Rodríguez et al. *EPJ Data Science* (2024) 13:23
https://doi.org/10.1140/epjds/s13688-024-00459-0

EPJ Data Science
a SpringerOpen Journal

RESEARCH

Open Access

# Identification of suspicious behavior through anomalies in the tracking data of fishing vessels

Jorge P. Rodríguez[1,2*], Xabier Irigoien[3], Carlos M. Duarte[4] and Víctor M. Eguíluz[5,6]

*Correspondence:
jorge@ifisc.uib-csic.es;
jorgeprodriguezg@gmail.com
[1]Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC), CSIC-UIB, Palma de Mallorca, 07122, Illes Balears, Spain
[2]Instituto Mediterráneo de Estudios Avanzados (IMEDEA), CSIC-UIB, Esporles, 07190, Illes Balears, Spain
Full list of author information is available at the end of the article

**Abstract**

Automated positioning devices can generate large datasets with information on the movement of humans, animals and objects, revealing patterns of movement, hot spots and overlaps among others. However, in the case of Automated Information Systems (AIS), attached to vessels, observed strange behaviors in the tracking datasets may come from intentional manipulation of the electronic devices. Thus, the analysis of anomalies can provide valuable information on suspicious behavior. Here, we analyze anomalies of fishing vessel trajectories obtained with the Automatic Identification System. The map of silent anomalies, those that occur when positioning data are absent for more than 24 hours, shows that they are most likely to occur closer to land, with 87.1% of anomalies observed within 100 km of the coast. This behavior suggests the potential of identifying silence anomalies as a proxy for illegal activities. With the increasing availability of high-resolution positioning of vessels and the development of powerful statistical analytical tools, we provide hints on the automatic detection of illegal activities that may help optimize the management of fishing resources.

**Keywords:** Automatic Identification System (AIS); Fishing vessels; Tracking data; Exclusive Economic Zones (EEZ); Marine Protected Areas (MPA)

## 1 Introduction

Illegal, unreported and unregulated (IUU) fishing represents a problem for actors, both nations and companies, that cooperate to sustainably exploit fishing resources [1, 2], leading for example to inaccuracies in the catch reports, hindering fisheries management. IUU displays a global spread [3], enhanced by the high benefit from this activity, in contrast to the low detection rates that may lead to penalties imposed on these actors [4]. Furthermore, other illegal activities, such as forced labor, occur within the fishing industry [5]. In this context, satellite tracking of fishing vessels has the potential to discover and quantify such behaviours [6], facilitating the management of the fishing resources.

Automatic Identification System (AIS) data provides valuable information about shipping activity in the oceans, revealing for example the appearance of new shipping routes [7]. In particular, concerning fishing vessels, global AIS analyses have helped quantify the global distribution of fishing intensity [8], the unequal share of the industrial fishing effort among countries [9], the economic impact of fishing in the high seas [10], the network of fishing ports supporting fishing activity in different ocean regions [11], or the overlap between vessels and marine animals [12].

Interestingly, AIS tracking data from fishing vessels (and in general any automated positioning system) reveals anomalous behaviors in the reported positions [13]. For example, vessels often reported consecutive locations that would represent traveling distances associated with non-feasible speeds, as well as trajectories displayed long periods with absence of reported locations. While the failure of the devices and processes involved in the transmission and reception of the data is a natural source of these anomalous positions, anomalies can also represent operational issues or intentional manipulation [14]. Complementary datasets have been already used to show that vessels can intentionally disconnect the AIS system to stop reporting their locations while performing illegal activities at sea. For example, data obtained from night light satellite pictures have reported illegal trajectories along North Korean waters [15].

Previous works have introduced statistical methods to detect anomalies or spoofed positions [16]. However, a global analysis of the distribution of AIS anomalies in fishing vessels that may provide cues to illegal or unreported activity is still pending. Here, we focus on silence anomalies, that is, long periods (compared to the typical AIS temporal resolution) without reporting AIS location. In particular, we aim to statistically identify geographical locations where fishing vessels show an excess of anomalous events that cannot be explained due to randomly distributed operational issues and are likely to represent intentional manipulation of these devices to hide the vessels' real position and trajectories.

## 2 Results

Considering the global fishing vessels tracking dataset in 2014, obtained from AIS, we measured the observed inter-event times between consecutive locations of the same vessel, together with the estimated speed (see Methods, Fig. 1). The observed values of inter-event times and estimated speeds displayed a high variability and, although the estimated speeds were mostly concentrated in values lower than 100 km/h, the inter-event times marginal distribution decayed slowly. Despite such a slow decay, 99.3% these times were shorter than 24 h. Considering the unlikelihood of these occurrences on the AIS system, we developed an analysis of these long periods of absence of reported locations, which we named silence anomalies.

We identified 770K silence anomalies in our tracking dataset. For this analysis, we discretized the world in grid cells of size 0.5° lat × 0.5° lon and ignored grid cells which include ports (see Methods), resulting in 25,795 grid cells with at least one observed anomaly, out of 87,853 total (non-port) grid cells with fishing vessels observations. Most anomalies occurred in grid cells with a high number of location estimates, such that these anomalies could be explained by random operational issues, happening more frequently in highly visited grid cells. To compare these anomalies and their expected geographical distribution if they were random, we extracted each vessel's trajectory and its set of inter-event times and randomized the inter-event times, such that the silence anomalies happened randomly along the trajectory, i.e., proportionally to the number of locations inside each grid cell. Several realizations of this random process led to the computation of a $p$-value (see Methods). To test the effect of non-uniform spatial distributions of the number of

