized using three variables: the value between the current and previous position (“prior”), the value between the current and next position (“post”) and the average value between these two (“mean”). Absolute turning angles, the magnitude of change in angle regardless of the direction were calculated from the headings of the legs to and from the current position. Changes in velocity and turn angles were calculated using “prior” values. Decimal hour of day was rounded to four decimal places. Lunar phase was calculated using the date as input, with ranges between −1 for new moon to +1 for full moon.

Outliers when speed was larger than 10 $m{\cdot}s^{-1}$ were deleted. Only data within the GoM were retained (23.5–31N, 78–98W). Nearshore and onshore VMS positions recorded within 1 km of land masses were not included in the analyses. Coastline was described using the Global Self-consistent, Hierarchical, and High-resolution Geography Database (Wessel and Smith 1996).

Observer data were preprocessed as described in the section *“3. Linking observer and VMS data”* above. An observer record was considered an outlier and deleted if the duration of the trip was longer than 30 days or soaking time was longer than 24 h. Observer data were matched to VMS data using a window-labelling technique, namely labelling as fishing all VMS positions within the start and end time of the set, plus those located half of the interval between positions (O’Farrell et al. 2017).

When datasets are large, regular tests of significance between two groups are not useful to detect meaningful differences because they frequently show a significant nonzero difference between the groups, albeit very small and meaningless (Nakagawa and Cuthill 2007). In these cases, it is better to focus on the magnitude of the effect rather than simply the significance of the test. Here, we looked at the differences in the predictor variables between fishing and not fishing using the Hedges test of effect size for standardized differences (Ben-Shachar et al. 2020).

Fishing activity for longline and bandit was identified using random forests (Breiman 2001), one of the most popular classifiers in movement pattern recognition. Random forests have the advantage of being robust to overfitting to training data and allowing nonlinear interactions to be captured (Strobl et al. 2009). About 30% of the labelled records are not used during classification but are used to cross-validate the performance of the random forest classifier. Performance was measured in terms of true positive, true negative rates, and balanced accuracy (mean of true positive and negative rates). We plotted the relative importance of each variable to the classification according to the mean decrease in accuracy metric (Liaw and Wiener 2002).

To spatially summarize the data, fishing positions were rasterized using a grid of 5 min. Catch and revenue information was distributed uniformly among VMS pings.

Four broad study areas were recently delineated as a basis to identify aquaculture opportunity areas in the GoM: West GoM, Central GoM, East GoM, and Southeast GoM (Riley et al. 2021). These areas have consistent bathymetry appropriate for aquaculture development, but distinct biogeographic relevance. Within these regions, potential opportunity areas can be identified using spatial analyses (e.g., Riley et al. 2021). We showed the link of revenues between fishing grounds and fishing communities in each of these four study areas and highlighted different issues that can be relevant during the marine spatial planning process for aquaculture in the GoM. In particular, for each Study Area, we calculated the number of US States and Counties that fish in its waters. We also calculated the equitability of revenue shares among GoM counties for each Study Area. Equitability was calculated using Pielou's evenness index (Dixon 2003). The index is constrained between 0 and 1, with lower values indicating more dominance and less equitability. The analyses focused on quantifying the current fisheries use of the area and did not contemplate the potential redistribution of the resource (e.g., Chollett et al. 2016).

All analyses were carried out in R (R Core Team 2022). Data handling was done using the R packages data.table (Dowle et al. 2015) and dplyr (Wickham et al. 2015), and bathymetry extraction using marmap (Pante and Simon-Bouhet 2013). Bearings and distances were calculated using geosphere (Hijmans 2015). Lunar phase was extracted using the package lunar (Lazaridis 2014). GIS operations were carried out with the aid of the packages raster (Hijmans 2017) and sp (Bivand et al. 2008). Effect sizes were calculated using the package effectsize (Ben-Shachar et al. 2020). Random forests classifications were performed with the aid of the package randomForest (Liaw and Wiener 2002). Evenness was calculated using the package vegan (Dixon 2003).

Results

Identifying the target fishery: subsetting VMS data

The VMS data for 2021 contained a total of 4 488 634 vessel positions from 494 different vessels. A total of 2 986 423 VMS records corresponded to vessels using target gears in the logbook data.

Extracting variables

We used 13 variables to describe fishing activity: depth, vessel length, bearing, prior leg distance, prior velocity, post velocity, mean velocity, velocity change, absolute velocity change, turn angle, change in turn angle, decimal hour of the day, and lunar phase. Depth had a skewed distribution with a mean of 55.83 m (±91.73 m standard deviation). Positions had bearings along all angles. Distances were skewed with a mean of 2.49 ± 4.06 km. Speeds were also skewed with a mean of 0.77 ± 1.19 $m \cdot s^{-1}$. Velocity changes were centred around zero (−0.01 ± 1.03) and turning angles had a mean of 84.12 ± 63.63 (Figs. 3, 4). Variables behave similarly for bandit and longline fishing, but longliners tend to show more variability in turning angles (Figs. 3, 4).

Linking observer and VMS data

The observer data for 2021 had 51 477 records for the bandit fishery and 11 683 for the longline fishery. Records were linked to VMS data, and a total of 213 220 VMS pings were labelled within the bandit (122 365 records) and the longline fishery (90 855 records). From those, 27% VMS pings were identified as fishing for bandit and 39% as fishing for the longline fishery.

Patterns of fishing behaviour between fishing and not-fishing activity are depicted in Figs. 3 and 4. For bandit fishing, only differences in decimal hour of the day are meaningful between fishing and not fishing activity (Fig. 3, Hedges' g effect size, 0.78, outside the confidence interval for the variable: [0.856, 0.857]). For longline, differences are meaningful for absolute change in velocity (Hedges' g 0.46, confidence interval [0.5097, 0.5098]), turning angle (0.45 [0.498, 0.499]), change in turning angle (0.53 [0.582, 0.583]), and decimal hour of the day (0.87 [0.955, 0.956], Fig. 4).

Identifying fishing activity

Random forest classification provided good overall accuracy for both gears. The balanced accuracy of the classification, that is, the mean of all true positive rates (i.e., 1—mean of the class errors) was 0.85 for bandit fishing and 0.9 for longline fishing, with not-fishing classes better classified than fishing classes (Table 2). In a binary classification, balanced accuracy effectively penalizes incorrect classification, as well as rewarding correct classification.

