Analysis for: Satellites reveal global extent of forced labor in the world's fishing fleet

Gavin McDonald - Environmental Market Solutions Lab (emLab)

10/30/2020

Contents

Analysis

Load data and necessary functions

Define pre-processing steps

Specify models

Cross-validation

Cross-validation performance

Choose optimized model

Final model building using full dataset and optimized model

Robustness checks

Use BigQuery

High risk fishing effort

High risk port visits

Time at sea statistic

This query generates the fraction of total time at sea by included vessels in the analysis. By running this query, we find that "These vessels represent 33% of the total time at sea spent by all fishing vessels operating in this time period tracked by Global Fishing Watch."

Known registry vessel characteristics

Figures

Figure 1: Training data model feature summary

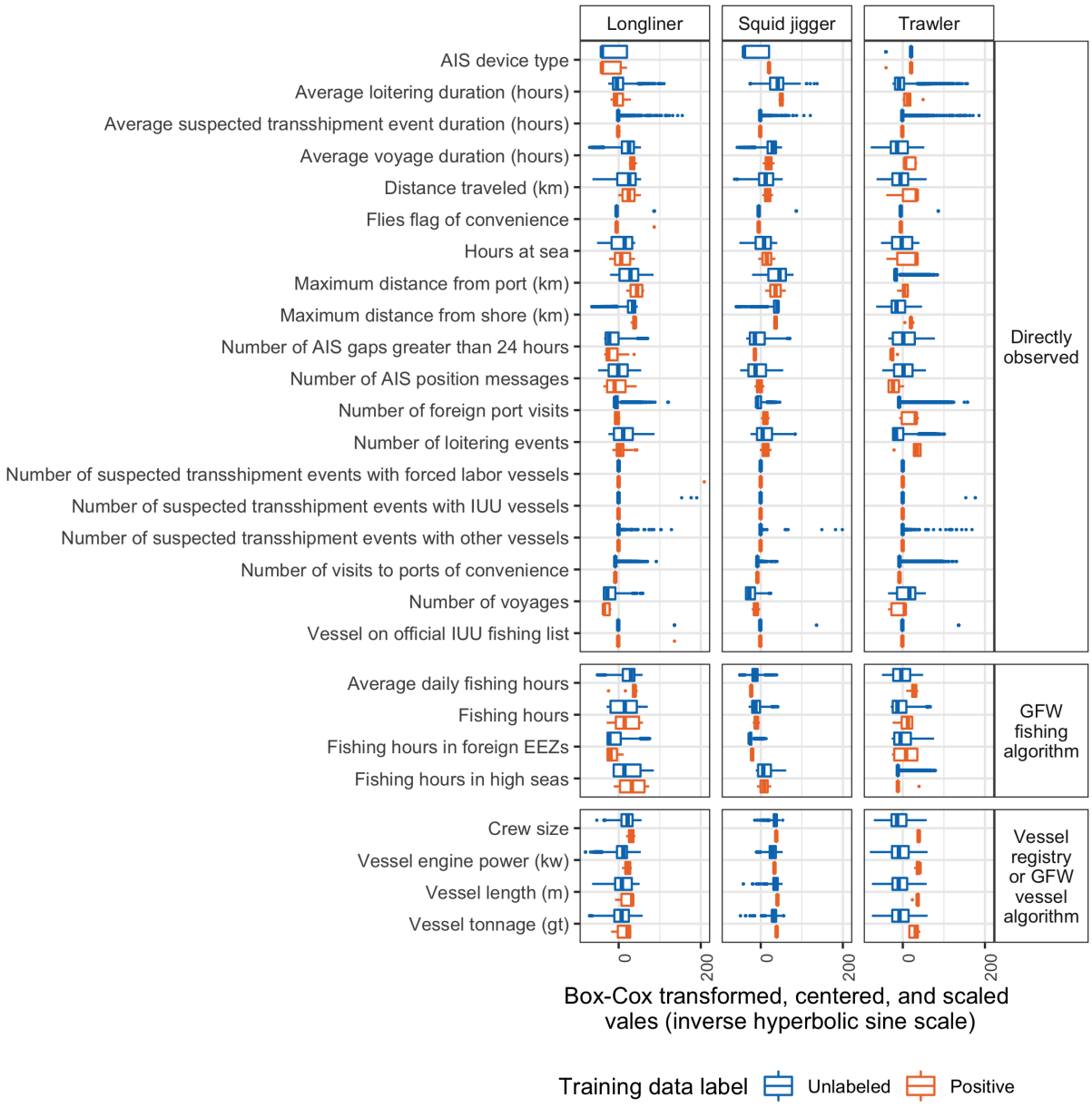


Figure 2 - Forced labor risk by fishing fleet

Figure 2 using point estimates (not shown in paper)

