# Using Remote Sensing Big Data to Combat Forced Labor Journal Article by Harry Newton, PhD

## Abstract

The research presented starts with modeling fishing ship behavior based on navigation signals received by satellites. Then, a risk framework is offered for which ships have indicators that are aligned with risk factors used in Human Rights research on Forced Labor practices. Next, imagery data is added to the sources to provide a way to start with imagery that indicates a problem and supplement with signals data, or vice versa. Future directions suggest additional remote sensing data that can be added to the model, a graph theoretic approach to adding non-remote sensing data, and a mobile app for data collection and results that could benefit local authorities and allow them to tag ships to track or inspect the next time they are in a port.

## Introduction

There are many complicating factors to investigating forced labor human rights violations in the Fishing Industry, including differences in maritime laws between countries and that fishing vessels often operate in waters other than their country of registration (known as their “flag”). These differences lead to a legal morass on how to interpret the laws. Additionally, most of the fishing is done at sea, far away from any attentive eyes & ears, except maybe through satellites, which is the angle that this paper takes.

In this paper, the satellite “eyes” brought to bear are from the Sentinel-2 satellite operated by the European Space Agency. The satellite “ears” are operated by Spire to detect the Automatic Identification System (AIS) emitted by the navigation system of most large ships. Taken together, satellite observations regardless of the sensor type are often referred to as Remote Sensing. Many more data sources exist and each one added can add additional insights of the operations of maritime vessels and fleets, thanks to the prevalence of Big Data platforms and the maturity of the Geographical Information Systems.

## Problem Statement

* Develop a tractable Big Data approach to aggregate millions of signals from ships to discern behavior across years of data and compare with typical operations of the same type of ship using Data Visualization
* Extend the data sources to include imagery data which can corroborate and supplement the signals data.

This paper first summarizes related literature (section 1) then develops an AIS Analytic Model for ship behavior across fishing seasons (section 2) and adds a Data Visualization (DV) of the results (section 3). Next, the advantages of adding imagery is discussed and added to the model (section 5). Conclusion and future directions follow (section 5).

Contributions of this paper: A big data model for analysis of ship voyages based on remote sensing of images and signals to detect risk factors for Human Trafficking.

## Literature Review

The research presented in this paper builds on recent Big Data analysis undertaken on data on ship registrations, satellite data, and national fishing activity databases, as described in this section and documented in the References.

(Park et al., 2020) provides a comprehensive data analytic approach to understanding the fishing industry of China, North Korea, and Russia. They built the analytic on top of an impressive data repository by the Global Fishing Watch, an international, nonprofit organization with the purpose of understanding and protecting fishing resources. This work leverages several different satellite data feeds that are part of the GFW data repository, in order to estimate the size of the “Dark Fleet” of fishing vessels that do not transmit navigational signals.

Research on Human Trafficking and Forced Labor in the fishing industry literature includes (Bonfanti & Bordignon, 2017) which was first published as a series of articles by the Associate Press under the heading “Seafood from Slaves” for which they won the Pulitzer Prize for Public Service on April 18, 2016. The articles and actions generated by them freed 2,000 crew members which local authorities ruled were Forced Labor victims, according to the AP Press Release that day. This work also highlights the difficulty countries have when trying in enforce laws that should pertain to their sovereign waters but often the violations are by vessels “flagged” by other nations.

The legal difficulties of enforcing fishing laws are further described in (Rowlands, Brown, Soule, Boluda, & Rogers, 2019), who note cases where if one country cracks down, the offending ships simply move outside of that country’s jurisdiction.

To further study how ships with Forced Labor crews could escape notice, (Miller, Roan, Hochberg, Amos, & Kroodsma, 2018) undertook research which demonstrated that Transshipment vessels are sometimes used to transfer a fish catch from a vessel and provide resupplies. Their work generated a global map showing these likely transhipments and analytic insights on the countries of registration for both the fishing vessel and the transport ship, as well as the next port visited to presumably offload the cargo. Many of these transhipments were in or near areas with fishing restrictions in place; therefore are conjectured to also be likely violations of the fishing laws. There is even conjecture that sometimes “slaves” are part of the resupply. The witness accounts in the “Seafood from Slaves” articles bear allegations that many of the forced labor crew died at sea which would give credence to the need for new crew members. This conjecture was suggested by a U.S. government official who had been responsible for U.S. efforts on human trafficking in an interview.