The variables that contribute the most to the classifications for both gears are decimal hour of the day and depth (Fig. 5). It is interesting to note that this contribution is not apparent in the univariate comparisons between fishing and not-fishing classes shown in the previous section, highlighting the strength of higher dimensional machine learning routines that can take nonlinearities and complex interactions into account. For example, in both the longline and bandit data, the patterns in decimal hour of day for fishing versus not-fishing are almost the inverse of each other (Fig. 4). Here, fishing occurs during the day and rarely/never at night, so this is a powerful variable for random forest to split pings into fishing versus not-fishing. Likewise depth, as the reef-fish fishery is most active on or near the reefs that occur within a particular depth range. Random forest is adept at discov-

Fig. 3. Density histograms for bandit reel-associated variables during fishing (A) and not fishing (B) activity, as identified using the observer data.

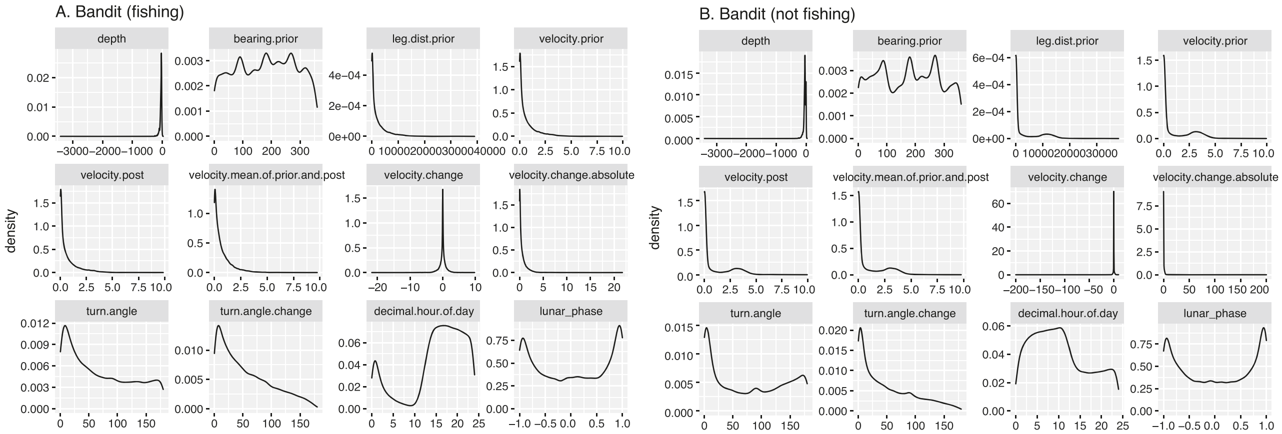


Fig. 4. Density histograms for longline-associated variables during fishing (A) and not fishing (B) activity, as identified using the observer data.

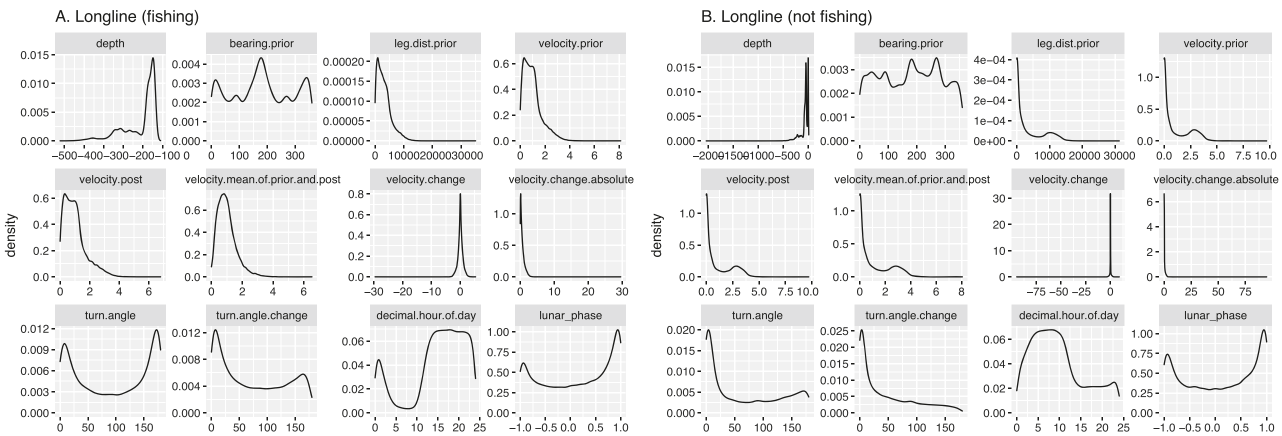


Table 2. Confusion matrix indicating the number of samples correctly classified, incorrectly classified, and the class error for both bandit and longline fisheries.

	Fishing	Not fishing	Class error
Bandit			
Fishing	17 056	4216	0.198
Not fishing	3744	33 956	0.099
Longline			
Fishing	22 640	2698	0.107
Not fishing	3481	31 108	0.101

ering interactions among variables without requiring those interactions to be specified a priori as in a linear regression model. So if a reef-fishery vessel is over the reef (i.e., within a particular depth range) during certain times of day (i.e., within a particular time window), the probability that the vessel is fishing increases. However, other not-fishing activities may also occur during the day over the reef (e.g., steaming, searching) and random forest further exploits the small differences in other predictor variables such as velocity metrics (linear speed and acceleration) and turning angles to further discriminate activities. The ability to uncover complex patterns in variable values and interactions is what makes agile machine-learning routines such as random forests or neural networks more effective at discriminating fishing from not-fishing than are simple univariate classifiers such as speed filters or even higher dimensional models that require putative interactions to be defined by the user.

Linking VMS and logbook data

Linking VMS to logbook data allows catch and revenue values to be assigned to each fishing trip in a spatially explicit manner. The trip-linking process is visualized in Fig. 6 for a hypothetical boat within the fleet. Some VMS trips have no matching logbook data. This is possible given that GoM vessels are sometimes used as charter boats or fish different targets, which are reported using a different logbook system or none at all. All 2021 logbook trips in the GoM were matched using this method. For the sample boat whose data are depicted in Fig. 6, there were seven logbook trips—all of which

Fig. 5. Relative contribution of variables used by the Random Forest classifier when identifying fishing in (A) bandit reel and (B) longline fisheries. The mean decrease accuracy plot shows how much accuracy the model loses by excluding each variable. This is calculated by permuting the values in each variable.

