



Satellites can reveal global extent of forced labor in the world's fishing fleet

Gavin G. McDonald^{a,b,1}, Christopher Costello^{a,b}, Jennifer Bone^{a,b}, Reniel B. Cabral^{a,b}, Valerie Farabee^c, Timothy Hochberg^d, David Kroodsmad, Tracey Mangin^{a,b}, Kyle C. Meng^{a,e}, and Oliver Zahn^f

^aBren School of Environmental Science & Management, University of California, Santa Barbara, CA 93106; ^bMarine Science Institute, University of California, Santa Barbara, CA 93106; ^cLiberty Shared, Washington, DC 20001; ^dGlobal Fishing Watch Inc., Washington, DC 20036; ^eDepartment of Economics, University of California, Santa Barbara, CA 93106; and ^fGoogle, Mountain View, CA 94043

Edited by James N. Sanchirico, University of California, Davis, CA, and accepted by Editorial Board Member Catherine L. Kling November 6, 2020 (received for review July 31, 2020)

While forced labor in the world's fishing fleet has been widely documented, its extent remains unknown. No methods previously existed for remotely identifying individual fishing vessels potentially engaged in these abuses on a global scale. By combining expertise from human rights practitioners and satellite vessel monitoring data, we show that vessels reported to use forced labor behave in systematically different ways from other vessels. We exploit this insight by using machine learning to identify high-risk vessels from among 16,000 industrial longliner, squid jigger, and trawler fishing vessels. Our model reveals that between 14% and 26% of vessels were high-risk, and also reveals patterns of where these vessels fished and which ports they visited. Between 57,000 and 100,000 individuals worked on these vessels, many of whom may have been forced labor victims. This information provides unprecedented opportunities for novel interventions to combat this humanitarian tragedy. More broadly, this research demonstrates a proof of concept for using remote sensing to detect forced labor abuses.

forced labor in fisheries | machine learning | satellite vessel monitoring data

Forced labor in fisheries, a type of modern slavery, is increasingly recognized as a human rights crisis. The International Labor Organization (ILO) defines forced labor as “all work or service which is exacted from any person under the menace of any penalty and for which the said person has not offered himself voluntarily” (1). The ILO provides a framework of 11 forced labor risk indicators (2) that have all been documented within the fisheries sector, including indicators representative of debt-bonded labor, as well as indicators representative of servitude or slave labor such as abusive working and living conditions. In 2015, reports emerged on forced labor in Thai fisheries (3) and the role of forced labor in producing seafood imported to the United States (4). More recent reports have described the global nature of the problem (5), and there has been a call to integrate social responsibility into ocean science (6). Despite widespread condemnation and ambitious commitments, forced labor remains poorly understood in the fisheries sector. Here we show that recently available high-frequency vessel monitoring of the global industrial fishing fleet can shed new light on forced labor at a much finer resolution. We combine expertise from on-the-ground human rights practitioners and satellite vessel monitoring data for over 16,000 industrial fishing vessels to estimate 1) the number of high-risk vessels and the number of crew who may be victims working on those vessels, 2) where these vessels fish, and 3) what ports these vessels visit. This information can inform new market, policy, and enforcement interventions to combat forced labor in global fisheries. This research more generally demonstrates how remote sensing can detect forced labor abuses by observing dynamic behavior.

Current estimates of forced labor in fisheries are coarse and are based on country-level statistics. Using country-level household surveys, the ILO estimated that 16 million people were victims of forced labor in 2016, with 11% of these in agriculture,

forestry, or fisheries (7). The Global Slavery Index reports that the seven countries with highest slavery risk in 2018 generated 39% of global fisheries catch (3, 8), and Tickler et al. found that the United States has slavery risks of 0.2 kg per metric ton for domestic seafood and 3.1 kg per metric ton for imported seafood (9). While these studies are important for broadly understanding which countries have risk, current methods are unable to detect this problem at the level of individual fishing vessels, which will be essential for targeted interventions.

We empirically examine whether vessels reported to exhibit any of the ILO indicators of forced labor behave in ways that are systematically different from other vessels, and then exploit this information using machine learning to discriminate between vessels that use forced labor from those that do not. We do so by measuring a suite of features that can be observed using satellite Automatic Identification System (AIS) vessel monitoring data made available by Global Fishing Watch (GFW) (10). There may be many behavioral correlates with forced labor that could help to differentiate between high-risk and low-risk vessels. To determine which model features to include, we first conducted a literature review of investigative journalism reports and looked for instances of forced labor case accounts that detailed specific behaviors that could be observed using vessel monitoring data. We next conducted informal phone interviews with experts from several nongovernmental organizations (NGOs) working in this field, during which we asked interviewees what observable vessel behaviors they would look for if they wanted to identify

Significance

Forced labor in fisheries is increasingly recognized as a human rights crisis. Until recently, its extent was poorly understood and no tools existed for systematically detecting forced labor risk on individual fishing vessels on a global scale. Here we use satellite data and machine learning to identify these high-risk vessels and find widespread risk of forced labor in the world's fishing fleet. This information provides new opportunities for unique market, enforcement, and policy interventions. This also provides a proof of concept for how remotely sensed dynamic individual behavior can be used to infer forced labor abuses.

Author contributions: G.G.M., C.C., R.B.C., V.F., T.H., D.K., T.M., K.C.M., and O.Z. designed research; G.G.M., R.B.C., T.H., D.K., and T.M. performed research; G.G.M. analyzed data; and G.G.M., C.C., J.B., R.B.C., V.F., T.H., D.K., T.M., K.C.M., and O.Z. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. J.N.S. is a guest editor invited by the Editorial Board.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: gmcdonald@bren.ucsb.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2016238117/-DCSupplemental>.

Published December 21, 2020.

suspicious activity. The machine-learning approach we use does not assume that vessels behave in any particular way; rather, it merely uses the features identified by literature review and expert insight to exploit any observed empirical differences between vessels that use forced labor and other vessels. NGO experts and investigative journalism suggest that gaps in AIS transmission, port avoidance, transshipment, and extended time at sea may indicate the presence of forced labor (11). Certain features, like information on catch and the species being targeted, could also be helpful in discriminating between high- and low-risk behavior by providing more context on the fishing taking

place. However, these data are not currently available at the vessel level on a global scale. Data on recruitment practices and vessel ownership and information on from where the crew originates could also be helpful, but, again, these data are not widely available. We arrived at a list of 27 vessel behavior and characteristic features for which we have globally available data at the vessel level (*SI Appendix, Table S1, and SI Appendix*).

