From plague to coronavirus: vessel trajectory data from ship automatic identification systems for epidemic modeling[1,2]

Katherine E. Hoffmann Pham, M. Phil.

Department of Technology, Operations, and Statistics,
NYU Stern School of Business

United Nations Global Pulse

Miguel Luengo-Oroz, Ph.D.

United Nations Global Pulse

[1] United Nations Global Pulse is supported by the Governments of the Netherlands, Sweden, and Germany and by the William and Flora Hewlett Foundation.

[2] Correspondence and reprint requests may be addressed to Miguel Luengo-Oroz (miguel@unglobalpulse.org).

Highlights

In addition to moving people and goods, ships can spread disease. Vessel trajectory data from ship Automatic Identification Systems (AIS) is available online and can be extracted and analyzed, as we illustrate in the case of the current coronavirus epidemic. This data should be included in epidemiological models of disease transmission to complement air traffic data and inform operational responses.

Faced with a global epidemic, there is a need to quickly estimate potential routes for disease transmission in order to prepare a public health response. In recent years, the growing availability of big data sources has facilitated the modeling of these potential routes. For example, previous studies have used mobile phone records, census data, and airline flight matrices to identify possible disease hotspots, predict contagion patterns, and/or estimate import risk.[1] In this research, we highlight another dataset – Automatic Identification System (AIS) data containing historical and real-time ship trajectories – as a complementary source of information on possible transmission routes.

AIS data forms a global database of maritime traffic. As of 2008, all passenger ships, international ships over 300 tons, and cargo ships over 500 tons must be equipped with an AIS transceiver, which reports dynamic information about the ship's position – such as its location, speed, and course over ground – as well as static information such as the ship's identifiers, vessel type, and flag.[2] This data is used to track ship positions, to help avoid collisions, and for general maritime situation awareness. It is collected by a network of land-, sea-, and satellite-based receivers and subsequently aggregated by third-party providers, which redistribute the data and/or render it for viewing in a graphical user interface. AIS data has been used in a diverse range of applications, ranging from detecting fishing behavior to studying search and rescue operations involving migrant and refugee boats.[3-4] However, we are not aware of any epidemiological models of disease transmission which use this data source.

Currently, a standard approach to calculating disease import risk at the international level involves epidemic modeling using flight network data. Such network-based risk models typically use an origin-destination matrix of geographic pairs (e.g., airports), populated with empirical data on real or expected flows between pairs (e.g., the number of passengers traveling between these airports), in order to estimate the spread of disease over this network. However, vessels can also be global carriers for infected individuals and disease vectors. For example, the “Black Death’’ plague outbreak in 1347, which killed an estimated 25 million Europeans, is believed to have been spread throughout Europe by merchant ships, possibly by rats on board and their parasites. Similarly, the travel of infected civilians and soldiers along shipping routes is credited with helping to spread the catastrophic 1918 Spanish flu epidemic, which killed an estimated 20 – 40 million people worldwide. More recently, cargo and passenger ships have been associated with over 100 outbreaks of infectious diseases including measles, chickenpox, and norovirus.[5]

While airlines represent the dominant mode for international travel, ships still carry large volumes of passengers and goods. Ferries serve an estimated 2.1 billion passengers each year, whereas commercial ships move an estimated 11 billion tons of international cargo and cruise ships host an estimated 30 million annual holiday travelers. Ship traffic patterns have transmission potentials that are distinct from airline traffic, and may therefore complement flight-based origin-destination matrices by highlighting additional possible transmission routes or alternative weights between existing routes. Passenger ship traffic may be particularly pertinent in the case of human-hosted respiratory illnesses, whereas cargo ship traffic might be more relevant in the case of vector-borne diseases for which the vectors may be carried along with the cargo.

Although maritime flows have occasionally been used to model the spread of vector-borne disease,[6] they are not typically taken into account in current network-based epidemic risk models. The growing availability of AIS data makes it feasible to incorporate ship traffic into the study of disease spread, and below we highlight two case studies in which AIS data on cargo and/or passenger ship traffic was relevant to the study of disease transmission: the 2017 Madagascar plague and the 2020 coronavirus outbreak.

