An Open Science Approach to Infer Fishing Activity Pressure on Stocks and Biodiversity from Vessel Tracking Data

Gianpaolo Coro[a,1,2,*], Anton Ellenbroek[b], Pasquale Pagano[a]

[a] *Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" – CNR, Pisa, Italy*
[b] *Food and Agriculture Organization of the United Nations, Viale delle Terme di Caracalla, 00153 Rome, Italy*

Abstract

Vessel tracking data help study the potential impact of fisheries on biodiversity and produce risk assessments. Existing workflows process vessel tracks to identify fishing activity and integrate information on species vulnerability. However, there are significant data integration challenges across the data sources needed for an integrated impact assessment due to heterogeneous nomenclatures, data accessibility issues, geographical and computational scalability of the processes, and confidentiality and transparency towards decision making authorities.

This paper presents an Open Science data integration approach to use vessel tracking data in integrated impact assessments. Our approach combines heterogeneous knowledge sources from fisheries, biodiversity, and environmental observations to infer fishing activity and risks to potentially impacted species. An Open Science e-Infrastructure facilitates access to data sources and maximises the reproducibility of the results and the method's reusability across several application domains.

Our method's quality is assessed through three case studies: The first demonstrates cross-dataset consistency by comparing the results obtained from two different vessel data

*Corresponding author
Email addresses: coro@isti.cnr.it (Gianpaolo Coro), anton.ellenbroek@fao.org (Anton Ellenbroek), pagano@isti.cnr.it (Pasquale Pagano)
[1] Telephone Number: +39 050 315 8210
[2] Fax Number: +39 050 621 3464

sources. The second performs a temporal pattern analysis of fishing activity and potentially impacted species over time. The third assesses the potential impact of reduced fishing pressure on marine biodiversity and threatened species due to the 2020 COVID-19 lockdown in Italy. The method is meant to be integrated with other systems through its Open Science-oriented features and can rapidly use new sources of findable, accessible, interoperable, and reusable (FAIR) data. Other systems can use it to (i) classify vessel activity in data-limited scenarios, (ii) identify bycatch species (when catchability data are available), and (iii) study the effects of fisheries on habitats and populations' growth.
Keywords: Vessel transmitted information, Vessel tracking data, Automatic identification system, Statistical Analysis, e-Infrastructures, Open Science, Biodiversity, Integrated Environmental Assessment

---

Monitoring fishery activity is integral for ecosystem approaches to resource planning that involve species vulnerability and complex social and economic factors (Bergh and Davies, 2002; Gianelli et al., 2018; Lockerbie et al., 2018; Muawanah et al., 2018; Koen-Alonso et al., 2019). Integrated Environmental Assessment (IEA) systems use this information to model casual links between driving forces (economic and human activities), pressures (emissions, waste), chemico-physical and biological states, and the impact and responses of ecosystems (Antunes and Santos, 1999; Kristensen, 2004) (DPSIR framework). In this context, fishery activity classification is integral to systems that support policymakers at understanding fishery activity patterns and the impact of regulations and management strategies on ecosystems' quality (Robards et al., 2016; Le Tixerant et al., 2018). Today, DPSIR frameworks and vessel-data processing systems have limitations due to the heterogeneity of nomenclatures, the accessibility of data resources, the interoperability of methodologies, and the scalability and reusability of models across ecosystems (Gari et al., 2015; Taconet et al., 2016; James et al., 2018). Furthermore, few of these systems guarantee the transparency of the results to decision making authorities through open repetition and reproduction (Jennings and Lee, 2012; Dunn et al., 2018; Song et al., 2018).

Fishery data processing is commonly based on information transmitted by vessels during navigation via an Automatic Identification System (AIS) or other satellite-based and radio systems (Chang, 2003; ITU, 2009; Previero and Gasalla, 2018; Kurekin et al., 2019). Typical vessel transmitted data include coordinates, speed, route, vessel identity, and day/-time. AISs can have a high reporting frequency (every few seconds) but may have limitations in range coverage due to the terrestrial receiver (in the case of radio frequency-based systems) and vessel type (e.g., they are usually installed on vessels with length overall above 15m) (European Parliament, 2008). Furthermore, technical and meteorological issues can compromise data quality (Taconet et al., 2019).

Several vessel data processing systems enhance information quality and coverage by integrating gear-specific information (Lee et al., 2010; Palmer and Wigley, 2009), logbook information (Gerritsen and Lordan, 2011; Muench et al., 2018), and inter-port shared data (Kia et al., 2000; Davis, 2001; Olesen et al., 2012; Shaw et al., 2017; Roberson et al., 2019). Moreover, modern analytical frameworks integrate and correlate vessel data with other knowledge sources of fisheries, biodiversity, and societal information to extract new knowledge (Campanis, 2008; Agapito et al., 2019; Dinesen et al., 2019; Farmanbar et al., 2019). Applications of these frameworks to maritime spatial planning include: (i) identifying fishing activity locations with the highest density and intensity in specific monitored regions (Bastardie et al., 2010; Gerritsen and Lordan, 2011; Le Guyader et al., 2017;

Belhabib et al., 2020), (ii) estimating the spatial overlap between large- and small-scale fisheries (Le Tixerant et al., 2018; Shepperson et al., 2018; Mullié, 2019), (iii) monitoring unregulated activities (Natale et al., 2015; Kurekin et al., 2019), (iv) studying species-vessel interaction (Robards et al., 2016; Lopes et al., 2019; Iacarella et al., 2020), and (v) monitoring maritime traffic (Tetreault, 2005; Eriksen et al., 2006, 2010; Pallotta et al., 2013; Yang et al., 2019). These approaches often use fishing activity classification algorithms based on AIS data, which are either rule-based processes (Coro et al., 2013) or machine learning models (de Souza et al., 2016).

This paper contributes to developing holistic approaches to IEA and introduces a new multi-source analytical workflow that contextualises vessel tracking data with indicators on stocks, biodiversity, and the geophysical conditions of an area. Our workflow (i) identifies fished locations, (ii) estimates fishing pressure per species, (iii) identifies possible target stocks in these locations, (iv) identifies non-commercial species that could be impacted by the fisheries because concentrated in high fishing-pressure locations. Unlike other approaches (Le Tixerant et al., 2018; Farmanbar et al., 2019; Galdelli et al., 2019), our process is flexible enough to be applied at multiple spatial and temporal scales and can work on user-provided data, while being fully integrated with fisheries and biodiversity knowledge sources. Our method can seamlessly work with satellite and AIS input data in near real-time. It uses an Open Science oriented e-Infrastructure (D4Science, Assante et al. (2019b)) to facilitate access to vast collections of stock and biodiversity information. In particular, this e-Infrastructure optimises access to data sources that enrich vessel tracking data, and that meet the principles of Findability, Accessibility, Interoperability, and Reusability (FAIR data).

In the presented methodology, the integrated FAIR data include (i) environmental data from the Copernicus marine environment monitoring service (CMEMS, Von Schuckmann et al. (2016)), (ii) species occurrence records retrieved from the Ocean Biodiversity Information System (OBIS, Grassle (2000)), (iii) species' risk level from the Red List of the International Union for Conservation of Nature (IUCN, 2001), (iv) taxonomic data accessed from the Aquatic Sciences and Fisheries Information System of FAO (ASFIS, Garibaldi et al. (2002)), and (v) global stocks' distributions from the Global Record of Stocks and Fisheries (GRSF, i-Marine (2020)). Our method is provided as two Open Science-oriented Web services deployed in D4Science (differing only for their input parameters), available through a Virtual Research Environment that guarantees the reproduction and replication of all experiments and fosters the reuse of the processes across several domains.

Three case studies demonstrate the validity and versatility of our method. The first one uses vessel transmitted data from the West coasts of Canada and U.S.A. to compare information extracted from two different large data sources, i.e., the Global Fishing Watch and the BOEM-Marine Cadastre. This case demonstrates the consistency of our results using the two datasets separately. The second case reports a pattern analysis over time (2012-2016) of the fishing activity in the same area as the first case, using Global Fishing Watch data. This case demonstrates how to use our method to monitor fishing activity and species-pressure change over time. The third case shows how to use our method to (i) evaluate the change of fishing pressure on known biodiversity indicators, (ii) develop vulnerability patterns following a particular event - in this case, the March-April 2020 COVID-19 lockdown in Italy -, and (iii) assess potential benefits to biodiversity.

The research addressed by this paper is also to explore the benefits and potential pitfalls of the EU-promoted FAIR approach (Collins et al., 2018), by assessing the requirements of an IEA system against the realities of an EU-supported *Blue* flagship e-Infrastructure. In particular, based on the obtained results, this paper discusses the advantages brought by the use of FAIR data in statistical analyses but also the limitations due to data incompleteness, low update rates, and low quality control.

As further applications, our workflow can also be used by other systems to (i) extend current vessel activity classification processes through a more flexible approach that works in data-limited scenarios, (ii) identify bycatch species when catchability and fishing gear data are available, (iii) study the effects of fisheries on habitats and populations' growth when species distribution and life-history-trait data are available. FAIR data and Open Science-oriented technology guarantee integrating new data sources rapidly and reusing the workflow in other systems.

1. Materials and methods

Our process is an Open Science workflow that combines vessel tracking data from satellite and AIS systems with global, high-quality, environmental, biodiversity, and fishery data (Figure 1). The core process is made up of three macro-steps:

1. *Classifying vessel activity*: Vessel activity is (i) classified as fishing/not-fishing using a rule-based algorithm that uses bathymetry data from CMEMS, (ii) spatially aggregated, and (iii) processed to estimate high/medium/low fishing activity cells;
2. *Enriching vessel data with biodiversity and stock information*: Fishing activity cells are enriched with (i) species occurrence records from OBIS, (ii) vulnerability level information from the IUCN Red List, and (iii) stock information from the GRSF;

3. *Estimating fishing pressure on stocks and other species*: Fishing hours per cell (pressure) are associated with all species, to estimate the number of hours each species likely undergoes in each cell. Furthermore, pressure on stocks is categorised as high-/medium/low through statistical analysis, and highly impacted *non-stock* species are identified.

In our methodology, fishing activity is aggregated at a user-defined spatial resolution to make results comparable between different data sources and more independent of the abundance of data (Section 3). Spatial aggregation makes results more robust to missing vessel data and sampling biases in the species-observation data sources, and to differences between the data collection networks used by different data providers.

In our method, *non-stock* species are those species that are not the target of any fisheries in the study area according to the GRSF. Thus, the GRSF is considered as our main reference of stock information and, for ease of notation, *non*-GRSF species are named *non-stock* species in this paper.

The main input to our method is a table dataset (in plain-CSV format) containing punctual vessel trajectory data from a specific area. This dataset should include at least the following minimal information:

- A vessel identification code (even anonymous);
- Speed per point;
- Longitude/latitude per point;
- Timestamp per point.