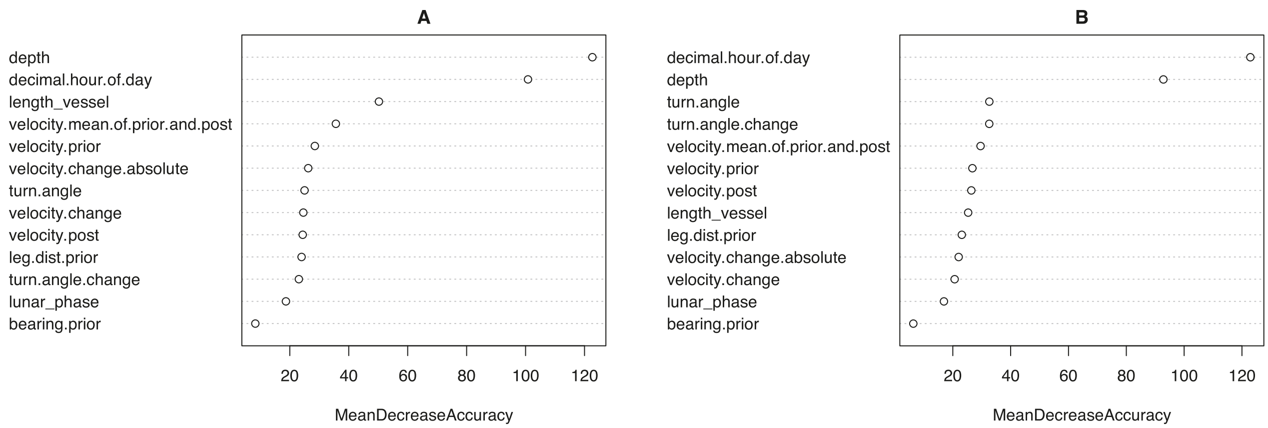
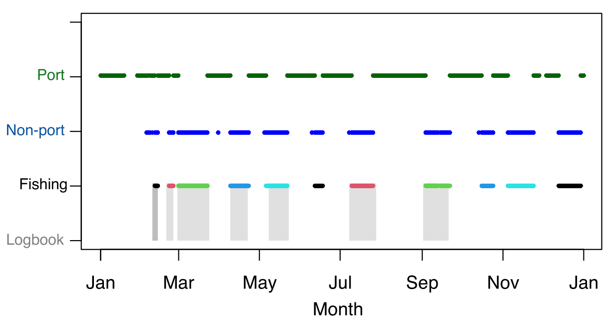


Fig. 6. Linking logbook and VMS data for the year 2021 for a hypothetical vessel. Vertical grey bars show the time windows of the seven trips recorded in the logbook data. Coloured horizontal dashes show the time windows of the 11 trips segmented in the VMS data, as inferred from the event of a vessel crossing and then returning across a spatial boundary between "nearshore" and "offshore". Unmatched VMS trips can result from voyages made by vessels that are not related to commercial reef-fishing activity and therefore do not require a logbook record, such as charter-boat hires.



were successfully linked to a corresponding VMS trip—and four residual VMS trips without corresponding logbook trips. Data visualizations of this nature are invaluable in graphically assessing the performance of trip matching routines.

Extracting derived variables

After cleaning and preprocessing, the VMS dataset held 2 371 158 records for 385 vessels. In the GoM, fishing for reef fish is distributed along the entire reef system, with effort concentrated mainly along the eastern portion of the Gulf (Fig. 7).

Fishing effort, measured as the number of VMS pings in each 5′ cell, was related to total catch and total revenue (Pearson correlations 0.72 and 0.73, Fig. 8). Catch and revenue of reef fish were highly related for this fishery (correlation 0.99,

Fig. 7. Gulf of Mexico map showing total revenue over a 5' grid (thousand USD). Only cells fished by more than three vessels have been included to comply with requirements of data confidentiality. Land data from the Global Self-consistent, Hierarchical, High-resolution Geography Database (Wessel and Smith 1996). EPSG:4326.

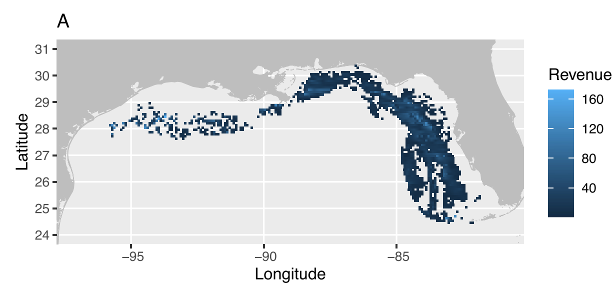


Fig. 8), indicating that either variable is useful in depicting spatial patterns of fishing intensity in the GoM at this spatial resolution. The relationship between fishing effort and revenue, however, gets weaker when increasing the spatial resolution of the analysis (not shown). The strength of the relationship between catch and revenue can also depend on variability in prices throughout season, ports, or species, and can be influenced by external disturbances if supply is constrained by fishery closures (e.g., Upton 2011).

Fishing information was extracted for each of the four aquaculture study areas in the GoM (Fig. 9). Again, patterns of catch and revenue are highly related (correlation 0.98), but effort provides different patterns (correlation 0.44 for catch and 0.46 for revenue). The distribution of fishing is uneven between study areas. The study area that is subjected to the most fishing effort is the Eastern GoM (Fig. 9A), while the one that generates the largest revenue is SouthEast GoM (Fig. 9C). Although the Western region accrues the least fishing effort, it provides the second highest level of revenues (Figs. 9A, 9C).

Closing areas to fishing in different study areas will have uneven effects on counties across the GoM (Table 3,

Fig. 8. Scatterplots showing the relationship between (A) fishing effort and total catch, (B) fishing effort and total revenue, and (C) total catch and total revenue.

