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

Jorge P. Rodríguez, Xabier Irigoien, Carlos M. Duarte, Víctor M. Eguíluz

Corresponding author: Jorge P. Rodríguez
Email: jrodriguez@ifisc.uib-csic.es / jorgeprodriguezg@gmail.com

**This PDF file includes:**

**Statistical filter in non-uniform events distributions**

We tested the filtering method (see Methods in main text) in a non-uniform distribution of events through the implementation of a toy model. We considered 10 grid cells with a different number of tracking location data points, which we called events. Specifically, they had $N_{ev}$={10,20,30,40,50,60,70,80,90,100} events. We set a base probability of an event being an anomaly across all the cells of $p_a$=0.1. Then, we increase the probability of observing an anomaly at one cell in $\Delta p_a$ and computed the p-value for that case (i.e., that cell having $p_a+\Delta p_a$ and the rest having $p_a$).

In this toy model, we computed analytically the p-value for any scenario. Considering that the cell *i* had $p_a+\Delta p_a$, leading to $a_i$ anomalies out of ei events, and with A and E being the total number of, respectively, anomalies and events in the system, the p-value was the probability that in the random case (distributing the anomalies uniformly across all the events), the number of anomalies in *i* was equal or higher than in the reality:

$$p_i = \sum_{j=a_i}^{\min(e_i,A)} \binom{A}{j} \left(\frac{e_i}{E}\right)^j \left(1-\frac{e_i}{E}\right)^{A-j}$$

We compared this analytical expression with the p-value obtained for $10^4$ random realizations in the cases of $\Delta p_a$ = {0,0.1}, showing agreement between the analytical prediction and the stochastic results (Fig. S1).

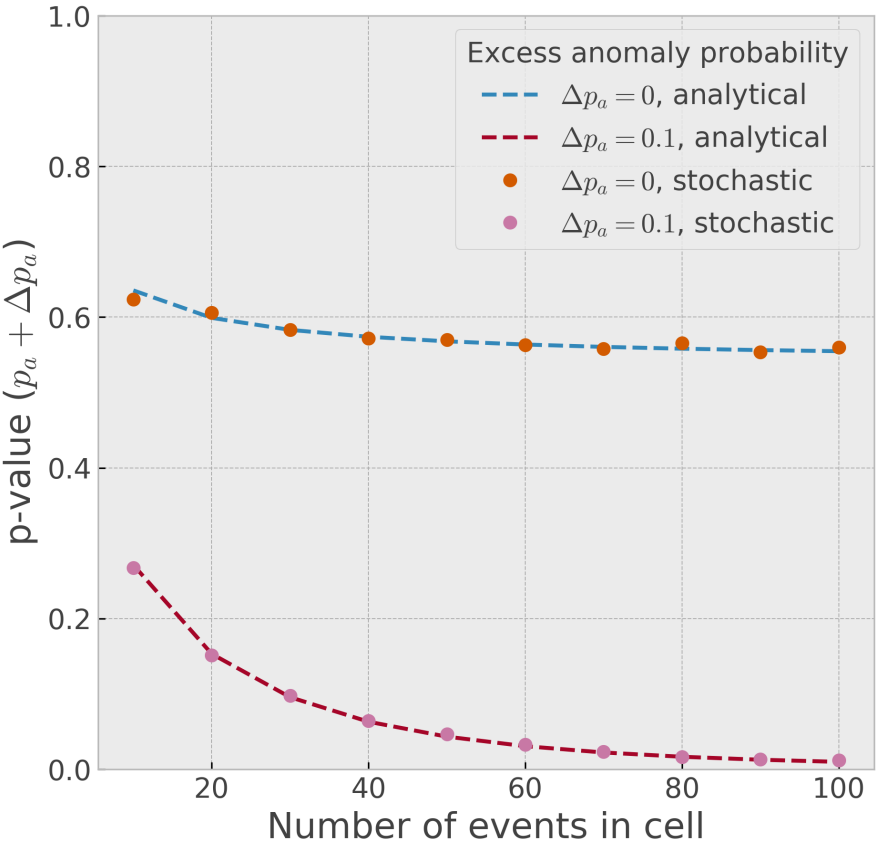


**Fig. S1.** Comparison of the analytical prediction (dashed curves) and the stochastic computation (dots) of the p-value for $\Delta p_a$ = {0,0.1}.

The results of our toy model showed that the behavior was practically uniform for $\Delta p_a$=0 and in this case no anomaly passed the filter p<0.01 (Fig. S11). As we increased $\Delta p_a$, we observed that

anomalies in cells with larger number of events were selected for lower values of $\Delta p_a$, requesting larger $\Delta p_a$ for cells with a low number of events.