To build a predictive model for identifying high-risk vessels, we developed a training dataset that includes the behavior and characteristics of known forced-labor vessels, as well as the behavior and characteristics of other vessels. We compiled a

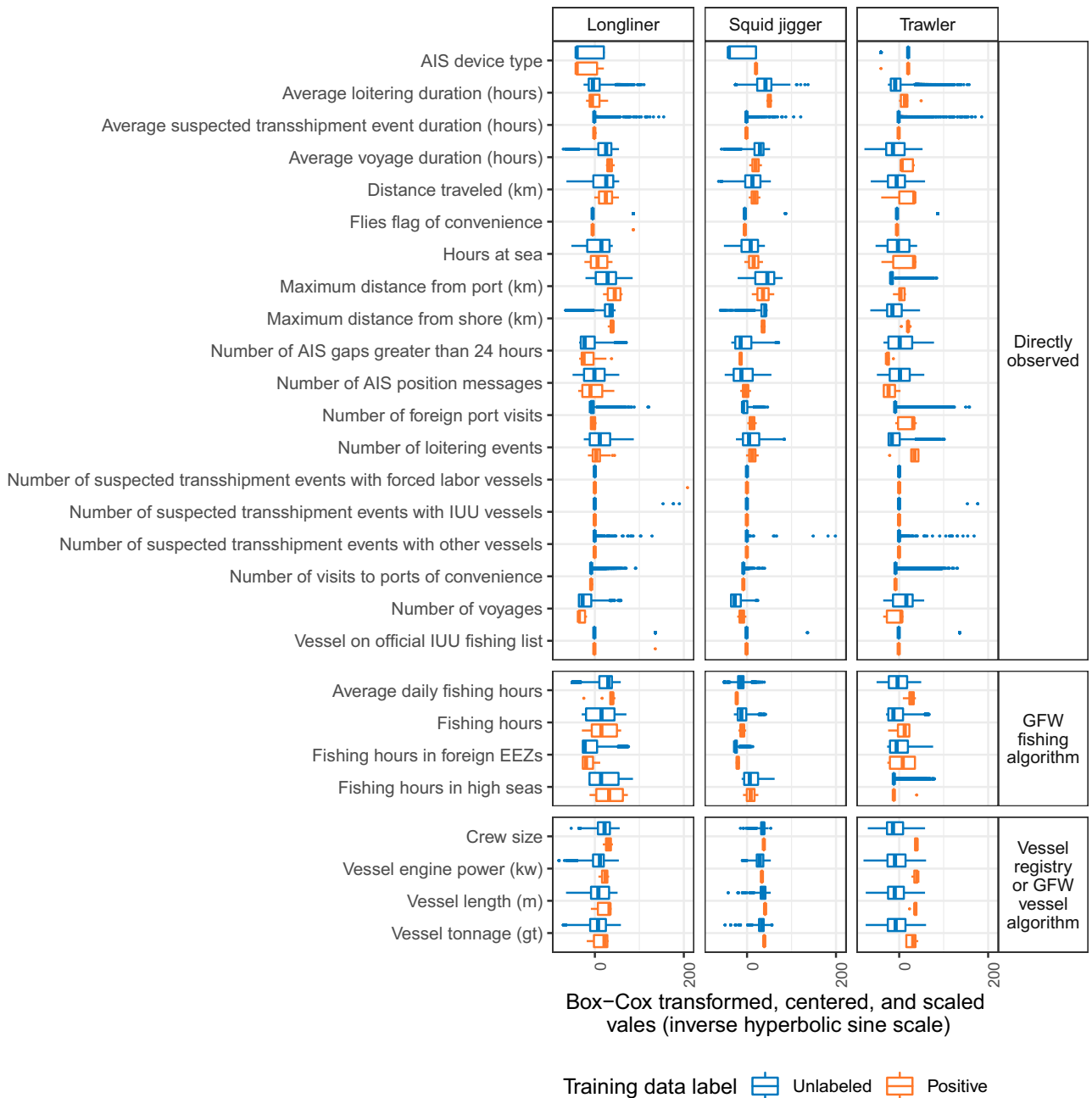


Fig. 1. Box-Cox-transformed, centered, and scaled annual feature values for positive vessels (those that were documented using forced labor) and unlabeled vessels (those that were not documented for forced labor or that use forced labor but have not been caught) for longliner, squid jigger, and trawler fleets. The x-axis is displayed using an inverse hyperbolic sine scale. Features are each directly observed using AIS data, inferred using the GFW fishing algorithm, or vessel characteristics that come from either a vessel registry or the GFW vessel characteristic algorithm.

comprehensive database of vessels that were reported to display one or more of the ILO forced labor indicators (2); these vessels are labeled as “positives.” We do not, however, know which vessels do not use forced labor (“negatives”). Rather, any vessel that we do not label as positive is “unlabeled,” and may in fact be a positive vessel that has not yet been identified or may truly be a negative vessel. This is an example of “positive-unlabeled (PU)” learning, a less straightforward problem than traditional supervised machine learning (12). We use PU learning to predict whether or not 16,261 longliner, trawler, and squid jigger fishing vessels were high-risk during each year they operated between 2012 and 2018 (“vessel-years”). We focus on this subset of vessels because they broadcasted sufficient and reliable AIS positions and because these are the only fishing gear types with documented cases of forced labor aboard vessels that broadcasted sufficient AIS data. These vessels represent 33% of the total time at sea spent by all fishing vessels operating in this time period tracked by GFW. Our PU approach leverages information from all positively labeled vessels ($n = 22$ unique vessels across 22 vessel-years using our baseline model assumption), but places less emphasis on unlabeled vessels given their uncertain nature ($n = 16,257$ unique vessels across 66,314 vessel-years using our baseline model assumption).

Results

We find that fishing vessels using forced labor behave differently than the rest of the global fishing fleet (Fig. 1). Longliners and trawlers using forced labor travel further from port and shore, fish more hours per day than other vessels, and have fewer voyages and longer voyage durations. 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 20% of the total vessel-years). Between 2,300 and 4,200 unique vessels were high-risk during at least 1 y (between 14% and 26% of the total unique vessels). We extrapolate the number of high-risk vessels to the number of crew working on those vessels to show that between 57,000 and 100,000 crew members were working on these boats and thus potential victims of forced labor during at

least 1 y. We do so by leveraging the GFW vessel characterization algorithm that infers the estimated number of crew on board individual vessels based on other vessel characteristics including gear, length, engine power, flag, and tonnage (13).

The most important model features for identifying high-risk vessels were engine power, maximum distance from port, number of voyages per year, and average daily fishing hours (*SI Appendix*, Fig. S5 and *SI Appendix*). The importance of engine power, a proxy for vessel size, suggests that forced labor appears to cluster in similarly sized vessels. The importance of maximum distance from port and average daily fishing hours is consistent with expert opinion, and also reflects the ILO indicators of isolation and excessive overtime. Number of voyages per year is a proxy for the number of port visits and the voyage duration, and its importance as a model feature indicates that vessels engaged in forced labor may actively avoid scrutiny and opportunities for enforcement.

While some of these newly discovered high-risk vessels may be false positives, the results are likely a conservative estimate of the global extent of forced labor in fisheries. This analysis only includes longliner, trawler, and squid jigger vessels for which we have sufficient and reliable AIS monitoring data, although we know the problem extends to other vessels and gear types. Out of 193 vessels that we identified in our database as having documented forced labor violations, only 58 of these vessels (30%) use AIS and transmitted sufficient AIS positions to be monitored. Additionally, in the PU setting, differences in observed behavior are between positive and unlabeled vessels. Because the set of unlabeled vessels contains both positive and negative vessels, our model is trained on differences that likely understate the true difference between positive and negative vessels, which may lead to underpredictions in the number of high-risk vessels.