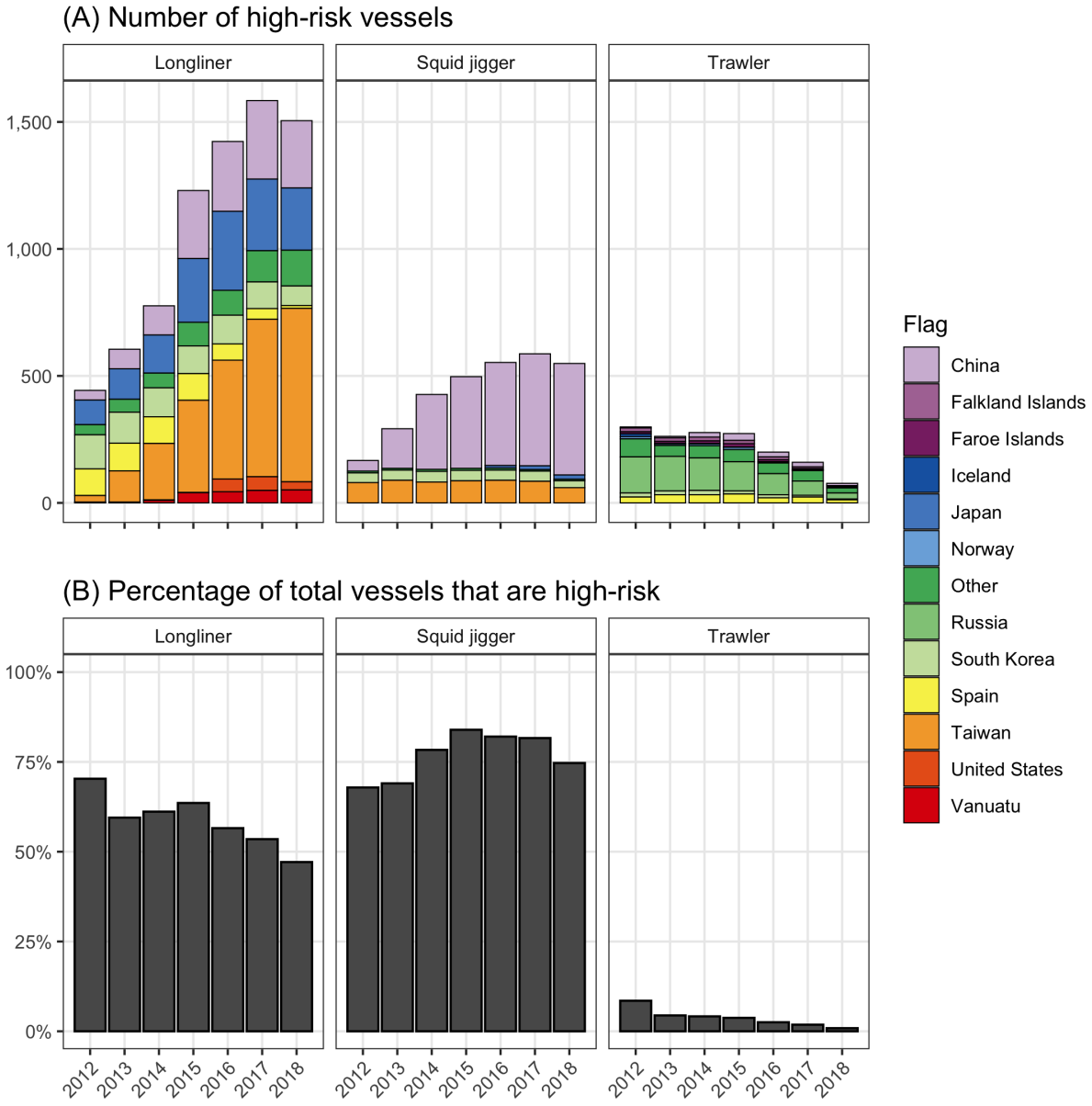


Figure 2 using range estimates from robustness checks

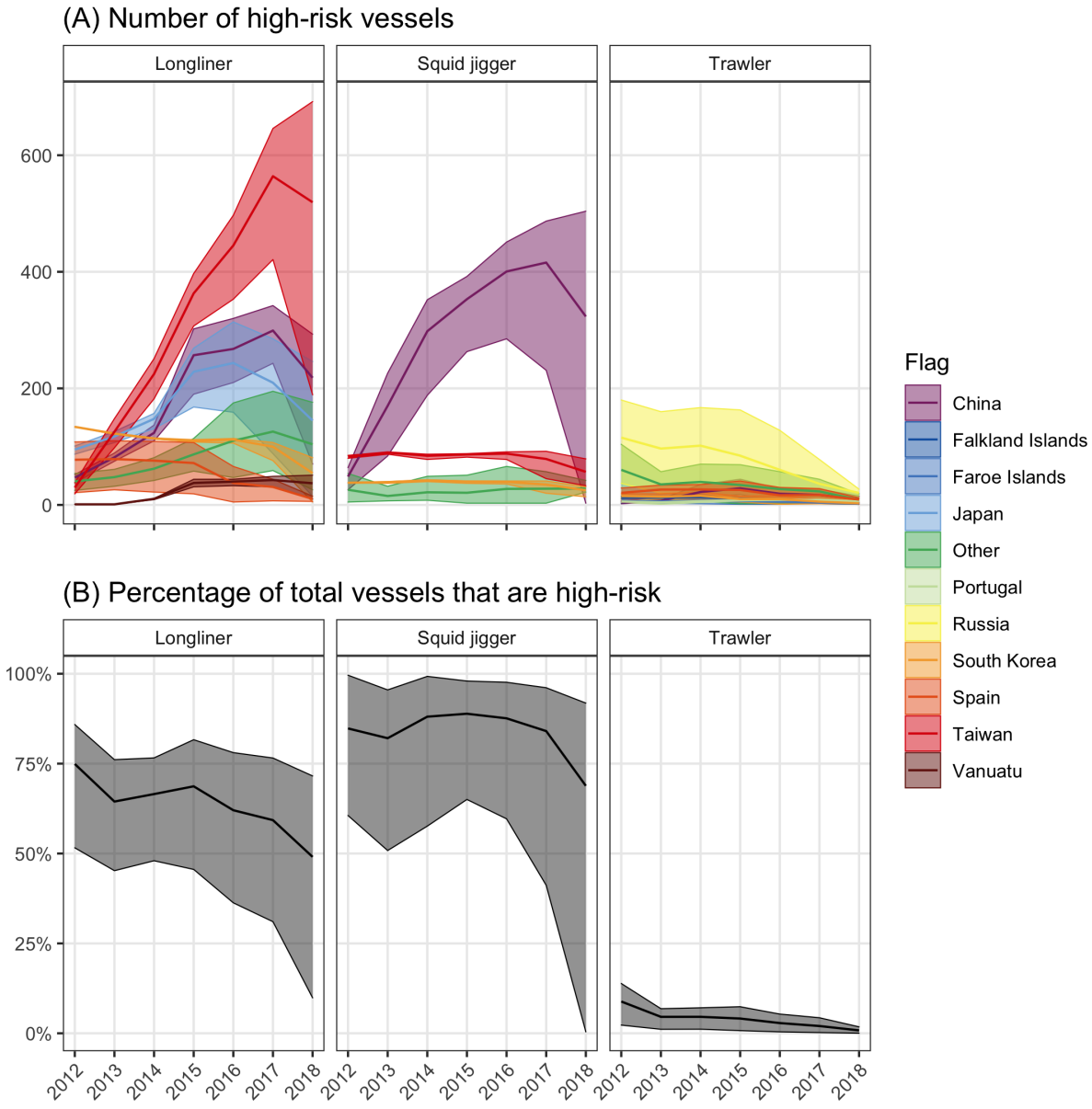