Three additional input parameters are required, which are used to keep the analysis consistent:

- *Spatial Resolution*: The spatial aggregation of the data analysis. Default value is 0.5°;
- *Minimum Number Of Records*: The minimum number of species occurrence records per aggregation cell to consider the species as being present there. Default value is 5 records;
- *Occurrence Time Range*: The additional time extent, around the dataset time frame, to retrieve species observation records from OBIS. Default value is 5 years.

These parameters are meant to make the statistical analysis consistent, to reduce potential sources of biases due to a non-uniform reporting of occurrence records in OBIS, and to conduct multi-scale spatial analyses, as discussed in Section 3.

The output of our process is made up of the following information:

1. Distribution of fishing activity locations at the input spatial resolution (in tabular and image formats);
2. Histograms reporting the distribution across fishing-pressure categories of (i) the number of different species retrieved from OBIS, (ii) the fishing hours, and (iii) the overall species' occurrence records in the fishing cells;
3. One table reporting the list of species possibly involved in the fishing activities. For each species, the table specifies: (i) taxonomic information, (ii) the number of occurrence records in the fishing locations, (iii) the total fishing hours in the observation locations, (iv) if the species is a stock in the study area, (v) if the species

is threatened according to the IUCN Red List, and (vi) if the species is potentially highly impacted by the fishing activity.

The next sections explain the computational details and the input parameter roles.

1.1. Classifying vessel activity

Classifying vessel trajectory points as corresponding to fishing or not-fishing activity is the first computational step. Classification is operated through a rule-based algorithm (Coro et al., 2013) that uses bathymetry and speed information:

Algorithm 1 Vessel fishing activity classifier

For each vessel trajectory:

  for each trajectory point:

    if (speed $\leq$ 2kn)$\rightarrow$ hauling (*Not-Fishing*)

    if (2kn < speed $\leq$ 4kn and depth $\geq$ 500m) $\rightarrow$ trawling (*Fishing*)

    if (2kn < speed $\leq$ 4kn and depth < 500m) $\rightarrow$ mid-water trawling (*Fishing*)

    if (speed > 5kn) $\rightarrow$ steaming (*Not-Fishing*)

This algorithm assumes that speed-bathymetry criteria reflect the spatial relationships of consecutive points (Murawski et al., 2005; Lee et al., 2010; Lambert et al., 2012; Russo et al., 2013). Bathymetry data are accessed on-the-fly through longitude/latitude querying of the CMEMS *Global Ocean 1/12° physics analysis and forecast updated daily* NetCDF file. One copy of this file is hosted on a Unidata Thredds (John Caron and Davis, 2006) instance of the D4Science e-Infrastructure (Supplementary material) and is remotely accessed through the OPeNDAP protocol (Cornillon et al., 2003).

Based on the point-by-point classification, our process calculates fishing hours (*fah*s) as the time difference (in hours) between consecutive *fishing* points in a trajectory (i.e., the locations classified as *Fishing* locations). Furthermore, any *fishing* point that differs from a previous *fishing* point with more than 4 hours is excluded and assigned 0 *fah*s. This heuristic is compliant with the one used by other vessel activity classification algorithms (Campanis, 2008; Galdelli et al., 2019).

As an additional step, the *Spatial Resolution* input parameter is used to spatially aggregate *fah*s into fishing cells. A statistical analysis classifies these cells as high/medium/low-fishing activity cells:

Algorithm 2 Fishing hours' classifier

For each cell of *Spatial Resolution* square size in the study area:

find all fishing points in the cell from all trajectories

sum the fishing hours

associate the total fishing hours to the cell

Fit all *fah*s to a log-normal distribution

calculate the geometric mean

calculate confidence limits

For each fishing cell:

if (*fah*s ≥ upper confidence limit) → *high fishing-activity cell*

if (lower confidence limit < *fah*s < upper confidence limit) → *medium fishing-activity cell*

if (*fah*s ≤ lower confidence limit) → *low fishing-activity cell*

The use of a log-normal distribution comes after the empirical observation - over the used repositories of satellite and AIS data - that larger *fah*s tend to be farther away from the geometric mean than smaller values. This is particularly evident for large areas and input datasets, where effort is fairly and log-normally distributed across the fishing-activity classes.

1.2. Enriching vessel data with biodiversity and stock information

In the next computational step, biodiversity information is attached to each cell. This operation retrieves species' occurrence records and taxonomic information from OBIS per fishing cell. Records are extracted from a time frame of *Occurrence Time Range* years, before and after the time period referred by the vessel dataset. For example, if the vessel trajectories refer to a particular period in 2018, and *Occurrence Time Range=2*, OBIS occurrences will be retrieved from 2016 to 2020 for each fishing cells. To this aim, a geospatial query to OBIS is made through the "robis" R package (Provoost et al., 2017). This package is also used to validate if a species is *threatened* (i.e., at least vulnerable) according to the IUCN Red List. Overall, the *Occurrence Time Range* parameter is used as a tolerance threshold to account for missing records due to under-sampling in the time period of the vessel activity.

The list of species retrieved from OBIS is further reduced to consider only species included in the FAO-ASFIS collection, which includes more than 12,000 species of particular interest to fisheries, aquaculture, and biodiversity. ASFIS is used to select species related with fisheries activities, which are the focus of our study. This dataset is accessed as a CSV table (updated to 2020) hosted by the D4Science platform storage system for computational use (Supplementary material). Association between cells and species is made after checking, for each species, which cells have a number of observation records higher than *Minimum Number Of Records*. As a final step, the species are checked against the GRSF to be *stocks* or *non-stocks* in the fishing cells. This operation is achieved through a direct query to the SPARQL endpoint of the GRSF semantic knowledge base (i-Marine, 2020)

Overall, this process can be summarised as follows:

Algorithm 3 Biodiversity and stock information extraction

Build a geospatial query to OBIS for the fishing cells

Query OBIS in the time frame of the vessel dataset ± *Occurrence Time Range* years, to retrieve:

- the list of species observed in the area
- all species occurrence records
- taxonomic information for these species
- their IUCN threatening status

Reduce the list of species by taking only those present in ASFIS, to maximise relation with fisheries activities

For each species:

- select and associate those cells with occurrence records above *Minimum Number Of Records*
- check if the species is a stock for the GRSF in these cells, and classify the species as *stock* or *non-stock* accordingly

In summary, this algorithm associates information on stocks, species variety, and threatening status to the fishing hours of each cell in the studied area.

1.3. Estimating fishing pressure on stocks and other species

Fishing *pressure* per stock is here defined as the number of fishing hours per cell where the stock occurs:

$$FP_{|s} = \frac{\sum_{c=1}^{C} fah_{|s}(c)}{\sum_{c=1}^{C}(associated(c,s))}$$

where C is the total number of cells; $fah_{|s}(c)$ is the number of fishing hours in cell c associated with stock s; $associated(c,s)$ is a function that returns 1 if cell c is associated with stock s and 0 otherwise.

Considering $FP_{|s}$ as a statistical variable over the stocks, its distribution can be heuristically approximated with a log-normal distribution. Thus, the confidence intervals of this distribution can be used to classify stocks as subject to a high/medium/low fishing pressure:

Algorithm 4 Stock pressure classifier

For each stock:

sum the *fah*s in the associated cells

count the associated cells

calculate fishing *pressure* $FP_{|s}$

Fit $FP_{|s}$ to a log-normal distribution

calculate the geometric mean

calculate confidence limits

For each stock:

if ($FP_{|s}$ ≥ upper confidence limit) → *high-pressured stock*

if (lower confidence limit < $FP_{|s}$ < upper confidence limit) → *medium-pressured stock*

if ($FP_{|s}$ ≤ lower confidence limit) → *low-pressured stock*

Calculating fishing pressure on a species is meaningful only if we know the species catchability, which is directly related with the gear deployed and the species' location. Introducing catchability data in our algorithm requires the availability of global FAIR stock assessment data (Thorsteinsson, 2002), which unfortunately do not exist yet. Instead, our approach approximates risk for *non-stock* species from observations in the catch (then assuming their catchability is high), or where there is a significant indirect consequence of the interference of human activity with the ecosystem (Tromeur and Doyen, 2019; Hilborn et al., 2020). This risk is higher for those species that are concentrated and frequent in the high fishing-activity cells.

One way to estimate this risk for *non-stock* species as an *impact score*, is to define the following weighted sum:

$$I_{|s} = \frac{\sum_h fah_{|s}(h) \cdot n.occurrences_{|s}(h)}{count(h_{|s})}$$

where h is a high fishing-activity cell; $fah_{|s}(h)$ is the number of fishing hours in cell h where also species s (a *non-stock* species) is present; $n.occurrences_{|s}(h)$ is the number of occurrence records of species s in cell h (related with its commonness in the area, Coro et al. (2015b)); $count(h_{|s})$ is the total number of high fishing-activity cells $h_{|s}$ where the species is present. It is worth stressing that the set s of species over which $I_{|s}$ is defined (i.e., *non-stocks*) is different from that of $FP_{|s}$ (i.e., *stocks*).

The rationale of this formula is the following: if a species is very concentrated where a high fishing activity is present, the species impact score should increase. Conversely, if the species is rare in the high fishing-activity zones, the impact score should decrease. Finally, if fishing hours are relatively low in an extensive area, the impact score should decrease. The $I_{|s}$ score can be seen as a statistical variable over the non-stock species and can be heuristically modelled with a log-normal distribution. The confidence intervals of this distribution can be used to assess if a species is *potentially impacted* by the fishing activity:

Algorithm 5 Potentially impacted species detector

For each non-stock species:

select the high fishing-activity cells where the species has occurrence records

weigh the *fah*s in these cells for the number of occurrence records

sum the weighted *fah*s and divide by the number of selected high fishing-activity cells

calculate the *impact score* $I_{|s}$

Fit $I_{|s}$ to a log-normal distribution

calculate the geometric mean

calculate confidence limits

For each non-stock species:

if ($I_{|s}$ ≥ upper confidence limit) → *potentially impacted species*

The use of the upper confidence limit of the log-normal distribution selects species that have an outstanding *impact score*. This approach aims at maximising the probability that a species with a high score is really impacted (i.e., is a true positive), and thus to increase the classification precision. Using a lower threshold would have been more precautionary but less precise. In order to explore the unclassified impacted species, more information should indeed be introduced in our system (Section 3).

This final step of our algorithm produces a list of potentially impacted species and the distribution of fishing pressure on stocks, which contribute to give an overview of the potential impact of the fishing activity on the study area.