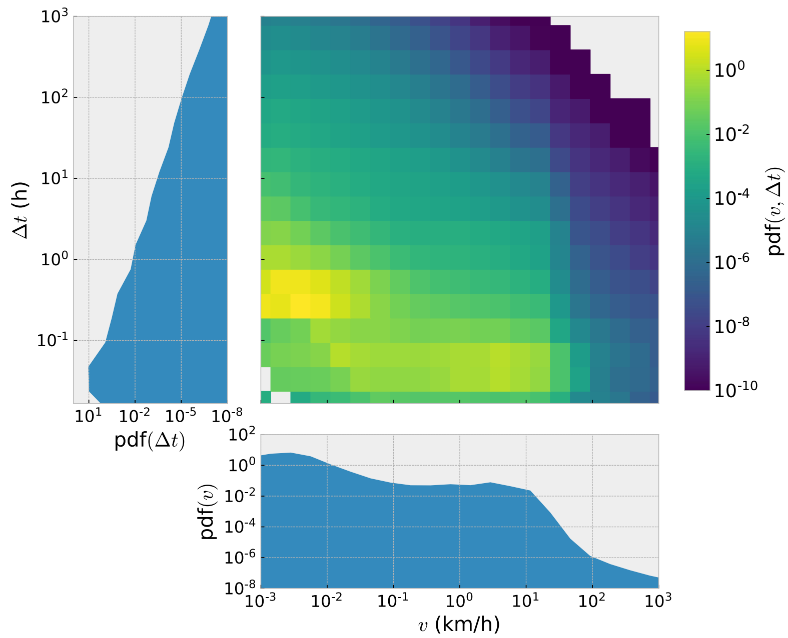


**Figure 1** Statistics of the inter-event time and estimated speed between consecutive AIS locations. The central heatmap represents the joint probability density function $\mathbf{pdf}(v, \Delta t)$ of observing a time delay $\Delta t$ and an estimated speed $v$, defined as the traveled distance divided by $\Delta t$, while the bottom and left plots represent the marginal distributions of, respectively $v$ and $\Delta t$

locations, emerging from different space uses of the vessels or different sampling rates across the ocean, we have tested this computation of the $p$-value on a null model. Particularly, we distributed anomalies through grid cells with different number of location data points (see Statistical filter in non-uniform events distributions in Additional file 1). In our null model, we obtained an analytical expression for the $p$-value that agreed with its stochastic computation, obtained from multiple realizations of the random process (Fig. S1). The model showed that in the case of uniform probabilities of observing anomalies, but different event densities, the anomalies of no location were selected under the filter $p < 0.01$, with minor differences in the $p$-value of cells with different number of events (Fig. S2). Moreover, we increased the likelihood of observing an anomaly in a grid cell with a specific number of locations, and our model showed that a lower increase is required in cells with a large number of data points, such that cells that have lower AIS coverage (and thus lower number of data points) would be requested to have higher fractions of anomalies in order to pass the filter, implying that anomalies on grid cells with poor AIS coverage would pass this filter less frequently.

To remove the anomalies that could be explained by these random distributions, we discarded, for each vessel in our dataset, the anomalies observed in grid cells with a $p$-value $\geq 0.01$, selecting for our analysis the anomalies in the rest of cells (i.e., those that could not be explained by a random distribution). After this statistical filter, 169K anomalies (22% of the anomaly observations) remained significant with respect to the random model (Figs. S3, S4; for a $p$-value $< 0.05$, 239K anomalies, i.e. 31% of the anomaly observations, remained significant). These significant anomalies, unlikely to be attributable to random events, were observed in 5758 unique grid cells (for $p < 0.05$, 10577 unique grid cells), and appeared more frequently in the Northeastern Atlantic Ocean, the Mediterranean Sea