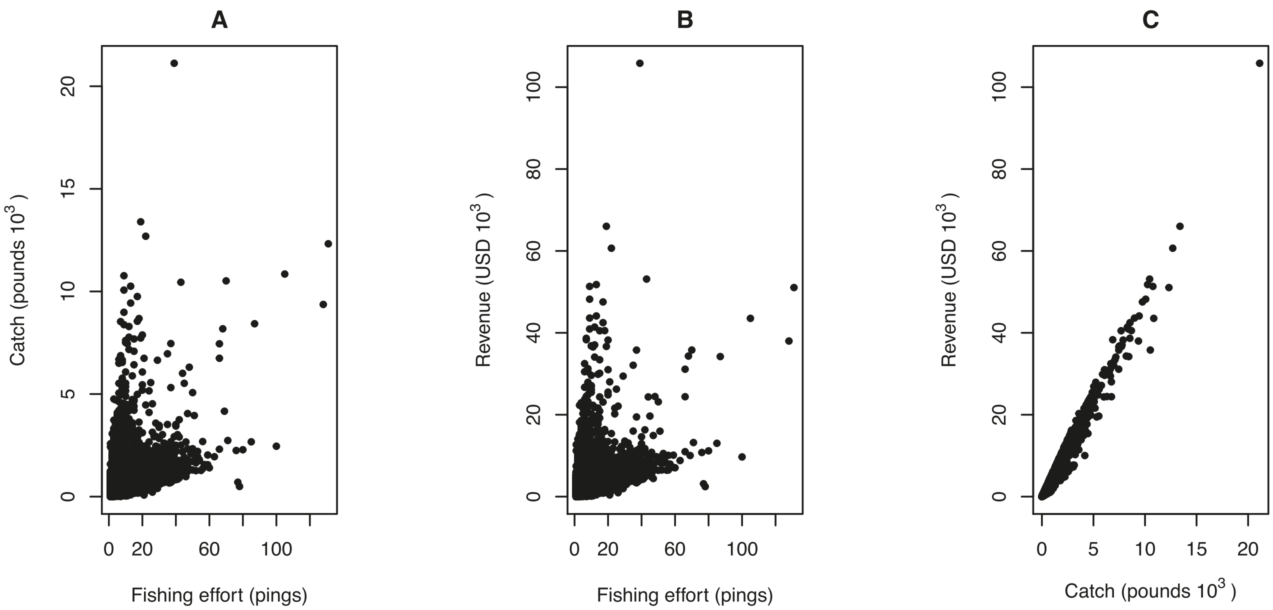


Fig. 9. Barplots showing (A) fishing effort; (B) total catch; and (C) total revenue, all normalized by area (km^2) in each potential aquaculture study area: "C" Central; "E" Eastern; "SE" Southeastern; and "W" Western GoM.

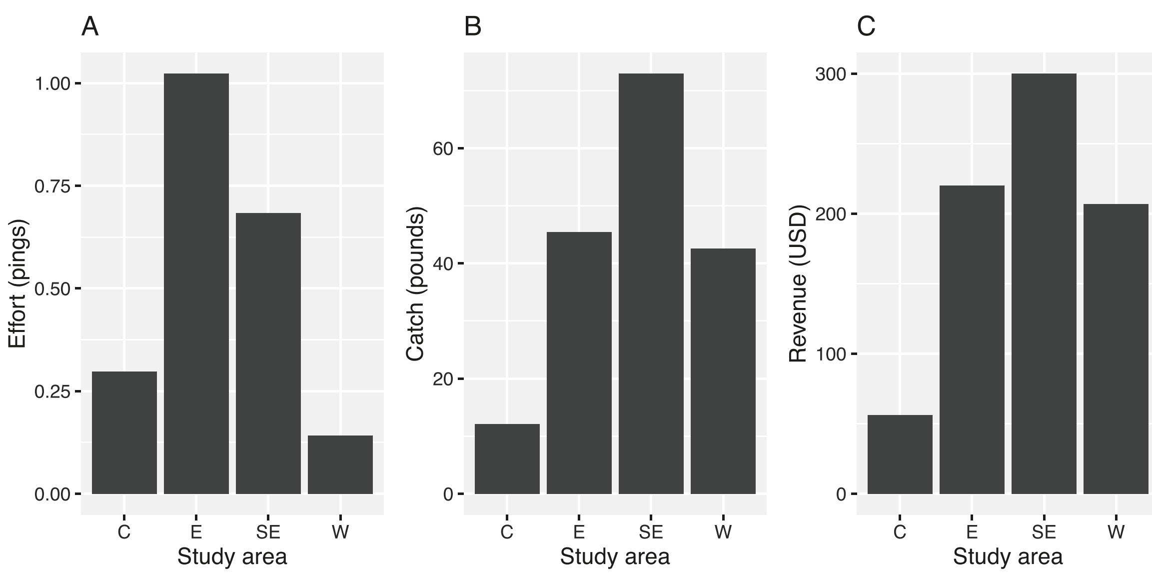
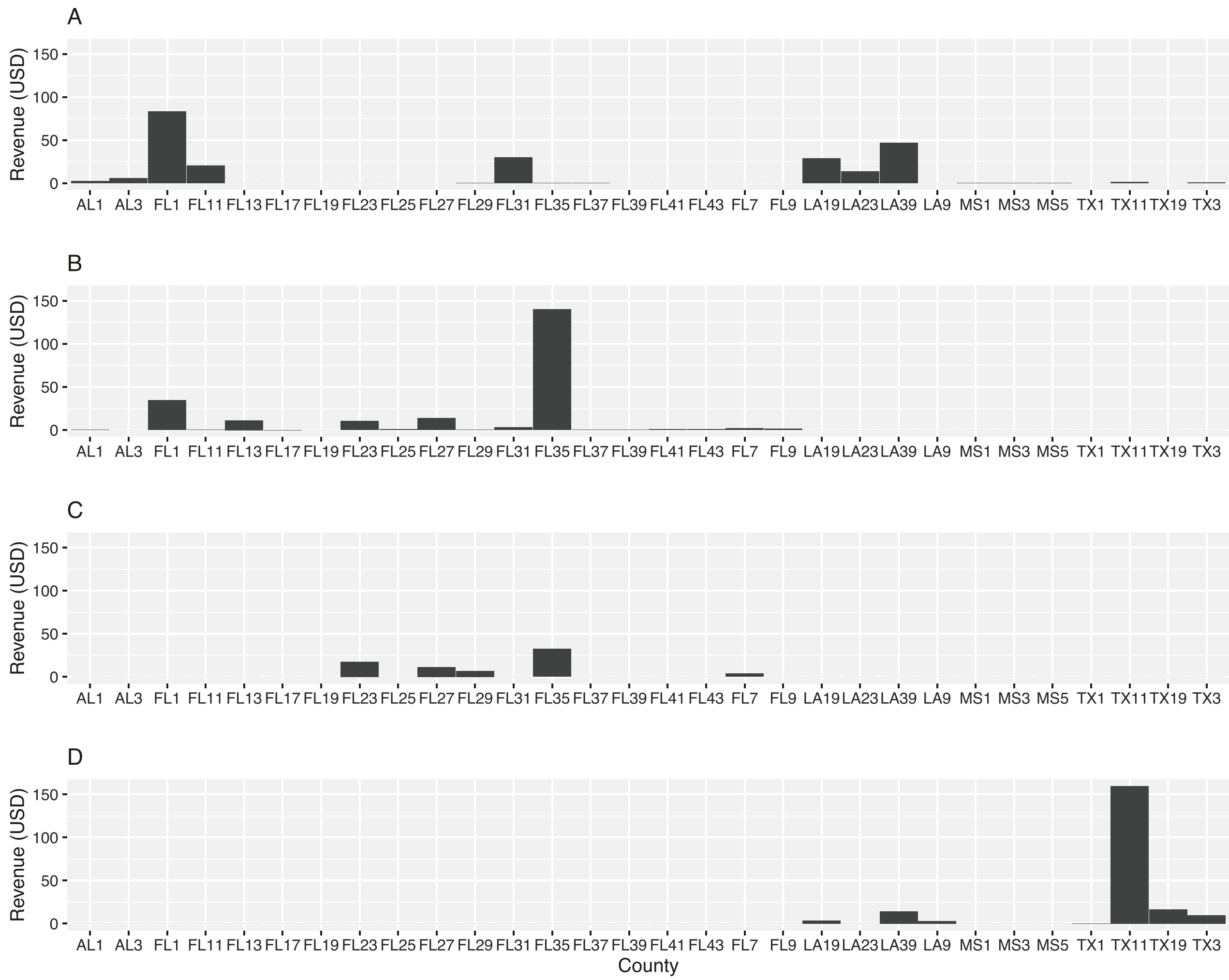