We find that risk is widespread across many fisheries and flags (Fig. 2). Looking across the range of model assumptions and across all years, Taiwanese longliners, Chinese squid jiggers, and Chinese, Japanese, and South Korean longliners are consistently the five fisheries with the largest number of unique high-risk vessels. This pattern is consistent with reports on the abuses seen within distant water fleets that receive little legal oversight

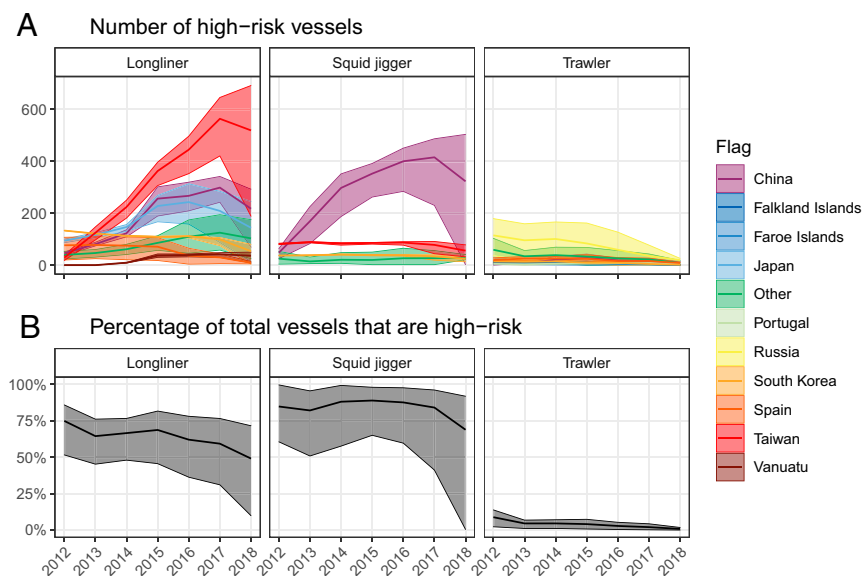


Fig. 2. (A) Number of model-identified high-risk vessels and (B) percentage of total vessels that are high-risk. Statistics are summarized by year within the longliner, squid jigger, and trawler fleets. The “other” flag category groups flags that represent less than 2.5% of vessels across years for a particular gear. The upper and lower bounds of each ribbon respectively represent the minimum and maximum values across all model robustness checks that include vessel characteristic model features, while the middle line of each ribbon represents the average value across all model robustness checks that include vessel characteristic model features.

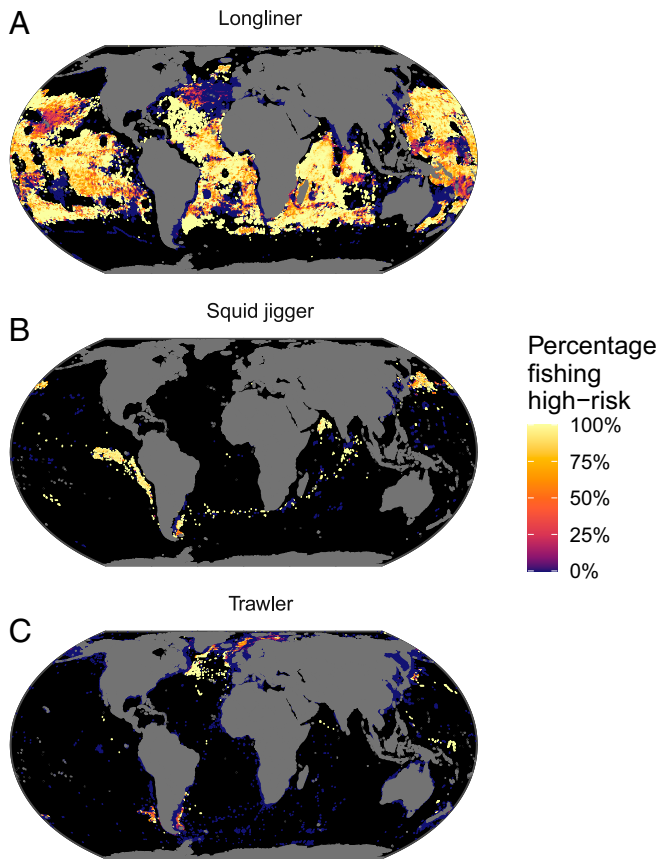


Fig. 3. Percentage of 2018 fishing effort (in kilowatt-hours) made by model-identified high-risk vessels out of the total fishing effort by all vessels included in the model, using baseline assumptions, within the (A) longliner, (B) squid jigger, and (C) trawler fleets. Fishing effort is calculated for $0.5 \times 0.5^\circ$ latitude/longitude gridded bins, and areas with no forced labor risk are shown in dark blue.

and often use marginalized migrant workers (5, 14). The number of high-risk longliner and squid jigger vessels increased from 2012 to 2017, with a decrease in 2018. 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%). In general, the percentage of each fleet that is high-risk has been declining, which may reflect increasing oversight but may also reflect the increasing number of vessels that carry AIS (*SI Appendix, Fig. S2 and SI Appendix*). Looking at 2018 data spatially and focusing on our baseline model assumptions, fishing by high-risk vessels occurs worldwide, both in the high seas and within national jurisdictions (Fig. 3). Longliner risk is widespread spatially, with hotspots including the western Indian Ocean, the coasts off West Africa and South Africa, and the central Atlantic, an area that had not previously received much media attention. Meanwhile, portions of the north Atlantic are hotspots for high-risk trawling activity, and areas to the west and southeast of South America, in the northwestern Pacific, and in the northern Indian Ocean are hotspots for high-risk squid jiggers. We also find that, in 2018 alone, model-identified high-risk vessels from the baseline model variation visited ports across 79 developed and developing countries (50% of all visited countries for these gear types), including 39 parties to the Port State Measures Agreement (PSMA; Fig. 4). Ports visited by high-risk vessels are predominantly in Asia, Africa, and South America, with notable exceptions being Canada, the United States, New Zealand, and several European countries.

Known positive vessels visited ports in 17 countries during the 2012 to 2018 time frame, while 64 of the countries visited by high-risk vessels in 2018 had not been visited by known positive vessels. This is reflective of our limited training data set but may also be reflective of the limited port oversight currently occurring in many countries. These ports represent both potential sources of exploited labor as well as transfer points for seafood caught using forced labor.

Discussion

Our approach to identifying individual vessels with a high risk of forced labor could improve existing monitoring efforts. Countries, enforcement bodies, and international agencies could use this model to conduct more targeted vessel inspections. Use of AIS or vessel monitoring system devices and detailed vessel registries could be further mandated to provide more accurate risk assessment for more vessels (15). Information from sanctions relating to forced labor could be made publicly available, alongside information from vessel inspections, in order to adaptively improve the model. Existing international legal frameworks should be leveraged to implement these policies, including United Nations Convention on the Law of the Sea Article 99 (Prohibition of the Transport of Slaves), ILO Work in Fishing Convention No. 188, and the 2012 Cape Town Agreement for safety aboard fishing vessels (16–18). The PSMA, which requires that parties implement port measures to prevent and deter illegal, unreported, and unregulated (IUU) fishing and