1.4. Open Science methodology and tools

Our workflow implements a FAIR approach that tests FAIR data principles' practicability. It is open-source (Supplementary material) and was integrated with the DataMiner Cloud computing platform of the D4Science e-Infrastructure (Coro et al., 2017), which allows accessing the mentioned knowledge sources on-the-fly during processing (Candela et al., 2016; Coro et al., 2015a). Data FAIRness is facilitated through the indexing of these resources in the D4Science *catalogue* (Assante et al., 2019b), which can be accessed by all processes via the *Catalogue Services for the Web* (CSW) standard of the Open Geospatial Consortium (OGC, 2020). Geospatial data are offered as standarised NetCDF files available on a distributed ISO/OGC compliant Spatial Data Infrastructure included in D4Science (Assante et al., 2019b). All other resources are stored on a distributed system based on MongoDB (mongodb.com) that ensures high availability and a fast access to the resources via direct HTTP connection (Assante et al., 2019a). The GRSF is hosted and managed by D4Science, which optimises access time and ensures a high service availability. Data external to D4Science (e.g., OBIS) are accessed by the processes via direct connection through provider-specific libraries.

DataMiner offers 15 machines with Ubuntu 18.04.5 LTS x86 64 operating system, 16 virtual cores, 32 GB of RAM, and 100 GB of disk, to run executions in parallel/distributed and multi-tenancy modes. Furthermore, this platform enables the repeatability, reproducibility, reusability, and interoperability of the processes, within a collaborative online environment (Assante et al., 2019b; Coro et al., 2021). To this aim, it offers a scriptto-service transformation tool and a provenance tracking feature (i.e., it records all input and output data, parameters, and metadata) (Coro et al., 2016c). The hosted services are published under the Web Processing Service standard (WPS, Schut and Whiteside (2007)) of the Open Geospatial Consortium (OGC) to maximise their reuse from other software. Moreover, a Web graphic interface is automatically generated based on the input/output definitions. Among the advantages of publishing our algorithms via D4Science are (i) low maintenance costs, (ii) the native support of an integrated distributed storage system in a cloud computing platform with online collaborative tools, and catalogues of metadata and geospatial data, and (iii) a long-term sustainability plan based on a large number of European projects (Assante et al., 2019b).

D4Science supports Virtual Research Environments (VREs), Web-based environments that foster the collaboration between users working on the same topic while managing data and services access policies (Candela et al., 2013). D4Science also includes security and accounting facilities that monitor the usage of all VRE resources (storage, computational services, etc), and prevent policy violations. Currently, D4Science hosts more than 150 VREs, with either free or moderated-access, which are the main means to foster the reuse of processes across application domains including our workflow. Overall, these features guaranteed a fast development of our workflow as a multi-source, parallel, secure, and Open Science process.

Our workflow was implemented on DataMiner as two different WPS Web services (Supplementary material), namely *Fishing Activity and Pressure from VTI data* (FAP-VTI) and *Fishing Activity and Pressure from Global Fishing Watch data* (FAP-GFW). The FAP-VTI service accepts user-provided vessel tracking data in CSV format, and focuses the analysis on the implicit time frame and spatial extent of these data. Instead, the FAP-GFW service embeds all daily vessel tracking data at 0.1° resolution, between 2012 and 2016, from the Global Fishing Watch. In this free-to-use dataset, fishing activity is pre-classified using a machine-learning model, which allows skipping our fishing activity classifier (Section 1.1). The FAP-GFW service asks the user to simply draw a bounding box on an ocean area, and to specify the analysis period, but is limited to the 2012-2016 time frame. The other input parameters and the produced output are the same between the two services (Section 1).

1.5. Benchmark data

As a first benchmark dataset, an area in the Northwest Atlantic off the coasts of Canada and North U.S.A. was used (Figure 2), because of its abundance of associated species and stock data. This 481,958.228 km^2 area is monitored by the Northwest Atlantic Fisheries Organization (NAFO) and is under the national jurisdiction of U.S.A. and Canada. In this area, vessel tracking data were collected from two main public sources:

- BOEM-Marine Cadastre (BMC) (BOEM, 2020): a collection of GIS mapping resources from Alaska to the Gulf of Mexico; started by the Bureau of Ocean Energy Management (BOEM) and the National Oceanic and Atmospheric Administration (NOAA) in U.S.A. to monitor offshore data and offer dissemination and analysis tools for policymakers and citizens. Marine Cadastre data are downloadable in CSV format and contain the information required by our methodology. Fishing vessel tracking data (types 30 and 1001) were downloaded for the Northwest Atlantic analysis area, and fishing activity was classified using the algorithm described in Section 1.1.

- Google Global Fishing Watch (GFW) (Merten et al., 2016): A Website managed by Google in partnership with Oceana and SkyTruth to give a global view of commercial fishing activities, based on satellite, AIS-terrestrial, and Infrared Imaging Radiometer Suite (VIIRS) data. At the time of writing, global-scale GFW data were downloadable for scientific purposes for the 2012-2016 period only. GFW distributes vessel data aggregated at 0.01° and 0.1° resolutions with fishing activity cells already classified through a machine learning model.

Our first benchmark dataset contained data from BMC and GFW in the fishing season from March to May 2015 for ease of result presentation. In the selected region and time frame, GFW data were much more abundant than BMC data: ~20,000 fishing hours (GFW) against ~5,500 hours (BMC).

As a second benchmark dataset, the complete time series of annual-aggregated GFW data from 2012 to 2016 in the Northwest Atlantic study area was used to show how our results can illustrate large fishing pattern changes over time.

As a third benchmark dataset, CSV data from the March-April 2020 COVID-19 lockdown period were produced from GFW raster images around Italy (GFW, 2020), and were reused to estimate the effects of the lockdown restrictions on the fishing pressure over biodiversity and threatened species.

2. Results

This Section reports three evaluation cases for our methodology. The first case demonstrates that the information retrieved from two different datasets is coherent in terms of stock and impacted species *composition*, i.e., in the number of different species retrieved for each year (Section 2.1). The second case demonstrates how our methodology can be used to highlight fishing-activity and stock composition patterns over time (Section 2.2). The third case, shows how our methodology can inform about the change of pressure on biodiversity and threatened species after a large socio-economic phenomenon like the COVID-19 lockdown in Italy. All experiments used the following input parameters: *Spatial Resolution* = 0.5°, *Minimum Number Of Records* = 5, *Occurrence Time Range* = 8. The complete output, the source code, and the links to the used Web services are reported in the Supplementary material.

2.1. Cross-data source consistency

Our analysis inferred 17 high fishing-activity cells from both the BMC and GFW datasets with independent statistical analyses in the selected Northwest Atlantic study area, but their distributions were different (Figure 2). Overlaps were present in the Southwest and central parts of the fishing area, but the fishing locations estimated from the BMC were generally closer to the coast than those estimated from the GFW. This discrepancy is due to (i) the different sizes of the two datasets, (ii) the use of our fishing activity classifier for BMC data against the Google's classification for GFW data, and (iii) a greater presence of AIS-terrestrial data in the BMC dataset. However, the fishing-activity cells were uniformly distributed over the entire area in both cases.

Despite the discrepancies between the two datasets, the extracted information on stocks and biodiversity had a great overlap (Table 1): 28 stocks (of which 12 highly-pressured stocks) were extracted from BMC, and 29 (of which 15 highly pressured stocks) from GFW, with an overlap of 27 stocks (90%). Only three stocks (10%) were not included in both the lists, and 11 stocks (68.8%) were jointly indicated as highly pressured. Furthermore, 92 *non-stock* species were inferred from BMC (2 *potentially impacted*), and 77 from GFW (3 *potentially impacted*), with an overlap of 70.7% between the two lists. The 2 BMC *potentially impacted* species were included in the 3 GFW *potentially impacted* species. As for threatened species (both stock and non-stock), 29 were detected from BMC and 27 from GFW, with 25 (80.6%) shared between the two analyses.

These overlaps indicate a large agreement between our two parallel analyses and thus enforce the validity of our approach.

2.2. Temporal analysis

In the second case study, our methodology was executed on annual-aggregated GFW data in the Northwest Atlantic study area (Figure 3). This analysis highlighted a change of fishing activity patterns and intensity over time (Figure 3-a). In particular, (i) an expansion of the fishing activity can be observed from 2012 to 2013, (ii) a partial shift of the intense fishing activity towards the South-East occurred in 2014, (iii) fishing patterns were very similar between 2014 and 2015, as also reflected by the low variation of both stock and non-stock compositions, (iv) an overall contraction of the fishing activity can be observed between 2015 and 2016 coupled with a significant increment of the fishing hours and fishing activity density in 2016. Generally, a non-linear growth of fishing hours occurred in the entire area and in the high fishing-activity cells (Figure 3-b and -c). The number of different species retrieved had a low variation (9.8% coefficient of variation) - since fishing activity covered a major part of study area - but decreased by 18% in 2016 (Figure 3-d). The number of inferred stocks and threatened species had a low variation too (8.6% and 6.7% coefficient of variation, respectively) (Figures 3-e and -f). There were generally few *potentially impacted* non-stock species (Figures 3-g). These always included the critically endangered *Squalus acanthias*, and also birds concentrated in high fishing-activity locations that could be accidentally captured (e.g., *Larus argentatus* and *Puffinus gravis*), which is a known issue (Eng and Pipkin, 2020).

The variation in the species composition, i.e., the percentage of different species involved between two years, was higher between 2013 and 2012 and between 2016 and 2015 than in the other years (Figure 3-h). This variation highlights a slight shift of the fishing patterns in these years, with potential different impacts on the biodiversity of the area. In particular, in 2013, the fishing area expanded to cover 30% more (reported) species than in 2012, and in 2016, the contracted fishing area covered 22% fewer species than in 2015. The consistent fishing patterns in intermediate years are related to a low species-composition variation. This similarity was reflected also by a stock composition overlap, but only when fishing patterns were very similar (e.g., between 2015 and 2014, Figure 3-i). Overall, the composition of targeted stocks hardly changed across the years, with a maximum of 5 stocks between 2013 and 2012. This observation may indicate that the changing patterns were caused by population shifts (Section 3).

2.3. Effect of COVID-19 lockdown on fisheries in Italy

The Global Fishing Watch recently published a comparative analysis between fishing activity around Italian coasts during the 2020 COVID-19 lockdown from 1 March to 30 April, and the activity of the same period in 2019 (GFW (2020), Figure 4-a). This time-frame is of particular importance for Italy, because it was the period of maximum lockdown and infection (Coro, 2020a).

