frontiers | Frontiers in Sustainable Food Systems

TYPE Original Research
PUBLISHED 04 May 2023
DOI 10.3389/fsufs.2023.1152226



OPEN ACCESS

EDITED BY
Tomaso Fortibuoni,
Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), Italy

REVIEWED BY
Easton R. White,
University of New Hampshire, United States
Domitila Kyule,
Kenya Marine and Fisheries Research Institute, Kenya

*CORRESPONDENCE
Gianpaolo Coro
✉ gianpaolo.coro@cnr.it

†These authors have contributed equally to this work

RECEIVED 27 January 2023
ACCEPTED 13 April 2023
PUBLISHED 04 May 2023

CITATION
Coro G, Sana L, Ferrà C, Bove P and Scarcella G (2023) Estimating hidden fishing activity hotspots from vessel transmitted data. *Front. Sustain. Food Syst.* 7:1152226. doi: 10.3389/fsufs.2023.1152226



# Estimating hidden fishing activity hotspots from vessel transmitted data

Gianpaolo Coro[1*†], Lorenzo Sana[1†], Carmen Ferrà[2,3†], Pasquale Bove[1†] and Giuseppe Scarcella[2†]

[1]Institute of Information Science and Technologies (ISTI), National Research Council of Italy (CNR), Pisa, Italy, [2]Institute for Biological Resources and Marine Biotechnology (IRBIM), National Research Council of Italy (CNR), Ancona, Italy, [3]Department of Biological, Geological and Environmental Sciences (BiGeA), University of Bologna, Bologna, Italy

Monitoring fishery activity is essential for resource planning and guaranteeing fisheries sustainability. Large fishing vessels constantly and continuously communicate their positions via Automatic Identification System (AIS) or Vessel Monitoring Systems (VMSs). These systems can use radio or Global Positioning System (GPS) devices to transmit data. Processing and integrating these big data with other fisheries data allows for exploring the relations between socio-economic and ecosystem assets in marine areas, which is fundamental in fishery monitoring. In this context, estimating actual fishing activity from time series of AIS and VMS data would enhance the correct identification of fishing activity patterns and help assess regulations' effectiveness. However, these data might contain gaps because of technical issues such as limited coverage of the terrestrial receivers or saturated transmission bands. Other sources of data gaps are adverse meteorological conditions and voluntary switch-offs. Gaps may also include hidden (unreported) fishing activity whose quantification would improve actual fishing activity estimation. This paper presents a workflow for AIS/VMS big-data analysis that estimates potential unreported fishing activity hotspots in a marine area. The workflow uses a statistical spatial analysis over vessel speeds and coordinates and a multi-source data integration approach that can work on multiple areas and multiple analysis scales. Specifically, it (i) estimates fishing activity locations and rebuilds data gaps, (ii) estimates the potential unreported fishing hour distribution and the unreported-over-total ratio of fishing hours at a 0.01° spatial resolution, (iii) identifies potential unreported fishing activity hotspots, (iv) extracts the stocks involved in these hotspots (using global-scale repositories of stock and species observation data) and raises an alert about their possible endangered, threatened, and protected (ETP) status. The workflow is also a free-to-use Web Service running on an open science-compliant cloud computing platform with a Web Processing Service (WPS) standard interface, allowing efficient big data processing. As a study case, we focussed on the Adriatic Sea. We reconstructed the monthly reported and potential unreported trawling activity in 2019, using terrestrial AIS data with a 5-min sampling period, containing ~50 million records transmitted by ~1,600 vessels. The results highlight that the unreported fishing activity hotspots especially impacted Italian coasts and some forbidden and protected areas. The potential unreported activity involved 33 stocks, four of which were ETP species in the basin. The extracted information agreed with expert studies, and the estimated trawling patterns agreed with those produced by the Global Fishing Watch.

KEYWORDS

**big data, vessel transmitted data, spatial analysis, statistical analysis, fisheries, vulnerable species, cloud computing, open science**

## 1. Introduction

Fishing is one of the most impacting human activities on marine resources (Kroodsma et al., 2018). Monitoring fishing activity pressure on marine resources, protected species and areas is essential to guarantee food availability from the sea and safeguard ecosystems (Bergh and Davies, 2002; Gianelli et al., 2018; Lockerbie et al., 2018; Muawanah et al., 2018; Koen-Alonso et al., 2019). Ecosystem approaches to marine resource management analyse the causal relations between economic and human activities, and their pressure (emissions, waste, intensity) on the ecosystems' chemical and biological states (Antunes and Santos, 1999; Kristensen, 2004). Integrated Environmental Assessment systems (IEAs) are computer science workflows that estimate these relations by extracting and combining information from fisheries data collections containing complex, heterogeneous, large-volume, and noisy data (Coro et al., 2021). It is primary for IEA to estimate fishing activity patterns, understand human-activity pressure distribution, and asses if regulations and management strategies correctly contributed to ecological and economic sustainability (Robards et al., 2016; Le Tixerant et al., 2018; Coro et al., 2022b). However, data gaps can hide much fishing activity (*unreported fishing activity*). Filling these gaps would improve fishing activity estimation accuracy. Following international regulations, large fishing vessels over 15 m in length overall must be equipped with transmission devices that communicate their positions to monitoring systems such as Automatic Identification System (AIS) and Vessel Monitoring Systems (VMSs) (European Parliament, 2002, 2011). Communication is generally based on satellite or radio communication devices (Chang, 2003; Agnew et al., 2009; ITU, 2009; Previero and Gasalla, 2018; Kurekin et al., 2019). AIS and VMS data collectors in the European seas receive billions of transmitted data records per month, which include information on coordinates, speed, route, date, time, and vessel identity. The data transmission period can be of a few seconds or minutes. However, records could get lost because of limited terrestrial receiver's range coverage (for radio systems), satellite communication hindrances, onboard technical issues, adverse meteorological conditions, or voluntary transmission device switch-offs (Taconet et al., 2019). Vessel-transmitted data are the basis of many IEAs that analyse fishing patterns (GFW, 2020). However, the heterogeneous nomenclatures, accessibility, process-interoperability, and scalability issues across the vessel-transmitted data collections make data and processes suitable only for specific areas and thus limit their re-usability (Gari et al., 2015; Taconet et al., 2016; James et al., 2018). Only a few systems publish data as findable, accessible, interoperable, and re-usable (FAIR) data (Jennings and Lee, 2012; Dunn et al., 2018; Song et al., 2018; Depellegrin et al., 2020).

Integrating vessel-transmitted data with gear, logbook, and port information can enhance the precision of fishing activity detection (Kia et al., 2000; Davis, 2001; Palmer and Wigley, 2009; Lee et al., 2010; Gerritsen and Lordan, 2011; Olesen et al., 2012; Shaw et al., 2017; Muench et al., 2018; Roberson et al., 2019). However, integration is rarely possible for extensive data collections containing anonymous data or involving vessels coming from far ports and different nations. Several studies have proposed data processing workflows, working on minimal transmitted information (such as AIS and VMS location, speed, and course data), to estimate the volumes of illegal, unreported, and unregulated (IUU) fishing activity (Tetreault, 2005; Bastardie et al., 2010; Eriksen et al., 2010; Gerritsen and Lordan, 2011; Pallotta et al., 2013; Natale et al., 2015; Le Guyader et al., 2017; Le Tixerant et al., 2018; Shepperson et al., 2018; Kurekin et al., 2019; Mullié, 2019; Tassetti et al., 2019; Yang et al., 2019; Belhabib et al., 2020). These workflows often use rule-based or machine-learning algorithms to automatically classify vessel activity based on coordinate, speed, direction, and geo-morphological information (Coro et al., 2013; de Souza et al., 2016). In some cases, they conduct analyses by vessel gear and thus preliminarily identify the used fishing gears (Coro et al., 2022b). Machine learning-based models learn to detect IUU fishing patterns as anomalous patterns in vessel tracks (Agnew et al., 2009; Ford et al., 2018; Singh and Heymann, 2020; Wolsing et al., 2022). However, the specificity of model training on particular areas makes them lowly re-usable for other areas. Moreover, they usually do not estimate potential IUU fishing activity that may occur in vessel track gaps. Some approaches combine synthetic-aperture radar (SAR) images with AIS data to gain better process scalability (Galdelli et al., 2021). These approaches focus on estimating the correspondence between transmitted data and actual vessel presence, and consequently estimate unreported vessel presence (Perez et al., 2013; Pew Trusts, 2015; HawkEye360, 2020; Cutlip, 2022). However, they do not distinguish unreported activity per fishing activity type (e.g., trawling, purse seine, etc.) and wrongly include small vessels that are not equipped with transmission devices.

This paper presents a workflow to estimate potential hidden fishing activity hotspots in a marine area from big data of temporal sequences of vessels' coordinates and speeds belonging to an AIS or VMS collection. With the expression "potential hidden fishing activity" we indicate that the detected hidden-activity hotspots very likely correspond to unreported fishing activity. However, for simplicity, hereafter we will just use the more general expression "unreported fishing activity."

The workflow is easily scalable from small to extensive data collections. It only requires prior indications about the fishing speed ranges on which the analysis should focus, which it revises through statistical data analysis. The workflow detects fishing activity in the data gaps and produces aggregated distributions of reported and unreported fishing activity and unreported activity hotspots. Moreover, it estimates the stocks potentially targeted in the hotspots and their "endangered, threatened, and protected" (ETP) status. We released our workflow as a cloud-computing Web service (Coro et al., 2015, 2017) invocable through the Web Processing Service standard (WPS) of the Open Geospatial Consortium (OGC) (Schut and Whiteside, 2007). This service supports Open Science-oriented features such as experiment reproduction, replication, re-use, and integration in other workflows and IEAs (Coro et al., 2021).

We tested our workflow effectiveness at reporting valuable information for fisheries management organizations and national governments in the Adriatic Sea. We analyzed terrestrial AIS data in this basin with a 5-min sampling period, and reconstructed the reported and unreported trawling vessels' activity in all months of 2019. This was a "regime" condition that approximated the full-potential volume of the fishery, unaltered by the COVID-19 pandemic (and consequent restrictions) and economic failures occurring from 2020 onwards (Coro et al., 2022b). Trawling activity in the Adriatic has the most significant volumes of catch and vessels in the Mediterranean (Mannini et al., 2005; FAO, 2020) and among the highest intensity globally (Amoroso et al., 2018). We evaluated if the estimated monthly fishing patterns of unreported trawling reflected known deviations from regulations and restrictions across several areas monitored by fisheries management organizations. We also verified the consistency of the extracted target and ETP stock-species in the unreported-fishing locations. Finally, we demonstrated that the detected trawling-activity patterns largely agree with those published by Global Fishing Watch (GFW) (Merten et al., 2016), estimated through a machine learning model.

In summary, this paper describes a multi-source statistical workflow that can be integrated with larger IEAs to estimate the volume of unreported fishing activity (due to technical issues or voluntary switch-offs) and contribute to improving the characterization of fishing activity in terms of compliance and impact on ecosystems. It also indicates the stocks subject to the most significant unreported fishing activity, along with their ETP status, to inform stock assessment models. The workflow application goes beyond the presented use case. It can scale from large to small areas because it integrates global-scale data sources, uses algorithms that adapt the analysis to the study-area data, and runs on a cloud computing platform.

## 2. Methods

This section first describes our workflow (Section 2.1), then the Web service version that complies with the Open Science paradigm (Section 2.2), and finally the evaluation case study and approach (Section 2.3).