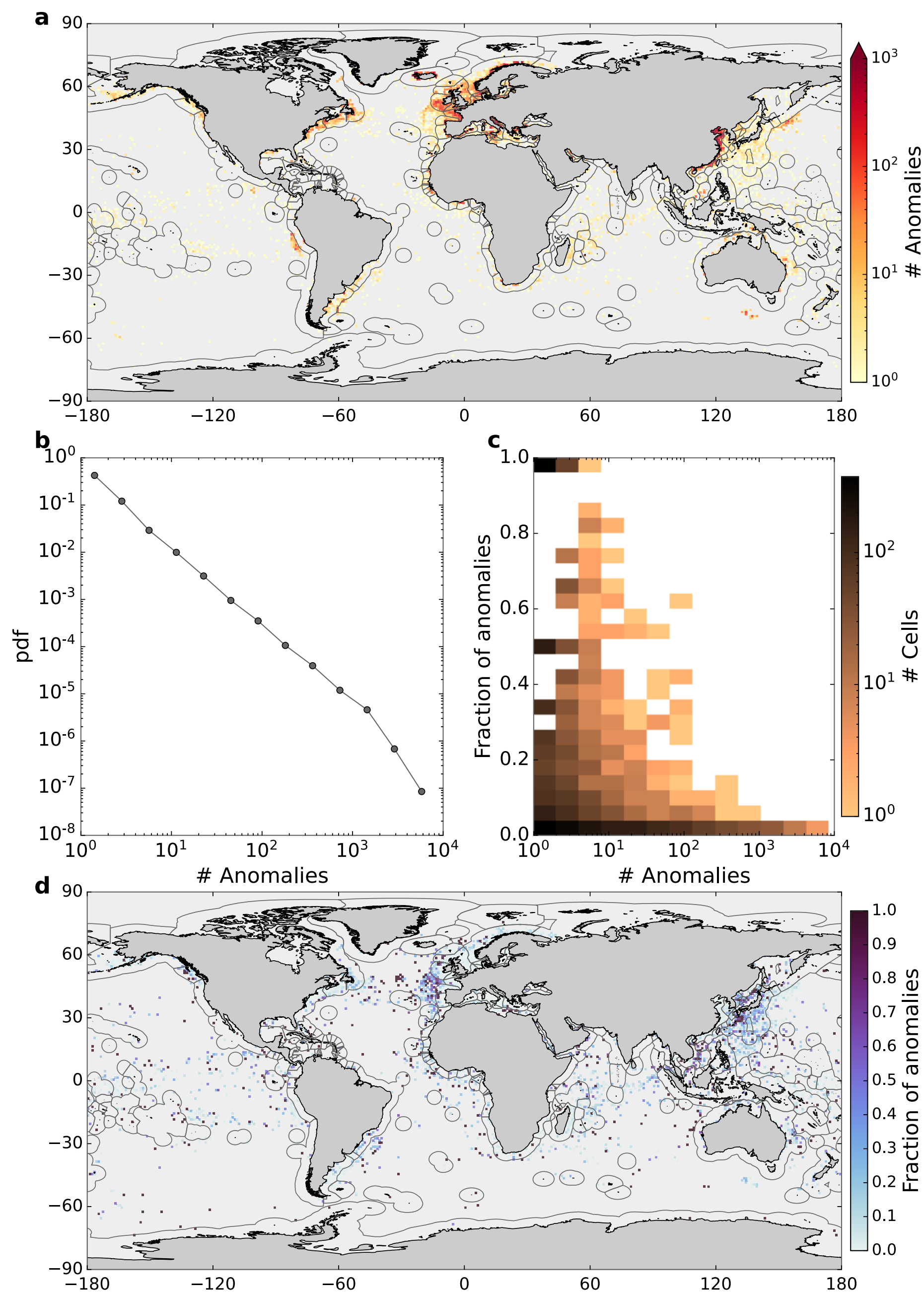


**Figure 2** A global dataset of silence anomalies observations. (a) Geographical extent of the number of significant anomalies starting at each 1° lat × 1° lon grid cell. (b) Distribution of the number of anomalies observed in each grid cell. c Density plot of the fraction of significant anomalies, computed as the sum of significant anomalies in each grid cell divided by the number of locations of vessels with at least one anomaly in that grid cell, as a function of the total number of anomalies. (d) Geographical extent of the fraction of anomalous locations. Grey contours on the maps represent the limits of Exclusive Economic Zones

and the Western Pacific (Eastern Asia shore) (Fig. 2a). The distribution of the number of significant anomalies per grid cell displayed a heavy tail, indicating the presence of a few grid cells with a number of anomalies much greater than the average number of anomalies per grid cell (Fig. 2b). A key indicator of how frequent visits to grid cells ended up in anomalies was the fraction of anomalies, computed as the number of significant anomalies in each grid cell divided by the total number of locations from vessels in that grid cell, including only those vessels with at least one anomaly along their trajectories (Fig. 1c). This indicator revealed cells where more than 50% of the tracking locations were anomalies in the Southwest of Ireland, included within the Sole Bank fishing ground, and the South of Japan (Fig. 2d). These results were robust after considering a global grid composed of cells with sides of 50 km length for the statistical filter, and the graphical representation on a grid with sides of 100 km length (Fig. S5). Interestingly, the inter-event time distribution associated with the significant anomalies was universal across different vessels' flags (Fig. S6).

Regarding the fishing vessels, 16K vessels displayed at least one significant anomaly, out of the 78K vessels included in our silence anomalies dataset. Although we observed at least one significant anomaly in only 20% of the analyzed vessels, there were geographical hotspots where most of the detected fishing vessels displayed at least one significant anomaly (Fig. S7). This result illustrates the presence of hotspots with diverse fishing vessels showing significant anomalies, but it does not inform about how often these vessels showed significant anomalies. Our statistical filter captured both locations with many anomalies as preferential hotspots for AIS disconnection, but also locations where there were few visits of vessels with anomalous behavior, but most visits led to an event of silence (Fig. 3). Focusing on the regions where there were remarkable combinations of number and fraction of anomalies, we observed spatial structures in the vicinity of the Exclusive Economic Zone (EEZ) limits of the United States, Ireland, Libya and Japan (Fig. 3a-h). In the region including the Bay of Biscay and the Atlantic coast of Ireland and Great Britain, the anomalies displayed a specific pattern following the limit of the West of Scotland Marine Protected Area (MPA) and appeared along the Bay of Biscay slope current (Fig. 3a, b). While the latter seems to be a geographical pattern linked to fishing pressure, the accumulation of anomalies at the border of the West of Scotland MPA provides high suspicion of illegal fishing activity. Another region of accumulation of anomalies was located close to the Northwest Atlantic shore (Fig. 3c, d), especially in the vicinity of the Northeast Canyons and Seamounts Marine Protected Area, again bordering an MPA. In the Mediterranean Sea, most of the anomalies were observed in the Italian EEZ, but interestingly the vessels that reached the border between Libyan and Greek EEZs displayed anomaly events in most of their visits (Fig. 3e, f), suggesting illegal fishing operations in these EEZs. The Northwest Pacific Ocean and the Sea of Japan displayed accumulations of anomalies at the border of the Russian EEZ and in the disputed control area adjacent to Japan, South Korea and North Korea (Fig. 3g, h). The replication of these results considering a global grid with sides led to similar results, confirming the robustness of our method (Fig. S8).

Considering only the distance from each anomaly to the nearest coast (see Methods), we observed that the frequency of observed significant anomalies decreased with distance from the coast, a behavior that we associate with the disconnection of the AIS system in the trips towards the ports, while this behavior displayed an increase of more than 100% in the vicinity of the expected border of the EEZ limit at 200 nautical miles (Fig. 3i). This limit separates the Areas Beyond National Jurisdiction from the national waters, while the EEZ limits between two nations may be located at closer distances from the shore. We observed quantitative and qualitative differences after classifying the vessels into short (length $< 40$ m) and long (length $\geq 40$ m) fishing vessels. For example, most of the anomalies at the