(Kroodsma et al., 2018) represents one of the most thorough studies of the fishing industry conducted with remote sensing. In it 22 billion AIS messages from 70,000 industrial fishing vessels were studied from 2012 to 2016. This study is where the public GFW datasets mentioned above were created. Results show that industrial fishing occurs 55% of the ocean—an area more than four times that of agriculture.

In the next sections, the data sources are summarized and the model evolution begins.

## Data Sources

A number of organizations have worked together to assemble data relevant to the fishing industry. In particular, Google and the main AIS provider, Spire, have partnered with the Global Fishing Watch to provide public datasets that are used in this research.

Table 2 Public Datasets for Remote Sensing

|  |  |
| --- | --- |
| Data | Source |
| 2011-2016 AIS data | Global Fishing Watch (GFW) Public Datasets in Google Cloud, which provides a sample of 28.5M AIS signals and anonymized information on the associated vessels. |
| 2018 AIS data | Global Fishing Watch (GFW) “Dark Fleet” Public Datasets in Google Cloud, which cover individual AIS signals and information on the categories of vessels. |
| Sentinel-2 Satellite Imagery | From Copernicus (European Space Agency) and where not found there, from Google Cloud Public Data |
| Sentinel-2 Index | Google Cloud Public Data |
| AIS Vessel Reports | Via the Spire AIS Vessel API |

## AIS Analytic Model

Each AIS signal includes the timestamp, the vessel ID, and the lat-long coordinates of the vessel, among other fields. By looking up the vessel ID in registration information, the ship type and other statistics can be obtained. There are often additional references which can be consulted to build additional information about the vessel. As shown in Table 1, Global Fishing Watch adds additional data about their analysis on the vessel type and whether the observation is likely associated with a time the vessel is fishing (based on its speed, location, and vessel type).

Modelling Approach:

* Correlate AIS observations for the same unique ship ID into voyages. This can be done based on anchorage datasets but since GFW has already provided the distance to port, a low value on that distance is taken as a port visit marking the beginning of a new voyage.
* From the gaps between AIS signals, note the deltas in the timestamp and compute the distance from the previous signal’s lat-lon position to the current one.
* For each ship, across each voyage it takes during the studied timeperiod, determine the distributions of the voyage lengths, time gaps in signals, and distances between reported locations.
* For each type of ship, determine the nominal behavior for comparison.

## Risk Assessment Model

Table 3 Risk Factors used

|  |  |
| --- | --- |
| Risk Factor | Analysis Approach |
| Long voyages (possibly to avoid inspections or contact in ports) | Based on the ships identification number, separate the AIS observations into voyages between port visits. Determined which of the ships have voyages in the top quartile for their type). |
| Time Gaps between signals (going dark) | Within each voyage computed the time delta between each signal and assessed a risk for ships with average gaps in the top quartile for their vessel type |
| Distance Gaps between signals (misleading locations) | Within each voyage computed the distance delta between each signal and assessed a risk for ships with average gaps in the top quartile for their vessel type |

From the analysis steps above, operating first on the 28.5M AIS signals from 354 ships over five years, a voyage statistics table is created for each voyage, and then a ship statistics table. The model runs on a laptop in about 15 mins to generate ship data and the reference set for comparison of all ships of the same fishing vessel type.

To identify specific ships, these risk factors can be reported out using a scheme such as demonstrated in the Table below, which assigns a risk score equal to the sum of the risk factors based on the behavior of each ship. Only the ships which had all three factors are presented.