In our third case study, the GFW data of this period were processed through our workflow to enrich them with information on stocks, biodiversity, and species' IUCN-threatening status, and to highlight possible consequences of lockdown restrictions. As a first step, our analysis aggregated the GFW data at a 0.5° resolution. Then, the areas with the highest fishing activity were calculated for 2019, and the same statistical confidence intervals were used to estimate high fishing-activity locations in 2020 (Figure 4-b and -c). As a result, this operation highlighted that the greatest loss of fishing hours occurred in the Adriatic Sea. Interestingly, a significantly lower number of fishing hours was reported in the northern Tyrrhenian Sea - within the Ligurian Sea Cetacean Sanctuary - and in the highly urbanised region of the gulf of Naples, which host several threatened species.

Out of this information, our analysis extracted data on the richness, presence, and vulnerability of ~100 species and 843 observations in the cells with the highest fishing activity. Species information was also aggregated per cell and classified into clusters of low/medium/high quantities using log-normal confidence limits (Figure 4-d, -e, and -f). As a result, some potential effects of the lockdown period on biodiversity (i.e., the number of different species per cell) and presence (i.e., the number of occurrence records per cell) became apparent. In particular, intense fishing activity decreased mainly in locations with *medium* biodiversity and presence levels. Thus, fishing pressure on the species living in these locations was lower in 2020 than in 2019. Unfortunately, locations with rich biodiversity and presence faced high fishing pressure also during the lockdown period. Interestingly, several locations with medium/high numbers of threatened species underwent a much lower fishing activity in 2020 (from 1000 to 3000 hours, i.e., 40-70% less).

3. Discussion and conclusions

This paper has presented a methodology to aggregate, classify, and extract new information from vessel-transmitted data through the analysis of heterogeneous data sources in an Open Science e-Infrastructure. The results confirm the feasibility of cross-domain analysis if FAIR data principles are considered when establishing data repositories. Potential applications, based on Open Science principles, have been demonstrated through three case studies.

The first case study has shown that our Open Science process can produce consistent information from two different and large input datasets. The extracted major fishing patterns differ because of (i) the heterogeneous data collection systems used, (ii) the order of magnitude difference of the dataset sizes, and (iii) the different fishing-activity classification algorithms used. However, the extracted information about stocks, species composition, vulnerability levels, and fishing pressure per species, largely overlaps and thus is cross-dataset consistent. The detected stocks are also monitored by the Northwest Atlantic Fisheries Organization in the study area (NAFO, 2020), and the highest *impacted* species have been reported as bycatch species in this area also by other studies. For example, *Puffinus gravis* and *Larus argentatus* are seabirds commonly captured in the Northwest Atlantic (Zhou et al., 2019; Kelleher, 2005); the IUCN-vulnerable *Squalus acanthias* is a benthopelagic and oceanic species that is frequently captured in Northwest Atlantic by commercial fisheries (Tallack and Mandelman, 2009). The few complementary species and stocks found by the two analyses depend on the different distributions of fishing hours across the study area, which consequently correspond to different observation records in OBIS. Overall, our first case study has shown how our method can infer the possible target stocks of the fisheries and the overlap between fishing activity and the threatened species present in the study area, just from a set of vessel trajectory data.

The second case study has shown a temporal analysis conducted on the GFW data, on another time scale (i.e., multi-annual) than the first case study. In the Northwest Atlantic study area, our analysis confirmed a general non-linear increase of the fishing effort and pressure also highlighted by other studies globally (Colloca et al., 2017; Froese et al., 2018; Rousseau et al., 2019). The analysis also highlighted expansions and contractions of fishing patterns between 2012 and 2016, which other studies have indicated as being the consequence of populations' shift due to climate change and fishing pressure increase (Greene and Pershing, 2000; Burgess et al., 2005; Merzouk and Johnson, 2011; Mills et al., 2013; Boudreau et al., 2017; Adams et al., 2018; McManus et al., 2018; Stanley et al., 2018). When used within an Integrated Environmental Assessment system, our analysis can inform about fishing-pressure change in time and its potential impact on threatened species (Piet et al., 2006; Mouillot et al., 2011; Coll et al., 2012). Furthermore, our output can be the input of other models that monitor and forecast fishing activity change, stock exploitation, and population shift (Coro et al., 2016a; Coro and Walsh, 2021). Finally, our Open Science services can flexibly manage different levels of temporal aggregations (i.e., seasonal, annual, etc.) to support studies of stock distribution change across fishing periods.

The third case study has demonstrated how our methodology can enrich the GFW analysis on the fishing-activity change in Italy due to the March-April 2020 COVID-19 lockdown. Our analysis has highlighted that a beneficial reduction of fishing pressure to ecosystems and biodiversity has potentially occurred in several sea areas of Italy. A major reduction of the potential impact of fishing activity is expected for vulnerable species in the Ligurian Sea and off the Campanian coasts, where a large variety of threatened species is concentrated. This observation agrees with the positive effects observed on Italian wildlife after the lockdown restrictions (Manenti et al., 2020), and generally with the expectations on other World areas (Chitra et al., 2020; McVeigh, 2020; Michael, 2020). However, high fishing pressure has persisted during the lockdown period in areas with a great variety of species. Fishing pressure did not reduce, especially in the Adriatic Sea, where already many stocks are at or above the maximum sustainable fishing pressure (Froese et al., 2018). This aspect is currently under study with localised investigations to evaluate its reflection on profitability (CNR, 2020).

One source of bias in our analysis is the non-uniform and scarce reporting of occurrence records in the OBIS database. Adjusting the *Occurrence Time Range* and the *Minimum Number Of Records* parameters can partially account for this bias because if a species is *uncommon* in a specific place, it unlikely has occurrence records in OBIS within a sufficiently large time frame (Coro et al., 2015b, 2016b; Claus et al., 2018). In the future developments of our methodology, this issue will be managed by enabling the possibility of using additional sources of occurrence records connected to the e-Infrastructure (e.g., the Global Biodiversity Information Facility, Lane and Edwards (2007)). Another approach to compensate for this bias would be to use multiple spatial resolutions - by changing the *Spatial Resolution* parameter - and check how consistent the list of species is across spatial aggregations. A multi-resolution decision approach is usually effective in these cases (Magliozzi et al., 2019). Another aspect of our approach is that the statistical analysis has a higher precision when vessel data are abundant, and the analysis resolution is suited to the study area. Thus, the user should provide statistically significant data and use the most appropriate spatial resolution for the analysis. These considerations relate to general issues with FAIR data and big data processing: easy access to a large amount of data comes at the expense of a low guarantee of data quality and completeness. The precision of our workflow's output depends on (i) the completeness of the input vessel data, (ii) the update rate of the GRSF, (iii) the completeness of the OBIS data in the selected time range, and (iv) the suitability of the selected spatial resolution for the analysis. However, our first and second case studies have demonstrated that our workflow can compensate for some of these biases - through data classification and spatio-temporal aggregation - mainly when large input datasets are used. Generally, it is worth noting that all big data processing methods are approximate, but they can discover general and valuable knowledge if the approximation is tolerated within the application context (Coro, 2020b).

One important aspect of our methodology is the use of Open Science, which required to release our process as OGC-compliant and open-access Web Services within an Open Science e-Infrastructure. The used platform supports reproducible and repeatable experimentation, thus all our results can be verified through simple WPS invocations via a Web browser or a compliant software (e.g., QGIS or ArcGIS). The accepted input data are plain CSV files, which allows for rapidly feeding the workflow with new vessel tracking data from private and public repositories while the e-Infrastructure guarantees the privacy of the data and of the experiments. Finally, the e-Infrastructure maximises the reuse of our processes across Virtual Research Environments, i.e., virtual laboratories for scientists focussing on different experiments related to Marine Science, COVID-19 (Coro, 2020a), or other disciplines (Coro and Trumpy, 2020c). Virtual Research Environments can be the backbone for instantiating cross-institute collaborations to process and share vessel data and to guarantee that data access policies are respected (Galdelli et al., 2019). Specific initiatives to investigate the effect of lockdown restrictions on marine resources through this technology have already started in Europe (Blue Cloud, 2019; CNR, 2020).

This paper has demonstrated how new knowledge can be generated out of FAIR fisheries data. Furthermore, newly available information (e.g., catchability) can be integrated with our methodology to enhance classification precision. For example, FAIR data with information on catchability, fishing gears, environmental parameters, and life-history traits can be used to identify bycatch species (Lewison et al., 2013), or to study the interaction between different fisheries (e.g., bottom, mid-water trawling, etc.) with the habitat preferences (e.g., benthic, epi-pelagic, and purse seine) and the size distribution of the species in the fished area (Armstrong and Falk-Petersen, 2008; Foley et al., 2012). The Open Science implementation of our methodology guarantees that these sources can be rapidly connected and integrated with the current implementation as soon as new FAIR data sources are available.

Acknowledgments

The reported work has been partially supported by the Blue Cloud project (H2020 framework of the European Commission, H2020-EU.3.2.5.1. Program grant agreement No. 862409).

References

Adams, C.F., Alade, L.A., Legault, C.M., O'Brien, L., Palmer, M.C., Sosebee, K.A., Traver, M.L., 2018. Relative importance of population size, fishing pressure and temperature on the spatial distribution of nine northwest atlantic groundfish stocks. PLOS ONE 13, 1–14. URL: `https://doi.org/10.1371/journal.pone.0196583`, doi:`10.1371/journal.pone.0196583`.

Agapito, M., Chuenpagdee, R., Devillers, R., Gee, J., Johnson, A.F., Pierce, G.J., Trouillet, B., 2019. Beyond the basics: improving information about small-scale fisheries, in: Transdisciplinarity for Small-Scale Fisheries Governance. Springer, pp. 377–395.

Antunes, P., Santos, R., 1999. Integrated environmental management of the oceans. Ecological Economics 31, 215–226.

Armstrong, C.W., Falk-Petersen, J., 2008. Habitat–fisheries interactions: a missing link? ICES Journal of Marine Science/Journal du Conseil 65.

Assante, M., Candela, L., Castelli, D., Cirillo, R., Coro, G., Frosini, L., Lelii, L., Mangiacrapa, F., Marioli, V., Pagano, P., et al., 2019a. The gcube system: delivering virtual research environments as-a-service. Future Generation Computer Systems 95, 445–453.

Assante, M., Candela, L., Castelli, D., Cirillo, R., Coro, G., Frosini, L., Lelii, L., Mangiacrapa, F., Pagano, P., Panichi, G., 2019b. Enacting open science by d4science. Future Generation Computer Systems 101, 555–563.

Bastardie, F., Nielsen, J.R., Ulrich, C., Egekvist, J., Degel, H., 2010. Detailed mapping of fishing effort and landings by coupling fishing logbooks with satellite-recorded vessel geo-location. Fisheries Research 106, 41–53.

Belhabib, D., Cheung, W.W., Kroodsma, D., Lam, V.W., Underwood, P.J., Virdin, J., 2020. Catching industrial fishing incursions into inshore waters of africa from space. Fish and Fisheries 21, 379–392.