Table 3. Aquaculture study areas in the GoM, number of states and counties that benefit from fishing in each area, and equitability of revenue shares between counties.

Study area	States	Counties	Equitability
Central GoM	5	16	0.65
Eastern GoM	2	18	0.44
Southeastern GoM	1	5	0.85
Western GoM	2	7	0.45

Figs. 10–11). Fishing in Eastern GoM benefits the most number of counties, followed by fishing in Central GoM (Table 3). Fishing in Eastern GoM is, however, less equitable and benefits mostly one county disproportionately (Pinellas in Florida). Central GoM fishing benefits are distributed widely among counties, reaching all of the five GoM states (Table 3). These results indicate that marine spatial planning for reef fish in the GoM could have very different social implications if focused on different regions of the Gulf.

Fig. 10. Revenue lost by different counties in the GoM for different study areas (A) Central GoM, (B) Eastern GoM, (C) Southeast GoM, and (D) Western GoM. Within counties, the first two letters indicate the state: "AL" Alabama; "FL" Florida; "LA" Louisiana; "MS" Mississippi; and "TX" Texas.

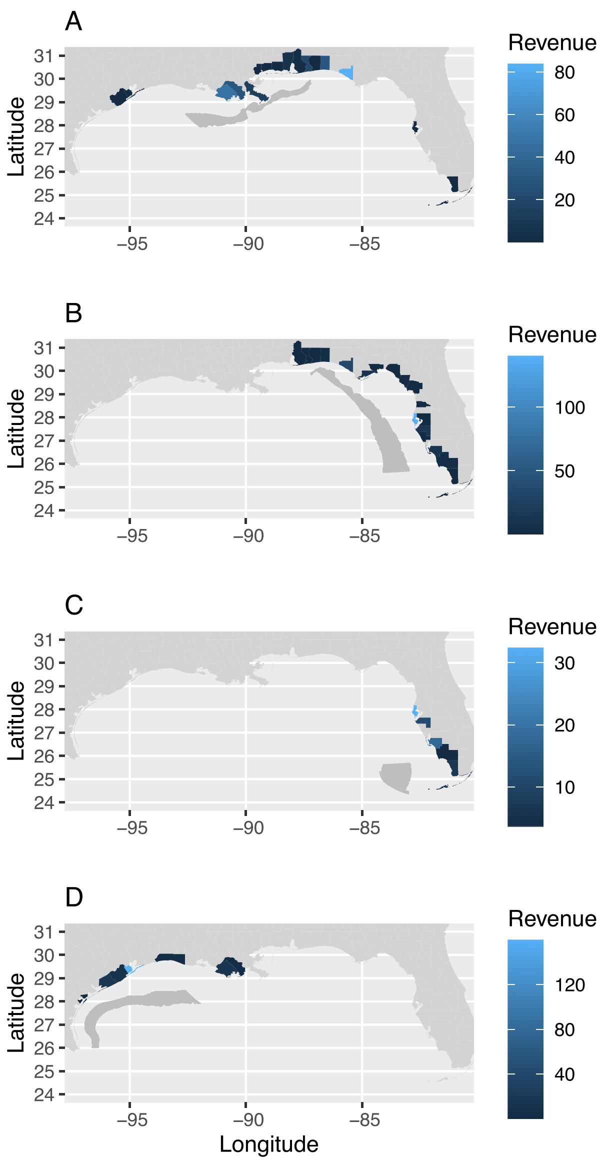


Discussion

A reliable depiction of fishing patterns is only possible when data analyses are tailored to the particular fishing practices. Simple steps to adjust the methodology used to analyse fisheries data allow researchers to deal efficiently with messy datasets, such as those characterized by short trips and lack of defined ports or unreliable self-reporting data, therefore promoting effective fisheries management.

Mapping effort, catches, and revenues provide different information that can be useful for resource management in different ways. The choice of variable to map and use in fisheries management is an important decision that needs to consider the objectives of the management task at hand. Several studies have shown that using different metrics of fishing activity can result in different spatial plans (e.g., Deas et al. 2014; Hamel et al. 2018); therefore the metric mapped and used in management should be chosen with care. Fishing effort data can be used as a proxy for environmental impacts of bottom and demersal fishing activities, which remove biomass, change ecological community dynamics, and damage habitat (Agbayani et al. 2015). Fishing effort also better reflects patterns of use of an area, and it is a more accurate metric of displacement (Chollett et al. 2016). Fishing is an economic activity and it is driven by profit; however, in practice, the relationship between fishing effort and catch is not straight-forward and relationships with buyers, quota holders, as well as vessel characteristics can all be factors informing the most-profitable fishing strategies, in some cases decoupling the relationship between these variables at the fishery level (Coccoli et al. 2018). Therefore, fisheries catch and revenue are a better measure of the economic value of a region compared to fishing effort. These two variables are more likely to be similar in monospecific fisheries or multi-species fisheries with targets that do not vary much in price, like in the reef fishery in the GoM. However, if the opposite occurs and fisheries target multiple species with different value, maps of catch and revenue could show very different spatial patterns. In general, most fisheries management requires maps of opportunity cost to quantify trade-offs in the use of different areas, and to this end maps of revenue are more useful (Jin et al. 2013).

Fig. 11. Revenue lost by different counties in the GoM for different aquaculture study areas, delineated as a grey polygon in the sea (A) Central GoM, (B) Eastern GoM, (C) Southeastern GoM, and (D) Western GoM. Land data from the Global Self-consistent, Hierarchical, High-resolution Geography Database (Wessel and Smith 1996). EPSG:4326.