Table 4 Ships with behavior on all risk factors

|  |  |  |  |
| --- | --- | --- | --- |
| Ship Index | MMSI | Type | risk-score |
| 4 | 5020143137211 | Trawlers | 3 |
| 33 | 23593130178765 | drifting\_longlines | 3 |
| 81 | 63716389910332 | purse\_seines | 3 |
| 94 | 77261928739173 | Trawlers | 3 |
| 120 | 89110306461984 | drifting\_longlines | 3 |
| 135 | 102111574707951 | drifting\_longlines | 3 |
| 153 | 110014060911622 | Trawlers | 3 |
| 171 | 125954407820672 | Trollers | 3 |
| 271 | 207595838999742 | drifting\_longlines | 3 |
| 284 | 218997104404214 | Trawlers | 3 |

Note that because the dataset from GFW was anonymized the MMSI (Maritime Mobile System Identification) cannot be relied on for checking for country of registration or additional vessel information. Normally, that would be possible and could be used for addition risk factors, such as a change in the country of registration or vessel name or ownership.

# Data Visualization

Using Python, the data for each voyage was aggregated for the risk factors, then a further aggregation was made across the voyages for each ship. The resulting trends can be observed in the Sankey diagram below which shows the pattern, especially the number of ships “guilty” of one risk factor that also have the next risk factor.

## Results

A picture containing screenshot

Description automatically generated

Figure 4 Sankey - risk factor proportions

From this visualization and Table 4 (Ships with behavior on all risk factors), the relatively small number of the 354 ships with all three risk factors becomes more apparent. The future directions at the end of the paper describe a few ways that these results could be used by authorities.

# Joint AIS & Imagery **Model**

By combining signals and imagery data, we can write a model that benefits from both types of data. As a quick summary of these advantages, the imagery is effective at detecting ships regardless of whether they are transmitting navigational signals; however, it relies on clear weather and daylight, so the collection of it is more sparse. On the other hand the signals data does not rely on clear weather and is equally prevalent at night; plus, in addition to detection, provides the identity of the ship from a combination of the signal and lookup tables of ship registries for each country.

Consequently, the signals data (for ships using AIS) can be used to re-establish the identity of ships between images and over long timeperiods of years to establish behaivor. The disadvantage of the AIS data is that some ships may not meet their country of registrations requirements for transmitting AIS (and so may not even have the equipment). Even for ships that do meet their countries requirements, there is little means to enforce that they transmit when outside of their country’s sovereign waters.

This section is motivated by two use cases: 1. Allow starting with a region & timeframe (for example for an image) and finding AIS data to provide vessel locations & identities and 2. The other way, for a selected vessel (perhaps based on results from the AIS model), find imagery which could shed light on it’s rigging configuration (to confirm it vessel type and the vessel information in registry systems) or to note nearby ships not transmitting AIS. In either use case, the two data sources will need to be correlated as explain below.

To extend the model to Joint AIS & Imagery, the time and location fields for both types of data can be used to align the data as shown in Table 4. Note that there are many other time and locations fields available in these data feeds, such as when the satellite received the signal or when the satellite was tasked to take an image.

Table 5 Aligning AIS & Imagery data

|  |  |  |  |
| --- | --- | --- | --- |
| **Data** | **Time** | **Location** | **Notes** |
| AIS | Timestamp signal sent | Lat-lon of ship at signal | Also contains ship identity which can be cross referenced to registration data and used to track vessels over time |
| Imagery | Sensing Time for image | Lat-lon corners of processed image tiles | Ability to detect and count ships does not rely on vessel systems, like AIS |
| Alignment used | AIS with closest timestamp to sensing time and not more than 1 hour different | AIS lat-lon within the polygon of the the image. | See Figure 3 below to see how results can be compared. |

For Use Case 1: Finding AIS data based on an area and time, the steps are:

Step 1 Using the Sentinel-2 index in Google Cloud, filter for near North Korea and less than 10% cloud, and with a sensing time in Dec 2018 (chosen because the constellation was operational then and we had matching AIS data for that month). That gives us the images who’s sensing time is along the x-axis of Figure 2

Step 2 Using the lat-lon box and sensing time of each image, find matching AIS signals. The blue vertical bars are a sum of the ships with one or more AIS messages in the image within 1 hour of the image.