### 2.1. Workflow description

Our workflow was entirely developed in R and is open source (see section Supplementary information). This section describes it, following the schema in Figure 1.

#### 2.1.1. Input data pre-processing

The workflow input data are assumed to come from one extensive AIS or VMS data collection. Our workflow requires that these data contain records with (at least) unique vessel identifiers and a temporal sequence of speed, longitude and latitude indications for each vessel. The data input format is Comma Separated Value (CSV), with each row containing at least one column for the required information (hereafter named *vessel-id*, *speed*, *x*, *y*, and *date-time*). The user should indicate the names of the CSV columns containing the required information. The input file should refer to a group of vessels acting in a limited time period (e.g., 1 month or 1 year). As a pre-filtering stage, our workflow deletes duplicate records and records with less than a second of mutual temporal distance. It also deletes records with coordinates falling on land, through the intersection with a global-scale land-coverage raster file. Records with implausible speeds [over 20 kn (Zhang et al., 2021)] are also deleted. The data are thus ordered temporally for each vessel, and the time differences (in min) between consecutive transmission records are calculated. The minimum average time difference of non-zero short-term gaps (<30 min), across all vessel tracks, is taken as the sampling period of the data. Finally, the workflow internally uses a global dataset of 5,729 harbor and port locations to exclude too close vessel positions (<1.2 nautical miles) that would unlikely correspond to fishing activity. This dataset was obtained by merging, cleaning, and harmonizing the harbor and port locations available on the EMODNET (EMODNET, 2022) and Marine Vessel Traffic (Marine Vessel Traffic, 2022) repositories. It is internally represented as a CSV file that could possibly be substituted with a more detailed dataset for local-scale analyses. The workflow internally has an option to skip the harbor and port proximity filter if not required.

Overall the pre-processing stage can be summarized as follows:

```
Read the input CSV file
Extract the vessel-id, speed, x, y, and date-time
columns (data records)
Delete duplicate records
Delete records with coordinates falling on land
Delete records with speed >20 kn
Delete records within 1.2 NM from harbors and
ports
For each vessel-id:

  Order transmission records temporally
  Calculate the time difference (time gap) in
  minutes between consecutive records
  Merge records with a time difference <1s
  Average the time differences that are >   0 and
  < 30 min;

Set the minimum average-time difference as the
data sampling period.
```

Algorithm 1. Data pre-processing.

As the result of this algorithm, the workflow produces a new table with clean and ordered records organized by vessel-id with a time gap indication for each record.

#### 2.1.2. Vessel activity classification and gap filling

Time gaps can be due to transmitter-receiver communication issues (e.g., saturated transmission band, temporary receiver or transmitter malfunction, poor receiver signal power) or voluntary switch-offs. Long gaps might indicate that a new vessel track started (e.g., after one night of resting), and the associated record sequence should be processed independently of the others. New vessel tracks should therefore be separated from other vessel tracks. Data gaps should be reconstructed only when the time gap is relatively short, otherwise the vessel movements' reconstruction would be highly

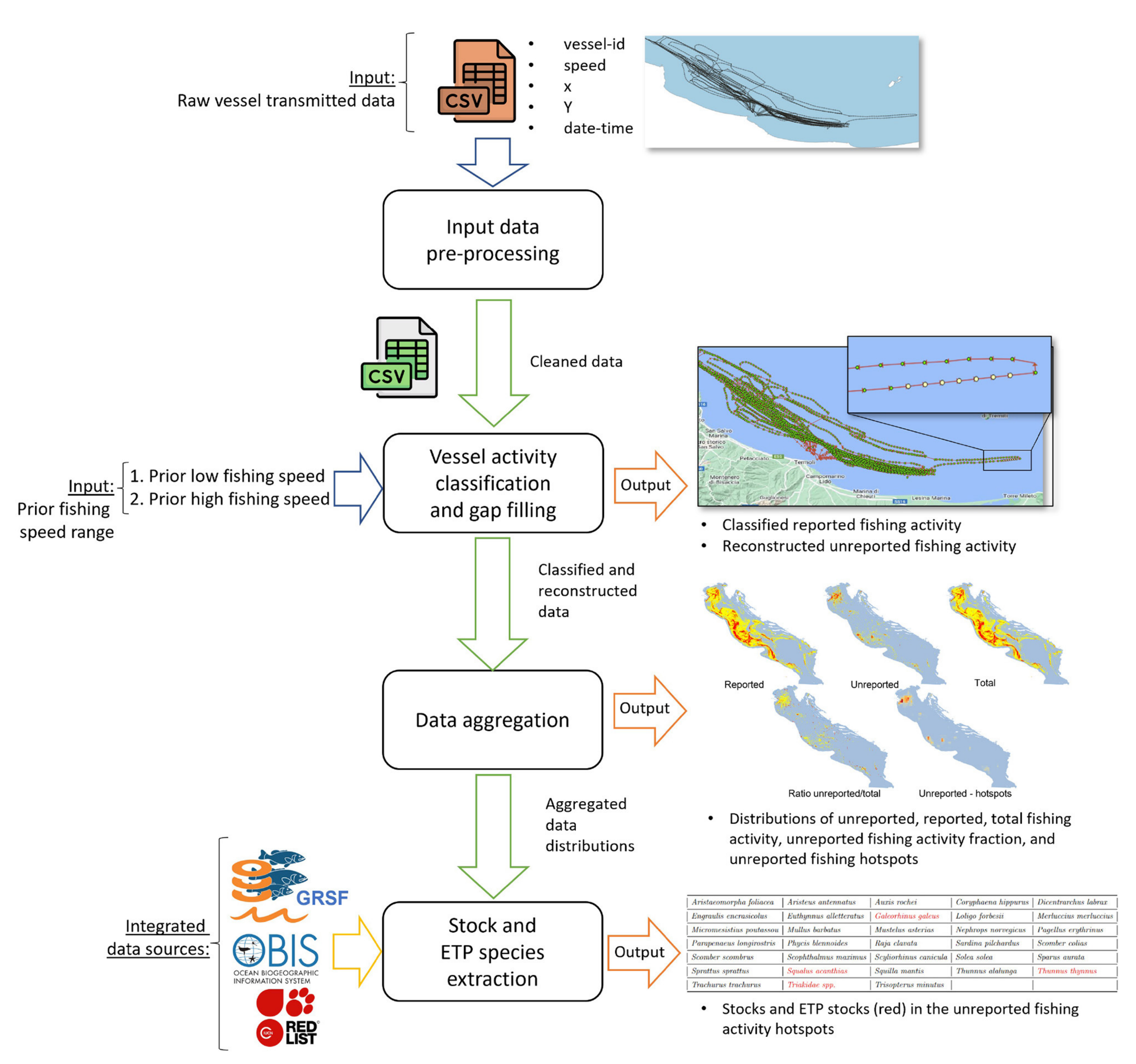


FIGURE 1
Conceptual schema of our workflow.

uncertain. It is thus crucial to estimate the time gap threshold that separates the data gaps that can be reasonably reconstructed from too-long gaps. We calculated this threshold by analyzing the time gap density distributions of the AIS data of our use case (Section 2.3), of the GFW training corpus (Global Fishing Watch, 2022), and of the BOEM Marine Cadastre (BOEM, 2020) repository used in other AIS and VMS data analyses (Coro et al., 2013, 2021). These distributions showed a clear separation between long and short gaps. Two distribution peaks could be separated for time gaps below 30 min (short gaps) and above 3.5 h (long gaps). Long gaps averagely corresponded to new tracks, whereas short gaps corresponded to temporary transmission loss. The time gaps between these two limits (*gaps-to-reconstruct*) were the data gaps to be filled. Tracks between these data gaps can be reconstructed through linear interpolation (straight line course assumption), considering the ordinarily low vessels' speed during fishing activity (Ferrà et al., 2020) (∼15 km in 3 h for a trawler). Our workflow linearly interpolates the gaps-to-reconstruct by generating a new point record (with associated speed) each sampling period.

An additional input to our workflow is a prior speed range of fishing activity (low and high prior fishing speed limits). This range can vary from one fishing activity type to another (e.g., it is typically between 2 and 4 kn for trawlers and 6 and 15 kn for tuna purse seiners) and usually identifies activity performed with a set of gears (e.g., bottom and mid-water trawling gears or purse seines). Indeed, a coarser precision on gear identification might extract more reliable results (Ferrà et al., 2020; Coro et al., 2021). Similarly to Bayesian approaches, our workflow revises the prior speed range based on a statistical analysis of speed distributions. It first cuts the total speed density distribution between the low and high prior limits (with a 50% tolerance). Then, it fits a logistic distribution to the data and uses the upper and lower confidence limits as new fishing speed limits. The use of a logistic distribution for this operation came after fitting several alternative distributions to our study case data and the GFW and Marine Cadastre data. This approach overcomes the issue of managing different speed ranges depending on the study area and analysis period (Reid et al., 2011; Coro et al., 2013) by re-adapting speed ranges to the data at hand. All data records with speeds falling within these limits are marked as fishing locations. The reconstructed gaps with speeds compatible with fishing activity are marked as *unreported fishing activity*. All other fishing locations are marked as *reported fishing activity* (re-using a common expression of VMSs). Figure 2 reports an example of one vessel's tracks in January 2019, off the Termoli harbor in Italy, with the indication of reported fishing locations and a zoom on one sequence of reconstructed unreported-fishing locations.

Overall, the vessel activity classification algorithm can be summarized as follows:

```
Mark the records with time gaps between 30 min
and 3.5 h as gaps-to-reconstruct
Linearly interpolate the gaps-to-reconstruct
at sampling-period steps and assign average
start-end speed to each reconstructed point
Read the low-high prior fishing speed range
Calculate the speed density distribution between
(low speed − 50% low speed) and (high speed + 50% high speed)
Fit a logistic distribution to the speed density
Record the logistic distribution lower and upper
confidence limits
For each record:

  If (lower confidence limit < speed < upper confidence limit), mark
  the record as fishing location

    If the record was reconstructed, mark it as
    unreported fishing location
    Else, mark it as reported fishing location

Return all marked vessel track records.
```

Algorithm 2. Vessel activity classification.

As the result of this algorithm, the workflow produces a new table with unreported and reported fishing location records.

### 2.1.3. Data aggregation

Based on the table produced by the previous step, the workflow creates a grid of spatial cells with a square kilometer resolution (∼0.01°). Four data aggregations are generated by summing the time gaps of the classified records onto this spatial grid: (i) reported fishing hours, (ii) unreported fishing hours, (iii) total fishing hours, and (iv) the ratio between unreported and total fishing hours. The hour ratio highlights the locations where unreported activity dominates reported activity, which is essential to detect potential IUU fishing activity or poor communication zones (Section 3.1). The grids are further processed to categorize the fishing hours as low/medium/high fishing activity intensity. To this aim, the workflow fits a log-normal distribution to each distribution. It uses lower and upper confidence limits to classify the fishing hours. The choice of using a log-normal distribution derives from a multi-distribution test analysis similar to the one used for vessel activity classification. Intuitively, it is reasonable because high-intensity fishing activity usually concentrates in a few hotspots, whereas many sparse cells in the study area present a much lower fishing intensity. Alternatively, the workflow internally allows to set classification thresholds to fixed value, e.g., to produce results categorized with respect to the ranges of another dataset (Section 3.1). As an additional step, the high-intensity unreported fishing locations are processed through a kernel density estimator to produce unreported activity hotspots. We use quartic kernel shape and automatic bandwidth estimation at 0.1° spatial resolution (∼11 km) for this operation [through the MASS R package Ripley (2022)]. The hotspot intensity is finally categorized as low/medium/high intensity through a linear separation of the kernel density range.

```
For each 0.01° spatial cell in the study
area:

  Sum the reported fishing hours in the cell
  Sum the unreported fishing hours in the cell
  Sum the reported + unreported fishing hours (total
  fishing hours) in the cell
  Calculate the summed unreported fishing hours/total fishing hours
  in the cell

For each spatial aggregation:

  Fit the fishing hour distribution to a
  log-normal distribution
  Calculate the lower and upper confidence limits
  Categorize each cell as one among:

    low-intensity (fishing hours < lower confidence limit)
    medium-intensity (lower confidence limit < fishing hours <
    upper confidence limit)
    high-intensity (fishing hours > upper confidence limit)

Save all classified aggregations as CSV files
Apply a 0.1°-resolution kernel density estimation
to the high-intensity unreported fishing cells to
produce unreported fishing hotspots
Classify hotspot intensity as low, medium, or
high through linear range subdivision
Save the raw hotspot and the classified hotspot
as two GeoTiff files.
```

Algorithm 3. Data aggregation.