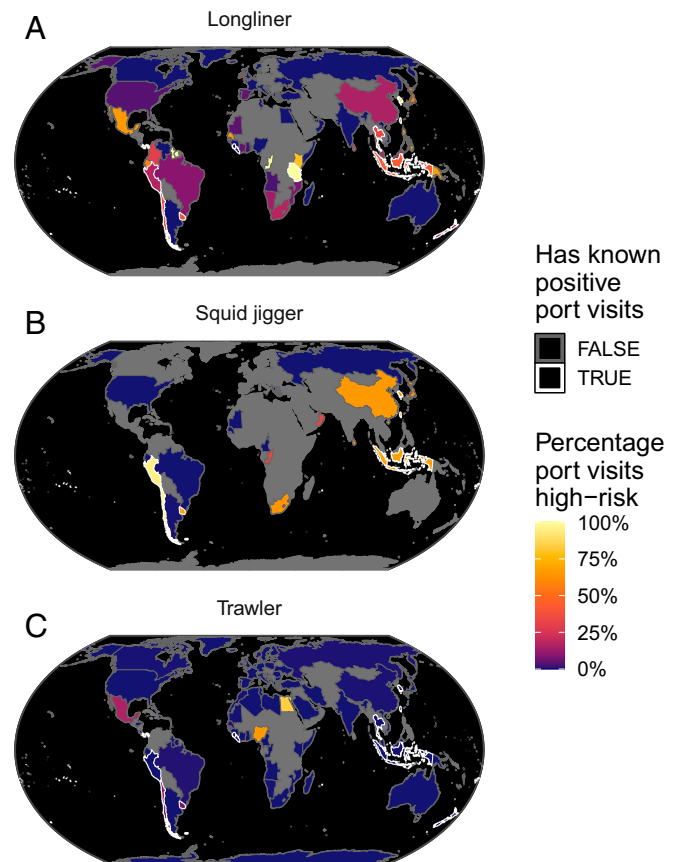


Fig. 4. Percentage of 2018 port visits made by model-identified high-risk vessels out of the total number of port visits by all vessels included in the model, by country, using baseline assumptions and within (A) longliner, (B) squid jigger, and (C) trawler fleets. Countries with no high-risk port visits by a particular fleet are shown in dark blue, while countries with no port visits are shown in gray. For countries with port visits by known positive vessels that occurred within the 2012 to 2018 time frame, the border of the country is highlighted in white.

allows for vessel inspections, may provide an additional opportunity for identifying forced labor and collecting information that could be used to adaptively improve the model (19).

In addition, the vessel-level forced labor risk information makes it possible to implement and improve market interventions in the sector. Model outputs could provide new information for market approaches aimed at informing businesses and consumers, and in turn could generate market pressure to reduce forced labor and improve working conditions in fisheries. Seafood distributors could use model outputs for targeted due diligence within their supply chains, adding a new piece of information to existing supply chain tools (20). Social network analysis, leveraging satellite-detected transshipment events (21), could be used to assess risk within at-sea supply chains to put pressure on vessels and companies to operate responsibly. Certification programs such as Marine Stewardship Council or Fair Trade and risk rating resources such as Seafood Slavery Risk Tool could use the model outputs as an additional data source for assessing risk within specific fisheries and to incentivize improved working conditions as a condition for certification. Furthermore, consumers may be willing to pay a premium for seafood that is free of forced labor, which would give seafood companies an incentive to demand clean seafood from its value chain. Producers that economically benefit from such premiums could direct funds to improve working conditions. Although existing literature suggests that there is limited evidence that seafood producers economically benefit from eco-label price premiums (22–25), some studies indicate that fair-trade programs have resulted in economic benefit for producers (26). Given the mixed results of such programs, preferential market access for vessels that are not engaged in forced labor practices may be a more powerful incentive for producers.

The potential applications of this model should be viewed within a broader context of policies aimed at addressing forced labor and very poor working conditions. Targeted end-of-pipe vessel-level interventions should not act in isolation, nor should they redirect attention away from addressing the underlying structural drivers of forced labor in fisheries. The processes that lead to labor abuses are complex (27), and the eradication of forced labor in fisheries will also require policies that address poverty, depleted fish stocks, and disenfranchisement of vulnerable populations such as migrant workers. Policies that aim to rebuild fish stocks and reduce subsidies could reduce the demand for cheap labor. Interventions should address challenges that migrant workers face, including the lack of access to formal credit, educational opportunities, social programs, and alternative economic opportunities (28). Governments should also provide legal labor rights protection for migrant workers who sometimes do not enjoy the same protections as domestic workers (29). By addressing underlying drivers, the risk of shifting vulnerable workers away from fisheries and into other high-risk sectors such as agriculture could also be mitigated. The appropriate vessel-level intervention will also depend on whether or not crew are actually working involuntarily. While ILO forced labor indicator presence does imply the use of forced labor, it does not guarantee involuntary or forced labor since crew may choose to work in poor conditions (2, 27). Indeed, a spectrum of human rights violations ranging from poor working conditions to servitude have been documented in the seafood industry (6, 27). In cases with forced labor violations, responses should be severe and in accordance with international policy; in cases when crew are working voluntarily but in abusive conditions, responses should aim to improve working conditions. These interventions should complement, but not substitute for, on-the-ground initiatives that provide victim assistance and promote ethical recruitment practices, investigative journalism, and on-vessel worker-voice monitoring that provides technology for crew to report real-time working conditions

(30). Leveraging multiple approaches can triangulate information and provide more effective interventions.

We emphasize that that this work should be seen as an initial proof of concept, and that model predictions must be used cautiously given the ethical and practical consequences of acting against certain vessels or fisheries, especially with the unavoidable presence of both false negatives and false positives. Because the training dataset of positive vessels is based on a limited number of documented forced labor cases in the fisheries that have received the most attention, the sample is not random and may not fully represent the range of vessel types that use forced labor, resulting in sample selection bias. The model may therefore understate the relative risk among vessels with characteristics underrepresented in the training set of positive cases, but also overstate relative risk among vessels with characteristics overrepresented in the training set of positive cases. Nevertheless, while there are inherent tradeoffs when using machine learning to classify risk, biases also exist in risk-detection systems based on expert knowledge and judgment (31). Finally, vessels may change their use of forced labor, and consequently their behavior, from year to year due to a number of factors. These factors may include changes in labor policy and enforcement, changes in vessel ownership or supply-chain oversight, supply of migrant laborers from marginalized areas, market conditions and demand for targeted species, condition of the underlying targeted fish stocks, and evolving *modus operandi* of the transnational organized crime networks that often support forced labor (32). Forced labor should therefore be viewed as a dynamic challenge that is constantly evolving, and which may cause the risk classification of vessels to change from year to year. Assessing vessel-level risk beyond this study's analysis period of 2012 to 2018 would ideally involve retraining the model using more recently reported cases of forced labor and generating updated classifications based on more recent vessel-level behavior data. While we hope that the initial results presented here can inform the broader discussion around forced labor in fisheries and perhaps even inform more targeted intervention design, we stress that ongoing work will be needed to continuously update, validate, and improve the model with better data in order to make it a more actionable tool for practitioners.

Our approach contributes to an emerging literature that uses remote sensing to shed light on social and human rights challenges. Remote sensing has been used to detect forced labor in other sectors, but that literature uses satellite imagery of static infrastructure such as brick kilns and fish processing plants known to be associated with forced labor (33, 34). Satellite imagery has also been used to map rural populations in marginalized communities (35) and detect poverty by using nighttime lighting as an indicator for household wealth (36). We complement this important work by detecting the dynamic behavior of individual fishing vessels induced by forced labor abuses. Finally, we posit that, if forced labor in fisheries can be detected from satellites, perhaps other forms of human rights abuses induce observable behavior that can also be remotely sensed.