Figure 3 - Spatial forced labor risk

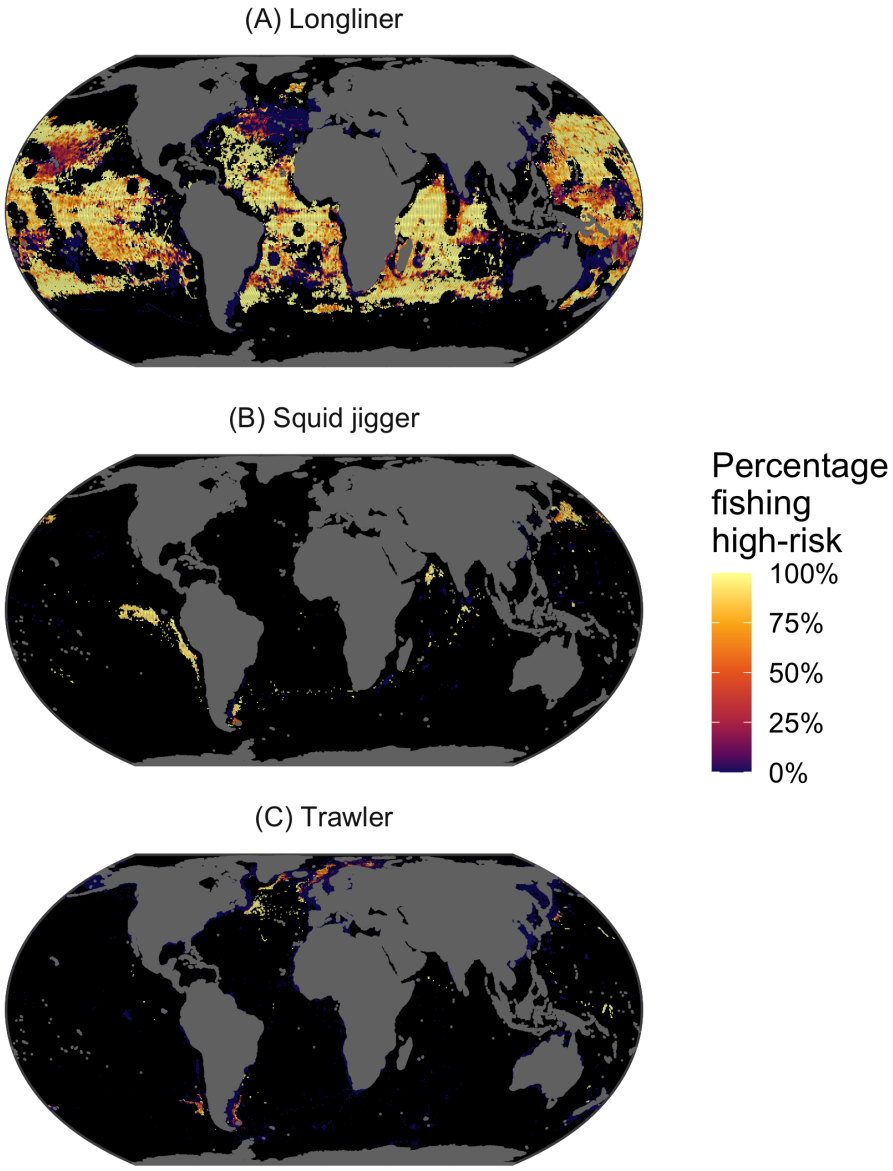


Figure 4: Port visits by high-risk vessels

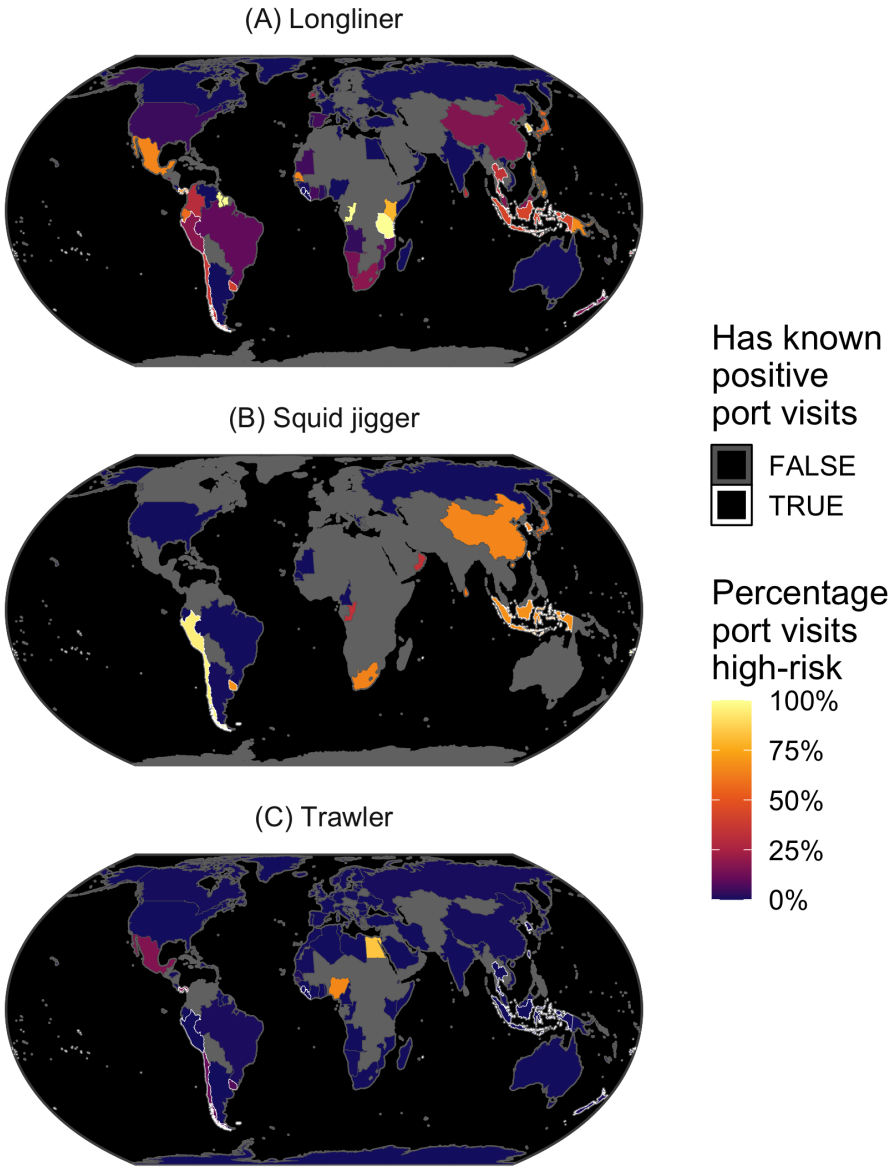