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

Blue Cloud, 2019. The Blue Cloud European Project. Project description accessible online at https://cordis.europa.eu/project/id/862409/it - accessed Sept. 2020.

BOEM, 2020. BOEM Marine Cadastre Initiative. Web site https://www.nafo.int/Science/Science-Advice/Species - accessed Sept. 2020.

Boudreau, S.A., Shackell, N.L., Carson, S., den Heyer, C.E., 2017. Connectivity, persistence, and loss of high abundance areas of a recovering marine fish population in the northwest atlantic ocean. Ecology and Evolution 7, 9739–9749. doi:10.1002/ece3.3495.

Burgess, G.H., Beerkircher, L.R., Cailliet, G.M., Carlson, J.K., Cortés, E., Goldman, K.J., Grubbs, R.D., Musick, J.A., Musyl, M.K., Simpfendorfer, C.A., 2005. Is the collapse of shark populations in the northwest atlantic ocean and gulf of mexico real? Fisheries 30, 19–26. URL: https://doi.org/10.1577/1548-8446(2005)30[19:ITCOSP]2.

0.CO;2, doi:10.1577/1548-8446(2005)30[19:ITCOSP]2.0.CO;2, arXiv:https://doi.org/10.1577/1548-8446(2005)30[19:ITCOSP]2.0.CO;2.

Campanis, G., 2008. Advancements in vms data analyses. NAFO Annual Report for 2008 Accessible online at https://www.nafo.int/Portals/0/PDFs/ar/ar08.pdf?ver=2016-02-10-101458-760.

Candela, L., Castelli, D., Coro, G., Pagano, P., Sinibaldi, F., 2016. Species distribution modeling in the cloud. Concurrency and Computation: Practice and Experience 28, 1056–1079.

Candela, L., Castelli, D., Pagano, P., 2013. Virtual research environments: an overview and a research agenda. Data Science Journal , GRDI–013.

Chang, S.J., 2003. Vessel identification and monitoring systems for maritime security, in: IEEE 37th Annual 2003 International Carnahan Conference onSecurity Technology, 2003. Proceedings., IEEE. pp. 66–70.

Chitra, J., Rajendran, S., Mercy, J.J., Jeyakanthan, J., 2020. Impact of covid-19 lockdown in tamil nadu: Benefits and challenges on environment perspective. online publication http://nopr.niscair.res.in/handle/123456789/54777.

Claus, S., Arvanitidis, C., Bailly, N., Deneudt, K., Lear, D., Oset, P., Vandepitte, L., 2018. Unlocking european marine biodiversity under emodnet biology data using the fair principles. Bollettino di Geofisica , 215.

CNR, 2020. The SNAPSHOT Project Virtual Research Environment of the National Research Council of Italy. Web portal to the VRE: https://snapshot.d4science.org/ - accessed Sept. 2020.

Coll, M., Piroddi, C., Albouy, C., Ben Rais Lasram, F., Cheung, W.W.L., Christensen, V., Karpouzi, V.S., Guilhaumon, F., Mouillot, D., Paleczny, M., Palomares, M.L., Steenbeek, J., Trujillo, P., Watson, R., Pauly, D., 2012. The mediterranean sea under siege: spatial overlap between marine biodiversity, cumulative threats and marine reserves. Global Ecology and Biogeography 21, 465–480. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1466-8238.2011.00697.x, doi:10.1111/j.1466-8238.2011.00697.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1466-8238.2011.

Collins, S., Genova, F., Harrower, N., Hodson, S., Jones, S., Laaksonen, L., Mietchen, D., Petrauskaitė, R., Wittenburg, P., 2018. Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data. Online publication: https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf.

Colloca, F., Scarcella, G., Libralato, S., 2017. Recent trends and impacts of fisheries exploitation on mediterranean stocks and ecosystems. Frontiers in Marine Science 4, 244.

Cornillon, P., Gallagher, J., Sgouros, T., 2003. Opendap: Accessing data in a distributed, heterogeneous environment. Data Science Journal 2, 164–174.

Coro, G., 2020a. A global-scale ecological niche model to predict sars-cov-2 coronavirus infection rate. Ecological Modelling 431, 109187.

Coro, G., 2020b. Open science and artificial intelligence supporting blue growth. Environmental Engineering and Management Journal 19, 1719–1729.

Coro, G., Candela, L., Pagano, P., Italiano, A., Liccardo, L., 2015a. Parallelizing the execution of native data mining algorithms for computational biology. Concurrency and Computation: Practice and Experience 27, 4630–4644.

Coro, G., Fortunati, L., Pagano, P., 2013. Deriving fishing monthly effort and caught species from vessel trajectories, in: 2013 MTS/IEEE OCEANS-Bergen, IEEE. pp. 1–5.

Coro, G., Large, S., Magliozzi, C., Pagano, P., 2016a. Analysing and forecasting fisheries time series: purse seine in indian ocean as a case study. ICES Journal of Marine Science 73, 2552–2571.

Coro, G., Magliozzi, C., Vanden Berghe, E., Bailly, N., Ellenbroek, A., Pagano, P., 2016b. Estimating absence locations of marine species from data of scientific surveys in obis. Ecological Modelling 323, 61 – 76. URL: http://www.sciencedirect.com/science/article/pii/S0304380015005761, doi:https://doi.org/10.1016/j.ecolmodel.2015.12.008.

Coro, G., Panichi, G., Pagano, P., 2016c. A web application to publish r scripts as-a-service on a cloud computing platform. Bollettino di Geofisica Teorica ed Applicata 57, 51–53.

Coro, G., Panichi, G., Pagano, P., Perrone, E., 2021. Nlphub: An e-infrastructure-based text mining hub. Concurrency and Computation: Practice and Experience 33, e5986.

Coro, G., Panichi, G., Scarponi, P., Pagano, P., 2017. Cloud computing in a distributed e-infrastructure using the web processing service standard. Concurrency and Computation: Practice and Experience 29, e4219.

Coro, G., Trumpy, E., 2020c. Predicting geographical suitability of geothermal power plants. Journal of Cleaner Production , 121874.

Coro, G., Walsh, M.B., 2021. An intelligent and cost-effective remote underwater video device for fish size monitoring. Ecological Informatics 63, 101311.

Coro, G., Webb, T.J., Appeltans, W., Bailly, N., Cattrijsse, A., Pagano, P., 2015b. Classifying degrees of species commonness: North sea fish as a case study. Ecological modelling 312, 272–280.

Davis, J.M., 2001. Monitoring control surveillance and vessel monitoring system requirements to combat iuu fishing. FAO FISHERIES REPORTS , 244–256.

Dinesen, G.E., Neuenfeldt, S., Kokkalis, A., Lehmann, A., Egekvist, J., Kristensen, K., Munk, P., Hüssy, K., Støttrup, J.G., 2019. Cod and climate: a systems approach for sustainable fisheries management of atlantic cod (gadus morhua) in coastal danish waters. Journal of Coastal Conservation 23, 943–958.

Dunn, D.C., Jablonicky, C., Crespo, G.O., McCauley, D.J., Kroodsma, D.A., Boerder, K., Gjerde, K.M., Halpin, P.N., 2018. Empowering high seas governance with satellite vessel tracking data. Fish and Fisheries 19, 729–739.

Eng, E., Pipkin, W., 2020. These Simple Fixes Could Save Thousands of Birds a Year From Fishing Boats. Smithsonian Magazine

Web site https://www.smithsonianmag.com/science-nature/these-simple-fixes-could-save-thousands-birds-year-fishing-boats-18095 - accessed Sept. 2020.

Eriksen, T., Høye, G., Narheim, B., Meland, B.J., 2006. Maritime traffic monitoring using a space-based ais receiver. Acta Astronautica 58, 537–549.

Eriksen, T., Skauen, A.N., Narheim, B., Helleren, Ø., Olsen, Ø., Olsen, R.B., 2010. Tracking ship traffic with space-based ais: Experience gained in first months of operations, in: 2010 International WaterSide Security Conference, IEEE. pp. 1–8.

European Parliament, 2008. Directive 2008/56/ec of the european parliament and of the council.

Farmanbar, M., Palanisamy, A., Høydal, A., Keprate, A., Haug, G., 2019. A web based solution to track trawl vessel activities over pipelines in norwegian continental shelf, in: IOP Conference Series: Materials Science and Engineering, IOP Publishing. p. 012037.

Foley, N.S., Armstrong, C.W., Kahui, V., Mikkelsen, E., Reithe, S., 2012. A review of bioeconomic modelling of habitat-fisheries interactions. International Journal of Ecology 2012.

Froese, R., Winker, H., Coro, G., Demirel, N., Tsikliras, A.C., Dimarchopoulou, D., Scarcella, G., Quaas, M., Matz-Lück, N., 2018. Status and rebuilding of european fisheries. Marine Policy 93, 159–170.

Galdelli, A., Mancini, A., Tassetti, A.N., Ferrà Vega, C., Armelloni, E., Scarcella, G., Fabi, G., Zingaretti, P., 2019. A cloud computing architecture to map trawling activities using positioning data. Volume 9: 15th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications URL: https://doi.org/10.1115/DETC2019-97779, doi:10.1115/DETC2019-97779, arXiv:https://asmedigitalcollection.asme.org/IDETC-CIE/proceedings-pdf/

Gari, S.R., Newton, A., Icely, J.D., 2015. A review of the application and evolution of the dpsir framework with an emphasis on coastal social-ecological systems. Ocean & Coastal Management 103, 63–77.

Garibaldi, L., Busilacchi, S., et al., 2002. Asfis list of species for fishery statistics purposes. FAO-AGRIS .

Gerritsen, H., 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 Journal of Marine Science 68, 245–252.

GFW, 2020. Global fisheries during COVID-19 - Global Fishing Watch. Web site GlobalfisheriesduringCOVID-19 - accessed Sept. 2020.

Gianelli, I., Horta, S., Martínez, G., de la Rosa, A., Defeo, O., 2018. Operationalizing an ecosystem approach to small-scale fisheries in developing countries: The case of uruguay. Marine Policy 95, 180–188.

Grassle, J.F., 2000. The ocean biogeographic information system (obis): an on-line, world-wide atlas for accessing, modeling and mapping marine biological data in a multidimensional geographic context. Oceanography 13, 5–7.