Materials and Methods

Data Description. We train the predictive model using vessel monitoring data from GFW (10). For 16,261 unique longliner, trawler, and squid jigger vessels, we calculate a number of features on an annual basis from 2012 to 2018 (*SI Appendix, Table S1* and *SI Appendix*). 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 66,336 vessel-years of observation. We limit the analysis to these three gear types because they are the only gears for which we have known or highly suspected cases of forced labor for vessels that carry AIS, could be matched to GFW data, and broadcast sufficient and reliable AIS positions. We exclude vessels that

broadcast less than 100 AIS messages per year, appear to be offsetting their vessel position latitude/longitude coordinates, broadcast multiple names per hour, or broadcast more than 95% of their messages from within the Taiwanese, Japanese, Chinese, Republic of Korea, or Democratic People's Republic of Korea exclusive economic zones (EEZs). The high level of vessel congestion within these EEZs greatly reduces AIS coverage and limits our ability to calculate features for these locations (10). The full training dataset is provided in [Dataset S1 \(SI Appendix\)](#).

To identify which vessel-years should be labeled as positive for using forced labor, we developed a comprehensive database of 193 reported cases of forced labor that occurred on specific fishing or refrigerated fish cargo ("reefer") vessels. We define a reported case of forced labor as a situation that displays one or more of the ILO forced labor indicators ([SI Appendix, Fig. S1](#) and [SI Appendix](#)). In addition, these cases must also have one or more of the following case features: 1) eyewitness account, 2) nonofficial investigation (e.g., by an NGO through investigative journalism), 3) official investigation (e.g., by a government enforcement body), 4) arrests made, 5) charges filed, 6) conviction made, or 7) penalties sanctioned. Recognizing that the presence of ILO forced labor indicators does not guarantee involuntary or forced labor, and also recognizing that cases often simultaneously exhibit a number of different forced labor indicators, we endeavor only to detect forced labor broadly as specified by any of the 11 ILO indicators of forced labor, and do not distinguish between whether vessels may be using bonded labor or slave labor or which of the 11 indicators a particular vessel may be exhibiting. We found reported cases through extensive gray and scientific literature review, and also through discussions with NGOs including Liberty Shared, Environmental Justice Foundation, and Greenpeace Asia. For each case, we collect information on vessel identity including vessel name, Maritime Mobile Service Identity (MMSI) number, IMO number, call sign, information on when the vessel was thought to be conducting that behavior, and the source for the information ([Dataset S2](#) and [SI Appendix](#)).

The three most commonly observed ILO indicators relate to servitude labor: 1) abusive working and living conditions, 2) restriction of movement, and 3) isolation ([SI Appendix, Fig. S1](#) and [SI Appendix](#)). Debt bondage was only reported in one vessel case. While the information provided in case reports is likely not comprehensive of the conditions aboard these vessels, these indicators suggest that the model is primarily trained on forced labor cases indicative of servitude rather than bonded labor. Additionally, only four cases reported that convictions had been made and penalties sanctioned. This may reflect insufficient reporting of case information in publicly available reports, but may also reflect insufficient policy and legal institutions for prosecuting cases of forced labor in fisheries. Either of these insufficiencies would indicate a lack of appropriate deterrence for preventing vessels from using forced labor in fisheries. According to AIS data, the MMSI numbers associated with several of these vessels continued to operate in years following when they were labeled as positive ([SI Appendix, Fig. S1](#) and [SI Appendix](#)). 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. While this indicates these vessels were not taken out of commission following forced labor reports, it remains unclear if this is because sanctions were not imposed or if the vessel simply changed ownership in order to continue operations.

We label vessel-years as positive if the vessel is contained in our database of reported forced labor cases and if the year is the single year prior to when the case was reported, yielding $n = 21$ unique vessel cases. Since most case reports do not specify the time period during which abuses took place, we assume the abuses took place in the year prior to the report. While we will refer to this as our baseline model variation, this is an assumption we also test through a robustness check, where we vary our assumption to be that vessels should be labeled as positive in the 2 y prior to the report, in the 3 y prior, etc. We detail this robustness check below. For vessels included in the forced labor vessel database, any year not labeled as positive is excluded from the training dataset since we hypothesize these vessel-years have a higher chance of existing than other vessels, although we are uncertain due to the dynamic nature of forced labor on board fishing vessels ([SI Appendix, Fig. S1](#) and [SI Appendix](#)). We later use the model to make risk predictions for these excluded vessel-years. Vessels for which we have no information and which are not included in the forced labor database are unlabeled, meaning they could be free of forced labor or could have forced labor that has not yet been detected ([SI Appendix, Fig. S2](#) and [SI Appendix](#)). These labels allow us to perform PU learning using the training dataset (described in the following section). To apply positive labels, we matched our database of forced labor vessel cases to the GFW training dataset using MMSI number, IMO

number, call sign, and/or vessel name, and for the year prior to which the case was reported. In cases where the only vessel identification information that could be matched was vessel name, we conservatively disregard matches that use common vessel names, including Viking, Lucky Star, and Greenstar. [SI Appendix, Fig. S2](#), summarizes the number of labeled and unlabeled vessel-years by fishing gear and year and positive training data set label year assumption. A total of 35 reefer vessels were also matched and were used to generate a model feature that describes the number of suspected transshipment events that a particular vessel had with other vessels in the forced labor database.

Importantly, since these forced labor cases exhibit varying levels of evidence and occurred within various areas of legal jurisdiction, we are not implicating these vessels with any specific crimes or actions. Rather, we are labeling these vessels as high-risk vessels that warrant further scrutiny according to the ILO forced labor indicators. Case information was often sparse and did not usually indicate whether labor was involuntary or not, so we endeavored to consistently capture which ILO forced labor indicators were present. Vessel identification information sometimes included MMSI number, IMO number, or call sign, but often just included vessel name. In these cases, we searched for names in online databases such as [MarineTraffic.com](#) and endeavored to find matching vessels from the same flag as was reported in the case. Since this may not always provide a perfect match to the vessel in question, we again are not implicating these vessels with any specific crimes or actions. We also acknowledge that there is an inherent time lag between when a vessel may be using forced labor, when that vessel gets caught or when witnesses emerge, and when the case is reported to the public. This means that we may be observing fewer cases in recent years, which could lead the model to underpredict risk in these recent years. Given the sparsity of case information and the difficulty in matching vessels, the authors call for increased transparency for publishing detailed forced labor case reports and for increased use of AIS devices, MMSI numbers, and IMO numbers.

The data for each model feature come from one of three sources: 1) directly observed using AIS data; 2) inferred using the GFW fishing classification algorithm, which classifies individual AIS messages by gear type and labels them as either fishing or not fishing; or 3) vessel characteristics that come from either a known vessel registry (where available) or from the GFW vessel characterization algorithm (where vessel registry information is not available). The GFW fishing classification algorithm determines which gear type a fishing vessel uses and whether or not that vessel is fishing. For longliners and trawlers, the algorithm uses a convolutional neural network that has classification F1 scores, which combine precision and recall, of 0.93 for drifting longlines and 0.96 for trawlers (10). For squid jiggers, the algorithm uses a heuristic that labels vessels as fishing for squid if the vessel is more than 10 nautical miles from shore and moving at less than 1.5 kn for more than 4 h at night. This heuristic is reliable since squid jiggers have a distinctive behavior in which they fish only at night and only while moving very slowly (37). The GFW vessel characterization algorithm is a separate convolutional neural network that is trained using data from known vessel registries and predicts vessel length ($R^2 = 0.9$ across all gear types), engine power ($R^2 = 0.83$), gross tonnage ($R^2 = 0.77$), and crew size ($R^2 = 0.73$) (10). 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; the remainder of the parameters not contained within registries are obtained from the GFW vessel characterization algorithm. We include vessel characteristics as model features in our baseline model variation, although we test the sensitivity of our results to the inclusion of these features in a robustness check detailed below.