Figure S1 - Positive vessel cases

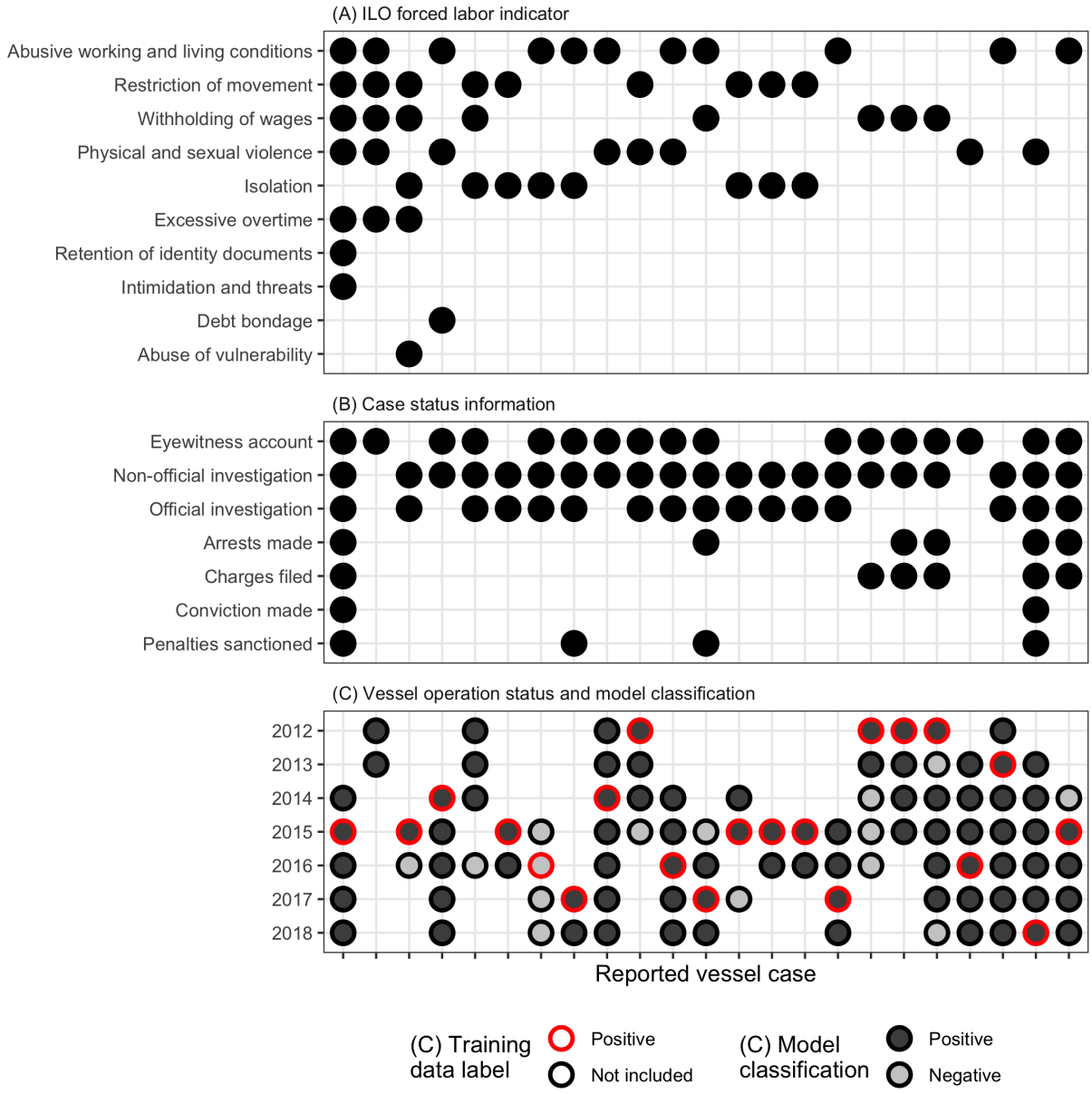


Figure S2 - Summary of training data labels