Greene, C.H., Pershing, A.J., 2000. The response of Calanus finmarchicus populations to climate variability in the Northwest Atlantic: basin-scale forcing associated with the North Atlantic Oscillation. ICES Journal of Marine Science 57, 1536–1544. URL: https://doi.org/10.1006/jmsc.2000.0966, doi:10.1006/jmsc.2000.0966, arXiv:https://academic.oup.com/icesjms/article-pdf/57/6/1536/2305779/57

Hilborn, R., Akselrud, C.A., Peterson, H., Whitehouse, G.A., 2020. The trade-off between biodiversity and sustainable fish harvest with area-based management. ICES Journal of Marine Science URL: https://doi.org/10.1093/icesjms/fsaa139, doi:10.1093/icesjms/fsaa139, arXiv:https://academic.oup.com/icesjms/advance-article-pdf/doi/10.1093/fsaa139.

i-Marine, 2020. The Global Record of Stocks and Fisheries. Accessible online at https://i-marine.d4science.org/web/grsf/data-catalogue - accessed Sept. 2020.

Iacarella, J.C., Lyons, D.A., Burke, L., Davidson, I.C., Therriault, T.W., Dunham, A., DiBacco, C., 2020. Climate change and vessel traffic create networks of invasion in marine protected areas. Journal of Applied Ecology 57, 1793–1805.

ITU, 2009. International telecommunication union - technical characteristics for an automatic identification system using time division multiple access in the vhf maritime mobile frequency band.

IUCN, 2001. IUCN Red List categories and criteria. IUCN.

James, M., Mendo, T., Jones, E.L., Orr, K., McKnight, A., Thompson, J., 2018. Ais data to inform small scale fisheries management and marine spatial planning. Marine Policy 91, 113–121.

Jennings, S., Lee, J., 2012. Defining fishing grounds with vessel monitoring system data. ICES Journal of Marine Science 69, 51–63.

John Caron, U., Davis, E., 2006. Unidata's thredds data server, in: 22nd International Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology, pp. 1–4.

Kelleher, K., 2005. Discards in the world's marine fisheries: an update. volume 470. FAO Web site `http://www.fao.org/3/y5936e/y5936e0d.htm#bm13.1` - Table 26.

Kia, M., Shayan, E., Ghotb, F., 2000. The importance of information technology in port terminal operations. International Journal of Physical Distribution & Logistics Management .

Koen-Alonso, M., Pepin, P., Fogarty, M.J., Kenny, A., 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. Marine Policy 100, 342–352.

Kristensen, P., 2004. The dpsir framework, european topic centre on water. European Environment Agency , 1–10.

Kurekin, A.A., Loveday, B.R., Clements, O., Quartly, G.D., Miller, P.I., Wiafe, G., Adu Agyekum, K., 2019. Operational monitoring of illegal fishing in ghana through exploitation of satellite earth observation and ais data. Remote Sensing 11, 293.

Lambert, G.I., Jennings, S., Hiddink, J.G., Hintzen, N.T., Hinz, H., Kaiser, M.J., Murray, L.G., 2012. Implications of using alternative methods of vessel monitoring system (vms) data analysis to describe fishing activities and impacts. ICES Journal of Marine Science 69, 682–693.

Lane, M.A., Edwards, J.L., 2007. The global biodiversity information facility (gbif). Biodiversity databases: Techniques, politics, and applications , 1–4.

Le Guyader, D., Ray, C., Gourmelon, F., Brosset, D., 2017. Defining high-resolution dredge fishing grounds with automatic identification system (ais) data. Aquatic Living Resources 30, 39.

Le Tixerant, M., Le Guyader, D., Gourmelon, F., Queffelec, B., 2018. How can automatic identification system (ais) data be used for maritime spatial planning? Ocean & Coastal Management 166, 18–30.

Lee, J., South, A.B., 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 Journal of Marine Science 67, 1260–1271.

Lewison, R., Wallace, B., Alfaro-Shigueto, J., Mangel, J.C., Maxwell, S.M., Hazen, E.L., 2013. Fisheries bycatch of marine turtles: lessons learned from decades of research and conservation, in: The Biology of Sea Turtles, Volume III. CRC Press, pp. 346–369.

Lockerbie, E.M., Lynam, C.P., Shannon, L.J., Jarre, A., 2018. Applying a decision tree framework in support of an ecosystem approach to fisheries: Indiseas indicators in the north sea. ICES Journal of Marine Science 75, 1009–1020.

Lopes, P.F., Verba, J.T., Begossi, A., Pennino, M.G., 2019. Predicting species distribution from fishers' local ecological knowledge: a new alternative for data-poor management. Canadian Journal of Fisheries and Aquatic Sciences 76, 1423–1431.

Magliozzi, C., Coro, G., Grabowski, R.C., Packman, A.I., Krause, S., 2019. A multiscale statistical method to identify potential areas of hyporheic exchange for river restoration planning. Environmental Modelling & Software 111, 311–323.

Manenti, R., Mori, E., Di Canio, V., Mercurio, S., Picone, M., Caffi, M., Brambilla, M., Ficetola, G.F., Rubolini, D., 2020. The good, the bad and the ugly of covid-19 lockdown effects on wildlife conservation: Insights from the first european locked down country. Biological conservation 249, 108728.

McManus, M.C., Hare, J.A., Richardson, D.E., Collie, J.S., 2018. Tracking shifts in atlantic mackerel (scomber scombrus) larval habitat suitability on the northeast u.s. continental shelf. Fisheries Oceanography 27, 49–62. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/fog.12233, doi:10.1111/fog.12233, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/fog.12233.

McVeigh, K., 2020. Silence is golden for whales as lockdown reduces ocean noise. The Guardian 27.

Merten, W., Reyer, A., Savitz, J., Amos, J., Woods, P., Sullivan, B., 2016. Global fishing watch: Bringing transparency to global commercial fisheries. arXiv preprint arXiv:1609.08756 .

Merzouk, A., Johnson, L.E., 2011. Kelp distribution in the northwest atlantic ocean under a changing climate. Journal of Experimental Marine Biology and Ecology 400, 90 – 98. URL: http://www.sciencedirect.com/science/article/pii/S0022098111000682, doi:https://doi.org/10.1016/j.jembe.2011.02.020. global change in marine ecosystems.

Michael, M., 2020. Which animals are benefitting from coronavirus lockdowns? New Scientist online article https://www.newscientist.com/article/2244359-which-animals-are-benefitting-from-coronavirus-lockdowns/ - accessed Sept. 2020.

Mills, K.E., Pershing, A.J., Sheehan, T.F., Mountain, D., 2013. Climate and ecosystem linkages explain widespread declines in north american atlantic salmon populations. Global Change Biology 19, 3046–3061. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.12298, doi:10.1111/gcb.12298, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.12298.

Mouillot, D., Albouy, C., Guilhaumon, F., Ben Rais Lasram, F., Coll, M., Devictor, V., Meynard, C., Pauly, D., Tomasini, J., Troussellier, M., Velez, L., Watson, R., Douzery, E., Mouquet, N., 2011. Protected and threatened components of fish biodiversity in the mediterranean sea. Current Biology 21, 1044 – 1050. URL: http://www.

sciencedirect.com/science/article/pii/S096098221100532X, doi:https://doi.org/10.1016/j.cub.2011.05.005.

Muawanah, U., Yusuf, G., Adrianto, L., Kalther, J., Pomeroy, R., Abdullah, H., Ruchimat, T., 2018. Review of national laws and regulation in indonesia in relation to an ecosystem approach to fisheries management. Marine Policy 91, 150–160.

Muench, A., DePiper, G.S., Demarest, C., 2018. On the precision of predicting fishing location using data from the vessel monitoring system (vms). Canadian Journal of Fisheries and Aquatic Sciences 75, 1036–1047.

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. Marine Policy 108, 103623.

Murawski, S.A., Wigley, S.E., Fogarty, M.J., Rago, P.J., Mountain, D.G., 2005. Effort distribution and catch patterns adjacent to temperate mpas. ICES Journal of Marine Science 62, 1150–1167.

NAFO, 2020. Northwest Atlantic Fisheries Organization - List of monitored stocks. Web site https://www.nafo.int/Science/Science-Advice/Species - accessed Sept. 2020.

Natale, F., Gibin, M., Alessandrini, A., Vespe, M., Paulrud, A., 2015. Mapping fishing effort through ais data. PloS one 10, e0130746.

OGC, 2020. Catalogue Services for the Web specifications. Open Geospatial Consortium Technical Reports .

Olesen, P.B., Dukovska-Popovska, I., Hvolby, H.H., 2012. Improving port terminal operations through information sharing, in: IFIP International Conference on Advances in Production Management Systems, Springer. pp. 662–669.

Pallotta, G., Vespe, M., Bryan, K., 2013. Vessel pattern knowledge discovery from ais data: A framework for anomaly detection and route prediction. Entropy 15, 2218–2245.

Palmer, M.C., 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. North American Journal of Fisheries Management 29, 928–942.

Piet, G.J., Quirijns, F.J., Robinson, L., Greenstreet, S.P.R., 2006. Potential pressure indicators for fishing, and their data requirements. ICES Journal of Marine Science 64, 110–121. URL: https://doi.org/10.1093/icesjms/fsl006, doi:10.1093/icesjms/fsl006, arXiv:https://academic.oup.com/icesjms/article-pdf/64/1/110/29126787/fs

Previero, M., 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 & Coastal Management 154, 83–95.

Provoost, P., Bosch, S., Appletans, W., 2017. robis: R client to access data from the obis api. Ocean Biogeographic Information System. Intergovernmental Oceanographic Commission of UNESCO. URL: https://cran. r-project. org/package= robis .

Robards, M., Silber, G., Adams, J., Arroyo, J., Lorenzini, D., Schwehr, K., Amos, J.,

2016. Conservation science and policy applications of the marine vessel automatic identification system (ais)—a review. Bulletin of Marine Science 92, 75–103.

Roberson, L.A., Kiszka, J.J., Watson, J.E., 2019. Need to address gaps in global fisheries observation. Conservation Biology 33, 966–968.

Rousseau, Y., Watson, R.A., Blanchard, J.L., Fulton, E.A., 2019. Evolution of global marine fishing fleets and the response of fished resources. Proceedings of the National Academy of Sciences 116, 12238–12243. URL: `https://www.pnas.org/content/116/25/12238`, doi:`10.1073/pnas.1820344116`, `arXiv:https://www.pnas.org/content/116/25/12238.full.pdf`.

Russo, T., Parisi, A., Cataudella, S., 2013. Spatial indicators of fishing pressure: Preliminary analyses and possible developments. Ecological indicators 26, 141–153.

Schut, P., Whiteside, A., 2007. OpenGIS Web Processing Service. Open Geospatial Consortium. Open Geospatial Consortium Technical Reports .

Shaw, D.R., Grainger, A., Achuthan, K., 2017. Multi-level port resilience planning in the uk: how can information sharing be made easier? Technological Forecasting and Social Change 121, 126–138.

Shepperson, J.L., Hintzen, N.T., Szostek, C.L., Bell, E., Murray, L.G., Kaiser, M.J., 2018. A comparison of vms and ais data: The effect of data coverage and vessel position recording frequency on estimates of fishing footprints. ICES Journal of Marine Science 75, 988–998.