Model Development and Testing. We use machine learning to develop a predictive model that discriminates between fishing vessels that are at high risk of using forced labor from those vessels that are not. We consider a number of predictive features that measure vessel behavior (e.g., average duration of voyages, average number of hours spent fishing per day, and number of suspected transshipment events with other vessels) as well as static vessel characteristics (e.g., vessel flag, gear type, and engine power; [SI Appendix, Table S1](#), and [SI Appendix](#)). The ideal training dataset to build this model would include a list of vessels that are known to have used forced labor and a list of vessels that are known to have not used forced labor, with both lists being randomly sampled from these two classes of vessels. However, although we compiled a list of vessels that were reported to use forced labor (i.e., "positive" cases), we do not have a list of vessels that we know did not use forced labor (i.e., "negative" cases). Upon discussion with a number of human rights experts, there is not currently a certification or transparency

scheme that can reliably guarantee that a specific vessel is not using forced labor. Even with existing schemes, vessels often self-select into these schemes and are therefore not necessarily representative of the larger fleet of vessels. Therefore, any vessel that was not reported to use forced labor is either truly clean of forced labor or simply was not caught (i.e., unlabeled cases). Our analysis therefore is an example of PU classification learning (12). If we had both positive and negative vessels in the training dataset, we would not need to use a PU learning approach and could instead use traditional supervised learning methods. More traditional supervised learning methods may more accurately be able to discriminate between positive and negative vessels because these methods could leverage behavior from known negative vessels, rather than relying on behavior from vessels that are unlabeled and thus could either be negative or positive (e.g., ref. 28). In future research, the model could potentially be improved by incorporating confirmed negative cases that could be obtained from improved social responsibility certification schemes, or from randomly sampled in-person vessel inspections. These types of inspections could also increase our number of positive training cases. However, these types of inspections would be expensive and logistically challenging to conduct, and could provide risks to both those conducting the inspections as well as to the crew aboard inspected vessels.

PU learning has a relatively nascent literature, but there are a number of methods for dealing with this problem that include 1) training a traditional classification model using positive cases and assuming unlabeled cases are negatives; 2) biased learning, where unlabeled cases are treated as negative cases with an unknown amount of class-label noise; and 3) training a traditional classification model using positive cases and assuming unlabeled cases are negatives, and then adjusting the predictions by a constant factor using the known true conditional probability of being positive (38). Because we do not know the true fraction of positive vessels, we focus on the first two of these PU approaches.

The first approach, building a traditional naïve classifier, is the most straightforward, but also leads to biased predictions. Although the probability of being positive should theoretically be biased by a constant factor across all cases and the relative probabilities across cases should be ranked correctly, this only applies if the observed positive cases were chosen at random from all positive cases (38). We know that this assumption does not hold in our case.

The second approach, biased learning, aims to address the class label noise due to unlabeled positive cases. One biased learned technique leverages bagging (bootstrap resampling with replacement), whereby many models are trained using all positive cases and a down-sampled number of unlabeled cases. Predictions are then calculated as the average across model runs (i.e., bags) (39). This reduces the relative importance of unlabeled cases that should actually be labeled positive, and has been shown to perform better than training a traditional classifier (40). Each iteration of the model leverages a base classifier that can be any supervised learning method (12). Much of the PU literature uses support vector machines (SVMs) as the base classifier, although random forests have been shown to perform as well as SVMs or better in some cases (41).

In order to determine the best method for our application, we test four model variations: 1) traditional naïve classifier using an SVM, 2) classifier trained using biased learning with SVM base classifier and the bagging approach, 3) traditional naïve classifier using a random forest, and 4) classifier trained using biased learning with random forest base classifier and the bagging approach.

For each random forest variation, we use the ranger R package and specify 1,000 trees and use the default values for the remaining hyperparameters (42). For each SVM model variation, we use the kernlab R package with a radial basis function and the default hyperparameter values (43). For variations that use the bagging technique, we build up to 100 different classifiers that are trained using up to 100 bags of the data that each have all positive cases and a random down-sampled subset of unlabeled cases resampled with replacement. The number of unlabeled cases included in each bag is a hyperparameter that is tuned (referred to as the down-sampling ratio) and varies from the number of positive cases up to five times the number of positive cases. We calculate the average model score across all bag classifiers and use this as the final model score.

Given our extremely limited training dataset of known positive cases, we do not have sufficient data to reserve a completely separate testing portion of the dataset for model validation. Instead, we use a 10-fold cross-validation (CV) analysis to tune hyperparameters for these four different model variations and to evaluate the performance of these variations to choose the best model. Importantly, to avoid data leakage, we do not split forced labor vessels or media sources across analysis and assessment folds. We suspect that all forced labor vessels described by a particular media source may

behave similarly, so we do not want to train and test the model on these similarly behaving vessels. Similarly, individual vessels may behave similarly across separate years, so we do not want to train and test the model using the same vessel across multiple years.

During CV, we optimize three hyperparameters simultaneously: 1) the threshold for determining the cutoff used to classify a vessel as a high risk; 2) for model variations with bagging, we tune the number of unlabeled vessels that are used in each bag, which is a down-sampling ratio of the number of positive vessels; and, 3) for model variations with bagging, we tune the number of bags. Hyperparameters are tuned by maximizing the mean of a modified F1 score across folds while minimizing the SD and aiming for model simplicity. The modified F1 score has been shown to be an appropriate model performance metric in the PU setting (44). It incorporates recall (the fraction of positive vessel-years that are correctly identified as positive) and detection prevalence (the percentage of vessel-years that are labeled as positives by the model), and can be calculated using only observed known positives and is thus appropriate for the PU learning environment. It is defined as recall squared divided by detection prevalence, and is proportional to the square of the geometric mean of recall and precision (precision is the fraction of correctly identified positives out of all identified positives) (45). The score therefore equally weights the importance of minimizing both type 1 and type 2 errors. We calculate the averages and SDs for recall, detection prevalence, and modified F1 score across folds for all model variations. Since we assess these metrics using CV, this gives us a sense of how the model will perform when predicting out of sample. We do not look at other traditional model performance metrics such as precision, area under the precision–recall curve, or area under the receiver operating characteristic curve, since these are known to be biased in the PU setting (46).

Within the model building procedure, we include a data-preprocessing step that is done both within the CV procedure and within the final model building procedure. During CV, the preprocessing is always completed using only the analysis data, and not using the assessment data, in order to avoid data leakage. The preprocessing is defined as follows: 1) impute missing numeric values using K nearest neighbors based on Gower's distance, with $k = 5$ neighbors, and imputing based on the gear, flag, and vessel length predictors (47, 48); 2) transform numeric predictors using Box–Cox transformation (49); 3) create dummy variables for all categorical predictors; 4) remove numeric predictors that have zero variance or are highly sparse (i.e., the number of unique values divided by the total number of samples is less than 10) and unbalanced (i.e., the frequency of the most prevalent value is higher than 19 times the frequency of the second most prevalent value); 5) remove numeric predictors that have a correlation with other predictors greater than 0.75; 6) center numeric predictors to have a mean of 0; and 7) scale numeric predictors to have an SD of 1. The preprocessed data for the full model training dataset are summarized in Fig. 1.