In Fall 2017 Madagascar was struck by a plague epidemic, affecting an estimated 2,348 individuals and resulting in 202 deaths.[7] Since Madagascar is an island nation, we examined travel patterns in the AIS data to estimate possible transmission routes through commercial and cruise ships. Our methodology consisted of gathering information from ports and cruise ships and constructing an origin-destination matrix that captured all trips between ports inside Madagascar (~ 40% of observed trips) and outside Madagascar (~ 60%).

From 26 September to 26 October 2017 we identified 126 vessel departures from the main ports in Madagascar, including cargo vessels, tankers, tugs, and fishing vessels. Interestingly from an epidemic risk perspective, long-range trips included destinations as diverse as Mayotte, Mauritius, Reunion, Mozambique, France, and China. A further analysis of cruise traffic specifically found 6 cruises that were scheduled to arrive in Madagascar over the course of November 2017. Generally, these cruises were scheduled to stop in more than one port, and the number of passengers and crew members on board was over 1,200 on average. These cruises therefore represent high-volume points of human contact which in this case would have subsequently visited South Africa, Mauritius, Mozambique, Reunion, Seychelles, and Italy.

Given the long incubation period of the coronavirus and the possibility of mild, asymptomatic cases, maritime traffic has played a significant role in the current epidemic. In particular, experience has shown that large cruise ships can become important hotspots due to their confined spaces and the high frequency of contacts between a diverse set of international travelers, with confirmed cases on at least 25 different ships. Below are three scenarios in which AIS information could be used to support further investigations and risk modeling.

- Wuhan, the city at the center of the outbreak, is located at the junction of the Yangtze and Han rivers. Therefore, shipping traffic is one pathway by which disease could have spread out of the city. Using AIS data from Exact Earth, we obtained all traffic from ships that passed within 50km of Wuhan with a speed of less than 2 knots for the period between 8 - 21 January 2020. All told, we identified 1,107 ships connecting 34 different ports across China and Japan (see Figure 1a).

- Spain's first case of the coronavirus was identified in the Canary Islands, in a traveler who traveled by ferry to La Gomera from Tenerife.[8] In addition to the islands' air traffic connections, AIS data analysis for 8 - 21 January 2020 showed that ships passing by the Canary Islands were associated with ports in 56 different countries, including frequent travel to peninsular Spain, Morocco, Senegal, Gibraltar, the UK, the Netherlands, and Portugal (a snapshot of regional traffic is shown in Figure 1b).

- After a disembarking passenger tested positive for the virus, the *Diamond Princess* cruise ship was quarantined off the coast of Japan with approximately 3,700 individuals on board; despite containment efforts which involved removing infected individuals and quarantining passengers, over 619 people on the ship were ultimately infected and at least six died.[9] A second cruise ship, the *MS Westerdam*, was repeatedly denied entry at port for fear that there were infected individuals among the 2,257 people on board, and a passenger tested positive for coronavirus only after the ship docked in Cambodia, leading to fears that other infected but asymptomatic individuals could spread new cases (although the passenger has since tested negative).[10] The locations of both ships are available in real time and their travel paths can be reconstructed using historical AIS traces (see Figure 1c). Similar data on positions, historical travel paths, maximum capacity, and expected ports of arrival could also be extracted for other cruise ships in the region.

Ultimately, ships represent a source of epidemic risk which might be particularly relevant in port cities and island nations, or for diseases with long incubation periods. Our analysis of data from Madagascar and the Canary Islands illustrates that island shipping traffic has a broad reach (in these cases, reaching three or more continents within a month) and may potentially carry high volumes of passengers. Even a landlocked city such as Wuhan saw a considerable volume of river traffic in the two-week period prior to quarantine.