Communities who fish for their livelihood are not only defined by the place where they live, but by the place where they work at sea (St. Martin and Olson 2017). When mapping fishing patterns and using this information for resource management, most studies do not allocate revenue back to the communities that extracted the resource. This fails to acknowledge the presence of fishing territories when multiple fishing communities target the same resource. Knowing which communities are using each area is key for understanding the potential impacts of management measures on fishers, and placing targeted measures to alleviate those impacts.

The choice of a classifier for identifying fishing behaviour is key for an accurate depiction of fishing activity. Machine learning tools such as random forests and neural networks have been used recently to map different fisheries across the globe using geopositional information (O'Farrell et al. 2019; Behivoke et al. 2021; Torres-Irineo et al. 2021). We used random forests because the routine provides accurate solutions for nonlinear discrimination in high-dimensional spaces with interacting variables (Cutler et al. 2007) and is a great choice for picking up information that would be impossible to harness using univariate, linear approaches.

The present modelling exercise allowed us to value reef fisheries in different regions in the GoM and produce maps that are useful for spatial planning for aquaculture and other uses in this important region. Fishing for reef fish is distributed throughout the Gulf. Although natural reefs are mainly located in the Eastern GoM along the coast of Florida, the Central and Western GoM contain many artificial reefs and oil platforms that serve as habitat for reef species (Villareal et al. 2007). There is an uneven distribution of fishing effort and revenues throughout the four aquaculture study areas in the GoM, and fishing in each area makes different contributions to the fishing communities in the region. Our framework could not only help identify which regions are less congested, would produce less displacement and the least loss of revenue if closed to fishing, but also aid determining which fishing communities will be more affected by potential closures. The analysis presented here, however, incorporates only 1 year of data and represents only a snapshot. Long-term patterns of fishing activity should be assessed for using this information in drafting robust spatial plans and management decisions (Chollett et al. in 2022).

We presented a general framework to analyse spatially explicit fisheries data, mapping revenues, and linking them from fishing grounds back to the fishing communities. The framework provides an overview of different processing steps involved in linking VMS data and self-reported logbook and observer data, and suggests different alternatives to tackle each step for the processing of messy, incomplete data. While mapping the total value of the ocean to fishing industry could be used to develop a trade-off analysis and evaluate alternative uses such as proposed aquaculture sites, the disaggregation of revenues across different counties could allow the assessment of the impacts of the introduction of alternative ocean uses on different fishing communities, and ensure socially equitable management outcomes.

Acknowledgements

We are very grateful to Cristopher Liese and two anonymous reviewers for their helpful comments.

Disclaimer

The views and opinions expressed or implied in this article are those of the authors and do not necessarily reflect the position of the National Marine Fisheries Service, NOAA.

Article information

History dates

Received: 3 May 2023
Accepted: 31 August 2023
Accepted manuscript online: 14 September 2023
Version of record online: 14 November 2023

Copyright

Data availability

The study data were obtained under a contractual agreement with the U.S. National Marine Fisheries Service (NMFS). The agreement prevents distribution of personally identifiable information, including variables directly included in the analysis. These data are archived at NOAA's SEFSC. Researchers under a contractual agreement with NMFS can access the data provided a nondisclosure agreement is signed.

Author information

Author ORCIDs

Iliana Chollett https://orcid.org/0000-0003-0536-499X

Author contributions

Conceptualization: SO, IC
Data curation: SO, LP
Formal analysis: SO, IC
Funding acquisition: LP, JNS
Methodology: SO, LP, JNS, IC
Writing – original draft: SO, IC
Writing – review & editing: SO, LP, JNS, IC

Competing interests

The authors declare there are no competing interests.

Funding information

This study was funded by NSF Coastal SEES Grant No. 1325452 (JNS); National Academy of Sciences, Gulf Research Program Data Synthesis Grant No. 2000007631 (JNS); and NOAA's Internal Competitive Aquaculture Funds grant program (LP).

References

Afonso-Dias, M., SimBes, J.M., Pinto, C., and Sousa, P. 2002. Use of satellite GPS data to map effort and landings of the Portuguese crustacean fleet (GeoCrust). EC-DGXIVISTUDYI991099. Universidade do Algarve, Faro, Portugal.

Agbayani, S., Picco, C.M., and Alidina, H.M. 2015. Cumulative impact of bottom fisheries on benthic habitats: a quantitative spatial assessment in British Columbia, Canada. Ocean Coast. Manage. **116**: 423–434.

Amante, C., and Eakins, B.W. 2009. ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis. US Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service, National Geophysical Data Center, Marine Geology and Geophysics Division, Boulder, CO.

Bastardie, F., Nielsen, J.R., Andersen, B.S., and Eigaard, O.R. 2013. Integrating individual trip planning in energy efficiency—building decision tree models for Danish fisheries. Fish. Res. **143**: 119–130. doi:10.1016/j.fishres.2013.01.018.

Bastardie, F., Nielsen, J.R., Eigaard, O.R., Fock, H.O., Jonsson, P., and Bartolino, V. 2015. Competition for marine space: modelling the Baltic Sea fisheries and effort displacement under spatial restrictions. ICES J. Mar. Sci. **72**(3): 824–840. doi:10.1093/icesjms/fsu215.

Bastardie, F., Nielsen, J.R., Ulrich, C., Egekvist, J., and Degel, H. 2010. Detailed mapping of fishing effort and landings by coupling fishing logbooks with satellite-recorded vessel geo-location. Fish. Res. **106**(1): 41–53. doi:10.1016/j.fishres.2010.06.016.

Behivoke, F., Etienne, M.P., Guitton, J., Randriatsara, R.M., Ranaivoson, E., and Léopold, M. 2021. Estimating fishing effort in small-scale fisheries using GPS tracking data and random forests. Ecol. indic. **123**: 107321. doi:10.1016/j.ecolind.2020.107321.

Ben-Shachar, M.S., Lüdecke, D., and Makowski, D. 2020. effectsize: estimation of effect size indices and standardized parameters. J. Open Source Softw. **5**(56): 2815. doi:10.21105/joss.02815.

Berenshtein, I., O'Farrell, S., Perlin, N., Sanchirico, J.N., Murawski, S.A., Perruso, L., and Paris, C.B. 2019. Predicting the impact of future oil-spill closures on fishery-dependent communities—a spatially explicit approach. ICES J. Mar. Sci. **76**(7): 2276–2285.