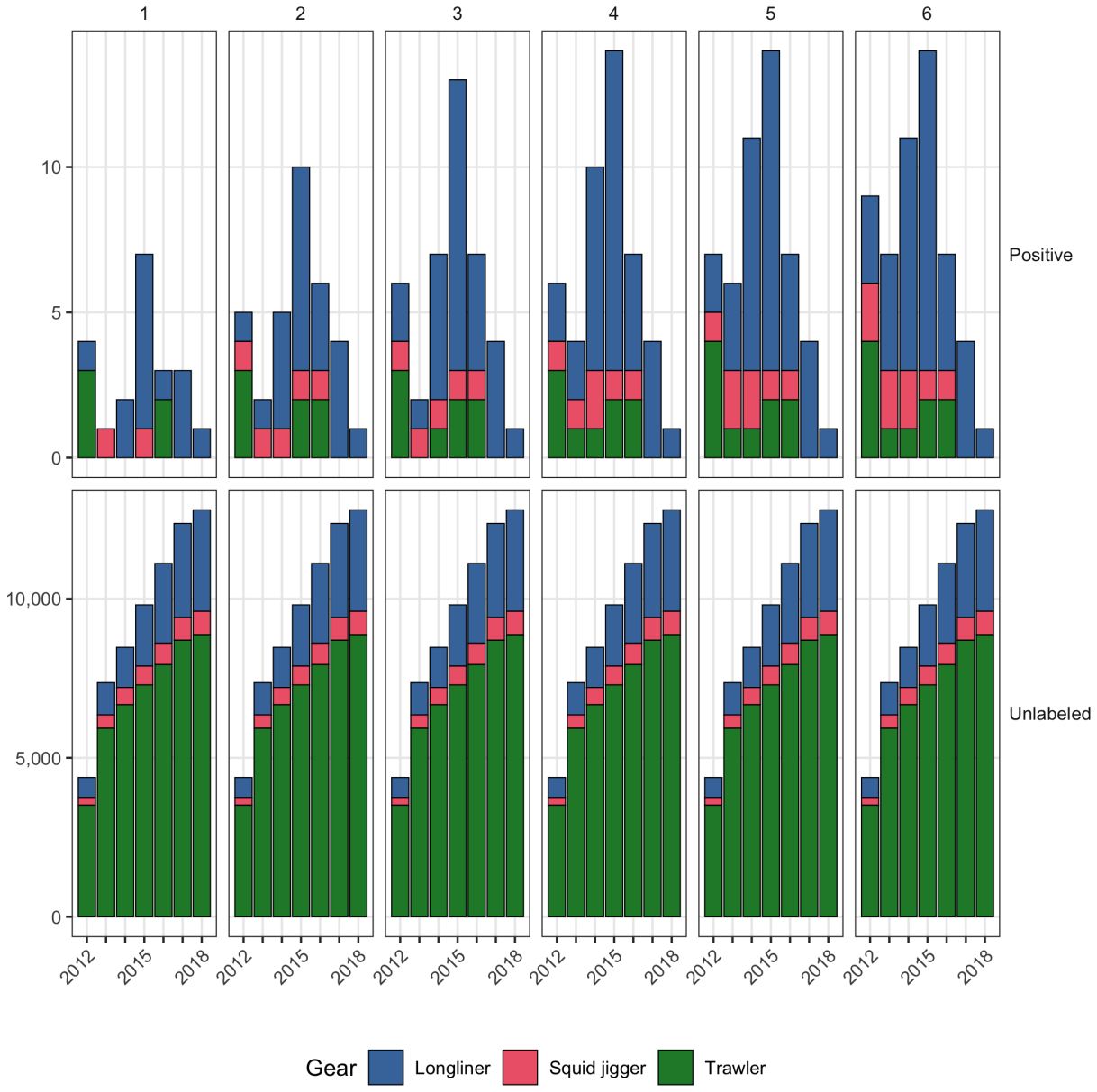


Figure S3 - Cross-validation performance

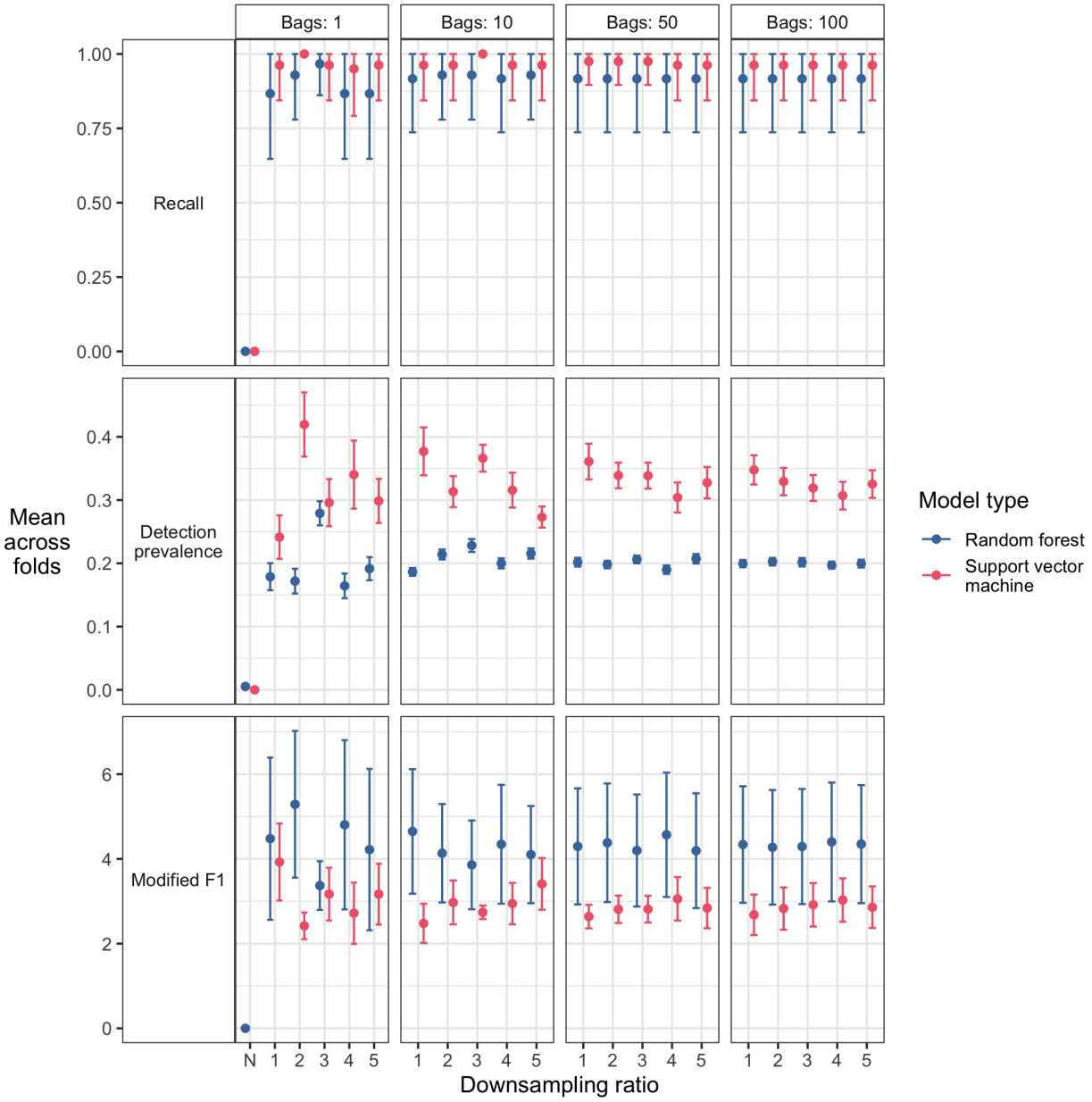