AIS data makes it possible to visualize individual ship trajectories, along with additional ship information such as vessel type and size. It is possible to automatically filter out, for instance, passenger ships that might have stopped in cities with high risk, and to estimate their expected arrival times in future ports. The ability to acquire such information in real time might be of critical relevance for operational responses to disease outbreaks.

The approach described in this communication represents a first step towards estimating the potential role of vessel traffic in global disease spread using AIS data. Ship origin-destination matrices can offer insights into travel routes which may be overlooked when focusing on flight path data alone, and should be considered when modeling the spread of disease. The availability of real-time and historical data from commercial providers makes timely analysis of AIS data feasible. Vessel trajectory data can be further enhanced with other sources of data on the number of passengers and crew aboard tourist or commercial ships. For example, data on ship passenger capacity can be obtained from the vessel information sheets that are available from AIS data providers, and data on cruise line capacity can be accessed from websites such as CruiseMapper and Wikipedia. These passenger capacity estimates provide an upper bound which can be adjusted based on expected occupancy rates; while data from ticket providers is needed to estimate more precise passenger counts, even flight network models often rely on such expected capacity models. Future work should explore how adding shipping traffic to existing flight network data changes the outcome of disease simulations and risk estimates.

Author contributions

M.L.O. conceived the idea. K.H.P. analyzed the data. M.L.O. and K.H.P. wrote the article.

Funding

This work was supported by the Governments of the Netherlands, Sweden, and Germany and by the William and Flora Hewlett Foundation.

Conflict of Interest

The authors have declared no conflicts of interest.

Acknowledgements

We thank Daniela Perrotta for her contributions to the analysis of the Madagascar data. We thank the UN Global Platform for access to its AIS database.

References

1. Tatem AJ. Mapping Population and Pathogen Movements. *Int Health* 2014; 6(1): 5-11.

2. Robards MD, Silber GK, Adams JD, *et al.* Conservation science and policy applications of the marine vessel Automatic Identification System (AIS)—a review. *Bulletin of Marine Science* 2016; 92(1): 75-103.

3. Natale F, Gibin M, Alessandrini A, *et al.* Mapping Fishing Effort through AIS Data. *PLoS ONE* 2015; 10(6).

4. Hoffmann Pham K, Boy J, and Luengo-Oroz M. Data Fusion to Describe and Quantify Search and Rescue Operations in the Mediterranean Sea. *IEEE 5th International Conference on Data Science and Advanced Analytics* 2018; 514–523.

5. Regan JJ, Vega JS, and Brown CM. Infectious illnesses on cruise and cargo ships. In Peterson E, Chen LH, and Schlagenhauf-Lawlor P, eds. *Infectious Diseases: A Geographic Guide.* 2nd Ed. Chichester: John Wiley & Sons Inc., 2017: 35 – 44.

6. Tatem AJ, Hay SI, and Rogers DJ. Global Traffic and Disease Vector Dispersal. *PNAS* 2006; 103(16): 6242-6247.

7. World Health Organization. *Plague – Madagascar*. http://www.who.int/csr/don/27-november-2017-plague-madagascar/en/ (last accessed 3 March 2020).

8. Paniagua R. En unas dos semanas el coronavirus se trasladó desde Wuhan hasta La Gomera. *El Mundo*. https://www.elmundo.es/salud/2020/02/01/5e356307fdddff2a348b464b.html (last accessed 3 March 2020) [language: Spanish].

9. Rocklöv J, Sjödin H, and Wilder-Smith A. COVID-19 outbreak on the Diamond Princess cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures. *J Travel Med* 2020.

10. BBC News. *How did Cambodia's cruise ship welcome go wrong?* https://www.bbc.com/news/world-asia-51542241 (last accessed 3 March 2020).