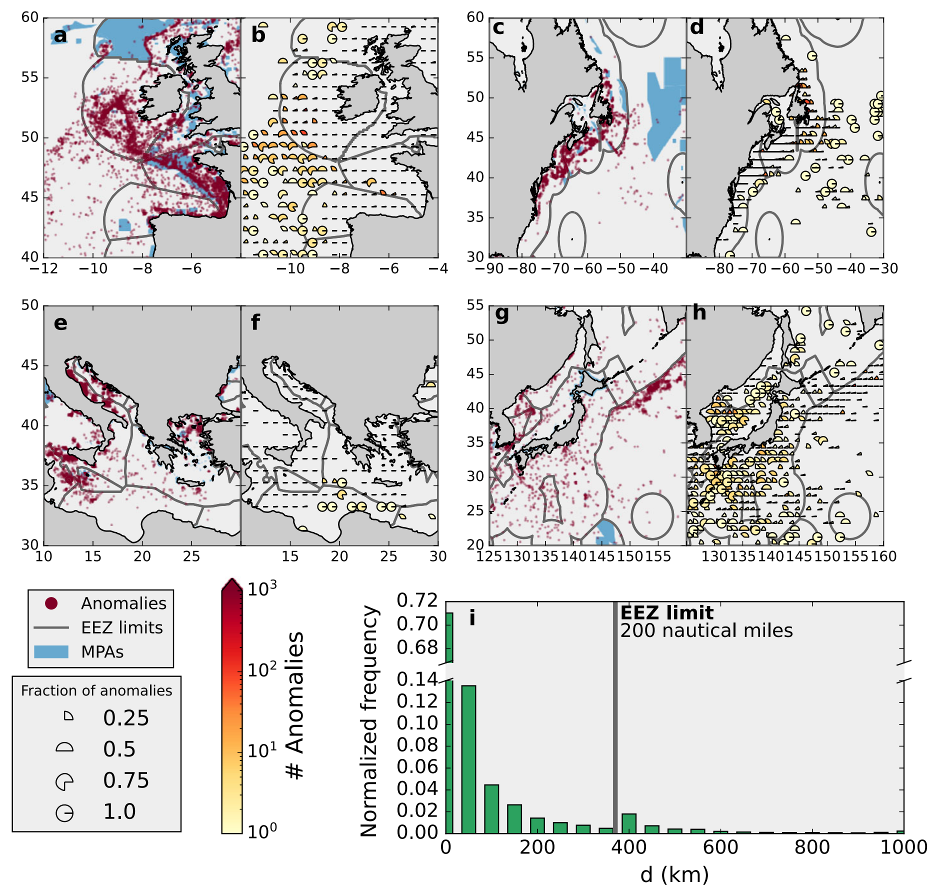


**Figure 3** Significant anomalous locations (green dots) in the proximity of Exclusive Economic Zones (EEZ) and Marine Protected Areas (MPAs). (a)-(h) Regional distribution of anomalous locations. Panels (a), (c), (e), (g) depict the location of the anomalies, while (b), (d), (f), (h) describe the number (color) and fraction (fraction of the pie) of observed silence anomalies in each 1° × 1° grid cell. Regions correspond to the Northeast Atlantic Ocean (a), (b), Northwest Atlantic Ocean (c), (d), Southern Mediterranean Sea (e), (f) and Western Pacific Ocean (g), (h). Grey contours and light blue areas represent, respectively, the limits of the Exclusive Economic Zones and the zones declared Marine Protected Areas. (i) Normalized frequency of observed anomalies (number of anomalies divided by the global number of anomalies) in cells located at a distance *d* from the shore

Great Sole Bank corresponded to short vessels, while the opposite behavior occurred in the South of Japan (Fig. S9). Additionally, short vessels displayed their anomalies mainly in the proximity to the shore, as expected, but we observed that long vessels showed fewer anomalies close to the coast, and a remarkable increase in the proximity of the EEZ limits, indicating that the suspicious behavior related to the borders of the high seas was mostly associated with longer vessels (Fig. S10).

62.8% of the anomalies had the same origin and destination port (see Methods). From the perspective of the vessels, (1) 47.5% of the vessels displayed all their anomalies on return trips from/to the same port (i.e., when an anomaly was observed, if A was the last visited port before the anomaly, A was also the first visited port after that anomaly) and, within them, 96.8% of the vessels traveled always from/to a unique port; (2) 31% of the vessels showing anomalies either with common or different origins and destinations (i.e., there were some anomalies located on return trips to/from a port and other anomalies located in the transit between different ports); and (3) 22% of the vessels with all their anoma-

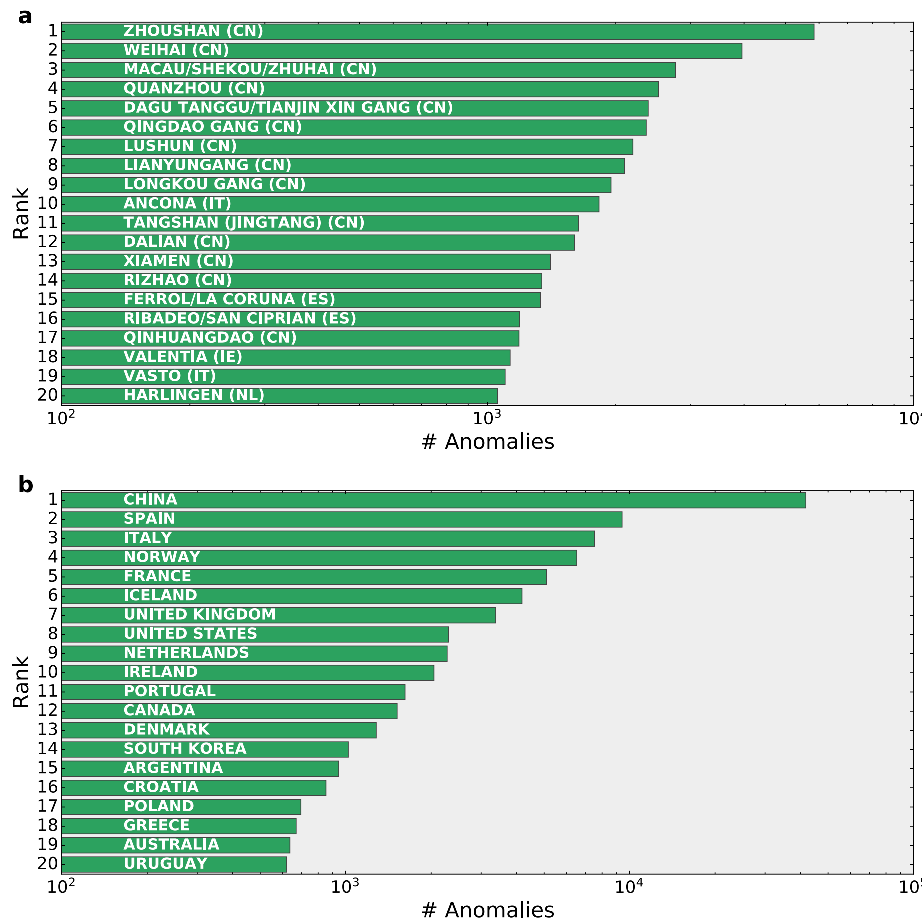


**Figure 4** Locations of the ports with anomaly detections before or after visiting them. (a) Rank of the 20 ports with the highest anomaly detections before or after visiting them. Countries are included within brackets. CN: China, IT: Italy, ES: Spain, IE: Ireland, NL: Netherlands. (b) Rank of the 20 countries whose ports are visited before or after the anomalies