Figure S4: Classification by training label

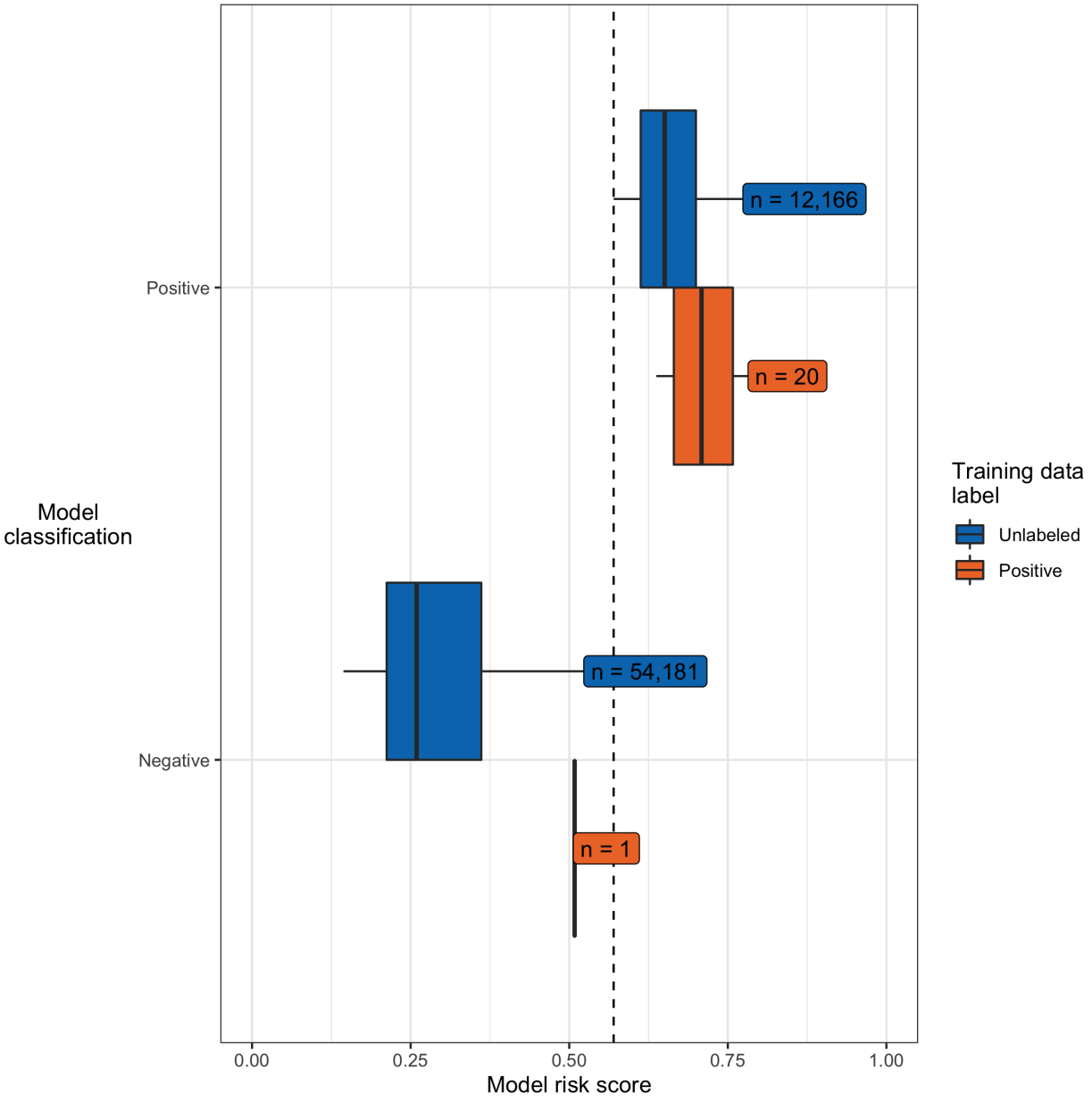


Figure S5: Variable importance

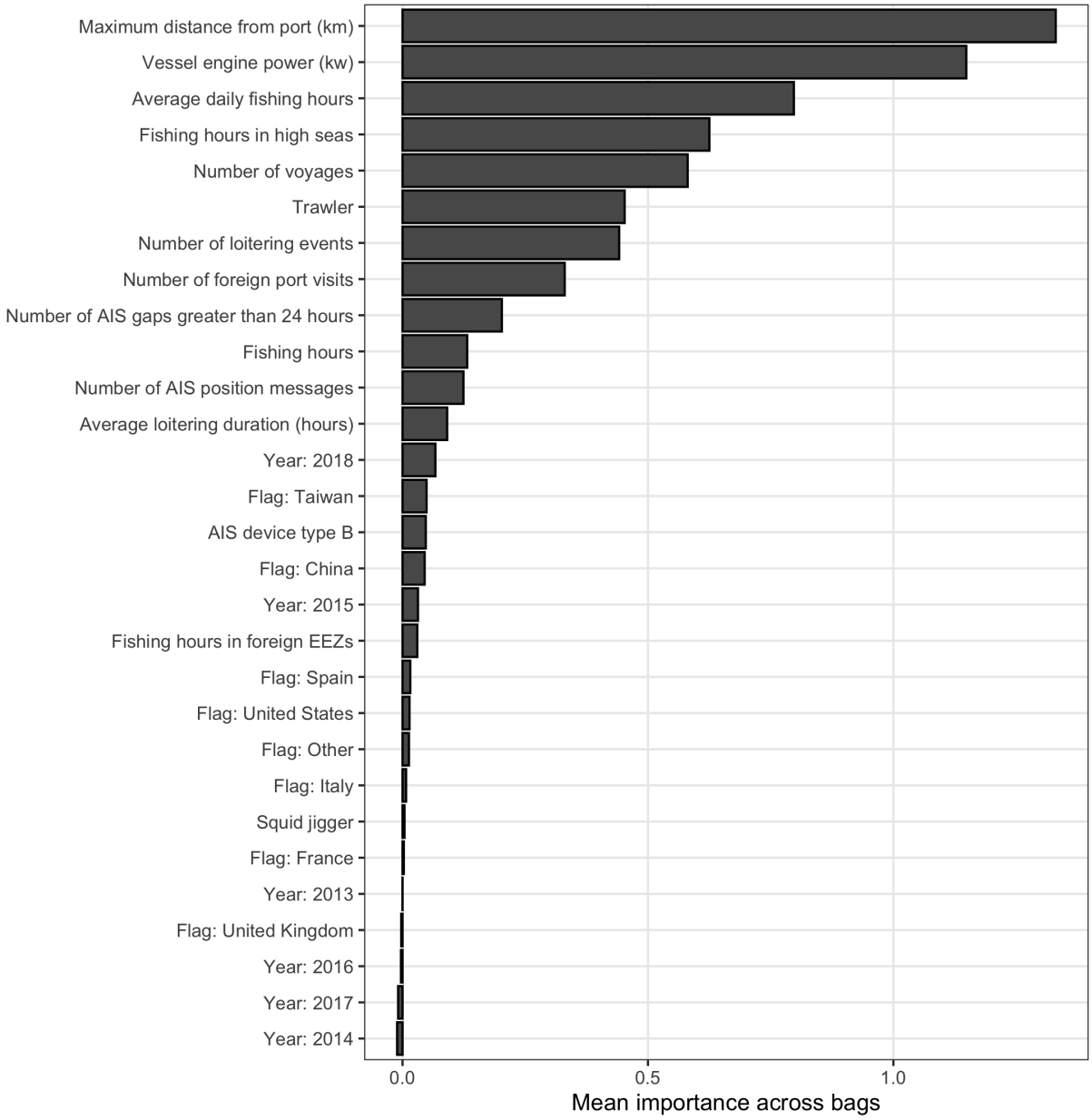