List of Figures

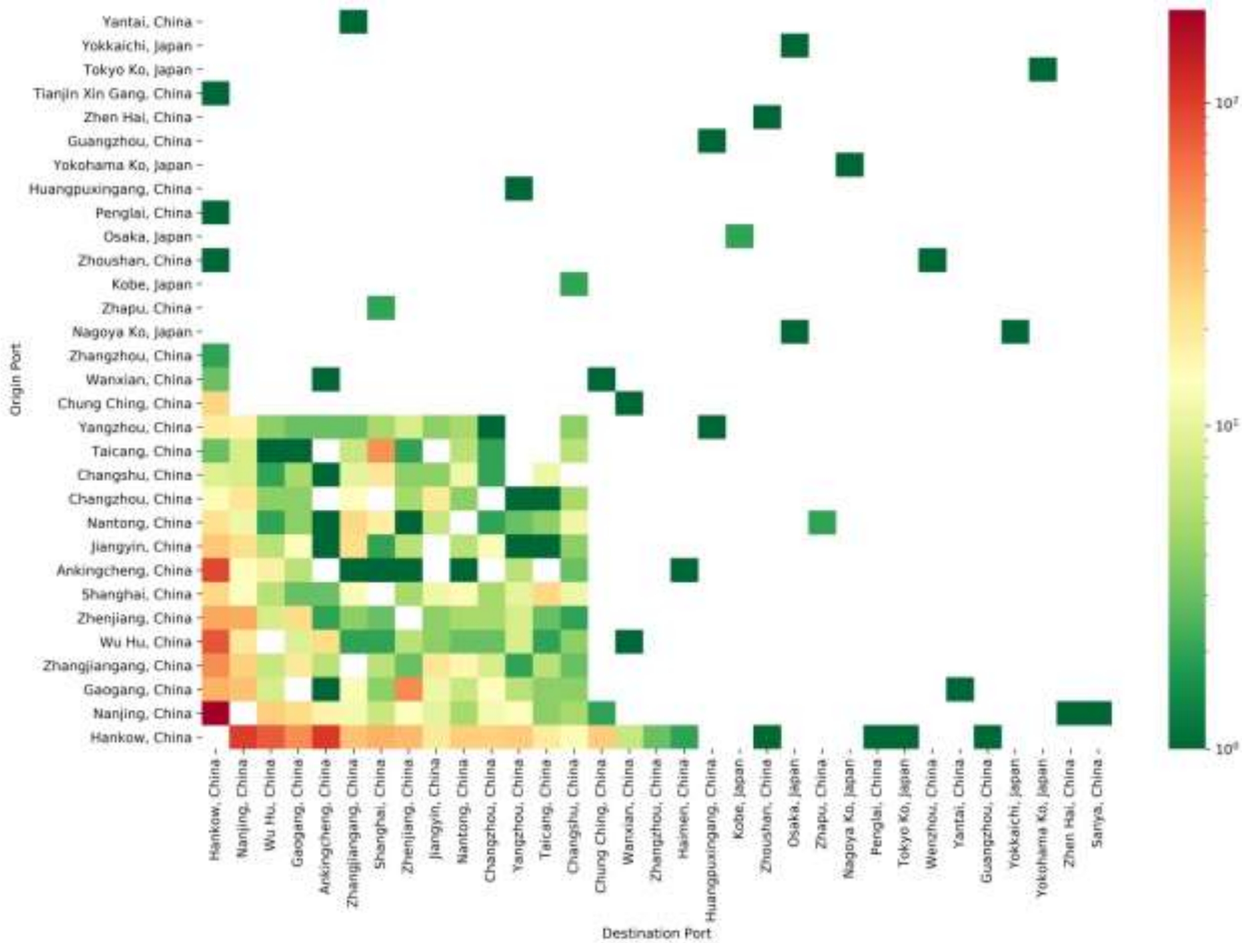

Figure 1a: Origin-destination matrix for ships visiting Wuhan from the 8th to 21st January 2020.

Data source: UN Global Platform AIS Database.