SI Appendix, Fig. S3, shows the recall, detection prevalence, and modified F1 score and recall of the various model variations during 10-fold CV and using the optimized threshold. We see that the base classifiers without bagging show large SDs for the modified F1 score, and so we eliminate these models as candidates. Next, we see that model performance stabilizes at 50 bags and above, although the mean and SD with 100 bags are the most stable across down-sampling ratios, with similar mean performance between both random forest and SVM. To maximize model simplicity, we therefore set the down-sampling ratio to 1. For 100 bags and a down-sampling ratio of 1, the random forest has the highest modified F1 score. We therefore select our optimized model variation to be random forest with 100 bags and a down-sampling 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 the optimized model building procedure determined using CV, we then train the final model using the entire training dataset and the optimized hyperparameters. *SI Appendix, Fig. S4*, summarizes the number of positive and negative classifications predicted by this model, broken apart by whether the original training dataset label was positive or unlabeled. Using the baseline model assumption, the model correctly classifies 20 of 21 positive vessel-years and also identifies 12,000 new high-risk vessel-years that were previously unlabeled. *SI Appendix, Fig. S5*, shows the feature importance, averaged across bags, for the final model. For each bag, the feature importance for that random forest classifier is the unbiased corrected Gini index, a measure of how well individual features do at reducing node impurity when they are used as the splitting feature in a decision tree. Importantly, feature importance is relative and does not provide information about the directionality of each feature's ability to accurately identify risk.

Robustness Checks. In order to test the sensitivity of our results, we perform two robustness checks: 1) given that vessel characteristics in the GFW database are often inferred rather than directly known from vessel registries, we run model variations that do not use any vessel characteristics as model features; 2) given that, for vessels with reported forced labor, it is usually unknown exactly which years forced labor may have been occurring prior to the report, we run model variations that make different assumptions about how many years prior to the report to label as positive. For each robustness check, we replicate the entire model training and prediction process as outlined above. We again use CV to tune the key hyperparameter of the cutoff threshold for determining which vessels should be labeled as high-risk. We focus on using the random forest model variation with 100 bags and an undersampling ratio of 1 in order to directly compare the results from these robustness checks with those from our baseline model assumptions.

For the first robustness check that tests the sensitivity of our results to the use of vessel characteristics as model features, we run variations of the model that omit the model features of engine power, tonnage, vessel length, crew size, and AIS device type. The model variation that includes vessel characteristics as model features is our baseline model variation. For the second robustness check, we vary the number of years in the training dataset that are labeled as positive for the vessels that are reported to have used forced labor. A value of 1 is the baseline model variation and means that only the year prior to a vessel being reported for forced labor is labeled as positive, a value of 2 means that the 2 y prior to being reported are labeled as positive, etc. Across the range of robustness checks we have 23 unique positive vessel cases, an increase from 21 unique cases in the baseline variation, since some vessels were not observed in the year prior to being reported (case information for all 23 cases vessels is provided in *SI Appendix, Fig. S1*, and *SI Appendix*). The CV model performance results are shown in *SI Appendix, Fig. S6*, which summarizes detection prevalence, modified F1 score, and recall across all model variations including the baseline variation. The predicted final model results are shown in *SI Appendix, Fig. S7*, which summarizes the sensitivity of our main results to different model variations: the fraction of correctly identified true positives, fraction of vessel-years identified as positives, fraction of vessels identified as positives, number of crew working on positive vessels, number of vessel-years identified as positives, and the number of vessels identified as positives across all model variations include the

baseline variation. Model variations that leverage vessel characteristics, even though these characteristics are inferred for many vessels, consistently have higher modified F1 scores than variations that do not use these characteristics. There is not a clear pattern for which year assumption yields the highest F1 score, although detection prevalence declines as the year assumption increases. As the year assumption increases and more positive vessel-years are used in model training, this may impose a more restrictive constraint on what similar unlabeled vessel-years should look like, which may decrease detection prevalence. Therefore, for the results presented in the *Results* and also in Fig. 2, we present the minimum and maximum value ranges from *SI Appendix, Fig. S7*, using model variations that include vessel characteristics and all year assumption model variations. We also present results using only two significant digits.

Given the sensitivity of our results to the assumption of which vessel-years to label as positive, we encourage reports that detail forced labor cases to provide as much information as possible regarding when suspected violations were taking place. We also encourage the use of more detailed public vessel registries that can provide known vessel length, engine power, tonnage, and crew size in order to further refine model features.

Data Availability. All CSV files necessary to reproduce this analysis are found in the supporting information. All CSV files and R code necessary to reproduce this analysis are available in GitHub (<https://github.com/emlab-ucsb/slavery-in-fisheries>) and Zenodo (DOI: [10.5281/zenodo.3635980](https://doi.org/10.5281/zenodo.3635980)).

ACKNOWLEDGMENTS. We thank the following organizations and individuals who provided feedback on our approach, insight into which features to include in the model, and information on publicly available cases of forced labor violations: Conservation International, Environmental Justice Foundation, Greenpeace Asia, Issara Institute, Katrina Nakamura, Liberty Shared, and Ian Urbina. We thank Grant McDermott for his original vision for this project, Diego Undurraga and Rachel Kenny for compiling cases of documented labor abuses, and Nate Miller, Dan Ovando, Lee Qi, and Brian Sullivan for providing thoughtful feedback. We thank the editor and two anonymous reviewers for their helpful comments on previous versions of the manuscript. We gratefully acknowledge Walmart Foundation for funding this research and Google for providing computational support.