Figure S6 - Cross-validation results from robustness checks

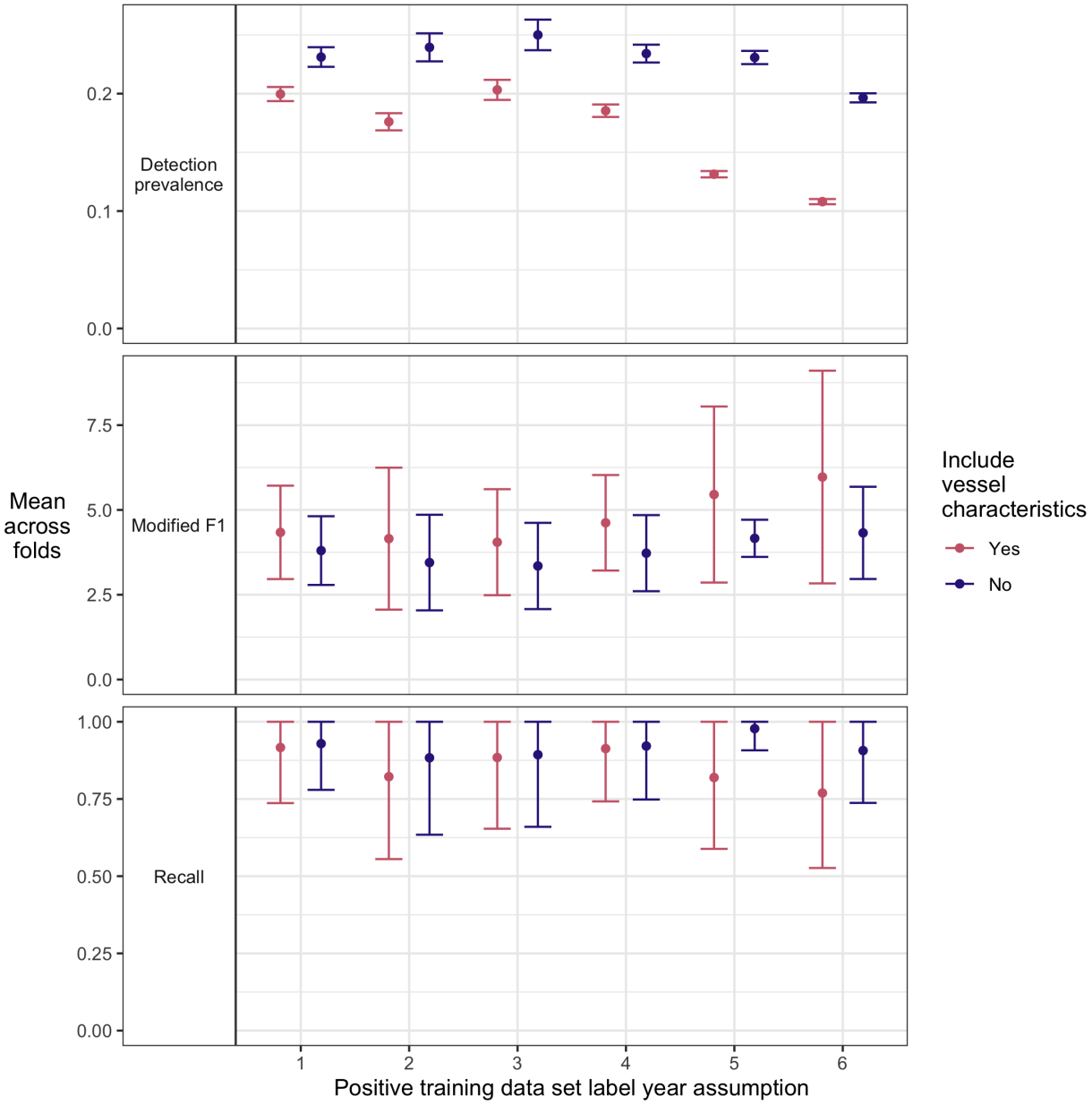
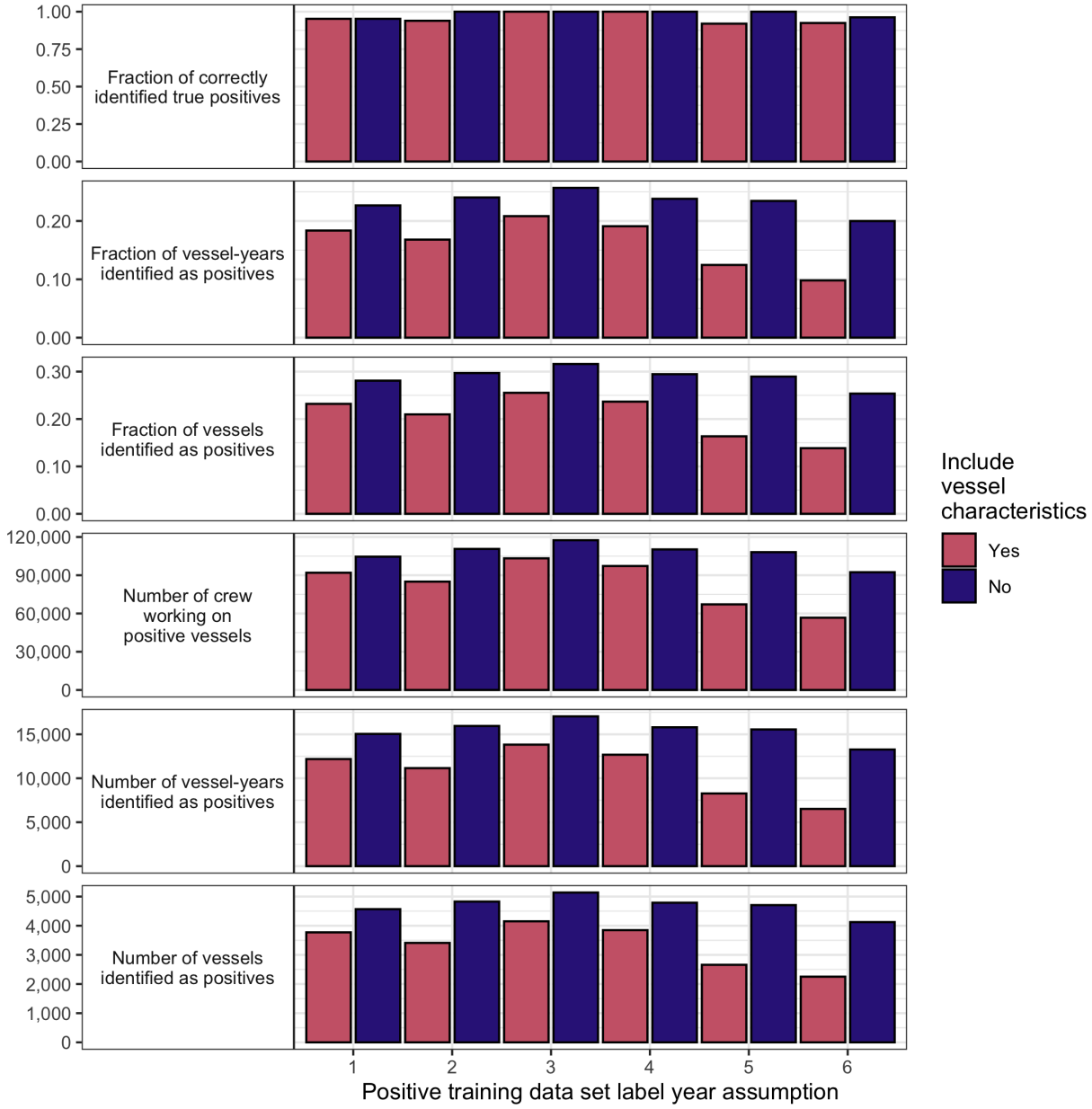