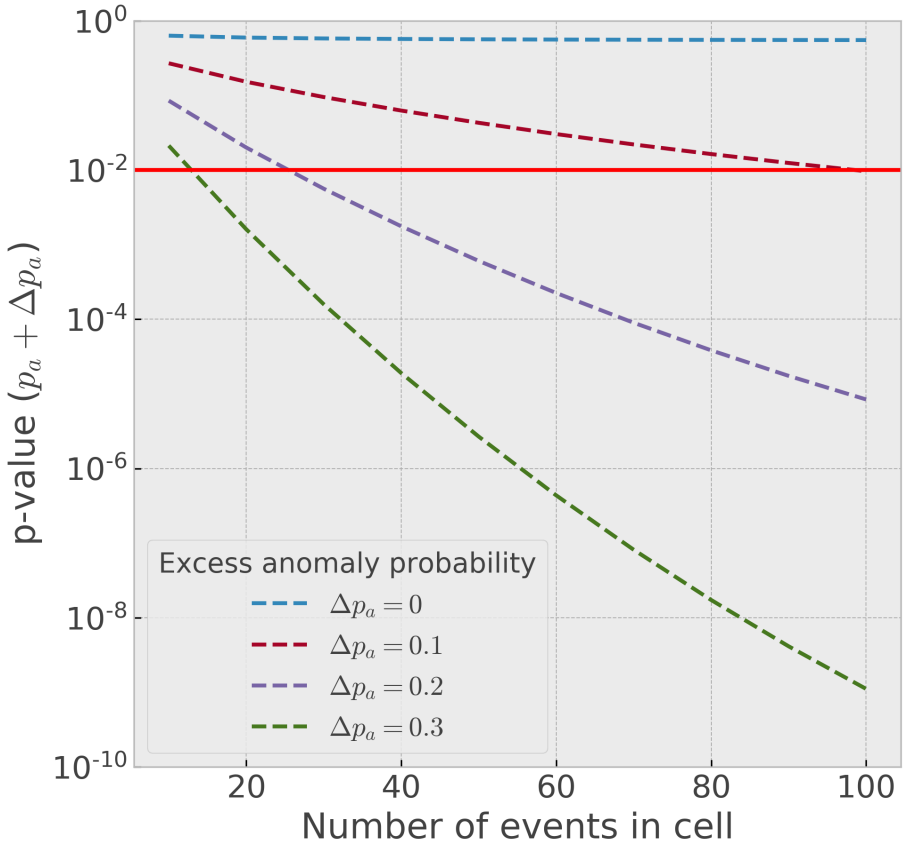


**Fig. S2.** Analytical prediction of the p-value (dashed curves) when each cell had an excess anomaly probability $\Delta p_a$ as a function of the number of events in the cell, for different values of $\Delta p_a$. The solid horizontal red line represents the selected limit p=0.01.

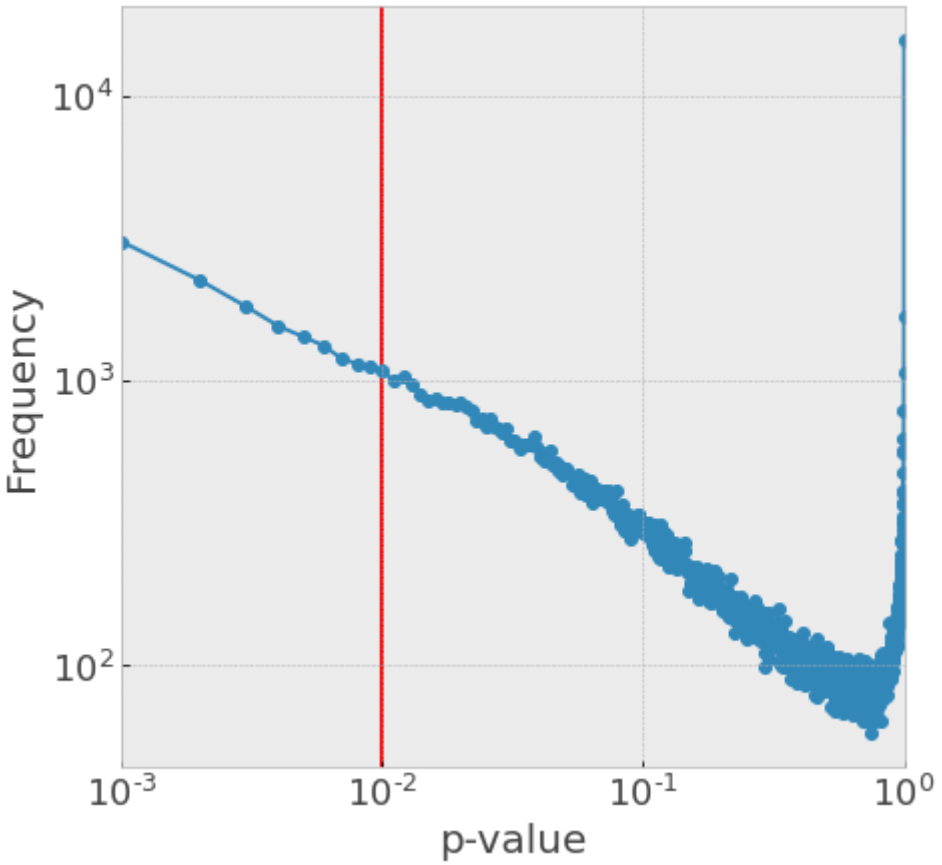


**Fig. S3.** Distribution of p-values. The red vertical line indicates the filter p<0.01 that we have used.

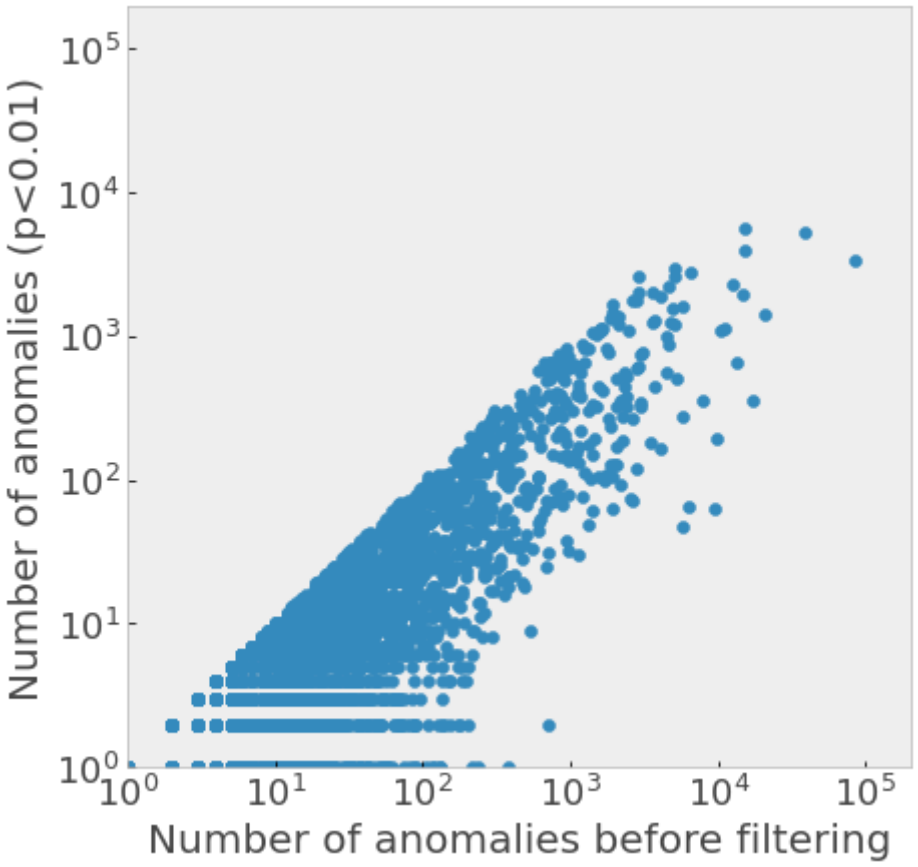


**Fig. S4.** Scatter plot of the remaining number of anomalies as a function of the original number of anomalies for the grid cells that keep at least one anomaly after filtering imposing p<0.01.