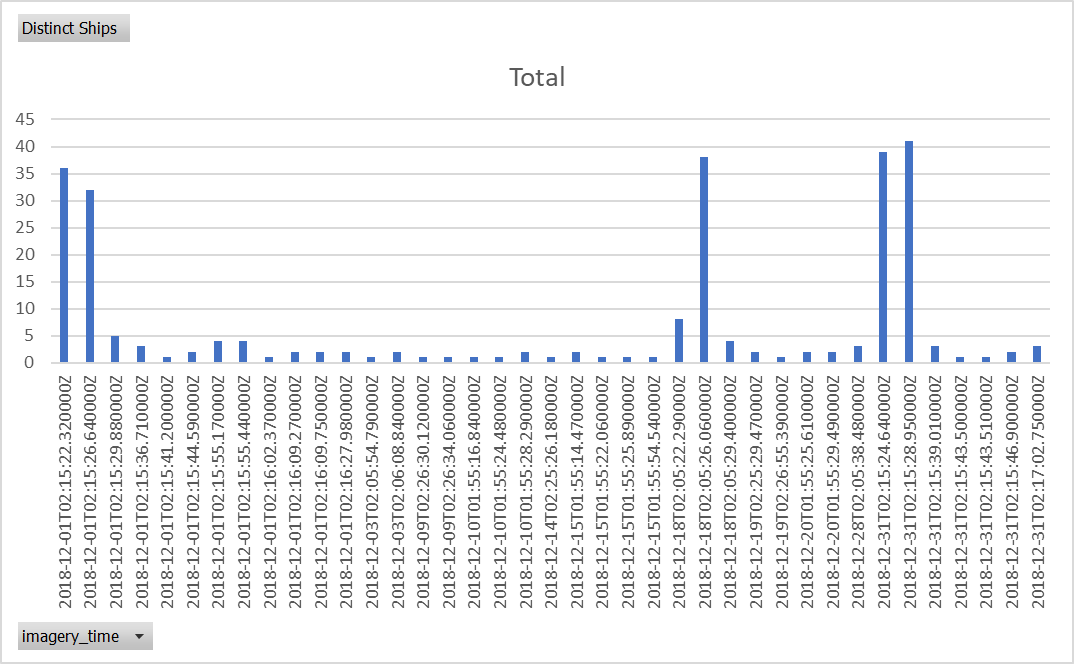


Figure 2 Simultaneous Signals & Imagery data in Dec 2018

These two steps were performed in Google Big Query with python code to implement SQL commands to pull the two datasets and find the intersection of the images and signals. Our code was run on a virtual machine in datalab and the “from” clauses of the SQL statements linked to the ESA provided Sentinel-2 data using an index which contains the information needed for Step 1 which is available as a Public Dataset in Google.

By performing an additional step to retrieve the processed tiles (pictures) from the image (also furnished by ESA and available in Google Cloud Platform, we can contruct Figure 3 showing the image collected with boxes to annotate the ships detected by the Object Detection Model provided in (Keeley, 2020) and red dots to reflect the signals data. In some cases there are multiple red-dots signifying continuous or rapid operations of the AIS. Also note that in Figure 3 that the object detection used is not effective near land where the shape of a ship does not stand out visually from the piers or nearby buildings.

The analysis can be repeated for the remaining images, but even the one image proves the point that the ship detection algorithm is validated by the signals data and vice versa. The other important point to make is that the combination of both imagery and signals means that the area that can be analyzed by this combination of models includes night and bad weather (when the imagery is not effective) and open-ocean where imagery is not typically tasked because most of the image would be water, yielding little information. This is particularly true of the Sentinel-2 satellite since its multi-spectrum sensor package and ability to yield color images is better spent on images where the spectrums can be used to study crop growth and other changes on land.

The results our joint model are shown in the table below for the images with the most ships shown by AIS data (the tallest blue bars in Figure 2).