Figure S7 - Final model results from robustness checks



Statistics

We select our optimized model variation to be random forest with 100 bags and a downsampling ratio of 1. This specification has the following mean performance across the 10 folds: modified F1 score: 4.3; recall: 0.92; and detection prevalence: 0.2.

Using base model assumptions

Our PU approach leverages information from all positively labeled vessels (n = 21 unique vessels across 21 vessel-years, using our base model assumption), but places less emphasis on unlabeled vessels given their uncertain nature (n = 16257 unique vessels across 66,315 vessel-years, using our base model assumption).

We train the predictive model using vessel monitoring data from Global Fishing Watch . For 16,261 unique longliner, trawler, and squid jiggers vessel vessels, we calculate a number of features on an annual basis from 2012 - 2018 (SI Table S1). We call the unit of observation a "vessel-year". These features represent aggregate annual observable vessel behavior features. We also include vessel characteristic features such as vessel flag and engine power. This training dataset includes 66,368 vessel-years of observation.

The model identifies 12,000 new high-risk vessel-years that were previously unlabeled.

The model correctly identifies 95% of positive vessel-years as being high-risk, while also identifying 12,000 total high-risk vessel-years (18% of the total vessel-years).

3,800 unique vessels were high-risk during at least one year (23% of the total unique vessels).

92,000 crew members were working on these boats and thus potential victims of forced labor during at least one year.

Taiwan longliners, China squid jiggers, Japan longliners, China longliners, South Korea longliners represent the five fisheries with the largest number of unique high-risk vessels.

While longliners have the largest number of high-risk vessel-years across years, squid jiggers have the highest percentage of high-risk vessels across all years (78%), followed by longliners (56%) and trawlers (3%).

We also find that known positive vessels visited ports in 17 countries during the 2012-2018-time frame using our base model assumptions.

In 2018 alone, model-identified high-risk vessels visited ports across 79 developed and developing countries in 2018 (50% of all visited countries for these gear types), including 39 Parties to the Port State Measures Agreement (Figure 4). The visited ports are predominantly in Asia, Africa, and South America, with notable exceptions being Canada, United States, New Zealand, and several European countries.

64 of the countries visited by high-risk vessels in 2018 had not been visited by known positive vessels, which is reflective of our limited training data set but may also be reflective of the limited port oversight currently occurring in many countries.

In 2018, 13 of 23 (57%) vessels with reported forced labor were still operating, with 11 of those 13 vessels (85%) being classified as high-risk.

Table 1: Countries that were visited by high-risk vessels in 2018

country
American Samoa
Angola
Argentina
Australia
Brazil
Canada

Cape Verde
Chile
China
Colombia
Congo - Brazzaville
Cook Islands
Côte d'Ivoire
Denmark
Ecuador
Egypt
El Salvador
Falkland Islands
Faroe Islands
Fiji
French Polynesia
Germany
Ghana
Guinea
Guinea-Bissau
Guyana
Iceland
Indonesia
Ireland
Japan
Kenya
Kiribati
Malaysia
Marshall Islands
Mauritania
Mauritius
Mayotte
Mexico
Micronesia (Federated States of)
Morocco
Mozambique
Namibia
Netherlands
New Caledonia
New Zealand
Nigeria
Norway
Oman
Palau
Panama
Papua New Guinea
Peru
Philippines
Poland
Portugal
Russia
Samoa

São Tomé & Príncipe
Senegal
Seychelles
Singapore
Solomon Islands
South Africa
South Korea
Spain
Sri Lanka
Suriname
Svalbard & Jan Mayen
Sweden
Taiwan
Tanzania
Thailand
Tonga
Trinidad & Tobago
Tuvalu
United States
Uruguay
Vanuatu
Western Sahara

Table 2: Countries that had ratified the PSMA and were visited by high-risk vessels in 2018

country
Australia
Cape Verde
Chile
Denmark
Faroe Islands
Germany
Ghana
Guinea
Guyana
Iceland
Indonesia
Ireland
Kenya
Mauritania
Mauritius
Morocco
Namibia
Netherlands
New Zealand
Norway
Oman
Palau
Panama
Peru

Philippines
Portugal
São Tomé & Príncipe
Senegal
Singapore
South Africa
South Korea
Spain
Sri Lanka
Sweden
Thailand
Tonga
United States
Uruguay
Vanuatu

For the training dataset used in this analysis, 59% of vessel-years have known vessel length from registries, 58% have known gross tonnage, 48% have known engine power, and 16% have known crew size.

Using range of results from robustness checks

The model correctly identifies between 92% and 100% of positive vessel-years as being high-risk, while also identifying between 6,500 and 14,000 total high-risk vessel-years (between 10% and 21% of the total vessel-years).

Between 2,300 and 4,200 unique vessels were high-risk during at least one year (between 14% and 26% of the total unique vessels).

Between 57,000 and 100,000 crew members were working on these boats and thus potential victims of forced labor during at least one year.

Looking across all model assumptions, Taiwan longliners, China squid jiggers, China longliners, Japan longliners, South Korea longliners represent the five fisheries with the largest number of unique high-risk vessels.

While longliners have the largest number of high-risk vessel-years across years, squid jiggers have the highest percentage of high-risk vessels across all years (between 45% and 94%), followed by longliners (between 33% and 60%) and trawlers (between 1% and 4%).