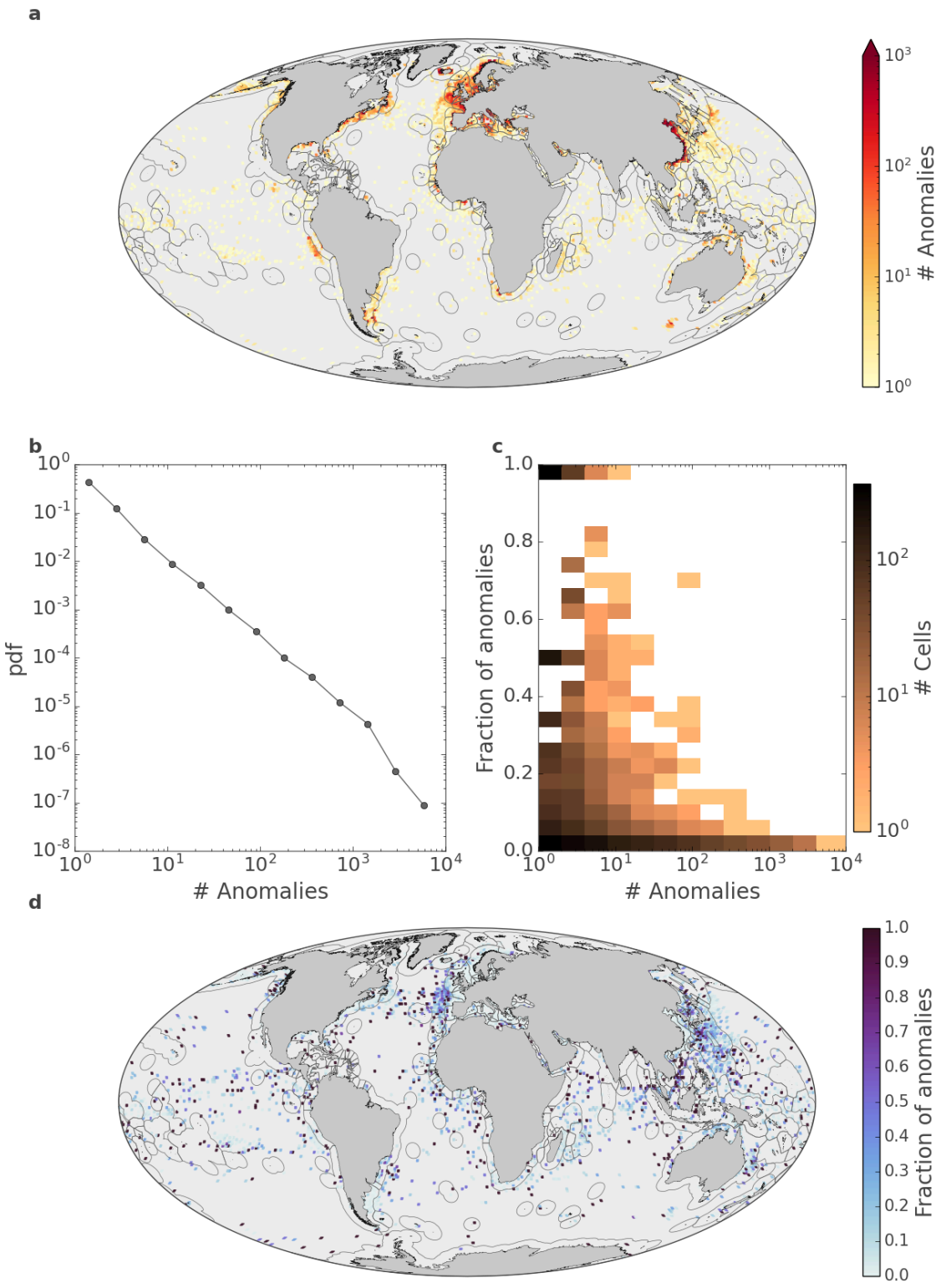


**Fig. S5.** Test of the detection and filter of silence anomalies using an equidistant side global grid (see Supplementary Methods). **a,** Geographical extent of the number of significant anomalies starting at each 100 km × 100 km grid cell. **b,** Distribution of the number of anomalies observed in 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.