As the result of this algorithm, the workflow produces four tables (as CSV files) with unreported, reported, total, and unreported/total fishing data at 0.01° spatial resolution, with a

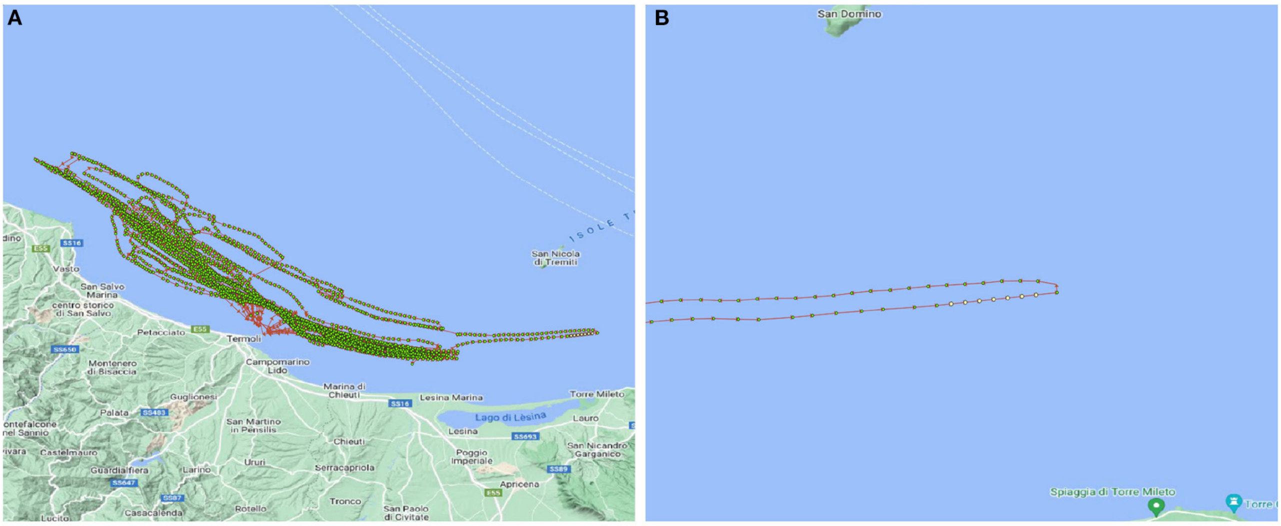


FIGURE 2
**(A)** Example of estimated trawling activity (green dots) for all trajectories of one vessel in January 2019 off the Termoli harbor (Italy). The arrows indicate the temporal direction of coordinate sequence. **(B)** Focus on unreported fishing activity points reconstructed between two reported fishing activity points (yellow dots).

fishing intensity categorization for each cell. It also produces two raster files (in GeoTiff format) with unreported fishing activity hotspots (at 0.1° resolution), one with the raw kernel density estimation and the other with hotspot intensity categorization as low/medium/high intensity.

### 2.1.4. Stock and ETP species extraction

As the final step of our workflow, the unreported fishing activity hotspots are intersected with open-access repositories of stock and species-observation data to detect the potential vessels' targets in those locations. This information is crucial to verify if species known to be the target of IUU fishing activity were involved in the unreported activity and give possible spatial references to this activity. For this task, our workflow uses the semantic knowledge base of the Global Record of Stocks and Fisheries (GRSF) (i-Marine, 2020). The GRSF is an authoritative open-access semantic knowledge base maintained and updated by the Food and Agriculture Organization of the United Nations (FAO). It integrates data from three authoritative sources: Fisheries and Resources Monitoring System (FIRMS), RAM Legacy Stock Assessment Database (RAM), and FishSource (Program of the Sustainable Fisheries Partnership). Each stock and fishery have unique identifiers and geospatial activity areas associated. A SPARQL endpoint allows sending queries to retrieve stock-related fisheries and fishing areas. Our workflow sends a SPARQL query for each 0.1° hotspot cell to retrieve the target stocks of the fisheries in that cell.

Although valuable for retrieving the potential list of target stocks in an area, the GRSF areas have a too-coarse resolution (e.g., at the entire basin scale for the Adriatic) to be used for hotspot-scale analyses. For this reason, our workflow checks if the GRSF stocks were actually observed in the extracted hotspots. It performs this operation by using the Ocean Biodiversity Information System (OBIS) (Grassle, 2000), a global, open-access, and authoritative database on marine biodiversity supported by the United Nations Educational, Scientific and Cultural Organization (UNESCO). Our workflow uses OBIS [through the robis R package Provoost et al. (2017)] to check if at least one expert-verified observation is available for the GRSF stock species in the unreported fishing activity hotspots. If a GRSF-stock presence was observed and certified by an expert, it is classified as a potential target stock in the hotspots. Using expert-verified records improves the reliability of this classification. OBIS also allows retrieving the stock's ETP status in the study area, according to the Red List of Threatened Species of the International Union for Conservation of Nature (IUCN) (IUCN, 2001). Stocks with vulnerable, endangered, or critically-endangered IUCN-status are marked as ETP species.

Overall, the stock and ETP species retrieval algorithm can be summarized as follows:

```
For each 0.1° cell in the unreported fishing
activity hotspots:

  Retrieve the list of stocks from the GRSF
  through a SPARQL query
  For each stock:

    Query OBIS to retrieve expert-verified
    observations for the stock-species in the
    cell
    If an observation exists, mark the species
    as potential target stock and report its ETP
    status

Return all found potential target stocks with
their ETP status.
```

Algorithm 4. Stock and ETP retrieval.

As the result of this algorithm, the workflow produces a table (in CSV format) with the list of potential target stocks in the unreported fishing activity hotspots, with their associated ETP status. It is important to highlight that the extracted stocks are related to the specific estimated unreported fishing activity hotspots. However, these stocks could also include those of other fishing activities having the same hotspots. Therefore, the stock list should be possibly revised by experts to discard stocks not belonging to the analyzed fishing activity. To enhance the stock-list reliability, information on the used gears should be available from both the fishing-activity classification algorithm and the GRSF. We have planned future actions for overcoming this issue (Section 4).

## 2.2. Web service

Our workflow was developed as an open-source R script internally combining functions to operate big data processing and aggregations (Section 2.1). We published the workflow as a multi-source, parallel, secure, and Open Science-oriented (Hey et al., 2009) Web service. To this aim, we integrated our workflow with the DataMiner cloud computing platform of the D4Science e-Infrastructure (Coro et al., 2015, 2017; Candela et al., 2016; Assante et al., 2019). This platform publishes the hosted processes under the WPS standard, which allows to directly integrate it in other geospatial data processing software supporting this standard (e.g., QGIS and ArcGIS) (Coro, 2020). Moreover, the DataMiner automatically produces a graphic user interface based on the workflow input/output data and parameter definitions (see section Supplementary information). Our process interface requires a CSV input file containing coordinate, speed, date/time, and vessel identifier records in each row. The user should also specify the names of the CSV columns containing this information (the interface includes a helper for this operation). Moreover, it requires specifying the prior fishing speed range. Input and output big data can be uploaded on an integrated distributed storage system (Assante et al., 2019).

The results produced for our study case are available as CSV files and in OGC-compliant formats (see section Supplementary information). The GRSF is hosted and managed by the same D4Science e-infrastructure, which ensures fast access and high service availability. The workflow accesses OBIS via direct connection through provider-specific libraries (Provoost et al., 2017). DataMiner supports our workflow with 15 machines equipped with Ubuntu 18.04.5 LTS x86 64 operating system, 16 virtual cores, 32 GB of Random Access Memory, and 100 GB of disk for each machine, and can manage executions in distributed and concurrent modes. Furthermore, the parameters, input and output data of each execution are tracked in a user's private data space as XML documents following the Prov-O ontological specifications (Lebo et al., 2013) for provenance tracking, which is integral for computational reproducibility and tracking experiment history (Koop et al., 2011; Freire et al., 2012). This way, the platform enables our workflow with repeatability, reproducibility, re-usability, and interoperability features. Integrating our workflow with D4Science also allowed us to process several big data collections concurrently, e.g., to study monthly activity change and long-term fishing activity quickly.

## 2.3. Case study

As a study case, we reconstructed the reported and unreported bottom and pelagic-pair trawling activity (hereafter indicated as *trawling* activity) in the Adriatic Sea in all months of 2019. We analyzed terrestrial-AIS data bought from the authoritative Astra Paging provider for the geographical sub-areas 17 and 18 (Adriatic Sea) of the General Fisheries Commission of the Mediterranean (GFCM), with a 5-min sampling period. We focused our analysis on *trawling* activity, which produces ∼70% of the catch in the basin (∼100k tons) and involves over ∼1,600 vessels fishing for ∼10k hours annually (Mannini et al., 2005; FAO, 2020). The analyzed dataset contained ∼50 million records transmitted by a consistent part of these vessels - most of them having at least 15 m length-overall - committed to equipping AIS transmitters according to international regulations. In agreement with other studies (Coro et al., 2013, 2021), the prior speed of trawling activity was set to 2–4 kn. The cloud computing platform was used to parallelise the process over the 2019 months using different parametrisations to verify the workflow sensitivity to prior speed ranges and reconstructed gap lengths, optimize the workflow code, and correct possible errors.

Adriatic Sea trawling is constantly monitored by national authorities and regulated by fisheries management organizations (especially the GFCM), which annually assess if the fishing effort (the total fishing hours) is commensurate with stock status and long-term planned sustainability goals, in accordance with the GFCM "Multi-annual Management Plan" for small pelagic and demersal species (General Fisheries Commission for the Mediterranean, 2013). Adriatic governmental authorities (Albania, Bosnia and Herzegovina, Croatia, Italy, Montenegro, and Slovenia) regulate the allowed fishing hours based on the annual estimated pressure on target stocks abundance, by limiting fishing days and applying spatial closures on vulnerable species, spawning, nursing, dangerous, and protected areas (Fonda et al., 1992; Froese et al., 2018b; Coro et al., 2022b). Over a year, fishing patterns change due to seasonal and monthly regulations and restrictions, but stocks are subject to moderate-high fishing pressure in all months of the year. Adriatic trawlers act as insatiable predators on their target resources (Coro et al., 2022b); therefore, estimating the actual trawlers total fishing hours is crucial to assess their sustainability and minimize the risk of stock depletion.