Bertrand, S., Díaz, E., and Lengaigne, M. 2008. Patterns in the spatial distribution of Peruvian anchovy (*Engraulis ringens*) revealed by spatially explicit fishing data. Prog. Oceanogr. **79**(2-4): 379–389. doi:10.1016/j.pocean.2008.10.009.

Bivand, R.S., Pebesma, E.J., Gómez-Rubio, V., and Pebesma, E.J. 2008. Applied spatial data analysis with R. Springer, New York.

Booth, A.J. 2000. Incorporating the spatial component of fisheries data into stock assessment models. ICES J. Mar. Sci. **57**(4): 858–865. doi:10.1006/jmsc.2000.0816.

Breiman, L. 2001. Random forests. Mach. Learn. **45**(1): 5–32.

Chollett, I., Box, S.J., and Mumby, P.J. 2016. Quantifying the squeezing or stretching of fisheries as they adapt to displacement by marine reserves. Conserv. Biol. **30**(1): 166–175. doi:10.1111/cobi.12573.

Chollett, I., Perruso, L., and O'Farrell, S. 2022. Toward a better use of fisheries data in spatial planning. Fish Fish. doi:10.1111/faf.12674.

Coccoli, C., Galparsoro, I., Murillas, A., Pınarbaşı, K., and Fernandes, J.A. 2018. Conflict analysis and reallocation opportunities in the framework of marine spatial planning: a novel, spatially explicit Bayesian belief network approach for artisanal fishing and aquaculture. Mar. Pol. **94**: 119–131. doi:10.1016/j.marpol.2018.04.015.

Cutler, D.R., Edwards, T.C. , Jr, Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., and Lawler, J.J. 2007. Random forests for classification in ecology. Ecology, **88**(11): 2783–2792. doi:10.1890/07-0539.1.

Deas, M., Andréfouët, S., Léopold, M., and Guillemot, N. 2014. Modulation of habitat-based conservation plans by fishery opportunity costs: a New Caledonia case study using fine-scale catch data. PLoS ONE, **9**(5): e97409. doi:10.1371/journal.pone.0097409.

de Souza, E.N., Boerder, K., Matwin, S., and Worm, B. 2016. Improving fishing pattern detection from satellite AIS using data mining and machine learning. PLoS ONE, **11**(7): e0158248. doi:10.1371/journal.pone.0158248.

Dixon, P. 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci. **14**(6): 927–930. doi:10.1111/j.1654-1103.2003.tb02228.x.

Dowle, M., Srinivasan, A., Short, T., and Lianoglou, S. 2015. data.table: extension of Data.frame. R package version 1.9.6. Available from https://CRAN.R-project.org/package=data.table [accessed January 2023].

Ducharme-Barth, N.D., Shertzer, K.W., and Ahrens, R.N. 2018. Indices of abundance in the Gulf of Mexico reef fish complex: a comparative approach using spatial data from vessel monitoring systems. Fish. Res. **198**: 1–3. doi:10.1016/j.fishres.2017.10.020.

eCFR. 2016. Electronic Code of Federal Regulations. US Government Publishing Office. Available from http://www.ecfr.gov/cgi-bin/text-idx?SID=86d3e4e21c5c4a3cd94b7f2 59d8700e1&node=50:12.0.1.1.2&rgn=div5#se50.12.622_128 [accessed 1 July 2016].

Gerritsen, H., and Lordan, C. 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 J. Mar. Sci. **68**(1): 245–252. doi:10.1093/icesjms/fsq137.

Gerritsen, H.D., Minto, C., and Lordan, C. 2013. How much of the seabed is impacted by mobile fishing gear? Absolute estimates from Vessel