lies happening in the transit between two different ports. The distribution of anomalies among origin/destination ports was highly heterogeneous, with 164 ports (15% of the ports with at least 0.5 anomalies) associated with 85% of the anomaly-port trips. In particular, the top-20 ports, linked to 39.5% of the anomalies, included 14 ports in China, 2 in Spain, 2 in Italy, 1 in Ireland and 1 in the Netherlands (Fig. 4a). The distribution was narrower when the anomaly-port links were reported by country, with 89% of the trips associated with only 16 countries (11% of the countries with at least 0.5 anomalies), and a top-20 list of countries whose ports support fishing vessels contributing 91% of those anomalous trips (Fig. 4b). Comparing the number of anomalies by ports with the fishing effort associated with those ports, we found a positive correlation (Pearson correlation of 0.67). This correlation was stronger when ports were joined into countries (Pearson correlation of 0.91, Fig. S11).

The structure connecting anomalies to countries can be better understood when all the grid cells are associated with one nation, either its EEZ for grid cells located in Areas Under National Jurisdiction or the nearest EEZ for those located on the high seas (Figs. S12, S13). This classification allowed us to create a network connecting the locations of the

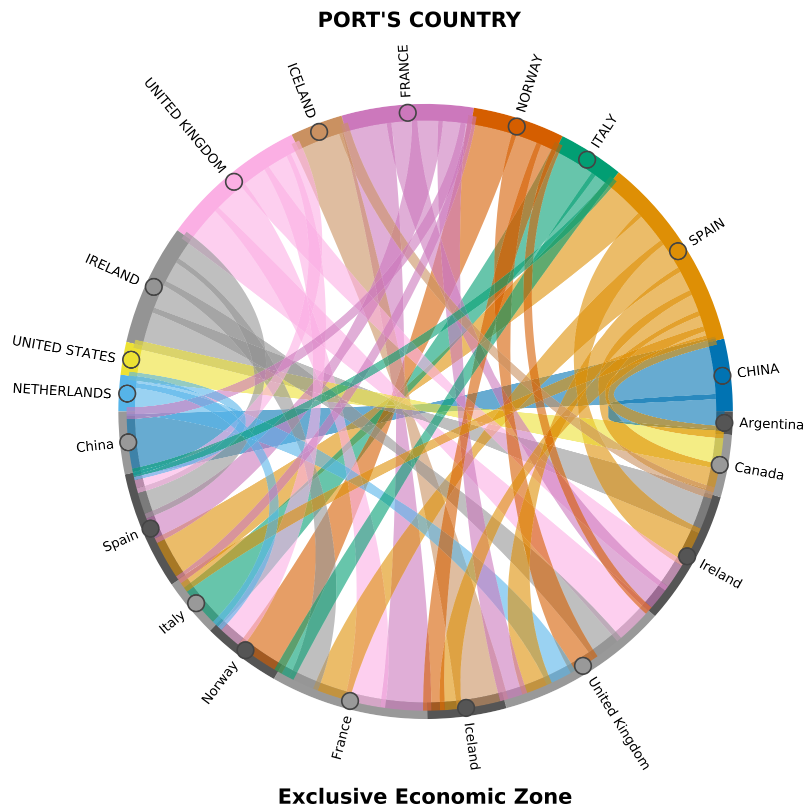


**Figure 5** Network between the top 10 countries (top, names in capital letters, colored nodes) with support from their ports to most of the observed anomalies, and the top 10 Exclusive Economic Zones (bottom, full names, gray nodes) where the anomalies were observed. Link widths are proportional (logarithmic scale) to the number of anomalies connecting the nation's ports to anomalies in Exclusive Economic Zones

anomalies to the countries whose ports were visited before and after the observation of those anomalies (Fig. 5). This network revealed a complex structure splitting into three regional groups (France-Spain-United Kingdom-Ireland, Norway-Iceland-United Kingdom, United States-Canada) and two nations with strong interactions within their own EEZs (China and Italy). We also assessed the role of the vessel flags on this network, revealing frequently the same EEZ, flag and country where the supporting port is, but also interesting links, for example, vessels with Chinese flag with anomalies in the Argentinian EEZ, receiving support from ports in Chile and Uruguay (Fig. S14).

## 3 Conclusion

A detailed analysis of the spatial and temporal properties of AIS data of fishing vessels revealed anomalous phenomena, including long periods of absence of tracking locations and non-feasible vessel traveling speeds. Particularly, we focused our research on such long periods of absence of tracking locations, named silence anomalies, finding that, although infrequent (0.3% of the number of tracking locations), their extent was global. Considering that those anomalies could arise from randomly operational issues of the tracking devices, we developed a statistical filter to remove the anomalies that could be explained by randomly distributed issues. Thus, keeping these significant anomalous observations, we found that their location was of particular interest to the fishing industry, suggesting the intentional manipulation of the tracking devices.

Tracking data of fishing vessels has advanced our understanding of when and where fishing activities occur, facilitating the quantitative estimation of the captures thanks to modern algorithms, for example, those using Artificial Intelligence for measuring fishing effort from the vessels' sequences of geolocated positions. However, intentional manipulation of the devices that report the fishing vessels' locations may bias these analyses, providing underestimated results, especially in the locations where low or even null fishing effort is expected due to their conservation status. The availability of alternative datasets may help contrast these trajectories and reveal behaviors that AIS tracking data does not highlight, such as the use of animals carrying devices that monitor fisheries [17], or those reported thanks to the nightlights from vessels [15, 18]. Nevertheless, there may be other data gaps, such as those related to the resolution of satellite imagery or the absence of lights due to the cloud cover. Such problems highlight, as a preliminary step before contrasting different datasets, the importance of new algorithms that detect anomalies within a single dataset (in parallelism with unsupervised learning), which subsequently additional datasets can validate.