We targeted the estimation of unreported trawling activity and effort in "regime" conditions that approximated the full-potential volume of the fishery. Therefore, we selected the 2019 conditions as representative of standard fishing effort levels unaltered by the COVID-19 pandemic, and the consequent restrictions and economic failures occurring after 2020. For each 2019 month, our workflow (i) identified trawling areas, (ii) estimated unreported trawling activity hotspots, and (iii) identified the stocks potentially involved in the unreported areas along with their vulnerability status. We projected the unreported activity hotspots onto zones monitored by fisheries management organizations (*sensitive* zones), such as marine protected areas and restricted areas. We used the density of hotspots falling in these zones as an impact measurement. Specifically, we focussed on seven sensitive zones currently monitored and regulated in the Adriatic: (i) coast-distance ban (between 1-6 NM, depending on the month and the


FIGURE 3
Sequence of monthly unreported trawling activity hotspots in the Adriatic in 2019, estimated through kernel density at 0.1° spatial resolution. The classification colors were assigned using the same classification thresholds for all months.

area), (ii) trawling-restricted areas, (iii) Pomo Pit fishery restricted area (FRA), (iv) marine protected areas (MPAs), (v) offshore platforms, (vi) areas with depth over 1,000 m, and (vii) Zones of Biological Protection (ZTB or BPA). We evaluated if the monthly hotspots and impact patterns corresponded to those reported by other expert studies and if the extracted target and ETP stocks were consistent with official reports.

As an additional evaluation, we compared our reported-trawling distribution with the one downloadable from the GFW Web portal at 0.1° resolution (Clavelle, 2022), to verify that they produced similar patterns. The GFW portal is managed by Google in partnership with Oceana and SkyTruth. It produces a global view of commercial fishing activities by collecting and analyzing VMS, AIS, and SAR data. The GFW data also report aggregated fishing activity for trawlers, estimated through a supervised machine learning model mainly based on speed information (de Souza et al., 2016). We calculated the fraction of 0.1° cells for which our distribution and the GFW data matched their classifications of fishing and non-fishing locations (accuracy). For this comparison, we up-scaled our reported-trawling hour distribution through nearest neighbor interpolation. Moreover, we calculated Cohen's kappa (Cohen, 1960) to estimate the agreement between the two distributions with respect to a chance agreement. We also calculated accuracy and kappa using four classes: low/medium/high fishing activity and non-fishing locations. This comparison aimed to demonstrate the effectiveness of our workflow's fishing activity classifier for trawlers, and, consequently, its potential reliability to classify unreported trawling activity (which focuses on gaps in this classification). However, it was impossible to compare unreported-trawling activity distributions because GFW does not report this information.

TABLE 1 Summary table of the main results obtained from the execution of our workflow on the Adriatic Sea case study data.

| | |
|---|---|
| Prior speed range of trawling activity | 2–4 kn |
| Statistically revised speed range | 1.6–4.8 kn |
| Sequential data processing time | 13 h |
| Cloud data processing time | 1 h |
| Highest-intensity unreported fishing activity region | Italian coasts |
| Lowest-intensity unreported fishing activity regions | Pomo Pit and 1,000 m areas |
| Unreported fishing activity decrease period | April–September |
| Month with the highest number of total trawling activity hours | April |
| Month with the lowest number of total trawling activity hours | August |
| Month with the highest number of unreported trawling activity hours | December |
| Month with the lowest number of unreported trawling activity hours | August |
| Regions with unreported fishing activity likely due to signal loss | Off the Po river's delta and far from the coasts |
| Regions with unreported fishing activity likely due to voluntary switch-offs | Northern prohibited areas (around hydrocarbon extraction platforms) and marine protected areas |
| Number of commercial stocks potentially involved in the unreported fishing activity | 33 |
| Number of ETP stocks potentially involved in the unreported fishing activity | 4 |

# 3. Results

## 3.1. Identification of Adriatic unreported trawling hotspots

Our workflow produced unreported trawling activity hotspots for all months of 2019 at a 0.1° resolution (Figure 3, with the main results summarized in Table 1). To enhance comparison, we used the fishing-hour classification range of the month with the most intense unreported trawling activity (December) for all months. The statistically revised trawling speed range went from ∼1.6 kn to ∼4.8, with up to a 20% variation across the months. The data processing required ∼1 h using cloud computing and ∼13 h using a sequential execution on one machine.

The analysis highlighted that Italian coasts were generally subject to the highest intensity of unreported trawling, even at a short distance from the coast. Here, the highest unreported trawling intensity was concentrated in the northwest (off Po river's delta), in central coasts (off Abruzzo and Marche coasts), and in a few localized spots in the south (off Apulian coasts). During warmer months (April–September), the unreported activity strongly decreased because of the summer fishing bans but was sparser and tended toward the middle of the basin. The month with the most intense total trawling activity was April, whereas the one with the lowest intensity was August.

Homogenizing trawling intensity categorization across the months enhanced inter-month result comparison. For example, it highlighted that unreported fishing activity always presented medium-high intensity locations across all months, even in the lowest-intensity month (Figures 4A, B; the "Supplementary information" section contains links to data and images for all months). These locations sensibly contributed to the total fishing effort. In fact, a high unreported/total activity ratio was often concentrated in the unreported activity hotspots, especially in the high intensity months (Figure 4A). These hotspots might correspond to systematic signal loss, e.g., off the Po river's delta where signal power loss and high traffic can interfere with transmission (Mantovani, 2019). However, they might also correspond to voluntary transmission switch-offs, especially in (or around) prohibited or protected areas (Ferrà et al., 2020). High-ratio locations could also be sparse outside unreported trawling hotspots (Figure 4B). These locations likely corresponded to communication issues, especially far from the coasts (Liping and Shexiang, 2012; Natale et al., 2015; Shepperson et al., 2018).

These observations can be further explored by analyzing the density of unreported fishing hotspots in the sensitive areas (Section 2.3). Since the definitions of these areas might change across the months, we distinguished and aggregated the month-data when the sensitive areas remained almost constant (Figures 5, 6). We used the same density scale and classification thresholds to compare the aggregations better. This analysis confirmed that Italian coasts were particularly subject to unreported trawling activity (Figure 5-Coast Distance Ban), in agreement with other studies (Scarcella et al., 2014; Ferrà et al., 2018). The hotspots intersected several MPAs (Figure 5-MPA Ban), suggesting a hidden negative pressure on these delicate ecosystems, especially off the northern and southern Croatian coasts (off Rijeka/Fiume and Split, and around Vis island), also highlighted by other studies (Chimienti et al., 2020; Ferrà et al., 2020). The Pomo Pit and the 1,000 m area (Figure 5-Pomo Ban and 1,000 m Ban) presented a low level of unreported trawling activity, slightly increasing in warmer months, in agreement with expert studies (de Juan and Lleonart, 2010; Elahi et al., 2018). This was probably due to too-deep waters. The prohibited and ZTB/BPA areas falling in unreported trawling hotspots were especially located in the North Adriatic (Figure 6-Prohibited Areas and BPA Ban), e.g., in the "Barbare" ZTB/BPA (at 30 NM from the Ancona coasts), and off the Croatian coasts. This observation finds confirmation in other studies (Galdelli et al., 2020, 2021). These areas contain several hydrocarbon extraction platforms (Figure 6-Platform Ban), where trawling is forbidden because of the inter-platform cables present on the seabed.

The unreported activity hotspots involved 33 commercial stocks, four of which (sharks and tuna) were also ETP species of the basin (Table 2). This list contains actual Adriatic trawlers' targets, and the detected ETP species are known to be subject to illegal fishing activity (Piroddi et al., 2015; Froese et al., 2018b; Armelloni et al., 2021).

## 3.2. Comparison with the Global Fishing Watch data

We compared the *reported* trawling hours estimated by our workflow with those available on the GFW Web portal (Global Fishing Watch, 2022). In particular, we compared the distributions

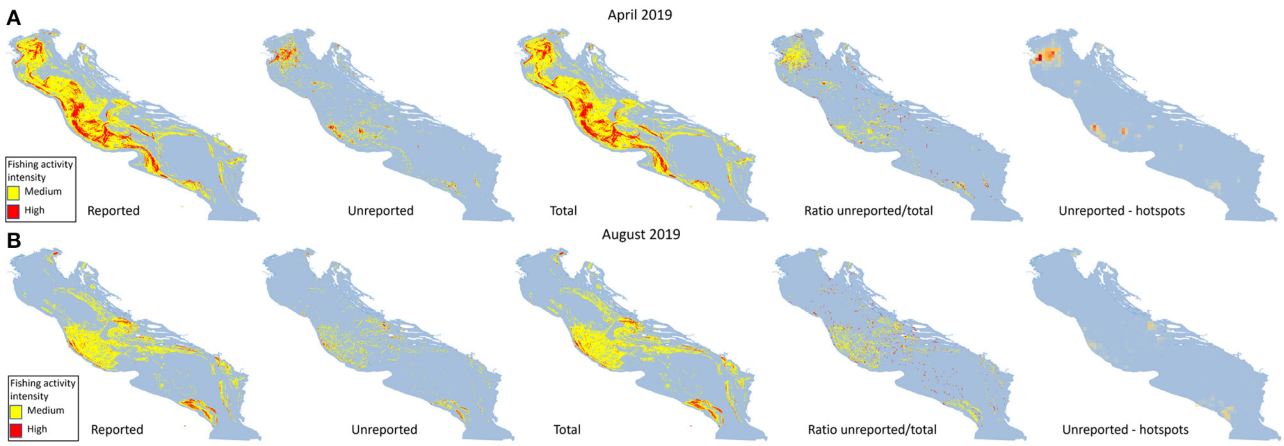


FIGURE 4
Comparison between the 0.01° resolution distributions of unreported, reported, total, unreported fraction, and unreported trawling hotspots for the **(A)** highest-intensity and **(B)** lowest-intensity total trawling-activity months (April and August). To enhance visual comparison, the same fishing-hour classification thresholds were used for the two months' distributions, and only medium and high locations were reported.

in the months of highest and lowest total activity (April and August 2019) (Figures 7A, B, Table 3). When comparing only two classes (fishing vs non-fishing activity), our workflow accuracy at reproducing the GFW distribution was 94.4% for April and 95.1% for August, with kappa agreements of 0.77 and 0.68 respectively [*good* according to Fleiss' interpretation (Fleiss, 1971)]. When using four classes (low/medium/high fishing activity and non-fishing activity), accuracy was 92.7% for April and 94.4% for August, with 0.70 and 0.64 kappa agreements (still *good*). Overall, these results indicate a high overlap between the distributions, which enforces our statistical approach reliability on the case study and consequently suggests coherence in detecting unreported fishing activity in the reported-activity gaps.

## 4. Discussion and conclusions

We have presented a workflow to estimate the distribution of reported and hidden/unreported fishing activity in a marine area. The workflow can process big data of sequences of vessels coordinate and speed information collected by AISs and VMSs. The workflow automatically revises a prior fishing speed range provided as input. It rebuilds gaps in the estimated fishing activity locations and estimates the potential unreported activity in the study area and the related hotspots. By integrating stock and biodiversity data sources, the workflow extracts the potential vessels target stocks in the unreported activity hotspots and their ETP status. Unlike other solutions (Merten et al., 2016), our workflow is also available as a standardized Web service running on a cloud computing platform, which supports the concurrent processing of big data flows.