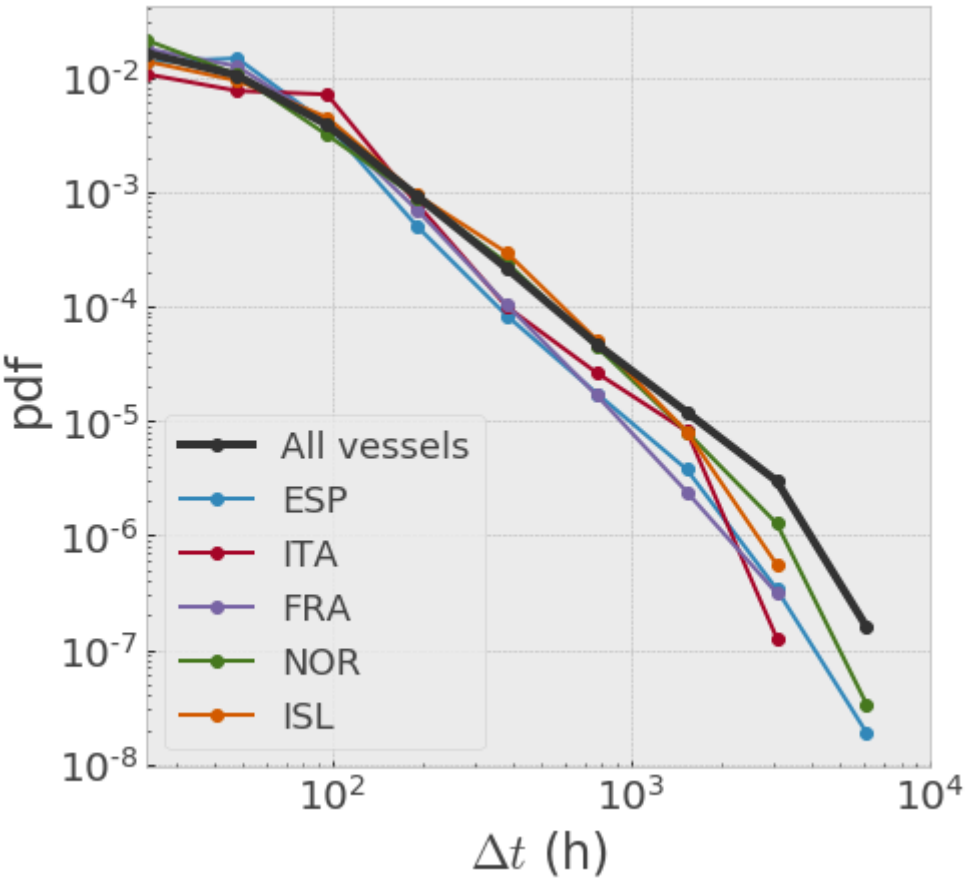


**Fig. S6.** Distribution of the inter-event times associated with significant anomalies. The black curve shows the distribution for all the vessels and the colored curves represent the vessels with flags (see Supplementary Methods) in the top-5 flags with most significant anomalies.

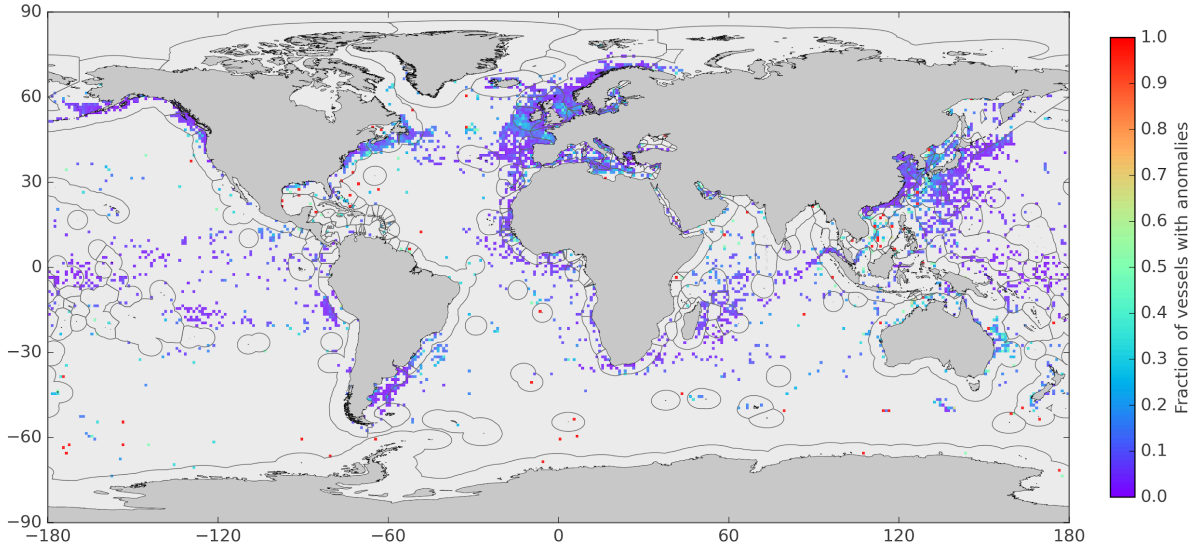


**Fig. S7.** Fraction of the detected vessels at each 1º × 1º grid cell displaying at least one significant anomaly within that cell.