Identifying anomalies from automatically retrieved vessel positions has applications beyond fishing vessels and AIS in marine shipping. In general, anomaly detection refers to the identification of entries in a dataset that do not conform to the expected value and different techniques and methods have been developed [19]. Social data for example is affected by biases, inaccuracies (e.g. at the source of the data, processing), and methodological limitations [20]. In particular, the analysis of human mobility can be hindered by missing and extraneous locations, for instance, 75% of check-in traces are extraneous and can be associated with the social reward system [21].

We have developed a method for extracting the locations where there is a statistically relevant accumulation of silence anomalies in fishing vessels' trajectories, as the combination of a high number of anomalies from specific vessels' trajectories with a low probability (< 1%) of occurring by chance. Using this approach, we detected the specific areas where these anomalies happened. We found four general cases that accounted for a large fraction of the oceanic hot spots of AIS anomalies, including (1) marine protected areas (Fig. 3a, c), (2) the edge of EEZs supporting productive fisheries (Fig. 3g)), (3) ocean areas disputed by different nations; for instance, the maritime border between Greece and Libya, which is known for experiencing territorial disputes, exhibits a low number of anomalies, but with a specific set of vessels that displayed an anomaly in most of their visits to that area (Fig. 3f); and (4) An additional case refers to anomalies by vessels within their own EEZ's, as exemplified by China and Italy. We hypothesize that this behavior may be due to fishers willing to hide their chosen location from possible competitors, a behavior deeply rooted in fishing culture. Fishermen are known not to share information freely and tend towards a culture of secrecy and deceit about the location of productive fishing sites [22], which is clearly at odds with the advent of AIS, where their position can be viewed by their competitors. Cases 1 and 2 above clearly point to illegal practices, suggesting that the approach developed here can effectively identify and monitor potential illegal fishing practices.

Our results may be impacted by biases emerging from the features of the AIS tracking dataset. First, the distribution of vessels carrying AIS is not uniform across different marine regions; we tackled this potential bias by performing the $p$-value filter at the individual vessel level, selecting the anomalous events that were significant within the trajectory of each vessel, without considering other vessels' influences. Second, other vessel's influence may be present due to the interference in highly transited regions [23]; we believe that the selection of 24 hours as the inter-event time without the reception of AIS messages is a long-enough period, such that a high level of traffic cannot explain this silence in most cases. Third, a single vessel can emit at different message rates in different locations, for example due to lower AIS coverage, which may impact the filtering method; we have tested this on a null model (see Statistical filter in non-uniform events distributions in Additional file 1), showing that, when the probability of finding an anomaly was uniform across different cells, our filter discarded all the anomalies, even though we introduced heterogeneity in the event densities across different cells. Finally, Fig. 1 illustrated anomalies in consecutive locations reported by AIS identified as delays longer than 24 h, but it also showed the presence of traveling distances between consecutive locations not feasible by fishing vessels. We believe that those speed anomalies should be investigated in the future, together with the silence anomalies, as we observed, although less frequently, locations simultaneously displaying both kinds of anomalies.

We have analyzed the case of anomalous silence events in the trajectories of fishing vessels in 2014, but we anticipate that the application of our methodology to more recent datasets, which include a higher number of vessels' trajectories, may reveal that our results are probably underestimates. In fact, advances in AIS instrumentation and reporting will help remove anomalies caused by technical malfunction and possibly help report attempts at tampering with AIS. Our approach also identifies the main ports where vessels that report suspicious anomalies use, which are not randomly distributed but are distributed in a limited subset of ports within specific nations, pointing at geographical targets for inspections. We suggest using game theory to design the right incentives to encourage the more reluctant fishermen to use AIS reliably, as punitive measures alone will not be able to improve current levels of anomalies.

In summary, our results provide an approach to single out excess anomalies in AIS positions of vessels that may represent deliberate manipulation of the reported position for reasons that may extend from maintaining confidentiality on productive fishing grounds from competitors to illegal fishing activity. We also tracked the vessels to the ports that support their operations and showed that a small number of ports supported the fishing vessels reporting most of the anomalies. This analysis, based on a complex systems analysis of port-fishing vessel networks, extends the uses of AIS to further provide a tool to combat illegal fishing and conserve fish stocks. In fact, our methodology can be adapted to the analyses of AIS reported locations of fishing vessels from the global (in the current study) to the regional scales, for example by modification of the grid cell size. Furthermore, we anticipate that the statistical filter that we have developed at the trajectory level can help to differentiate random operational issues from intentional manipulation as digital fingerprints of other phenomena or behaviors.

## 4 Methods

*Automatic Identification System (AIS) data* The AIS system reports, among other information, the location of the vessels carrying it with high temporal resolution. We use AIS satellite data of vessels categorized as 'fishing vessels' in 2014 provided by ExactEarth, with a total of $2.44 \times 10^8$ locations from 112,535 unique vessels, distinguishable by their

Maritime Mobile Service Identity (MMSI) numbers. Among the fields specified in each tracking location, we considered the MMSI number, latitude, longitude and time. The trajectories were cleaned before the analysis, removing both locations on land and distant locations with the same time value, as well as keeping a single location when we detected a duplicate record.

*Inter-event statistics* Considering a trajectory as the series of time-sorted locations corresponding to the same MMSI number, we computed the inter-event time $\Delta t_i$ as the time difference between the locations $i$ and $i + 1$, with $i$ running through the tracking locations of each trajectory, except for its last location. We also estimated the speed associated with each pair of consecutive locations of each vessel, computed as the traveling distance $\Delta x_i$ (through the great-circle distance between locations $i$ and $i + 1$, considering the Earth as a sphere with radius $R = 6371$ km), divided by the inter-event time $\Delta t_i$.

*Presence of vessels in ports* We obtained the global coordinates of ports from the World Port Index (2019), available from the National Geospatial-Intelligence Agency (https://msi.nga.mil/Publications/WPI). According to that database, we obtained the 0.5° lat × 0.5° lon grid cells $\{c_p\}$ that contained any port. We considered a location $i$ as the visit to a port if its associated grid cell $c_i$ belonged to $\{c_p\}$.