We have demonstrated the effectiveness of our workflow through a case study focussing on Adriatic trawling vessels. The workflow correctly extracted trawling patterns in agreement with an alternative model. Additionally, it detected unreported-activity hotspots and their potential impact on sensitive areas in agreement with expert studies. These hotspots contain valuable information for monitoring authorities, because they indicate areas with systematic communication problems and illegal fishing activity. The extracted target and ETP stocks in these locations were correct and included stocks known to be subject to major illegal fishing. Through cloud computing, we could quickly process ∼50 million records, experiment different workflow parametrisations, and produce monthly analyses for the case study. The generality of the workflow and its standardized Web service interface make it easily re-usable for other areas, in compliance with the Open Science paradigm (Hey et al., 2009).

Compared to other solutions, our workflow does not use gear/logbook data and spatially explicit catch information from Regional Fisheries Management Organizations (RFMO) (Palmer and Wigley, 2009; Lee et al., 2010; Gerritsen and Lordan, 2011; Olesen et al., 2012; Muench et al., 2018; Roberson et al., 2019; Burns et al., 2023). This information would likely increase our workflow accuracy in identifying unreported fishing activity hotspots. However, it would also increase the workflow dependency on the RFMO regions for which data are available, consequently lowering the current cross-region applicability. Moreover, our workflow does not consider the environmental effects on stock presence, which can enhance vessel activity prediction accuracy in multi-source integration models (Chang and Yuan, 2014; Coro et al., 2022a; Burns et al., 2023). The principal reason is that our workflow assesses the potentially involved stocks after identifying unreported fishing activity hotspots. Improving the detection of the involved stocks is among our prior directions of improvement, which will require including environmental aspects. Another limitation of our workflow is that processing AIS/VMS trajectory data alone does not allow for distinguishing between technical issues and voluntary transmitter switch-offs. Using supervised machine learning instead of statistical analysis would have helped identify voluntary switch-offs as anomalous patterns in vessel trajectories (Marzuki et al., 2015; Ford et al., 2018; Singh and Heymann, 2020; Wolsing et al., 2022). Some machine learning-based approaches search for these patterns also in the data gaps, using SAR image processing to

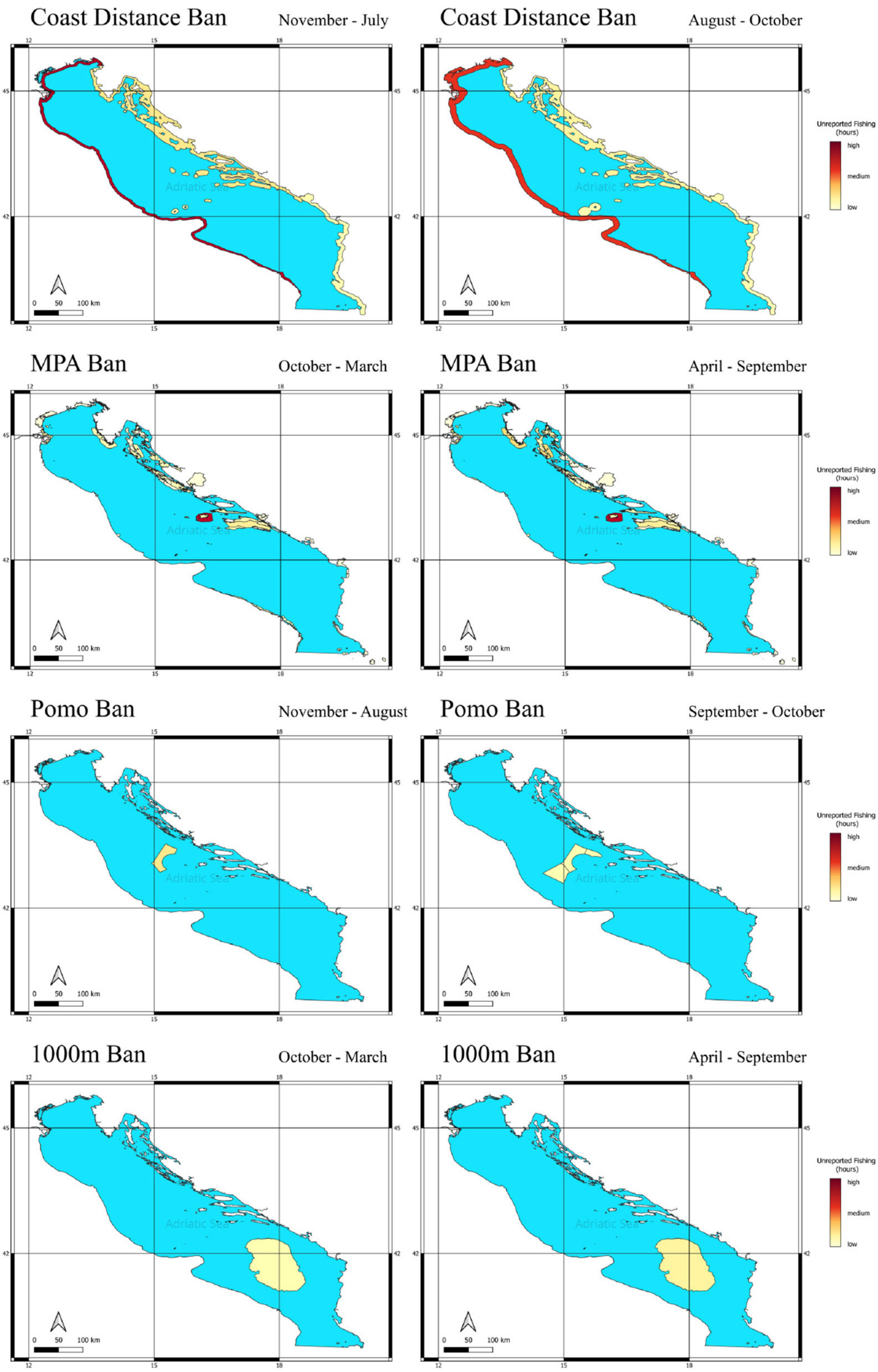


FIGURE 5
Classification of unreported hotspot density over 4 monitored Adriatic subdivisions: coast-distance ban (within 3 and 6 NM, respectively for Italian coasts) (Coast Distance Ban), marine protected areas (MPAs), Pomo Pit restricted area (Pomo Ban), areas with depth over 1,000 m (1,000 m Ban). The same classification thresholds were used for each chart pair to enhance comparison. The aggregation months were chosen as those during which the subdivisions remained almost constant.

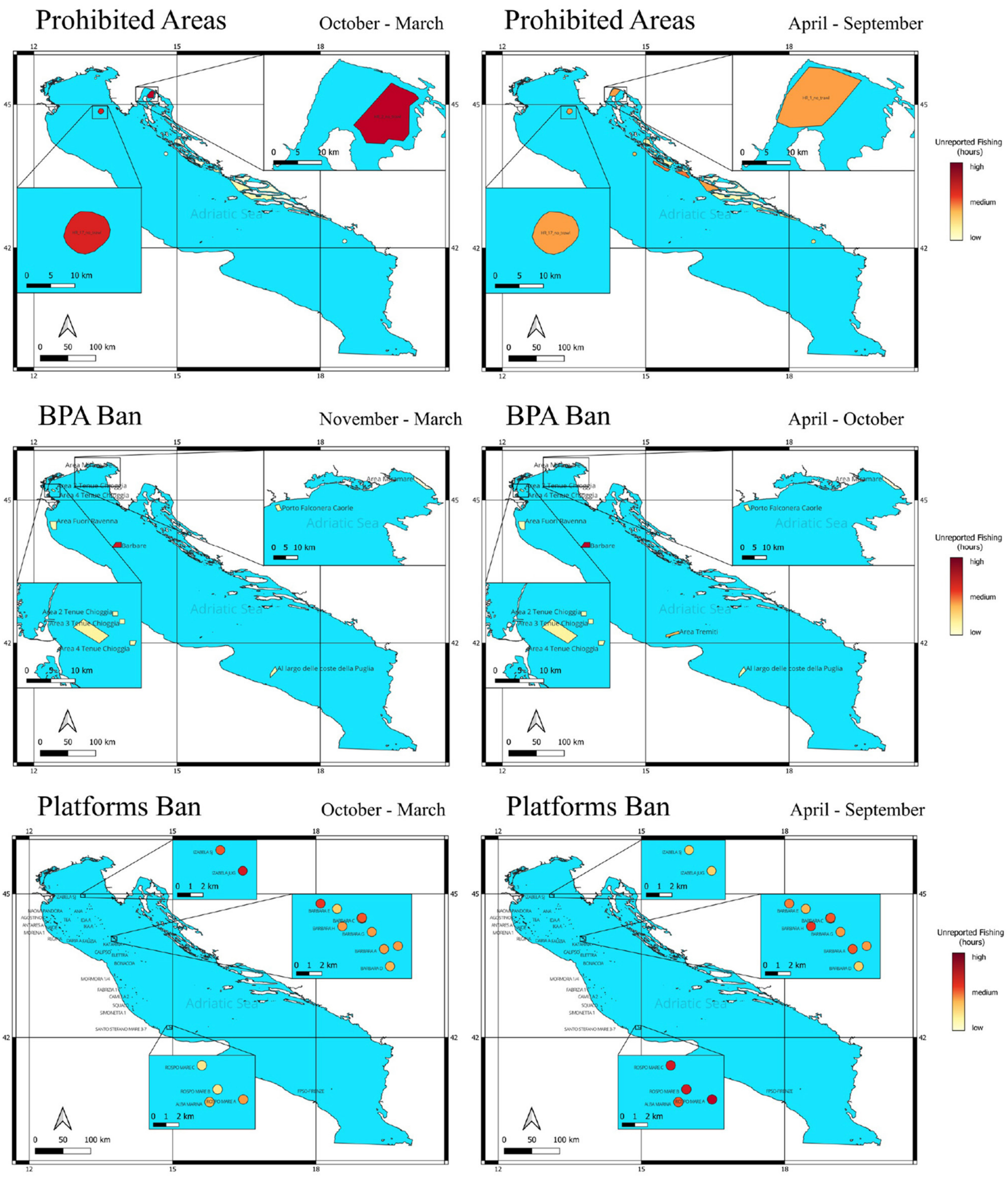


FIGURE 6
Classification of the unreported hotspot density over 3 monitored Adriatic location types: trawling-restricted areas (Prohibited Areas), Zones of Biological Protection (BPA Ban), offshore platforms (Platforms Ban). The same classification thresholds were used for each chart pair to enhance comparison. The aggregation months were chosen as those during which the areas remained almost constant.

estimate ship presence (Perez et al., 2013; Galdelli et al., 2021; Cutlip, 2022). Unlike supervised machine learning approaches, our workflow does not use (expensive) annotated corpora for model training. Moreover, since our workflow operates fishing activity classification and focusses on fishing vessels only, integrating SAR images would require reliable, currently unavailable (Cutlip, 2022), algorithms to identify fishing activity type and vessel size from the images.

TABLE 2 List of stock/species detected in the unreported trawling activity hotspots.

| | | | | |
|---|---|---|---|---|
| *Aristaeomorpha foliacea* | *Aristeus antennatus* | *Auxis rochei* | *Coryphaena hippurus* | *Dicentrarchus labrax* |
| *Engraulis encrasicolus* | *Euthynnus alletteratus* | *Galeorhinus galeus* | *Loligo forbesii* | *Merluccius merluccius* |
| *Micromesistius poutassou* | *Mullus barbatus* | *Mustelus asterias* | *Nephrops norvegicus* | *Pagellus erythrinus* |
| *Parapenaeus longirostris* | *Phycis blennoides* | *Raja clavata* | *Sardina pilchardus* | *Scomber colias* |
| *Scomber scombrus* | *Scophthalmus maximus* | *Scyliorhinus canicula* | *Solea solea* | *Sparus aurata* |
| *Sprattus sprattus* | *Squalus acanthias* | *Squilla mantis* | *Thunnus alalunga* | *Thunnus thynnus* |
| *Trachurus trachurus* | *Triakidae* spp. | *Trisopterus minutus* | | |

Red colors indicate endangered, threatened, and protected (ETP) species.