Song, A.M., Johnsen, J.P., Morrison, T.H., 2018. Reconstructing governability: How fisheries are made governable. Fish and Fisheries 19, 377–389.

de Souza, E.N., Boerder, K., Matwin, S., Worm, B., 2016. Improving fishing pattern detection from satellite ais using data mining and machine learning. PloS one 11, e0158248.

Stanley, R.R.E., DiBacco, C., Lowen, B., Beiko, R.G., Jeffery, N.W., Van Wyngaarden, M., Bentzen, P., Brickman, D., Benestan, L., Bernatchez, L., Johnson, C., Snelgrove, P.V.R., Wang, Z., Wringe, B.F., Bradbury, I.R., 2018. A climate-associated multispecies cryptic cline in the northwest atlantic. Science Advances 4. doi:10.1126/sciadv.aaq0929.

Taconet, M., Kroodsma, D., 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/.

Taconet, P., Chassot, E., Guitton, J., Fiorellato, F., Anello, E., Barde, J., 2016. Data toolbox for fisheries: the case of tuna fisheries. Accessible online at https://www.iotc.org/sites/default/files/documents/2018/04/IOTC-2016-WPDCS12-27_-_TUNA_DATA_TOOLBOX.pdf.

Tallack, S.M., Mandelman, J.W., 2009. Do rare-earth metals deter spiny dogfish? a feasibility study on the use of electropositive "mischmetal" to reduce the bycatch of squalus acanthias by hook gear in the gulf of maine. ICES Journal of Marine Science 66, 315–322.

Tetreault, B.J., 2005. Use of the automatic identification system (ais) for maritime domain awareness (mda), in: Proceedings of Oceans 2005 Mts/Ieee, IEEE. pp. 1590–1594.

Thorsteinsson, V., 2002. Tagging methods for stock assessment and research in fisheries. Report of concerted action FAIR CT 96, 179.

Tromeur, E., Doyen, L., 2019. Optimal harvesting policies threaten biodiversity in mixed fisheries. Environmental Modeling & Assessment 24, 387–403.

Von Schuckmann, K., Le Traon, P.Y., Alvarez-Fanjul, E., Axell, L., Balmaseda, M., Breivik, L.A., Brewin, R.J., Bricaud, C., Drevillon, M., Drillet, Y., et al., 2016. The copernicus marine environment monitoring service ocean state report. Journal of Operational Oceanography 9, s235–s320.

Yang, D., Wu, L., Wang, S., Jia, H., Li, K.X., 2019. How big data enriches maritime research–a critical review of automatic identification system (ais) data applications. Transport Reviews 39, 755–773.

Zhou, C., Jiao, Y., Browder, J., 2019. Seabird bycatch vulnerability to pelagic longline fisheries: ecological traits matter. Aquatic Conservation: Marine and Freshwater Ecosystems 29, 1324–1335.

All stocks	High-pressured stocks	Threatened non-stock species in the area	Impacted non-stock species
Clupea harengus	**Brosme brosme**	**Alca torda**	**Larus argentatus**
Brosme brosme	**Chaceon quinquedens**	**Alosa aestivalis**	**Puffinus gravis**
Chaceon quinquedens	**Cynoscion regalis**	**Balaenoptera borealis**	**Squalus acanthias**
Cynoscion regalis	**Glyptocephalus cynoglossus**	**Balaenoptera physalus**	
Gadus morhua	**Hippoglossus hippoglossus**	**Caretta caretta**	
Glyptocephalus cynoglossus	**Malacoraja senta**	**Cetorhinus maximus**	
Hippoglossoides platessoides	**Morone saxatilis**	**Clangula hyemalis**	
Hippoglossus hippoglossus	**Mustelus canis**	**Dermochelys coriacea**	
Homarus americanus	**Pandalus borealis**	**Eubalaena glacialis**	
Illex illecebrosus	**Pollachius virens**	**Fratercula arctica**	
Limanda ferruginea	**Pomatomus saltatrix**	**Lepidochelys kempii**	
Lophius americanus	**Raja eglanteria**	**Melanitta fusca**	
Malacoraja senta	**Stenotomus chrysops**	**Mobula tarapacana**	
Melanogrammus aeglefinus	**Thunnus thynnus**	**Mola mola**	
Merluccius bilinearis	**Urophycis tenuis**	**Physeter macrocephalus**	
Morone saxatilis	**Xiphias gladius**	**Pterodroma hasitata**	
Mustelus canis		**Rissa tridactyla**	
Pandalus borealis		**Sebastes fasciatus**	
Paralichthys dentatus		**Somateria mollissima**	
Peprilus triacanthus		**Sphyrna lewini**	
Pollachius virens		**Squalus acanthias**	
Pomatomus saltatrix			
Prionace glauca			
Pseudopleuronectes americanus			
Raja eglanteria			
Scophthalmus aquosus			
Stenotomus chrysops			
Thunnus thynnus			
Urophycis tenuis			
Xiphias gladius			

Table 1: Report of all stocks, highly pressured stocks, threatened species (both stocks and non-stocks), and possibly impacted species, retrieved by our analysis from the data of the Global Fishing Watch (GFW) and the BOEM-Marine Cadastre (BMC) altogether, in a Northwest Atlantic study region (bounding box: longitude [-72 ; -66], latitude [38 ; 45]). Red names refer to species retrieved from the GFW data but not from the BMC data, and vice-versa for yellow names. Blue names refer to species detected from both datasets.

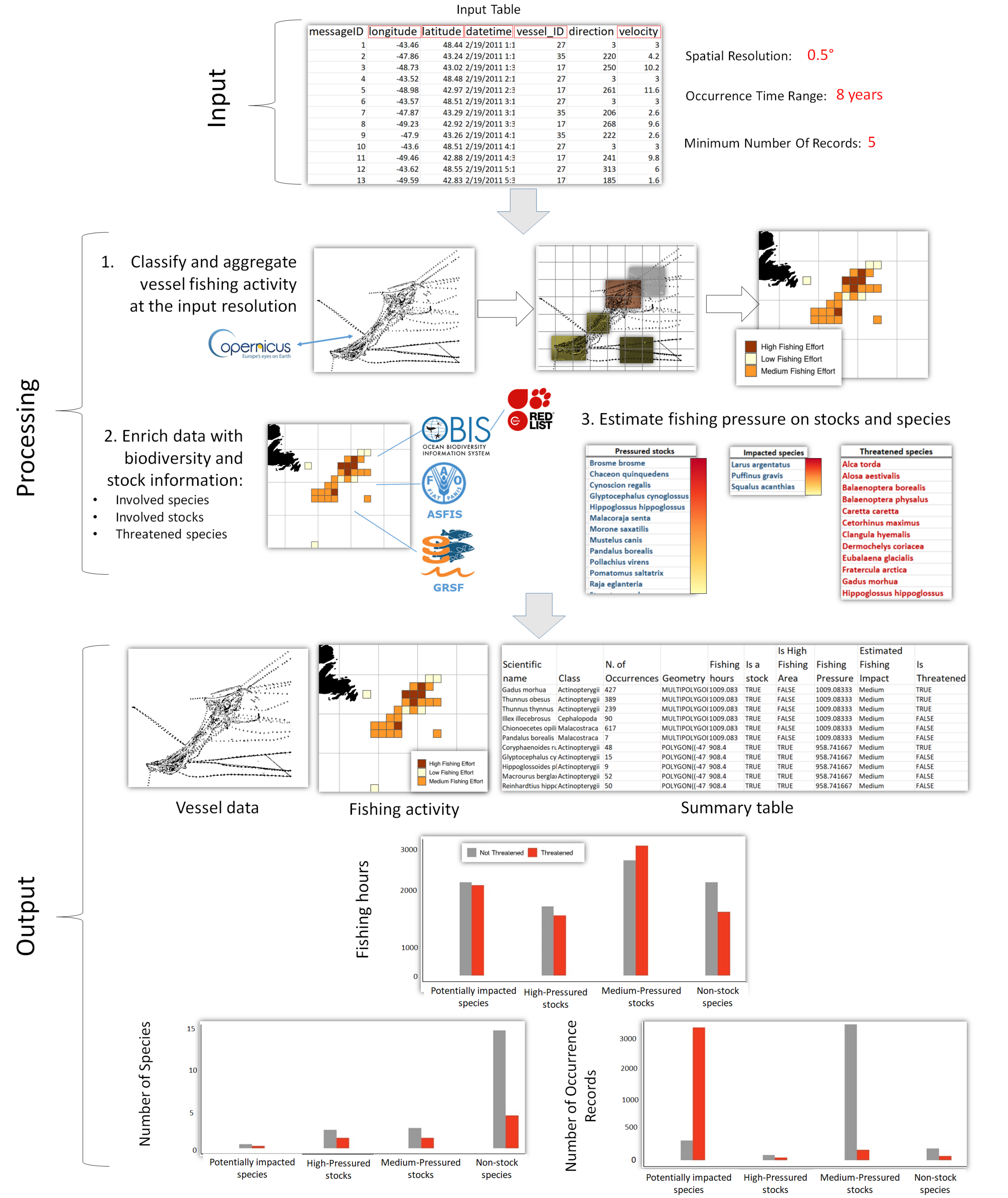

Figure 1: Schema of our methodological workflow, with processing divided into three separate steps. Red boxes in the input vessel tracking data table highlight mandatory fields of our process.