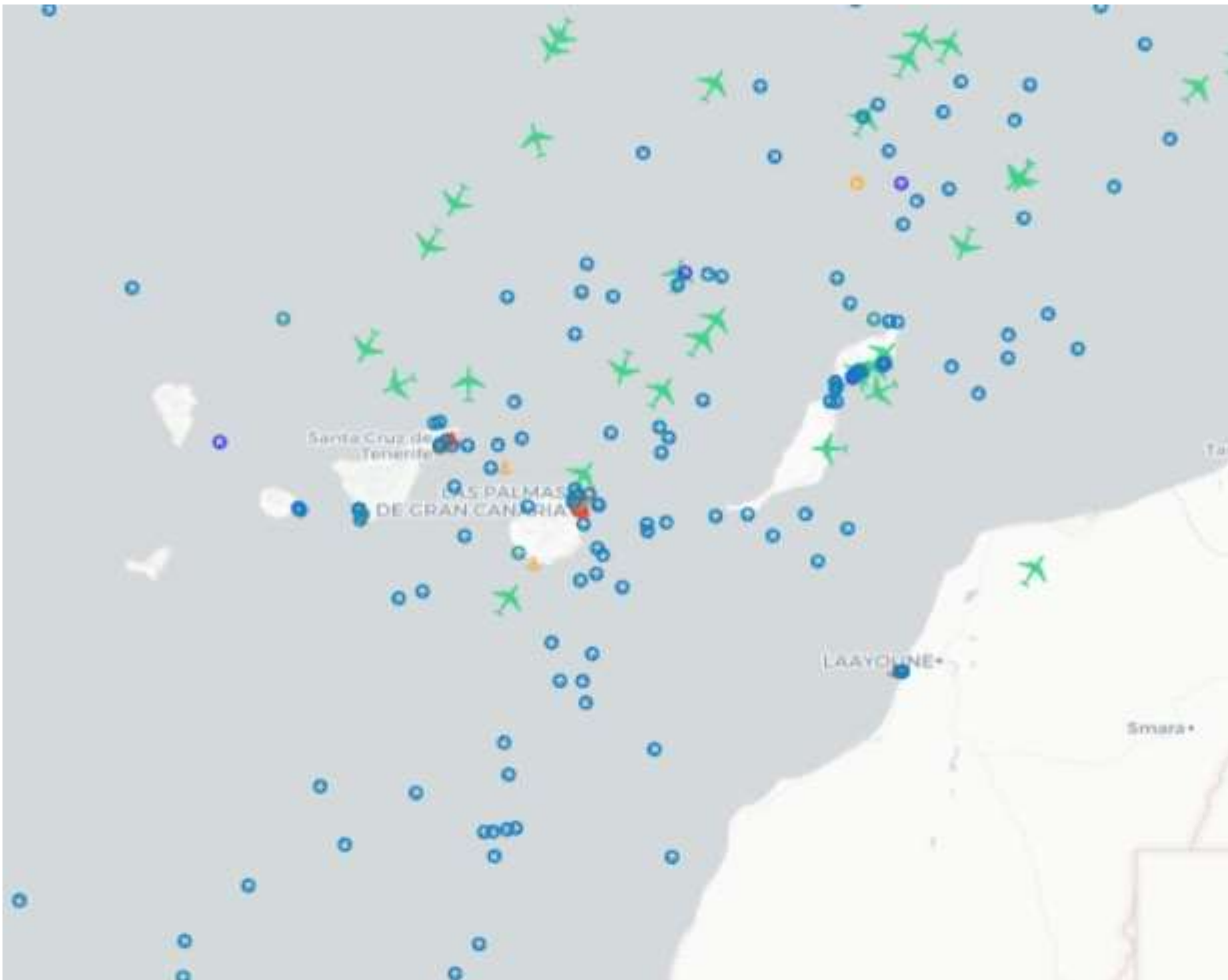

Figure 1b: A snapshot of flight and ship traffic over the Canary Islands from February 2020.

Data source: UN Global Platform AIS Database.

Figure 1c: The tracks of the cruise ship MS Westerdam from 1st January to 19th February 2020. *Data source: UN Global Platform AIS Database. Basemap source: Stamen Design and OpenStreetMap contributors.*