*Detecting anomalous behavior of silence* A location $i$ was considered anomalous if $\Delta t_i >$ 24 h and it was not located in a port (see Presence of vessels in ports).

*Statistical filter to remove random operational anomalies* Given the total number of anomalies $N_{i,j}$ and the total number $D_{i,j}$ of data points from vessel $j$ at grid cell (0.5° lat × 0.5° lon) $i$ (excluding those assigned to ports), we distributed the total number of anomalies randomly among the grid cells, with probability proportional to $D_{i,j}$, obtaining $R^k_{i,j}$, the number of randomly located anomalies of vessel $j$ in grid cell $i$ for each realization $k$ of the random distribution (note that $\sum_i N_{i,j} = \sum_i R^k_{i,j}$ for any $k$). Then, we considered the p-value as the fraction of realizations where a grid cell had a higher number of anomalies in the random case than in the observed data:

$$p_{i,j} = \frac{1}{J} \sum_k \mathrm{H}\left(R^k_{i,j} - N_{i,j}\right) \tag{1}$$

where $\mathrm{H}(x)$ is the Heaviside function, taking the value 0 if $x$ is lower than 0 and 1 otherwise, and $J$ is the number of realizations of the random distribution. We considered that the random distribution of anomalies could not explain the cells with p-value lower than 0.01, after generating $J = 10^3$ independent realizations. Hence, the cells with $p < 0.01$ displayed a systematic excess of anomalies that could not be explained by a random distribution. Consequently, we selected, for each trajectory, the anomalies located in grid cells with $p < 0.01$ and defined them as significant anomalies.

*Distance to shore* We computed the distance to shore from each significant anomalous tracking location using the dataset provided by the Pacific Islands Ocean Observing System in https://pae-paha.pacioos.hawaii.edu/thredds/dist2coast.html?dataset=dist2coast_4deg_ocean, which reported the distance from the oceanic locations to the nearest shore in a global grid with a resolution of 0.04° lat × 0.04° lon.

*Connecting anomalies to ports and countries* Considering the trajectory of a given vessel and a particular significant anomaly, we extracted the last port cell that was visited by the vessel before the observed anomaly (origin port) and the first port cell visited after it (destination port). To compute the number of anomalies linked to a port $p$, we averaged the number of anomalies linked to $p$ as their origin and the number of anomalies linked to $p$ as their destination, such that an anomaly with the same origin and destination port accounted as one for that port. We computed the number of anomalies at national level by the aggregation of the number of anomalies linked to the ports in each nation.

*Exclusive Economic Zones* We extracted the Exclusive Economic Zones from MarineRegions (https://www.marineregions.org) and assigned to each 0.5° lon × 0.5° lat grid cell the EEZ that covers most of its area. However, the Areas Beyond National Jurisdiction, where intentional manipulation of AIS tracking devices was highly expected, were not included in this dataset. Hence, we assigned each grid cell in the high seas to the closest EEZ, considering the distance on a regular lattice where each cell was connected to its closest 8 neighbors (East, West, North, South, Northeast, Northwest, Southeast, Southwest, Fig. S5).

*Marine Protected Areas* The information describing these areas in shapefile format was downloaded from protectedplanet.net.

## Supplementary information

**Supplementary information** accompanies this paper at https://doi.org/10.1140/epjds/s13688-024-00459-0.

**Additional file 1.** This article has an accompanying supplementary file including a section entitled 'Statistical filter in non-uniform events distributions', Figures S1-S13, Supplementary Methods and Supplementary References (PDF 5.2 MB)

**Acknowledgements**

J.P.R. received support from Juan de la Cierva Formación program (Ref. FJC2019-040622-I) funded by MCIN/AEI/10.13039/501100011033, the Spanish Research Agency MCIN/AEI/10.13039/501100011033 via project MISLAND (PID2020-114324GB-C22), and the Vicenç Mut program from Govern de les Illes Balears. This research is supported by María de Maeztu Excellence Unit 2023-2027 Refs. CEX2021-001201-M and CEX2021-001164-M, funded by MCIN/AEI/10.13039/501100011033.

**Funding**

Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. J.P.R. received support from Juan de la Cierva Formación program (Ref. FJC2019-040622-I) funded by MCIN/AEI/10.13039/501100011033, the Spanish Research Agency MCIN/AEI/10.13039/501100011033 via project MISLAND (PID2020-114324GB-C22), and the Vicenç Mut program from Govern de les Illes Balears. This research is supported by María de Maeztu Excellence Unit 2023-2027 Refs. CEX2021-001201-M and CEX2021-001164-M, funded by MCIN/AEI/10.13039/501100011033.

**Abbreviations**

IUU, Illegal, unreported and unregulated (fishing); AIS, Automatic Identification System; EEZ, Exclusive Economic Zone; MPA, Marine Protected Area; MMSI, Marine Mobile Service Identity.

**Data availability**

The datasets generated and/or analysed during the current study are available in the 'Fishing anomalies' repository, doi:10.5281/zenodo.10817048.

## Declarations

**Competing interests**

The authors declare that they have no competing interests.

**Author contributions**
Conceptualization: JPR, VME, CMD; Methodology: JPR, VME; Software: JPR; Formal analysis: JPR; Writing-Original draft: JPR, VME; Writing-Review and Editing: XI, CMD; Visualization: JPR. All authors read and approved the final manuscript.

**Author details**
[1]Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC), CSIC-UIB, Palma de Mallorca, 07122, Illes Balears, Spain. [2]Instituto Mediterráneo de Estudios Avanzados (IMEDEA), CSIC-UIB, Esporles, 07190, Illes Balears, Spain. [3]AZTI Marine Research, Pasaia, 20110, País Vasco, Spain. [4]Red Sea Research Center (RSRC), and Computational Biosciences Research Center (CBRC), King Abdullah University of Science and Technology (KAUST), Thuwal, 23955, Saudi Arabia. [5]Basque Centre for Climate Change (BC3), Leioa, 48940, País Vasco, Spain. [6]IKERBASQUE, Basque Foundation for Science, Plaza Euskadi 5, Bilbao, 48009, País Vasco, Spain.