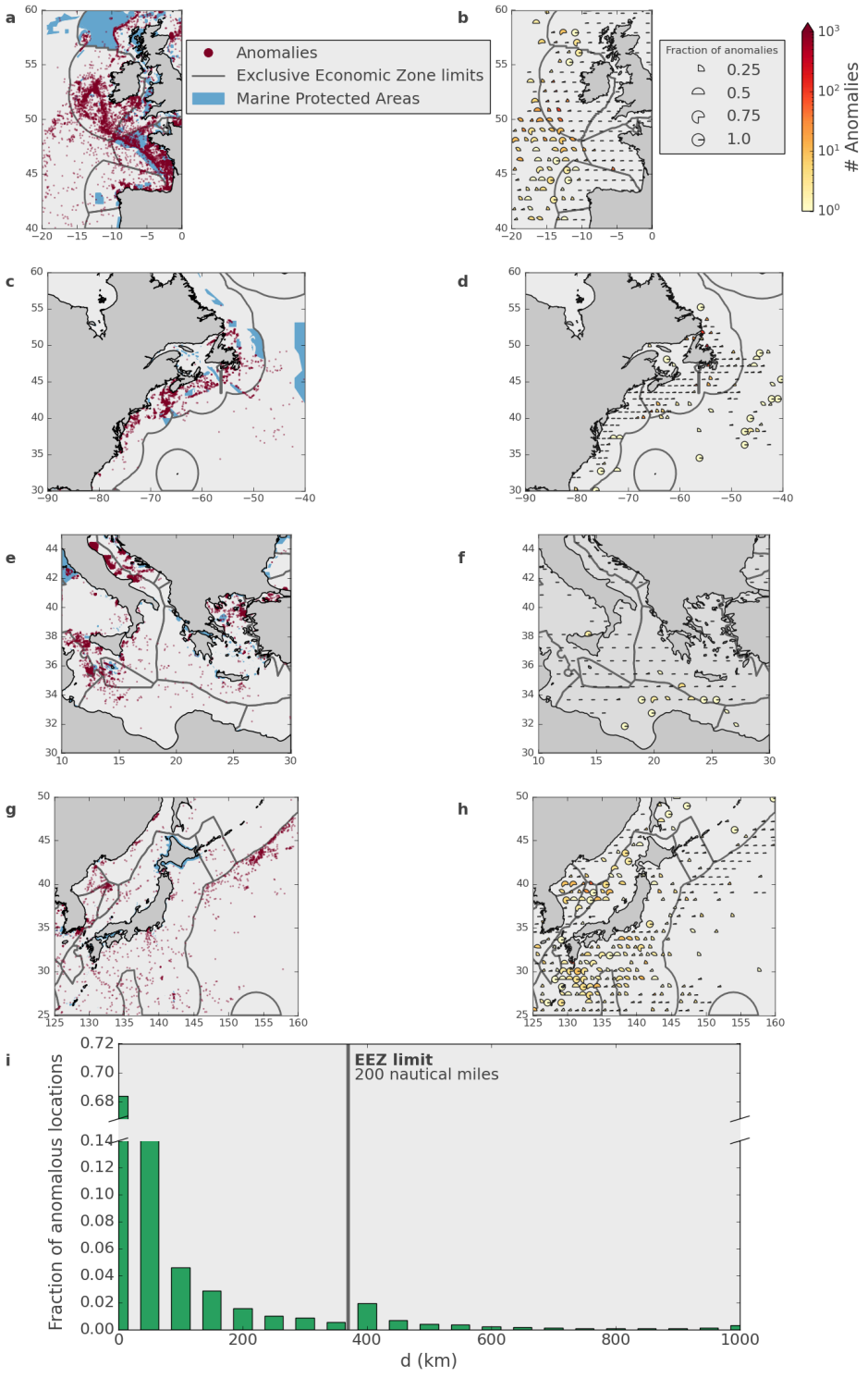


**Fig. S8.** Test of the spatial distribution of significant silence anomalies using an equidistant side global grid (see Supplementary Methods). **a-h,** Regional distribution of anomalous locations. Panels **a,c,e,g** depict the location of the anomalies, while **b,d,f,h** describe the number (color) and fraction (fraction of the pie) of observed silence anomalies in each 100 km × 100 km 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,** Fraction of the global number of anomalies observed in cells located at a distance *d* from the shore.

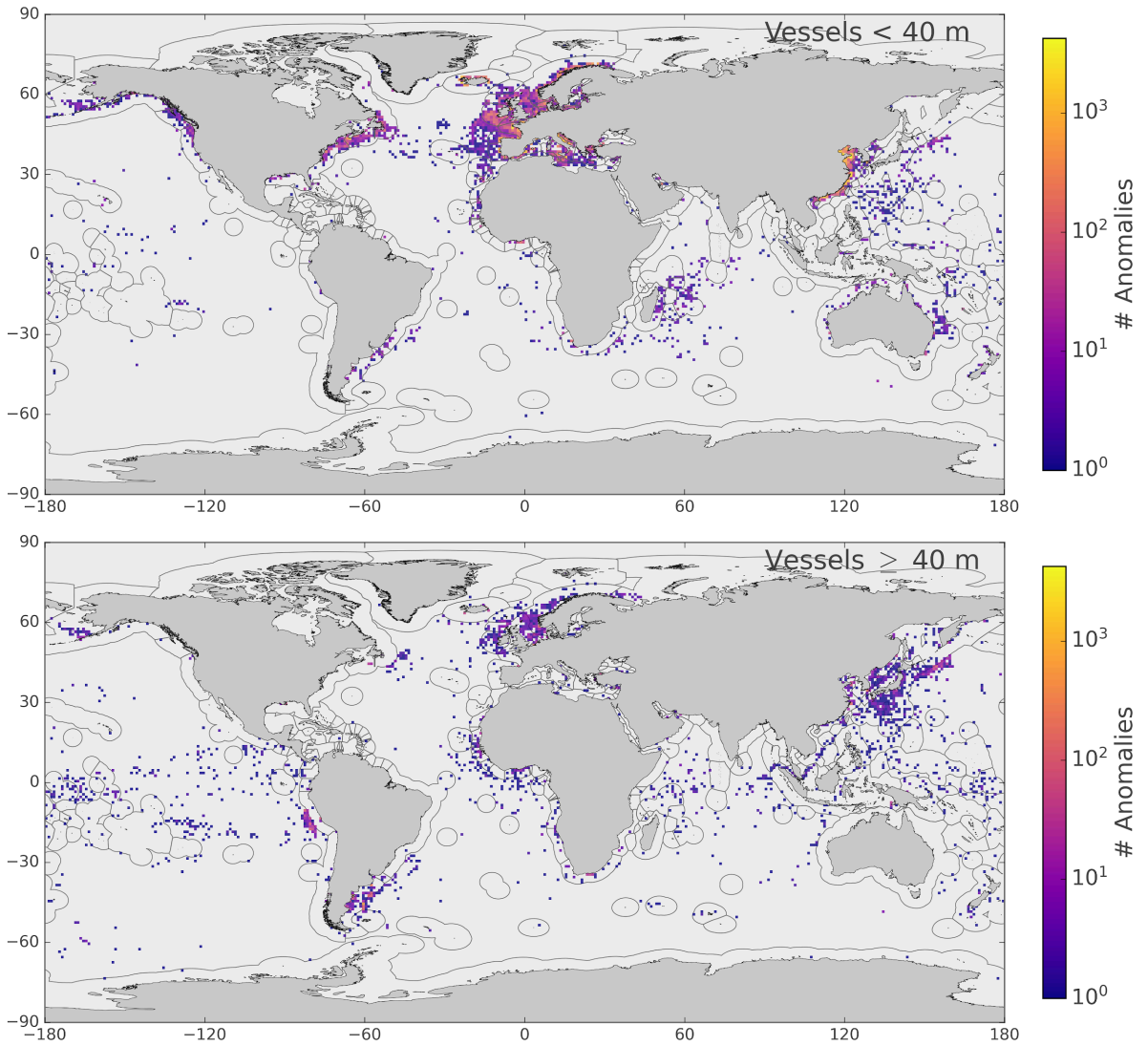


**Fig. S9.** Geographical extent of the number of significant anomalies starting at each 1° lat × 1° lon grid cell, for short (top, length < 40 m) and long (bottom, length ≥40 m) fishing vessels.