1. International Labour Organization, Convention C029–Forced Labour Convention, 1930 (No. 29). https://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0:NO::P12100_ILO_CODE:C029. Accessed 26 November 2019.
2. International Labour Organization, ILO indicators of Forced Labour. https://www.ilo.org/global/topics/forced-labour/publications/WCMS_203832/lang-en/index.htm. Accessed 26 October 2019.
3. Environmental Justice Foundation, “Pirates and slaves: How overfishing in Thailand fuels human trafficking and the plundering of our oceans” (Environmental Justice Foundation, 2015).
4. I. Urbina, ‘Sea slaves’: The human misery that feeds pets and livestock. *NY Times*, 21 June 2015. <https://www.nytimes.com/2015/07/27/world/outlaw-ocean-thailand-fishing-sea-slaves-pets.html>. Accessed 4 December 2020.
5. Environmental Justice Foundation, “Blood and water: Human rights abuse in the global seafood industry” (Environmental Justice Foundation, 2019).
6. J. N. Kittinger et al., Committing to socially responsible seafood. *Science* **356**, 912–913 (2017).
7. International Labour Organization, “Global estimates of modern slavery: Forced labour and forced marriage” (ILO, 2017). (May 30, 2019).
8. The Minder Foundation, “Global slavery index 2018: Fishing” (Minder Foundation, 2018).
9. D. Tickler et al., Modern slavery and the race to fish. *Nat. Commun.* **9**, 4643 (2018).
10. D. A. Kroodma et al., Tracking the global footprint of fisheries. *Science* **359**, 904–908 (2018).
11. Oceana, Illegal fishing and human rights abuses at sea. *Oceana USA*, 13 June 2019. <https://usa.oceana.org/publications/reports/illegal-fishing-and-human-rights-abuses-sea>. Accessed 4 December 2020.
12. J. Bekker, J. Davis, Learning from positive and unlabeled data: A survey. arXiv:1811.04820 (21 October 2019).
13. E. Sala et al., The economics of fishing the high seas. *Sci. Adv.* **4**, eaat2504 (2018).
14. Greenpeace, “Misery at sea: Human suffering in Taiwan’s distant water fishing fleets” (Greenpeace, 2018).
15. Environmental Justice Foundation, Out of the shadows. <https://ejfoundation.org/reports/out-of-the-shadows-improving-transparency-in-global-fisheries-to-stop-illegal-unreported-and-unregulated-fishing>. Accessed 22 October 2019.
16. International Maritime Organization, Cape Town Agreement Of 2012, 2018 Edition (IMO, 2018).
17. International Labour Organization, Convention C188–Work in Fishing Convention, 2007 (No. 188). https://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0:NO::P12100_ILO_CODE:C188. Accessed 3 December 2019.
18. United Nations, United Nations Convention on the Law of the Sea (UNCLOS). https://www.un.org/depts/los/convention_agreements/texts/unclos/UNCLOS-TOC.htm. Accessed 3 December 2019.
19. A. J. Ortiz, Agreement on port state measures to prevent, deter and eliminate illegal, unreported and unregulated fishing. *Int. Legal Mater.* **55**, 1157–1179 (2016).
20. K. Nakamura et al., Seeing slavery in seafood supply chains. *Sci. Adv.* **4**, e1701833 (2018).
21. N. A. Miller, A. Roan, T. Hochberg, J. Amos, D. A. Kroodma, Identifying global patterns of transshipment behavior. *Front. Mar. Sci.* **5**, 240 (2018).
22. H. Wakamatsu, The impact of MSC certification on a Japanese certified fishery. *Mar. Resour. Econ.* **29**, 55–67 (2014).
23. J. Blomquist, V. Bartolino, S. Waldo, Price premiums for providing eco-labelled seafood: Evidence from MSC-certified cod in Sweden. *J. Agric. Econ.* **66**, 690–704 (2015).
24. J. Blomquist, V. Bartolino, S. Waldo, Price premiums for eco-labelled seafood: Effects of the MSC certification suspension in the Baltic sea cod fishery. *Eur. Rev. Agric. Econ.* **47**, 50–70.
25. A. Stemle, H. Uchida, C. A. Roheim, Have dockside prices improved after MSC certification? Analysis of multiple fisheries. *Fish. Res.* **182**, 116–123 (2016).
26. A. Blackman, J. Rivera, Producer-level benefits of sustainability certification. *Conserv. Biol.* **25**, 1176–1185 (2011).
27. M. Marschke, P. Vandergest, Slavery scandals: Unpacking labour challenges and policy responses within the off-shore fisheries sector. *Mar. Policy* **68**, 39–46 (2016).
28. S. Kara, *Modern Slavery: A Global Perspective* (Columbia University Press, 2017).
29. A. Shen, Choppy waters–Forced labor and illegal fishing in Taiwan’s distant water fisheries. <https://www.greenpeace.org/usa/wp-content/uploads/2020/03/b87c6229-2020-choppy-waters-en.pdf>. Accessed 8 July 2020.
30. L. R. Taylor, E. Shih, Worker feedback technologies and combatting modern slavery in global supply chains. *J. Br. Acad.* **7**, 131–165 (2019).
31. J. Kleinberg, S. Mullainathan, M. Raghavan, Inherent trade-offs in the fair determination of risk scores. arXiv:1609.05807 (2 December 2019).
32. D. Liddick, The dimensions of a transnational crime problem: The case of IUU fishing. *Trends Organ. Crime* **17**, 290–312 (2014).
33. D. S. Boyd et al., Slavery from space: Demonstrating the role for satellite remote sensing to inform evidence-based action related to UN SDG number 8. *ISPRS J. Photogramm. Remote Sens.* **142**, 380–388 (2018).
34. C. McGoogan, M. Rashid, Satellites reveal “child slave camps” in UNESCO-protected park in Bangladesh. *The Telegraph*, 30 May 2019. <https://www.telegraph.co.uk/technology/2016/10/23/satellites-reveal-child-slave-camps-in-unesco-protected-park-in/>. Accessed 4 December 2020.
35. W. Hu et al., Mapping missing population in rural India: A deep learning approach with satellite imagery. https://dl.acm.org/doi/abs/10.1145/3306618.3314263?casa_token=W0l72zotxTMAAAAA-M-OavSjzfigaK_x-m5AmG6r5i3kX6bWckg94A1wVwXP6iQagD77x-TqrJrBYQ99H6MuQU5j375XWcA. Accessed 22 October 2019.
36. N. Jean et al., Combining satellite imagery and machine learning to predict poverty. *Science* **353**, 790–794 (2016).

37. D. Kroodsma, Data from "Updated algorithm for squid jiggers." GitHub. https://github.com/GlobalFishingWatch/paper-global-footprint-of-fisheries/blob/master/data_production/updated-algorithm-for-squid-jiggers.md. Accessed 9 October 2020.
38. C. Elkan, K. Noto, "Learning classifiers from only positive and unlabeled data" in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08* (Association for Computing Machinery, New York, NY, 2008), pp. 213–220.
39. F. Mordelet, J.-P. Vert, A bagging SVM to learn from positive and unlabeled examples. *Pattern Recognit. Lett.* **37**, 201–209 (2014).
40. M. Claesen, F. De Smet, J. A. K. Suykens, B. De Moor, A robust ensemble approach to learn from positive and unlabeled data using SVM base models. *Neurocomput.* **160**, 73–84 (2015).
41. D. Pancaroglu, M. Tan, "Improving positive unlabeled learning algorithms for protein interaction prediction" in *8th International Conference on Practical Applications of Computational Biology & Bioinformatics (PACBB 2014): Advances in Intelligent Systems and Computing*, J. Saez-Rodriguez, M. P. Rocha, F. Fdez-Riverola, J. F. De Paz Santana, Eds. (Springer International Publishing, 2014), pp. 81–88.
42. M. N. Wright, A. Ziegler, Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **77**, 1–17 (2017).
43. A. Karatzoglou, A. Smola, K. Hornik, A. Zeileis, Kernlab—An S4 package for kernel methods in R. *J. Stat. Softw.* **11**, 1–20 (2004).
44. P. Norlin, V. Paulsrud, Identifying new customers using machine learning: A case study on B2B-sales in the Swedish IT consulting sector. <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1118107&dsid=2134>. Accessed 22 October 2019.
45. W. S. Lee, B. Liu, "Learning with positive and unlabeled examples using weighted logistic regression" in *Proceedings of the Twentieth International Conference on Machine Learning*, pp. 448–455 (AAAI Press, Menlo Park, CA, 2003).
46. S. Jain, M. White, P. Radivojac, Recovering true classifier performance in positive-unlabeled learning. arXiv:1702.00518 (21 October 2019).
47. T. M. Cover, P. E. Hart, Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **13**, 21–27 (1967).
48. J. C. Gower, A general coefficient of similarity and some of its properties. *Biometrics* **27**, 857 (1971).
49. R. M. Sakia, The box-cox transformation technique: A review. *Statistician* **41**, 169–178 (1992).
50. S. Nembrini, I. R. König, M. N. Wright, The revival of the Gini importance? *Bioinformatics* **34**, 3711–3718 (2018).