Monitoring System (VMS) point data. ICES J. Mar. Sci. **70**(3): 523–531. doi:10.1093/icesjms/fst017.
Gilliland, P.M., and Laffoley, D. 2008. Key elements and steps in the process of developing ecosystem-based marine spatial planning. Mar. Pol. **32**(5): 787–796. doi:10.1016/j.marpol.2008.03.022.
Hamel, M.A., Pressey, R.L., Evans, L.S., and Andréfouët, S. 2018. The importance of fishing grounds as perceived by local communities can be undervalued by measures of socioeconomic cost used in conservation planning. Conserv. Lett. **11**(1): e12352. doi:10.1111/conl.12352.
Hijmans, R.J. 2015. geosphere: Spherical Trigonometry. R package version 1.5-1. Available from https://CRAN.R-project.org/package=geosphere [accessed January 2023].
Hijmans, R.J. 2017. Raster: Geographic Data Analysis and Modeling. R package.
Hintzen, N.T., Bastardie, F., Beare, D., Piet, G.J., Ulrich, C., Deporte, N., et al. 2012. VMStools: open-source software for the processing, analysis and visualisation of fisheries logbook and VMS data. Fish. Res. **115**: 31–43. doi:10.1016/j.fishres.2011.11.007.
Jennings, S., Lee, J., and Hiddink, J.G. 2012. Assessing fishery footprints and the trade-offs between landings value, habitat sensitivity, and fishing impacts to inform marine spatial planning and an ecosystem approach. ICES J. Mar. Sci. **69**(6): 1053–1063. doi:10.1093/icesjms/fss050.
Jin, D., Hoagland, P., and Wikgren, B. 2013. An empirical analysis of the economic value of ocean space associated with commercial fishing. Mar. Pol. **42**: 74–84. doi:10.1016/j.marpol.2013.01.014.
Joo, R., Bertrand, S., Tam, J., and Fablet, R. 2013. Hidden Markov models: the best models for forager movements?. PLoS ONE, **8**(8): e71246. doi:10.1371/journal.pone.0071246.
Lazaridis, E. 2014. lunar: lunar phase & distance, seasons and other environmental factors (Version 0.1-04). Available from http://statistics.lazaridis.eu [accessed January 2023].
Liaw, A., and Wiener, M. 2002. Classification and regression by random Forest. R news, **2**(3): 18–22.
Mamula, A., Thomas-Smyth, A., Speir, C.L., Kosaka, R., and Pearson, D.E. 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. 76p. doi:10.25923/7ag1-af45.
McCluskey, S.M., and Lewison, R.L. 2008. Quantifying fishing effort: a synthesis of current methods and their applications. Fish Fish. **9**(2): 188–200. doi:10.1111/j.1467-2979.2008.00283.x.
Mills, C.M., Townsend, S.E., Jennings, S., Eastwood, P.D., and Houghton, C.A. 2007. Estimating high resolution trawl fishing effort from satellite-based vessel monitoring system data. ICES J. Mar. Sci. **64**(2): 248–255. doi:10.1093/icesjms/fsl026.
Muench, A., DePiper, G.S., and Demarest, C. 2018. On the precision of predicting fishing location using data from the vessel monitoring system (VMS). Can. J. Fish. Aquat. Sci. **75**(7): 1036–1047. doi:10.1139/cjfas-2016-0446.
Nakagawa, S., and Cuthill, I.C. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol. Rev. **82**(4): 591–605. doi:10.1111/j.1469-185X.2007.00027.x.
NMFS. 2016. National Observer Program, National Marine Fisheries Service. Available from http://www.st.nmfs.noaa.gov/st4/nop/[accessed 11 July 2016].
O'Farrell, S., Sanchirico, J.N., Chollett, I., Cockrell, M., Murawski, S.A., Watson, J.T., et al. 2017. Improving detection of short-duration fishing behaviour in vessel tracks by feature engineering of training data. ICES J. Mar. Sci. **74**(5): 1428–1436. doi:10.1093/icesjms/fsw244.
O'Farrell, S., Sanchirico, J.N., Spiegel, O., Depalle, M., Haynie, A.C., Murawski, S.A., et al. 2019. Disturbance modifies payoffs in the explore-exploit trade-off. Nat. Commun. **10**(1): 1–9. doi:10.1038/s41467-019-11106-y.
Palmer, M.C., and Wigley, S.E. 2009. Using positional data from vessel monitoring systems to validate the logbook-reported area fished and the stock allocation of commercial fisheries landings. N. Am. J. Fish. Manage. **29**(4): 928–942. doi:10.1577/M08-135.1.
Pante, E., and Simon-Bouhet, B. 2013. marmap: a package for importing, plotting and analyzing bathymetric and topographic data in R. PLoS ONE, **8**(9): e73051. doi:10.1371/journal.pone.0073051.
Pedersen, S.A., Fock, H.O., and Sell, A.F. 2009. Mapping fisheries in the German exclusive economic zone with special reference to offshore Natura 2000 sites, Mar. Pol. **33**: 571–590. doi:10.1016/j.marpol.2008.12.007.
Perruso, L., Chollett, I., Michaelis, A., and O'Farrell, S. in prep. Fisheries data needs for better spatial planning in the Gulf of Mexico.
Piet, G.J., and Quirijns, F.J. 2009. The importance of scale for fishing impact estimations. Can. J. Fish. Aquat. Sci. **66**(5): 829–835. doi:10.1139/F09-042.
R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/ [accessed January 2023].
Riley, K.L., Wickliffe, L.C., Jossart, J.A., MacKay, J.K., Randall, A.L., Bath, G.E., et al. 2021. An aquaculture opportunity area Atlas for the US Gulf of Mexico. NOAA Technical Memorandum NOS NCCOS 299. 545p. doi:10.25923/8cb3-3r66.
Russo, T., D'Andrea, L., Parisi, A., and Cataudella, S. 2014. VMSbase: an R-package for VMS and logbook data management and analysis in fisheries ecology. PLoS ONE, **9**(6): e100195. doi:10.1371/journal.pone.0100195.
Russo, T., Parisi, A., Prorgi, M., Boccoli, F., Cignini, I., Tordoni, M., and Cataudella, S. 2011. When behaviour reveals activity: assigning fishing effort to métiers based on VMS data using artificial neural networks. Fish. Res. **111**(1-2): 53–64. doi:10.1016/j.fishres.2011.06.011.
Russo, T., Morello, E.B., Parisi, A., Scarcella, G., Angelini, S., Labanchi, L., et al. 2018. A model combining landings and VMS data to estimate landings by fishing ground and harbor. Fish. Res. **199**: 218–230. doi:10.1016/j.fishres.2017.11.002.
Saunders, F.P., Gilek, M., and Tafon, R. 2019. Adding people to the sea: conceptualizing social sustainability in maritime spatial planning. *In* Maritime spatial planning. *Edited by* J. Zaucha and K. Gee. Palgrave Macmillan, Cham. pp. 175–199.
Scales, K.L., Hazen, E.L., Maxwell, S.M., Dewar, H., Kohin, S., Jacox, M.G., et al. 2017. Fit to predict? Eco-informatics for predicting the catchability of a pelagic fish in near real time. Ecol. Appl. **27**(8): 2313–2329. doi:10.1002/eap.1610.
St. Martin, K.S., and Olson, J. 2017. Creating space for community in marine conservation and management: Mapping "Communities-at-Sea." *In* Conservation for the Anthropocene Ocean. Academic Press, Cambridge, MA. pp. 123–141.
Strobl, C., Malley, J., and Tutz, G. 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol. Methods. **14**(4): 323. doi:10.1037/a0016973.
Torres-Irineo, E., Salas, S., Euán-Ávila, J.I., Palomo, L.E., Quijano Quinones, D.R., Coronado, E., and Joo, R. 2021. Spatio-temporal determination of small-scale vessels' fishing grounds using a vessel monitoring system in the Southeastern Gulf of Mexico. Front. Mar. Sci. **8**: 542. doi:10.3389/fmars.2021.643318.
Upton, HF. 2011. The deepwater horizon oil spill and the Gulf of Mexico fishing industry. Congressional Research Service. Congressional Research Service, InLibrary of Congress.
Villareal, T.A., Hanson, S., Qualia, S., Jester, E.L., Granade, H.R., and Dickey, R.W. 2007. Petroleum production platforms as sites for the expansion of ciguatera in the northwestern Gulf of Mexico. Harmful Algae, **6**(2): 253–259. doi:10.1016/j.hal.2006.08.008.
Watson, J.T., and Haynie, A.C. 2016. Using vessel monitoring system data to identify and characterize trips made by fishing vessels in the United States North Pacific. PLoS ONE, **11**(10): e0165173. doi:10.1371/journal.pone.0165173.
Watson, J.T., Haynie, A.C., Sullivan, P.J., Perruso, L., O'Farrell, S., Sanchirico, J.N., and Mueter, F.J. 2018. Vessel monitoring systems (VMS) reveal an increase in fishing efficiency following regulatory changes in a demersal longline fishery. Fish. Res. **207**: 85–94. doi:10.1016/j.fishres.2018.06.006.
Wessel, P., and Smith, W.H. 1996. A global, self-consistent, hierarchical, high-resolution shoreline database. J. Geophys. Res.: Solid Earth. **101**(B4): 8741–8743. doi:10.1029/96JB00104.
Wickham, H., Francois, R., Henry, L., and Müller, K. 2015. dplyr: a grammar of data manipulation. R package version 0.4., **3**: p.156.