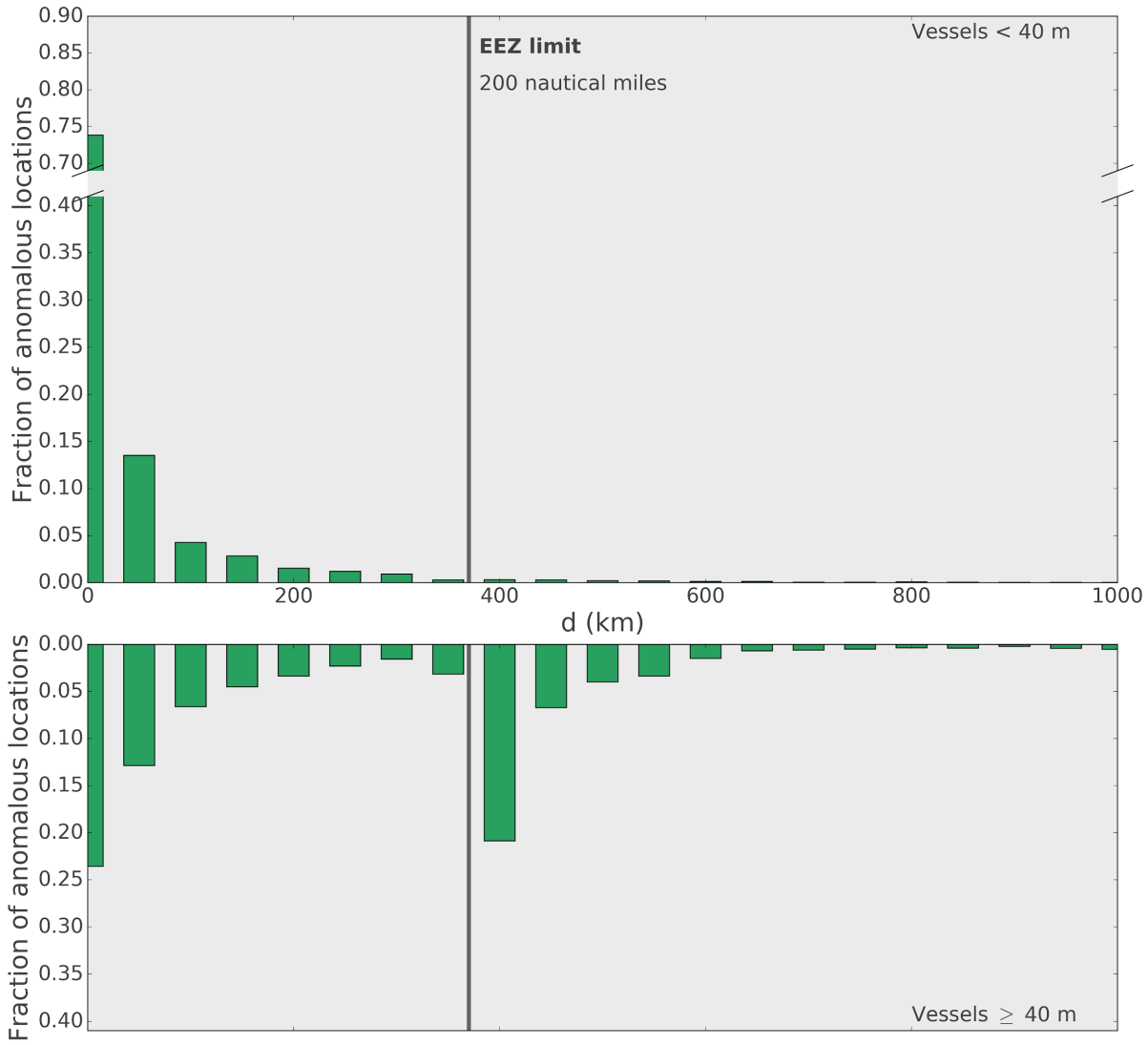


**Fig. S10.** Fraction of the global number of significant anomalies observed in cells located at a distance d from the shore, for short (top, length < 40 m) and long (bottom, length ≥40 m) fishing vessels.