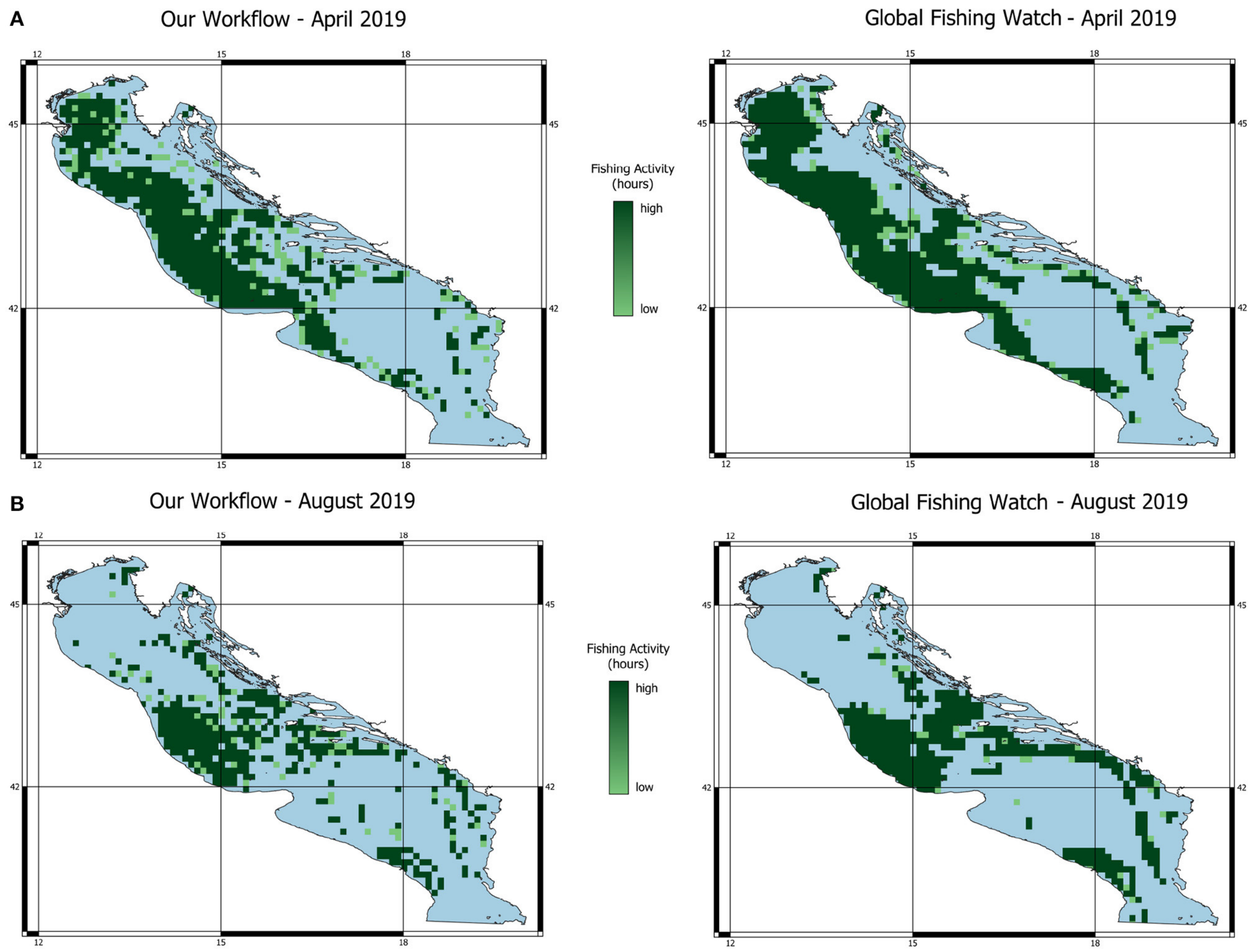


FIGURE 7
Comparison (at 0.1° resolution) between distributions of reported fishing activity hours produced by our workflow and by the Global Fishing Watch (GFW) in the months of **(A)** highest and **(B)** lowest intensity (April and August). To enhance comparison, fishing activity was classified as low to high in each distribution, using statistically calculated thresholds for our workflow and GFW-provided thresholds for the GFW data.

The information produced by our workflow is complementary to the one used by RFMOs to monitor fishing activity and conduct stock assessments. It is also valuable for IEAs to better estimate human-related driving forces on ecosystems, and suited for being integrated in marine spatial planning workflows, e.g., to prioritize closures and controls that guarantee fisheries sustainability while reducing unreported fishing activity (Agnew et al., 2009; Klein et al., 2010; Agardy et al., 2011; Plumptre et al., 2014; Metcalfe et al., 2015; Coro et al., 2021). The fishing activity classification algorithm is suited for applications where specific speed ranges characterize the fishing activity [e.g., 2–4 kn for trawlers and 6–15 kn for tuna purse seiners, Zhang et al. (2021)]. In a big data context, where data are noisy and untrustworthy (Coro et al., 2021), a more specific fishing-activity classification (e.g., with trawling

TABLE 3 Comparison between our reported fishing activity distribution and the Global Fishing Watch's distribution for the Adriatic Sea.

| | Accuracy | Kappa | Interpretation |
|---|---|---|---|
| April—two classes | 94.4% | 0.77 | good agreement |
| August—two classes | 95.1% | 0.68 | good agreement |
| April—four classes | 92.7% | 0.70 | good agreement |
| August—four classes | 94.4% | 0.64 | good agreement |

The comparison is reported for the highest (April) and lowest (August) total fishing-intensity months at a 0.1° spatial resolution, using two agreement classes (fishing vs non fishing) and four agreement classes (low, medium, high, non fishing). Performance is reported in terms of accuracy (number of agreed cells over total cells) and Cohen's kappa. Fleiss' kappa interpretation is reported for all comparisons.

type specification) might end in less reliable results or a less re-usable process for other areas (Coro et al., 2022b). Nevertheless, our workflow allows for using other, more precise, fishing-activity classifications. For example, the input data could come from the GFW-classified AIS data, with include fishing gear indication. In this case, our fishing activity classification algorithm would be skipped. We indeed plan to integrate GFW data with our workflow to support gear-specific analyses, which will require buying the GFW-AIS data. Alternatively, GFW provides open data sets to train supervised machine learning models and build alternative fishing-activity classification models for other AIS data (e.g., the Astra Paging data).

As for the GRSF, through an existing agreement between FAO and D4Science for knowledge base access (FAO, 2019), we will ask for accessing the data of the frequently used gears per stock. Alternatively, similarly to other approaches (Froese et al., 2018a), a strategy to exclude stocks from gear-specific captures could be developed, based on average stock weight, length, and habitat. Separating the analysis by fishing gear, would allow, for example, to differentiate the lower impact of bottom trawling on pelagic species with respect to overall trawling activity, and consequently refine the results. We plan to in-depth explore the balance between gear-specific activity classification and result reliability in future work.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

## Author contributions

All authors equally contributed to this work from conceptualization to development and paper review.

## Funding

This work has been funded by the EcoScope European project within the H2020-EU.3.2.-SOCIETAL CHALLENGES programme, with Grant Agreement ID 101000302.

## Acknowledgments

The authors wish to thank they colleagues of the EcoScope European project for their support to this research.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary information

The results on the use case data are freely available on the Zenodo open repository as CSV, raster (GeoTiff), and image files: https://zenodo.org/record/7402415#.Y46EvXbMJPY.

The Workflow source code is freely available as a GitHub repository at https://github.com/cybprojects65/VesselAnalytics, which also contains a one-month example of input data.

The Workflow Web service is freely available on the D4Science e-Infrastructure after registration to the RProtopypingLab Virtual Research Environment: https://services.d4science.org/group/d4science-services-gateway/explore.

It is accessible through the DataMiner cloud computing platform at https://services.d4science.org/group/rprototypinglab/data-miner?OperatorId=org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.transducerers.UNREPORTED_FISHING_ACTIVITY_HOTSPOTS.

Programmatic Web service calls can follow the WPS specifications (https://wiki.gcube-system.org/gcube/How_to_Interact_with_the_DataMiner_by_client).

## References

Agardy, T., Di Sciara, G. N., and Christie, P. (2011). Mind the gap: addressing the shortcomings of marine protected areas through large scale marine spatial planning. *Mar. Policy* 35, 226–232. doi: 10.1016/j.marpol.2010.10.006

Agnew, D. J., Pearce, J., Pramod, G., Peatman, T., Watson, R., Beddington, J. R., et al. (2009). Estimating the worldwide extent of illegal fishing. *PLoS ONE* 4, e4570. doi: 10.1371/journal.pone.0004570

Amoroso, R. O., Pitcher, C. R., Rijnsdorp, A. D., McConnaughey, R. A., Parma, A. M., Suuronen, P., et al. (2018). Bottom trawl fishing footprints on the world's continental shelves. *Proc. Nat. Acad. Sci.* 115, E10275–E10282. doi: 10.1073/pnas.1802379115

Antunes, P., and Santos, R. (1999). Integrated environmental management of the oceans. *Ecol. Econ.* 31, 215–226. doi: 10.1016/S0921-8009(99)00080-4

Armelloni, E. N., Scanu, M., Masnadi, F., Coro, G., Angelini, S., Scarcella, G., et al. (2021). Data poor approach for the assessment of the main target species of rapido trawl fishery in adriatic sea. *Front. Mar. Sci.* 8, 681. doi: 10.3389/fmars.2021.552076

Assante, M., Candela, L., Castelli, D., Cirillo, R., Coro, G., Frosini, L., et al. (2019). Enacting open science by d4science. *Future Gener. Comput. Syst.* 101, 555–563. doi: 10.1016/j.future.2019.05.063

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, 41–53. doi: 10.1016/j.fishres.2010.06.016

Belhabib, D., Cheung, W. W., Kroodsma, D., Lam, V. W., Underwood, P. J., Virdin, J., et al. (2020). Catching industrial fishing incursions into inshore waters of africa from space. *Fish Fish.* 21, 379–392. doi: 10.1111/faf.12436

Bergh, P. E., and Davies, S. (2002). Fishery monitoring, control and surveillance. *FAO Fisheries Technical Paper* (Rome: FAO), 175–204.

BOEM (2020). *BOEM Marine Cadastre Initiative*. Available online at: https://www.nafo.int/Science/Science-Advice/Species (accessed September 2020).

Burns, E. S., Bradley, D., and Thomas, L. R. (2023). Global hotspots of shark interactions with industrial longline fisheries. *Front. Mar. Sci.* 9, 2666. doi: 10.3389/fmars.2022.1062447

Candela, L., Castelli, D., Coro, G., Pagano, P., and Sinibaldi, F. (2016). Species distribution modeling in the cloud. *Concurr. Comput. Pract. Exp.* 28, 1056–1079. doi: 10.1002/cpe.3030

Chang, S.-J. (2003). "Vessel identification and monitoring systems for maritime security," in *IEEE 37th Annual 2003 International Carnahan Conference onSecurity Technology* (Piscataway, NJ: IEEE), 66–70.