Received: 3 November 2023 Accepted: 4 March 2024 Published online: 21 March 2024

**References**

1. Pitcher TJ, Watson R, Forrest R, Valtỳsson H, Guénette S (2002) Estimating illegal and unreported catches from marine ecosystems: a basis for change. Fish Fish 3(4):317–339
2. Sumaila UR, Alder J, Keith H (2006) Global scope and economics of illegal fishing. Mar Policy 30(6):696–703. https://doi.org/10.1016/j.marpol.2005.11.001
3. Agnew DJ, Pearce J, Pramod G, Peatman T, Watson R, Beddington JR, Pitcher TJ (2009) Estimating the worldwide extent of illegal fishing. PLoS ONE 4(2):4570
4. Temple AJ, Skerritt DJ, Howarth PE, Pearce J, Mangi SC (2022) Illegal, unregulated and unreported fishing impacts: a systematic review of evidence and proposed future agenda. Mar Policy 139:105033
5. McDonald GG, Costello C, Bone J, Cabral RB, Farabee V, Hochberg T, Kroodsma D, Mangin T, Meng KC, Zahn O (2021) Satellites can reveal global extent of forced labor in the world's fishing fleet. Proc Natl Acad Sci 118(3):2016238117. https://doi.org/10.1073/pnas.2016238117
6. Dunn DC, Jablonicky C, Crespo GO, McCauley DJ, Kroodsma DA, Boerder K, Gjerde KM, Halpin PN (2018) Empowering high seas governance with satellite vessel tracking data. Fish Fish 19(4):729–739
7. Eguíluz VM, Fernández-Gracia J, Irigoien X, Duarte CM (2016) A quantitative assessment of Arctic shipping in 2010–2014. Sci Rep 6:30682. https://doi.org/10.1038/srep30682
8. Kroodsma DA, Mayorga J, Hochberg T, Miller NA, Boerder K, Ferretti F, Wilson A, Bergman B, White TD, Block BA et al (2018) Tracking the global footprint of fisheries. Science 359(6378):904–908. https://doi.org/10.1126/science.aao5646
9. McCauley DJ, Jablonicky C, Allison EH, Golden CD, Joyce FH, Mayorga J, Kroodsma D (2018) Wealthy countries dominate industrial fishing. Sci Adv 4(8):2161. https://doi.org/10.1126/sciadv.aau2161
10. Sala E, Mayorga J, Costello C, Kroodsma D, Palomares ML, Pauly D, Sumaila UR, Zeller D (2018) The economics of fishing the high seas. Sci Adv 4(6):2504. https://doi.org/10.1126/sciadv.aat2504
11. Rodríguez JP, Fernández-Gracia J, Duarte CM, Irigoien X, Eguíluz VM (2021) The global network of ports supporting high seas fishing. Sci Adv 7(9):3470. https://doi.org/10.1126/sciadv.abe3470
12. Queiroz N, Humphries NE, Couto A, Vedor M, Da Costa I, Sequeira AM, Mucientes G, Santos AM, Abascal FJ, Abercrombie DL et al (2019) Global spatial risk assessment of sharks under the footprint of fisheries. Nature 572(7770):461–466. https://doi.org/10.1038/s41586-019-1444-4
13. Welch H, Clavelle T, White TD, Cimino MA, Van Osdel J, Hochberg T, Kroodsma D, Hazen EL (2022) Hot spots of unseen fishing vessels. Sci Adv 8(44):2109. https://doi.org/10.1126/sciadv.abq2109
14. Ford JH, Peel D, Kroodsma D, Hardesty BD, Rosebrock U, Wilcox C (2018) Detecting suspicious activities at sea based on anomalies in automatic identification systems transmissions. PLoS ONE 13(8):0201640. https://doi.org/10.1371/journal.pone.0201640
15. Park J, Lee J, Seto K, Hochberg T, Wong BA, Miller NA, Takasaki K, Kubota H, Oozeki Y, Doshi S et al (2020) Illuminating dark fishing fleets in North Korea. Sci Adv 6(30):1197. https://doi.org/10.1126/sciadv.abb1197
16. Katsilieris F, Braca P, Coraluppi S (2013) Detection of malicious ais position spoofing by exploiting radar information. In: Proceedings of the 16th international conference on information fusion. IEEE, Los Alamitos, pp 1196–1203
17. Weimerskirch H, Collet J, Corbeau A, Pajot A, Hoarau F, Marteau C, Filippi D, Patrick SC (2020) Ocean sentinel albatrosses locate illegal vessels and provide the first estimate of the extent of nondeclared fishing. Proc Natl Acad Sci 117(6):3006–3014. https://doi.org/10.1073/pnas.1915499117
18. Geronimo RC, Franklin EC, Brainard RE, Elvidge CD, Santos MD, Venegas R, Mora C (2018) Mapping fishing activities and suitable fishing grounds using nighttime satellite images and maximum entropy modelling. Remote Sens 10(10):1604. https://doi.org/10.3390/rs10101604
19. Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: a survey. ACM Comput Surv 41(3):1–58. https://doi.org/10.1145/1541880.1541882
20. Olteanu A, Castillo C, Diaz F, Kıcıman E (2019) Social data: biases, methodological pitfalls, and ethical boundaries. Front Big Data 2:13. https://doi.org/10.3389/fdata.2019.00013
21. Zhang Z, Zhou L, Zhao X, Wang G, Su Y, Metzger M, Zheng H, Zhao BY (2013) On the validity of geosocial mobility traces. In: Proceedings of the twelfth ACM workshop on hot topics in networks, pp 1–7. https://doi.org/10.1145/2535771.2535786
22. Evans KS, Weninger Q (2014) Information sharing and cooperative search in fisheries. Environ Resour Econ 58:353–372. https://doi.org/10.1007/s10640-013-9701-8
23. Taconet M, Kroodsma D, Fernandes J (2019) Global atlas of AIS-based fishing activity. Challenges and opportunities

## Publisher's Note