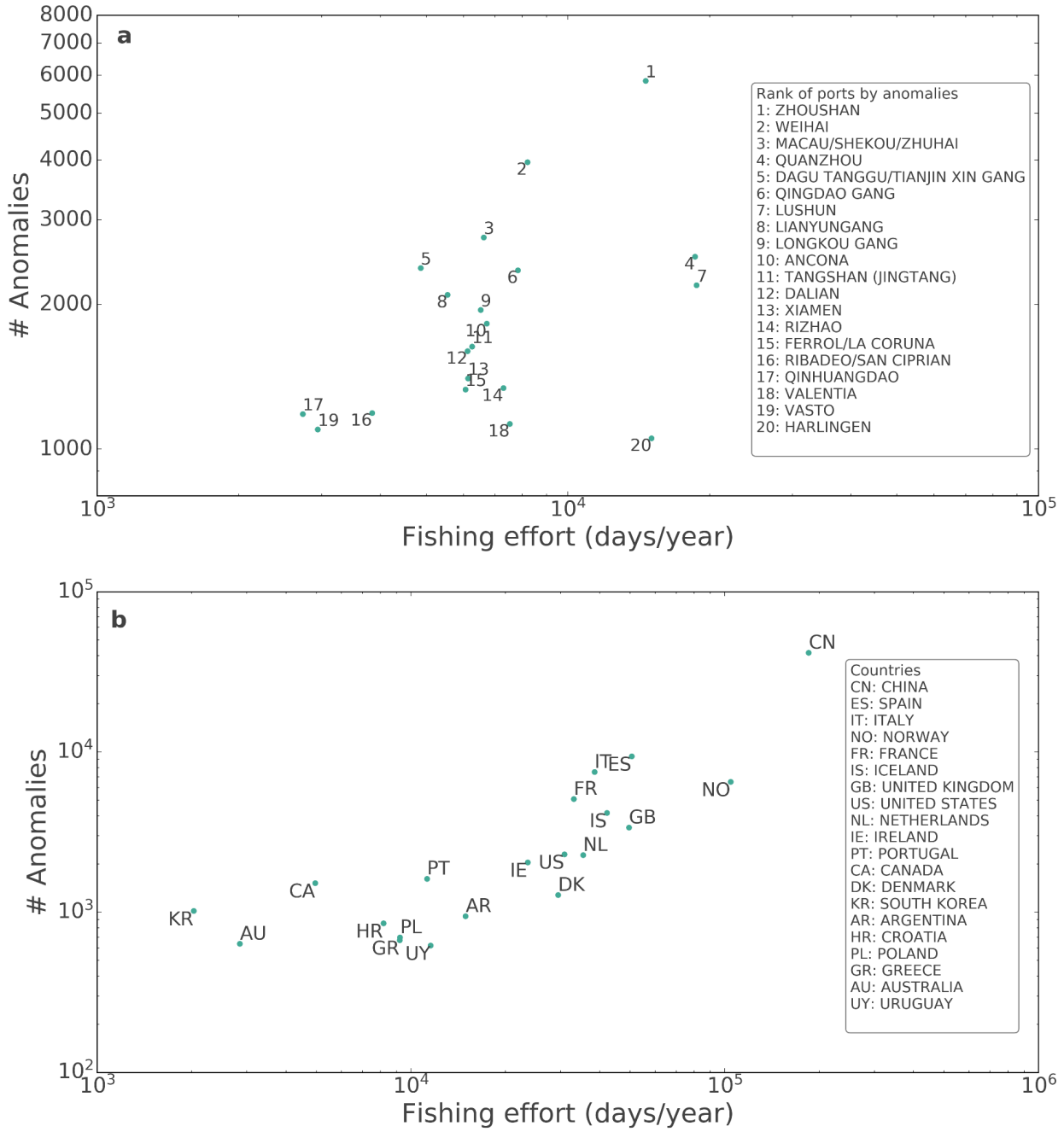


**Fig. S11.** Comparison between the number of anomalies linked to each port and the fishing effort associated to that port in **a,** the top 20 ports with most anomalies, and **b,** joining ports by countries and selecting the top 20 countries with most anomalies. Fishing effort was computed with AIS data from 2014, extending the methodology used in [S1] to consider fishing effort in high seas and within Exclusive Economic Zones.

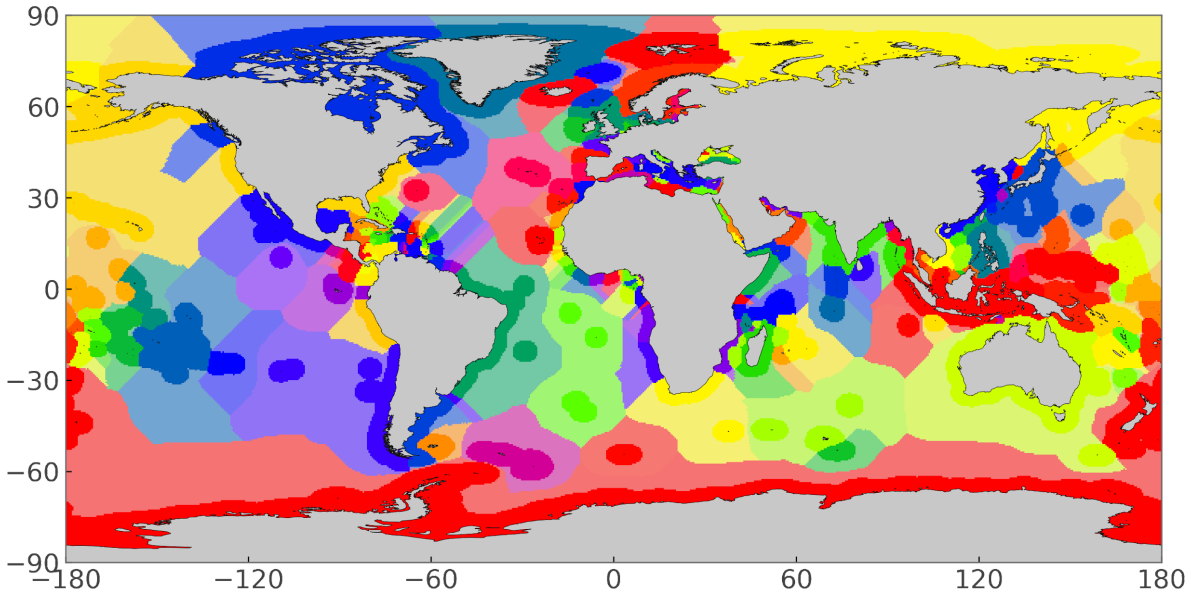


**Fig. S12.** Exclusive Economic Zones (bright colors) and their closest areas in the Areas Beyond National Jurisdiction (pale colors).