Chang, S.-K., and Yuan, T.-L. (2014). Deriving high-resolution spatiotemporal fishing effort of large-scale longline fishery from vessel monitoring system (VMS) data and validated by observer data. *Can. J. Fish. Aquat. Sci.* 71, 1363–1370. doi: 10.1139/cjfas-2013-0552

Chimienti, G., De Padova, D., Mossa, M., and Mastrototaro, F. (2020). A mesophotic black coral forest in the adriatic sea. *Sci. Rep.* 10, 1–15. doi: 10.1038/s41598-020-65266-9

Clavelle, T. (2022). *gfwr: Access data from Global Fishing Watch APIs*. Available online at: https://github.com/GlobalFishingWatch/gfwr (accessed April 20, 2023).

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104

Coro, G. (2020). Open science and artificial intelligence supporting blue growth. *Environ. Eng. Manag. J.* 19, 1719–1729. doi: 10.30638/eemj.2020.162

Coro, G., Bove, P., Armelloni, E. N., Masnadi, F., Scanu, M., Scarcella, G., et al. (2022a). Filling gaps in trawl surveys at sea through spatiotemporal and environmental modelling. *Front. Mar. Sci.* 9, 1106. doi: 10.3389/fmars.2022.919339

Coro, G., Candela, L., Pagano, P., Italiano, A., and Liccardo, L. (2015). Parallelizing the execution of native data mining algorithms for computational biology. *Concurr. Comput. Pract. Exp.* 27, 4630–4644. doi: 10.1002/cpe.3435

Coro, G., Ellenbroek, A., and Pagano, P. (2021). An open science approach to infer fishing activity pressure on stocks and biodiversity from vessel tracking data. *Ecol. Inform.* 64, 101384. doi: 10.1016/j.ecoinf.2021.101384

Coro, G., Fortunati, L., and Pagano, P. (2013). "Deriving fishing monthly effort and caught species from vessel trajectories," in *2013 MTS/IEEE OCEANS-Bergen* (Bergen: IEEE), 1–5. doi: 10.1109/OCEANS-Bergen.2013.6607976

Coro, G., Panichi, G., Scarponi, P., and Pagano, P. (2017). Cloud computing in a distributed e-infrastructure using the web processing service standard. *Concurr. Comput. Pract. Exp.* 29, e4219. doi: 10.1002/cpe.4219

Coro, G., Tassetti, A. N., Armelloni, E. N., Pulcinella, J., Ferrà, C., Sprovieri, M., et al. (2022b). Covid-19 lockdowns reveal the resilience of adriatic sea fisheries to forced fishing effort reduction. *Sci. Rep.* 12, 1–14. doi: 10.1038/s41598-022-05142-w

Cutlip, K. (2022). *IUU Illegal, Unreported, Unregulated Fishing*. Available online at: https://globalfishingwatch.org/fisheries/iuu-illegal-unreported-unregulated-fishing/ (accessed April 20, 2023).

Davis, J. M. (2001). Monitoring control surveillance and vessel monitoring system requirements to combat iuu fishing. *FAO Fisheries Reports* (Rome: FAO), 244–256.

de Juan, S., and Lleonart, J. (2010). A conceptual framework for the protection of vulnerable habitats impacted by fishing activities in the mediterranean high seas. *Ocean Coast. Manag.* 53, 717–723. doi: 10.1016/j.ocecoaman.2010.10.005

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, e0158248. doi: 10.1371/journal.pone.0158248

Depellegrin, D., Bastianini, M., Fadini, A., and Menegon, S. (2020). The effects of covid-19 induced lockdown measures on maritime settings of a coastal region. *Sci. Total Environ.* 740, 140123. doi: 10.1016/j.scitotenv.2020.140123

Dunn, D. C., Jablonicky, C., Crespo, G. O., McCauley, D. J., Kroodsma, D. A., Boerder, K., et al. (2018). Empowering high seas governance with satellite vessel tracking data. *Fish Fish.* 19, 729–739. doi: 10.1111/faf.12285

Elahi, R., Ferretti, F., Bastari, A., Cerrano, C., Colloca, F., Kowalik, J., et al. (2018). Leveraging vessel traffic data and a temporary fishing closure to inform marine management. *Front. Ecol. Environ.* 16, 440–446. doi: 10.1002/fee.1936

EMODNET (2022). *Human Activities Database*. Available online at: https://www.emodnet-humanactivities.eu/view-data.php (accessed April 20, 2023).

Eriksen, T., Skauen, A. N., Narheim, B., Helleren, Ø., Olsen, Ø., and Olsen, R. B. (2010). "Tracking ship traffic with space-based AIS: experience gained in first months of operations," in *2010 International WaterSide Security Conference* (Italy: IEEE), 1–8. doi: 10.1109/WSSC.2010.5730241

European Parliament (2002). *Council Regulation (EC) No 2371/2002 of 20 December 2002 on the conservation and sustainable exploitation of fisheries resources under the Common Fisheries Policy*. Available online at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32002R2371 (accessed April 20, 2023).

European Parliament (2011). *Commission Directive 2011/15/eu Establishing a Community Vessel Traffic Monitoring And Information System Text with EEA Relevance*. Available online at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32011L0015 (accessed April 20, 2023).

FAO (2019). *FAO and Italy's National Research Council strengthen iMarine ties*. Available online at: https://www.fao.org/news/story/en/item/1251563/icode/ (accessed April 20, 2023).

FAO (2020). *The State of Mediterranean and Black Sea Fisheries 2020*. Rome: Italy. doi: 10.4060/cb2429en

Ferrà, C., Tassetti, A. N., Armelloni, E. N., Galdelli, A., Scarcella, G., and Fabi, G. (2020). Using AIS to attempt a quantitative evaluation of unobserved trawling activity in the mediterranean sea. *Front. Mar. Sci.* 7, 580612. doi: 10.3389/fmars.2020.580612

Ferrà, C., Tassetti, A. N., Grati, F., Pellini, G., Polidori, P., Scarcella, G., et al. (2018). Mapping change in bottom trawling activity in the mediterranean sea through AIS data. *Mar. Policy* 94, 275–281. doi: 10.1016/j.marpol.2017.12.013

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychol. Bull.* 76, 378. doi: 10.1037/h0031619

Fonda, S., Franco, P., Ghirardelli, E., and Malej, A. (1992). *Outline of Oceanography and the Plankton of the Adriatic sea*. Available online at: https://arts.units.it/handle/11368/1747095 (accessed April 20, 2023).

Ford, J. H., Peel, D., Kroodsma, D., Hardesty, B. D., Rosebrock, U., Wilcox, C., et al. (2018). Detecting suspicious activities at sea based on anomalies in automatic identification systems transmissions. *PLoS ONE* 13, e0201640. doi: 10.1371/journal.pone.0201640

Freire, J., Bonnet, P., and Shasha, D. (2012). "Computational reproducibility: state-of-the-art, challenges, and database research opportunities," in *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data* (New York, NY: ACM), 593–596. doi: 10.1145/2213836.2213908

Froese, R., Winker, H., Coro, G., Demirel, N., Tsikliras, A. C., Dimarchopoulou, D., et al. (2018a). A new approach for estimating stock status from length frequency data. *ICES J. Mar. Sci.* 75, 2004–2015. doi: 10.1093/icesjms/fsy078

Froese, R., Winker, H., Coro, G., Demirel, N., Tsikliras, A. C., Dimarchopoulou, D., et al. (2018b). Status and rebuilding of European fisheries. *Mar. Policy* 93, 159–170. doi: 10.1016/j.marpol.2018.04.018

Galdelli, A., Mancini, A., Ferrà, C., and Tassetti, A. N. (2020). "Integrating AIS and SAR to monitor fisheries: a pilot study in the adriatic sea," In *Proceedings of the 2020 IMEKO TC-19 International Workshop on Metrology for the Sea*, Naples, Italy, 5–7.

Galdelli, A., Mancini, A. Ferrà, C., and Tassetti, A. N. (2021). A synergic integration of AIS data and SAR imagery to monitor fisheries and detect suspicious activities. *Sensors* 21, 2756. doi: 10.3390/s21082756

Gari, S. R., Newton, A., and Icely, J. D. (2015). A review of the application and evolution of the dpsir framework with an emphasis on coastal social-ecological systems. *Ocean Coast. Manag.* 103, 63–77. doi: 10.1016/j.ocecoaman.2014.11.013

General Fisheries Commission for the Mediterranean (2013). *Management Plans*. Available online at: https://www.fao.org/gfcm/activities/fisheries/management-measures/management-plans/en/ (accessed April 20, 2023).

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, 245–252. doi: 10.1093/icesjms/fsq137

GFW (2020). *Global Fisheries During COVID-19 - Global Fishing Watch*. Available online at: https://globalfishingwatch.org/data/global-fisheries-during-covid-19/ (accessed September 2020).

Gianelli, I., Horta, S. Martínez, G., de la Rosa, A., and Defeo, O. (2018). Operationalizing an ecosystem approach to small-scale fisheries in developing countries: the case of Uruguay. *Mar. Policy* 95, 180–188. doi: 10.1016/j.marpol.2018.03.020

Global Fishing Watch (2022). *Datasets and Code*. Available online at: https://globalfishingwatch.org/datasets-and-code/ (accessed April 20, 2023).

Grassle, J. F. (2000). The ocean biogeographic information system (OBIS): an on-line, worldwide atlas for accessing, modeling and mapping marine biological data in a multidimensional geographic context. *Oceanography* 13, 5–7. doi: 10.5670/oceanog.2000.01

HawkEye360 (2020). *Chinese Fishing Fleet Encroaches on the Galapagos Islands*. Available online at: https://www.he360.com/resource/potential-illegal-fishing-seen-from-space/ (accessed April 20, 2023).

Hey, T., Tansley, S., Tolle, K., and Gray, J. (2009). *The Fourth Paradigm: Data-intensive Scientific Discovery*, Volume 1. Redmond, WA: Microsoft Research.

i-Marine (2020). *The Global Record of Stocks and Fisheries*. Available online at: https://i-marine.d4science.org/web/grsf/data-catalogue (accessed September 2020).

ITU (2009). International Telecommunication Union - Technical Characteristics for an Automatic Identification System Using Time Division Multiple Access in the VHF Maritime Mobile Frequency Band. Available online at: https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.1371-5-201402-I!!PDF-E.pdf (accessed April 20, 2023).

IUCN (2001). *IUCN Red List Categories and Criteria*. Gland: IUCN.

James, M., Mendo, T., Jones, E. L., Orr, K., McKnight, A., Thompson, J., et al. (2018). AIS data to inform small scale fisheries management and marine spatial planning. *Mar. Policy* 91, 113–121. doi: 10.1016/j.marpol.2018.02.012

Jennings, S., and Lee, J. (2012). Defining fishing grounds with vessel monitoring system data. *ICES J. Mar. Sci.* 69, 51–63. doi: 10.1093/icesjms/fsr173

Kia, M., Shayan, E., and Ghotb, F. (2000). The importance of information technology in port terminal operations. *Int. J. Phys. Distrib. Logist. Manag.* 30, 331–344. doi: 10.1108/09600030010326118

Klein, C. J., Steinback, C., Watts, M., Scholz, A. J., and Possingham, H. P. (2010). Spatial marine zoning for fisheries and conservation. *Front. Ecol. Environ.* 8, 349–353. doi: 10.1890/090047

Koen-Alonso, M., Pepin, P., Fogarty, M. J., Kenny, A., and Kenchington, E. (2019). The northwest atlantic fisheries organization roadmap for the development and implementation of an ecosystem approach to fisheries: structure, state of development, and challenges. *Mar. Policy* 100, 342–352. doi: 10.1016/j.marpol.2018.11.025

Koop, D., Santos, E., Mates, P., Vo, H. T., Bonnet, P., Bauer, B., et al. (2011). A provenance-based infrastructure to support the life cycle of executable papers. *Proc. Comput. Sci.* 4, 648–657. doi: 10.1016/j.procs.2011.04.068

Kristensen, P. (2004). The DPSIR framework, european topic centre on water. *Eur. Environ. Agency* 1, 1–10.