Figure 3 Combined View of Signals and Imagery

## Conclusions and Future Enhancements

Like a couple of the references, this paper is based on remote sensing data in

Our model has used only two remote sensing data sources (Sentinel-2 imagery and AIS messages). Additional sources are already available, such as imagery from Planet “Doves” constellation which cover the land surface daily; however, it would need to be tasked over open water for many of the fishing applications, which is where it currently recharges its batteries. Similarly, the European Space Agency also offers radar data to the public from the Sentinel-1 pair of satellites. Like AIS, radar imagery is also available all-weather and day-night. However, like many other systems, this one is not typically tasked for open-ocean surveillance. Use of radar is further described in (Rowlands et al., 2019)

Whatever satellites are used, it would be helpful to have a budget for tasking them for precise collection needed, instead of relying on serendipitous collection because of their existing tasking is useful and relying on the data owners to quickly make the data publicly available in a consumable format.

A network approach to the forced labor problem could be undertaken which starts with reports of human smuggling and looks for evidence of the trade route and financial transactions involved (likely on the dark web). A similar approach has been taken for the related Human Rights problem of Sex Trafficking. One example of an approach based on Elastic Search that seems applicable to Forced Labor as well is (Szekely et al., 2015). That approach or others using Natural Language Processing would allow use of the semi-structured Twitter data to also be used.

Since the computational work has to be done in data centers anyway because of the vast size of the datasets, it would be possible to connect the results and additional data collection in a mobile app. QGIS for example, would be a good application to display the findings and collect data since much of the results and the data is geography based.

# References

Bonfanti, A., & Bordignon, M. (2017). ‘Seafood from Slaves’: The Pulitzer Prize in the Light of the UN Guiding Principles on Business and Human Rights. *Global Policy*. https://doi.org/10.1111/1758-5899.12495

Kroodsma, D. A., Mayorga, J., Hochberg, T., Miller, N. A., Boerder, K., Ferretti, F., … Worm, B. (2018). Tracking the global footprint of fisheries. *Science*. https://doi.org/10.1126/science.aao5646

Miller, N. A., Roan, A., Hochberg, T., Amos, J., & Kroodsma, D. A. (2018). Identifying global patterns of transshipment behavior. *Frontiers in Marine Science*. https://doi.org/10.3389/fmars.2018.00240

Park, J., Lee, J., Seto, K., Hochberg, T., Wong, B. A., Miller, N. A., … Kroodsma, D. A. (2020). Illuminating dark fishing fleets in North Korea. *Science Advances*. https://doi.org/10.1126/sciadv.abb1197

Rowlands, G., Brown, J., Soule, B., Boluda, P. T., & Rogers, A. D. (2019). Satellite surveillance of fishing vessel activity in the Ascension Island Exclusive Economic Zone and Marine Protected Area. *Marine Policy*. https://doi.org/10.1016/j.marpol.2018.11.006

Szekely, P., Knoblock, C. A., Slepicka, J., Philpot, A., Singh, A., Yin, C., … Ferreira, L. (2015). Demonstration paper: Using a knowledge graph to combat human trafficking. *CEUR Workshop Proceedings*.

Additional references which are not refereed:

McDowell, R., M.Mendoza, and M. Mason. 2015. AP tracks slave boats to Papua New Guinea available at https://www.ap.org/explore/seafood-from-slaves/ap-tracks-slave-boats-to-papua-new-guinea.html

Greenpeace. 2016. Turn the tide. Human Rights Abuses and Illegal Fishing in Thailand’s Overseas Fishing Industry. <http://www.greenpeace.org/australia/Global/australia/reports/Turn-The-Tide.pdf>

Cutlip, K. 2017. Skytruth Blog. <http://skytruth.org/2017/01/satellites-leave-no-place-to-hide-forrogue-thai-fishing-fleet/>

Investigating how North Korean fishermen plunder foreign waters, *France 24*, a French state-owned international news television network, May 8, 2020 available at <https://www.youtube.com/watch?v=ytbEt8do3r8>