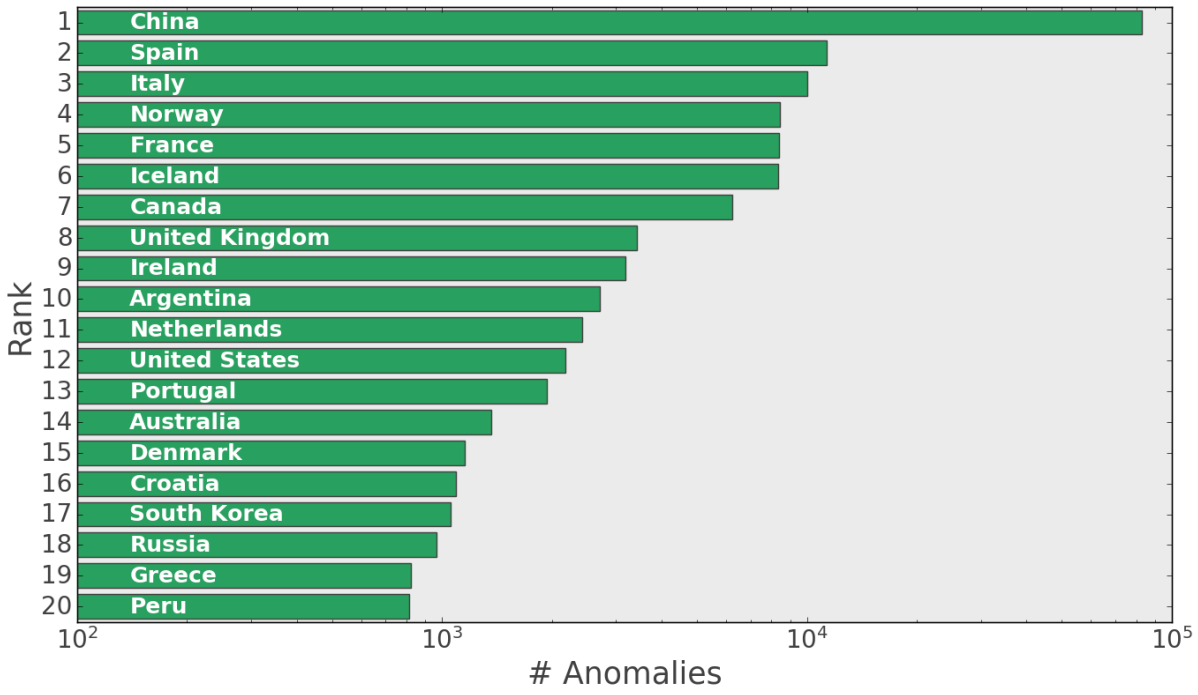


**Fig. S13.** Rank of the Exclusive Economic Zones (EEZ) with most of the anomalies associated with them. The closer cells to an EEZ than to any other are assigned to that EEZ (Fig. S11). The 20 EEZs represented here are associated with 94% of the observed significant anomalies.

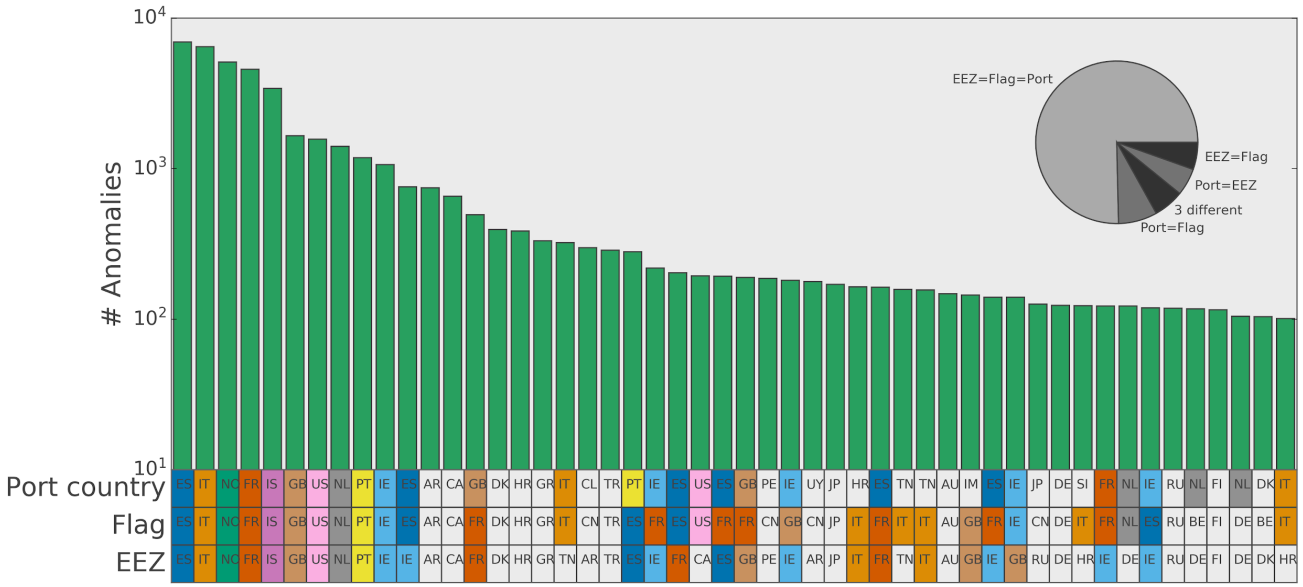


**Fig. S14.** Number of anomalies observed for each unique combination (Port country, Vessel flag, EEZ where the anomaly occurred). The anomalies account as 0.5 for the country before and after the anomaly, observed for the subset of MMSIs with available flag registry information (68 K anomalies). Colored cells correspond to the countries represented in the top-10 triplets. Country codes are ES: Spain, IT: Italy, NO: Norway, FR: France, IS: Iceland, GB: United Kingdom, US: United States, NL: Netherlands, PT: Portugal, IE: Ireland, AR: Argentina, CA: Canada, DK: Denmark, HR: Croatia, GR: Greece, TN: Tunisia, CL: Chile, CN: China, TR: Turkey, PE: Peru, UY: Uruguay, JP: Japan, AU: Australia, IM: Isle of Man, RU: Russia, DE: Germany, SI: Slovenia, BE: Belgium, FI: Finland.

## Supplementary Methods

**Vessels' features.** To extract the information needed for Figs. S6, S9 and S10, we searched the MMSI numbers recorded in our dataset in the Global Fishing Watch vessel database [S2], extracting the vessel's length and flag.

**Equidistant grid.** To select the cell ports, filter the significant anomalies (see Methods in main text) and develop the analyses in Figs. S5 and S8, we created a global grid composed of cells of size 50 km × 50 km. To assign each location to a grid cell, we computed as the vertical coordinate the great-circle distance to the equator, keeping the longitude constant, and as the horizontal coordinate the great-circle distance to the 0º meridian, being positive for positive longitudes and negative otherwise. Afterwards, we obtained the discrete values of these components by computing the integer part of their division by 50 km. With these discrete values, we assigned each location to a grid cell. Similarly, for visual purposes in the maps representation, we considered grid cells of size 100 km × 100 km.

## Supplementary References

[S1] Rodríguez, J. P., Fernández-Gracia, J., Duarte, C. M., Irigoien, X., Eguíluz, V. M. (2021). The global network of ports supporting high seas fishing. Science Advances, 7(9), eabe3470.

[S2] Kroodsma, D. A., Mayorga, J., Hochberg, T., Miller, N. A., Boerder, K., et al. (2018). Tracking the global footprint of fisheries. Science, 359(6378), 904-908.