Kroodsma, D. A., Mayorga, J., Hochberg, T., Miller, N. A., Boerder, K., Ferretti, F., et al. (2018). Tracking the global footprint of fisheries. *Science* 359, 904–908. doi: 10.1126/science.aao5646

Kurekin, A. A., Loveday, B. R., Clements, O., Quartly, G. D., Miller, P. I., Wiafe, G., et al. (2019). Operational monitoring of illegal fishing in ghana through exploitation of satellite earth observation and AIS data. *Remote Sens.* 11, 293. doi: 10.3390/rs11030293

Le Guyader, D., Ray, C., Gourmelon, F., and Brosset, D. (2017). Defining high-resolution dredge fishing grounds with automatic identification system (AIS) data. *Aquat. Living Resour.* 30, 39. doi: 10.1051/alr/2017038

Le Tixerant, M., Le Guyader, D., Gourmelon, F., and Queffelec, B. (2018). How can automatic identification system (AIS) data be used for maritime spatial planning? *Ocean Coast. Manag.* 166, 18–30. doi: 10.1016/j.ocecoaman.2018.05.005

Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., et al. (2013). *Prov-o: The Prov Ontology*. Biot: W3C Recommendation 30.

Lee, J., South, A. B., and Jennings, S. (2010). Developing reliable, repeatable, and accessible methods to provide high-resolution estimates of fishing-effort distributions from vessel monitoring system (VMS) data. *ICES J. Mar. Sci.* 67, 1260–1271. doi: 10.1093/icesjms/fsq010

Liping, L., and Shexiang, M. (2012). "Analysis and simulation of slot collision and reception performance of AIS," in *Advances in Electric and Electronics* (Berlin: Springer), 661–669. doi: 10.1007/978-3-642-28744-2_86

Lockerbie, E. M., Lynam, C. P., Shannon, L. J., and Jarre, A. (2018). Applying a decision tree framework in support of an ecosystem approach to fisheries: indiseas indicators in the north sea. *ICES J. Mar. Sci.* 75, 1009–1020. doi: 10.1093/icesjms/fsx215

Mannini, P., Massa, F., and Milone, N. (2005). Adriatic sea fisheries: outline of some main facts. *Population* 199, 47.

Mantovani, F. (2019). *La rete AIS Nazionale - Componenti*. Available online at: https://slideplayer.it/slide/16200416/

Marine Vessel Traffic (2022). *Seaports by Country*. Available online at: https://www.marinevesseltraffic.com/2013/09/number-of-commercial-ports-by-country.html (accessed April 20, 2023).

Marzuki, M. I., Garello, R., Fablet, R., Kerbaol, V., and Gaspar, P. (2015). "Fishing gear recognition from VMS data to identify illegal fishing activities in indonesia," in *OCEANS 2015-genova* (Genova: IEEE), 1–5. doi: 10.1109/OCEANS-Genova.2015.7271551

Merten, W., Reyer, A., Savitz, J., Amos, J., Woods, P., Sullivan, B., et al. (2016). Global fishing watch: bringing transparency to global commercial fisheries. *arXiv* [*Preprint*]. doi: 10.48550/arXiv.1609.08756

Metcalfe, K., Vaz, S., Engelhard, G. H., Villanueva, M. C., Smith, R. J., Mackinson, S., et al. (2015). Evaluating conservation and fisheries management strategies by linking spatial prioritization software and ecosystem and fisheries modelling tools. *J. Appl. Ecol.* 52, 665–674. doi: 10.1111/1365-2664.12404

Muawanah, U., Yusuf, G., Adrianto, L., Kalther, J., Pomeroy, R., Abdullah, H., et al. (2018). Review of national laws and regulation in indonesia in relation to an ecosystem approach to fisheries management. *Mar. Policy* 91, 150–160. doi: 10.1016/j.marpol.2018.01.027

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, 1036–1047. doi: 10.1139/cjfas-2016-0446

Mullié, W. C. (2019). Apparent reduction of illegal trawler fishing effort in ghana's inshore exclusive zone 2012-2018 as revealed by publicly available AIS data. *Mar. Policy* 108, 103623. doi: 10.1016/j.marpol.2019.103623

Natale, F., Gibin, M., Alessandrini, A., Vespe, M., and Paulrud, A. (2015). Mapping fishing effort through AIS data. *PLoS ONE* 10, e0130746. doi: 10.1371/journal.pone.0130746

Olesen, P. B., Dukovska-Popovska, I., and Hvolby, H.-H. (2012). "Improving port terminal operations through information sharing," in *IFIP International Conference on Advances in Production Management Systems* (Berlin: Springer), 662–669. doi: 10.1007/978-3-642-40361-3_84

Pallotta, G., Vespe, M., and Bryan, K. (2013). Vessel pattern knowledge discovery from AIS data: a framework for anomaly detection and route prediction. *Entropy* 15, 2218–2245. doi: 10.3390/e15062218

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. Manag.* 29, 928–942. doi: 10.1577/M08-135.1

Perez, J. C., Alvarez, M. A., Heikkonen, J., Guillen, J., and Barbas, T. (2013). The efficiency of using remote sensing for fisheries enforcement: application to the mediterranean bluefin tuna fishery. *Fish. Res.* 147, 24–31. doi: 10.1016/j.fishres.2013.04.008

Pew Trusts (2015). *Project Eyes on the Seas*. Available online at: https://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2015/03/project-eyes-on-the-seas (accessed April 20, 2023).

Piroddi, C., Gristina, M., Zylich, K., Greer, K., Ulman, A., Zeller, D., et al. (2015). Reconstruction of italy's marine fisheries removals and fishing capacity, 1950-2010. *Fish. Res.* 172, 137–147. doi: 10.1016/j.fishres.2015.06.028

Plumptre, A. J., Fuller, R. A., Rwetsiba, A., Wanyama, F., Kujirakwinja, D., Driciru, M., et al. (2014). Efficiently targeting resources to deter illegal activities in protected areas. *J. Appl. Ecol.* 51, 714–725. doi: 10.1111/1365-2664.12227

Previero, M., and Gasalla, M. A. (2018). Mapping fishing grounds, resource and fleet patterns to enhance management units in data-poor fisheries: the case of snappers and groupers in the abrolhos bank coral-reefs (south atlantic). *Ocean Coast. Manag.* 154, 83–95. doi: 10.1016/j.ocecoaman.2018.01.007

Provoost, P., Bosch, S., and Appletans, W. (2017). *robis: R Client to Access Data from the Obis Api. Ocean Biogeographic Information System*. Intergovernmental Oceanographic Commission of UNESCO. Available online at: https://cran.r-project.org/package=robis

Reid, D., Graham, N., Rihan, D., Kelly, E., Gatt, I., Griffin, F., et al. (2011). Do big boats tow big nets? *ICES J. Mar. Sci.* 68, 1663–1669. doi: 10.1093/icesjms/fsr130

Ripley, B. (2022). *MASS 7.3–58.1*. Available online at: https://cran.r-project.org/web/packages/MASS/index.html (accessed April 20, 2023).

Robards, M., Silber, G., Adams, J., Arroyo, J., Lorenzini, D., Schwehr, K., et al. (2016). Conservation science and policy applications of the marine vessel automatic identification system (AIS)–a review. *Bull. Mar. Sci.* 92, 75–103. doi: 10.5343/bms.2015.1034

Roberson, L. A., Kiszka, J. J., and Watson, J. E. (2019). Need to address gaps in global fisheries observation. *Conserv. Biol.* 33, 966–968. doi: 10.1111/cobi.13265

Scarcella, G., Grati, F., Raicevich, S., Russo, T., Gramolini, R., Scott, R. D., et al. (2014). Common sole in the northern and central adriatic sea: spatial management scenarios to rebuild the stock. *J. Sea Res.* 89, 12–22. doi: 10.1016/j.seares.2014.02.002

Schut, P., and Whiteside, A. (2007). *OpenGIS Web Processing Service. OGC Project Document*. Available online at: http://www.opengeospatial.org/standards/wps (accessed April 20, 2023).

Shaw, D. R., Grainger, A., and Achuthan, K. (2017). Multi-level port resilience planning in the uk: how can information sharing be made easier? *Technol. Forecast. Soc. Change* 121, 126–138. doi: 10.1016/j.techfore.2016.10.065

Shepperson, J. L., Hintzen, N. T., Szostek, C. L., Bell, E., Murray, L. G., Kaiser, M. J., et al. (2018). A comparison of VMS and AIS data: the effect of data coverage and vessel position recording frequency on estimates of fishing footprints. *ICES J. Mar. Sci.* 75, 988–998. doi: 10.1093/icesjms/fsx230

Singh, S. K., and Heymann, F. (2020). "Machine learning-assisted anomaly detection in maritime navigation using AIS data," in *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS)* (Portland, OR: IEEE), 832–838. doi: 10.1109/PLANS46316.2020.9109806

Song, A. M., Johnsen, J. P., and Morrison, T. H. (2018). Reconstructing governability: how fisheries are made governable. *Fish Fish*. 19, 377–389. doi: 10.1111/faf.12262

Taconet, M., Kroodsma, D., and Fernandes, J. (2019). *Global Atlas of AIS -Based Fishing Activity-Challenges and Opportunities*. Available online at: http://www.fao.org/documents/card/en/c/ca7012en/ (accessed April 20, 2023).

Taconet, P., Chassot, E., Guitton, J., Fiorellato, F., Anello, E., Barde, J., et al. (2016). *Data Toolbox for Fisheries: The Case of Tuna Fisheries*. Available online at: https://www.iotc.org/sites/default/files/documents/2018/04/IOTC-2016-WPDCS12-27_-_TUNA_DATA_TOOLBOX.pdf (accessed April 20, 2023).

Tassetti, A., Ferrà, C., and Fabi, G. (2019). Rating the effectiveness of fishery-regulated areas with AIS data. *Ocean Coast. Manag*. 175, 90–97. doi: 10.1016/j.ocecoaman.2019.04.005

Tetreault, B. J. (2005). "Use of the automatic identification system (AIS) for maritime domain awareness (MDA)," in *Proceedings of Oceans 2005 MTS/IEEE* (Washington, DC: IEEE), 1590–1594. doi: 10.1109/OCEANS.2005.1639983

Wolsing, K., Roepert, L., Bauer, J., and Wehrle, K. (2022). Anomaly detection in maritime AIS tracks: a review of recent approaches. *J. Mar. Sci. Eng.* 10, 112. doi: 10.3390/jmse10010112

Yang, D., Wu, L., Wang, S., Jia, H., and Li, K. X. (2019). How big data enriches maritime research-a critical review of automatic identification system (AIS) data applications. *Transp. Rev*. 39, 755–773. doi: 10.1080/01441647.2019.1649315

Zhang, H., Yang, S.-L., Fan, W., Shi, H.-M., and Yuan, S.-L. (2021). Spatial analysis of the fishing behaviour of tuna purse seiners in the western and central pacific based on vessel trajectory data. *J. Mar. Sci. Eng.* 9, 322. doi: 10.3390/jmse9030322