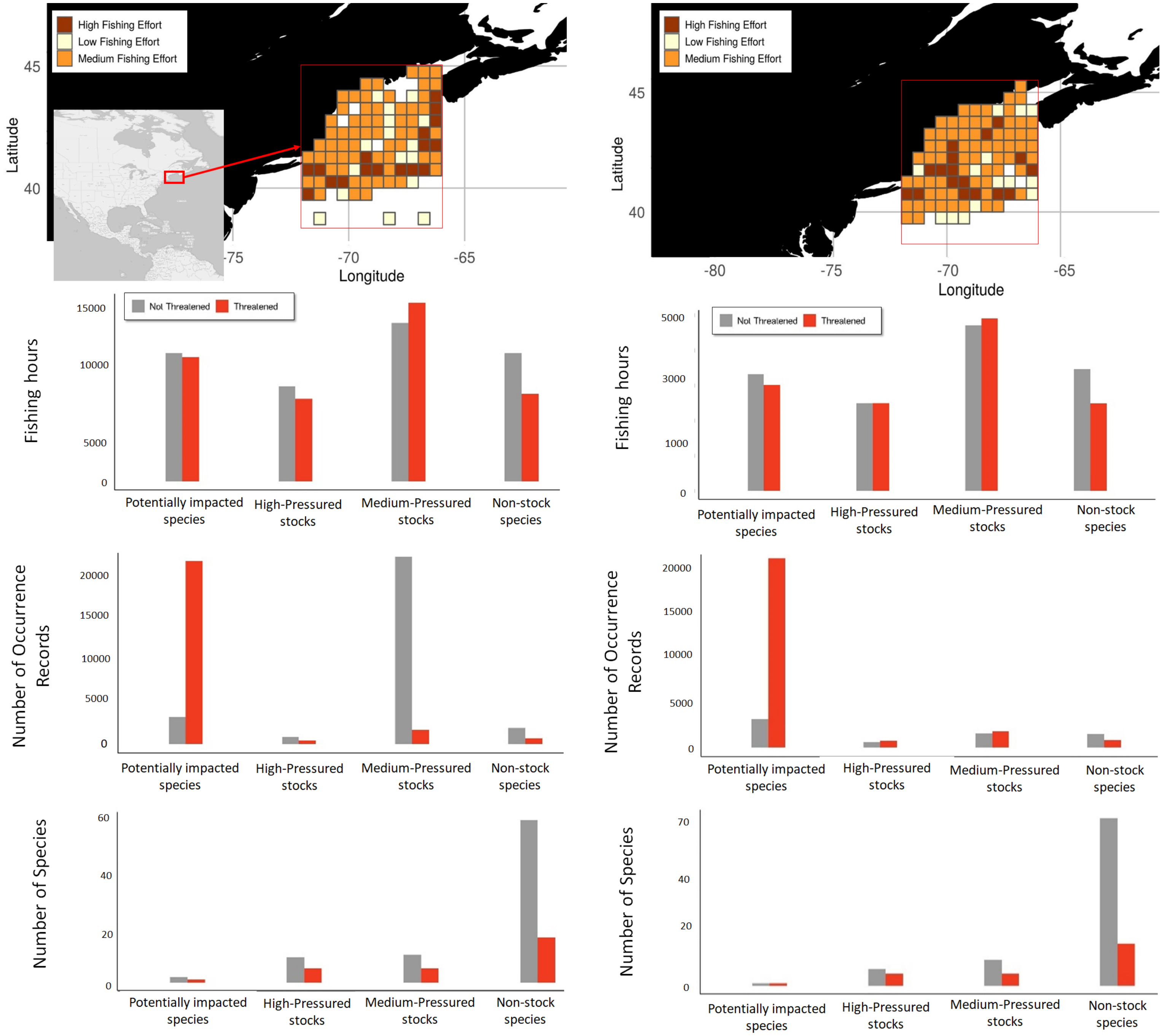

Figure 2: Comparison between the spatial distributions, the amounts of fishing hours, occurrence records, and different detected species, extracted by our analysis from the Global Fishing Watch data (left-hand side) and the BOEM-Marine Cadastre data (right-hand side) data, in a Northwest Atlantic study region (bounding box: longitude [-72 ; -66], latitude [38 ; 45]). The histograms report statistics for threatened and non-threatened species separately. Different chart scales are used because the two dataset sizes differ of one order of magnitude.

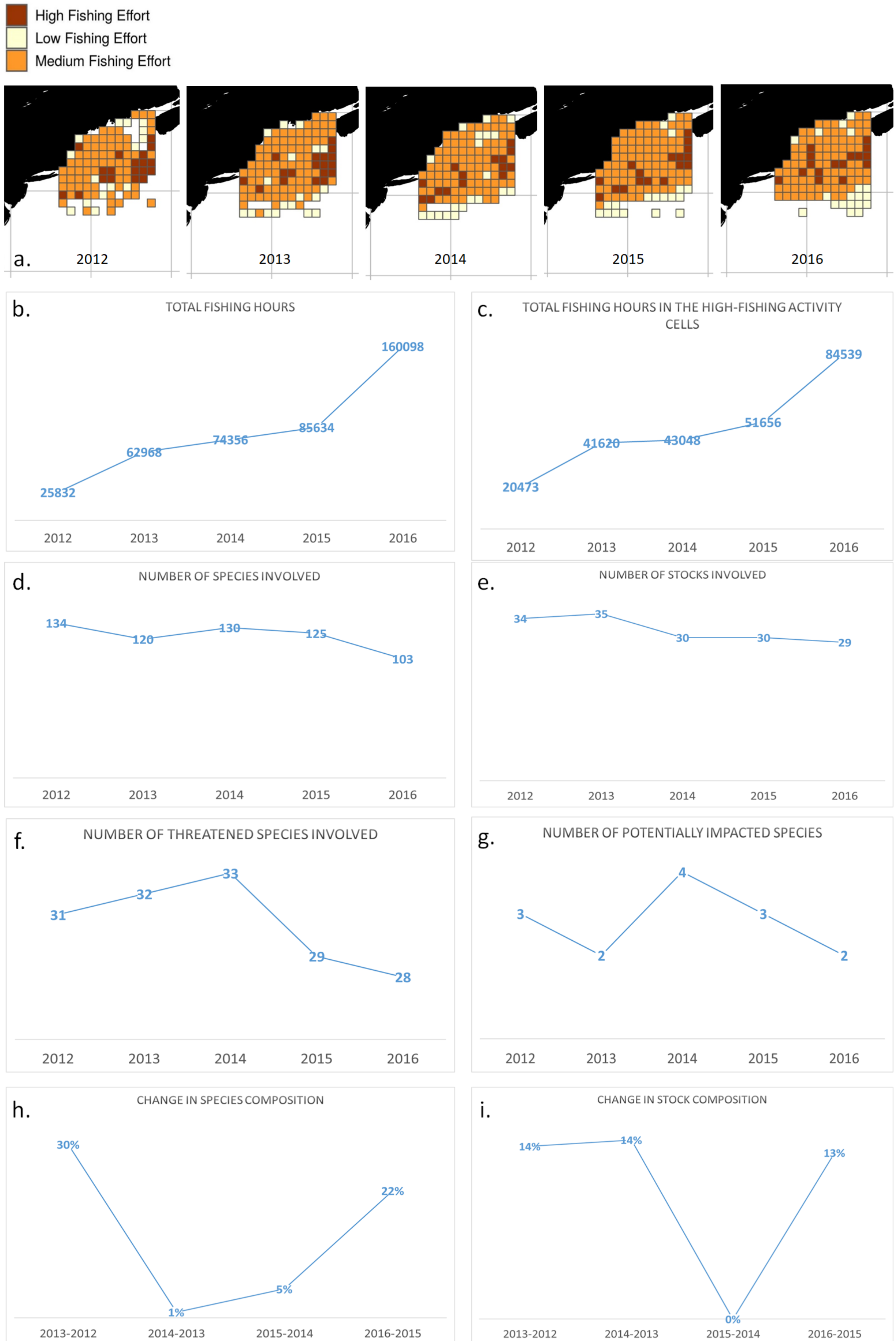

Figure 3: Temporal analysis of the daily data of the Global Fishing Watch from 2012 to 2016, annually-aggregated at a 0.5° resolution in a Northwest Atlantic study region (bounding box: longitude [-72 ; -66], latitude [38 ; 45]). The charts report the time series of a) the distribution of high/medium/low fishing-activity cells; b) the total fishing hours in the whole area and c) in the high fishing-activity cells; d) the number of different species, e) target stocks, and f) threatened species detected by our analysis; g) the number of non-stock species that are possibly ecologically impacted by the fisheries; h) the number of different species and i) stocks retrieved by our analysis from two consecutive datasets.

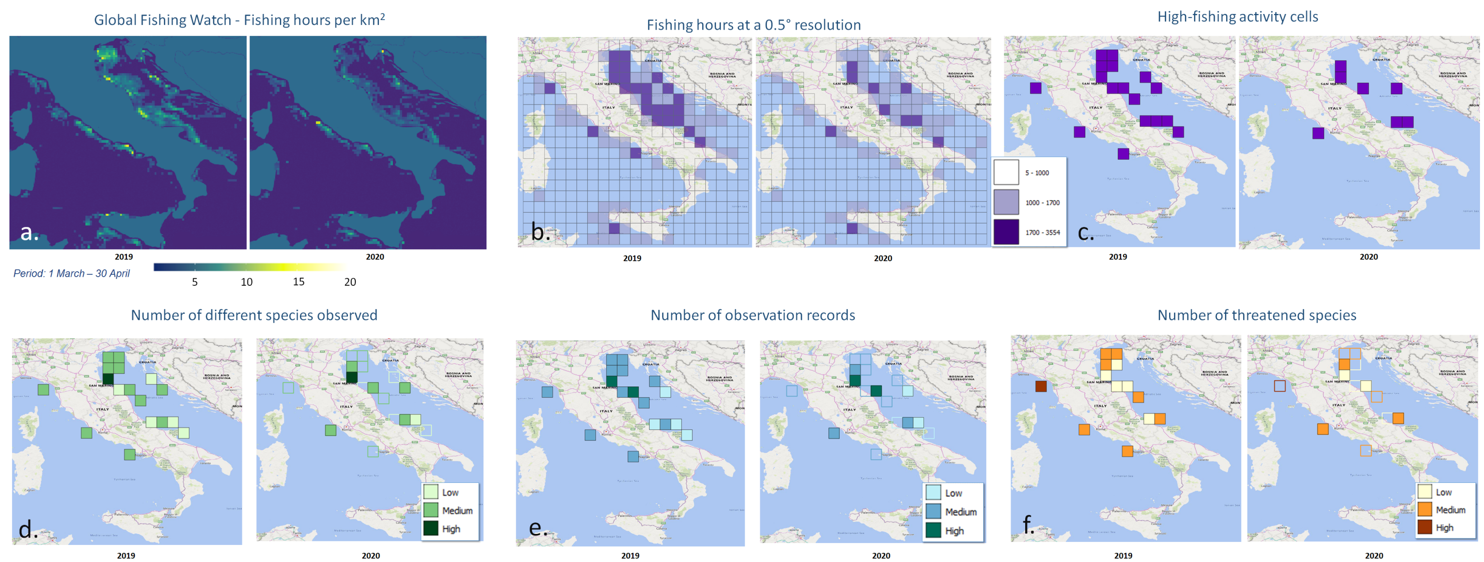

Figure 4: Comparison between the total fishing activity hours in the periods March-April 2019 and 2020 around Italian coasts, based on Global Fishing Watch (GFW) data: a) original raster data produced and published by GFW, b) aggregation and classification of fishing hours in cells at 0.5° resolution, c) highlight of the high fishing-activity cells, d) number of different species in OBIS in the high fishing-activity cells, e) number of OBIS-species observation records per cell, f) number of threatened species per cell according to the IUCN Red List. The 2019 cells are reported as contour lines in 2020 to highlight spatial distribution differences.