Impacts from Major Events on U.S Vessel Traffic

Peyton Moore
pmoore34@vols.utk.edu

Andrew Mueller
amuell11@vols.utk.edu

***Abstract*—Major events heavily impact the traffic flow of multiple different types of vessels. This paper focuses on the effect COVID-19 had on vessel traffic in the year 2020 and how it differs from the previous year, 2019. The dataset used contains relevant vessel data, such as identification numbers, time codes, vessel type, and position, that was used to analyze the difference and patterns between these two years.**

***Index Terms*—MMSI, UTM Zone, SOG, COG**

I. Introduction

U.S. ports receive a constant flow of traffic, which includes cargo ships, tankers, passenger ships, and pleasure craft [1]. On average, around 60,000 vessels will arrive at a U.S. port annually [2]. The traffic flow can be disrupted by numerous different sources, which can cause delays as well as increased costs. With the amount of vessels that go through the U.S., these vessels must arrive promptly. The most common factors that can cause these disruptions are pandemics, weather-related events, and some trade policy changes. The 3 U.S. coasts, the West, East, and Gulf, are the main points of interest that need to be examined when analyzing the effects of the events on vessel traffic. The main focus for this project is going to be the West Coast and the effect COVID-19 had on the flow of traffic coming out of its ports during 2020 [3]. The West Coast, mainly the ports of Los Angeles and Long Beach, handles most of the Asian trade import/export, which was primarily affected by the closures and restrictions caused by the COVID-19 pandemic. This project will analyze data from 2019 and compare it to the data obtained from 2020 to see how great the difference was between those two years.

Furthermore, we want to find a compression technique since we deal with large data files. Compressing the data is the first major hurdle. Our goal is to write a compression algorithm such that we can consider the compression lossless, meaning we should be able to reconstruct the full dataset from the compressed version.

The following will present our compression technique and the data analysis for ship movement on the West Coast of the United States.

II. Dataset

In our research, we utilize the Vessel Traffic dataset, sourced from the Automatic Identification System (AIS), to analyze vessel movements along the U.S. western seaboard. The dataset contains granular, point-based data that includes vessel identification numbers, timestamps, geographic coordinates (latitude and longitude), speed over ground, course over ground, heading, and vessel-specific details such as type, length, and width. This rich and structured dataset provides a comprehensive record of maritime activity, allowing for precise spatial and temporal analyses. Its segmentation by geographic regions and periods makes it particularly suited for targeted studies, such as understanding patterns of vessel traffic over specific years or in specific zones.

Given the vast amount of data spanning multiple years, this dataset also presents a valuable opportunity to explore efficient data management techniques. As part of our research, we aim to design a lossless compression algorithm tailored to this dataset, ensuring that all original information is retained while significantly reducing the storage size. This would enable us to handle the data more effectively and provide a resource for others who may wish to utilize this dataset in their own research. Once compressed, the dataset will support our primary focus: analyzing vessel traffic from 2019 to 2022 to investigate the impacts of COVID-19 on maritime activity along the U.S. western seaboard. This dual approach ensures that the data is both manageable and deeply informative, making it ideal for our study.

III. Method

A. Data Processing

The first step in this project will be to process the data from 2019 and 2020 from the Marine Cadastre Vessel Traffic dataset [1]. These files contain roughly 90 GB each year, and the data is inputted daily and filtered by every minute [1]. This leads to plenty of overlap, which means that the files must be compressed to process the desired information properly.

B. Lossless compression

To attempt a lossless compression of the original dataset, we delve into how the data is laid out. For the years 2019-2020, the data points collected are separated by day. The problem with this is that for every datapoint collected from any one vessel, a new line is entered into a CSV file. This new line contains duplicate information, such as the vessel type, MMSI number, vessel name, callsign, and IMO. The only fields that change pertain to the location and movement of the vessel itself. Our initial compression will deal with combining the data points for each vessel into its entry, in our case, a line per vessel in the CSV file. The first stage in our lossless compression is to compress each daily file. We do this by parsing out each vessel into a dictionary and adding only the data points to an array. In our first stage, we can drastically reduce the size of the file since we do not need much of the redundant information on each line.

In our second stage, we would like to condense vessel data points across the daily files. We key in on the fact that there should be overlap with vessels from the night of day A, and the morning of day B. Looking at the data points, each data point has a DayTime field, which means to condense the data points between days, we need to convert this into seconds since the epoch, which starts on January 1, 1970 (midnight UTC/GMT). In doing this, we should be able to track vessel movements much more fluidly, as we can convert the data points into a sorted array using seconds since epoch as our key. We can further assume that since vessels do not move fast, we should see enough overlap from day to day such that we can see just new ships seen from the next day added to a master list of ships. Furthermore, we can assume that there is a finite number of ships. Using this assumption, we should be able to condense the information down into a dataset, where each entry is a ship, and the subentry for each vessel is a unique trip. We define a unique trip as all the data points being linear in time with no significant jumps. The threshold for a significant jump in time is not known, but from our assumption, anything jump in time longer than at least a day would qualify as a unique trip.

IV. Data Analysis

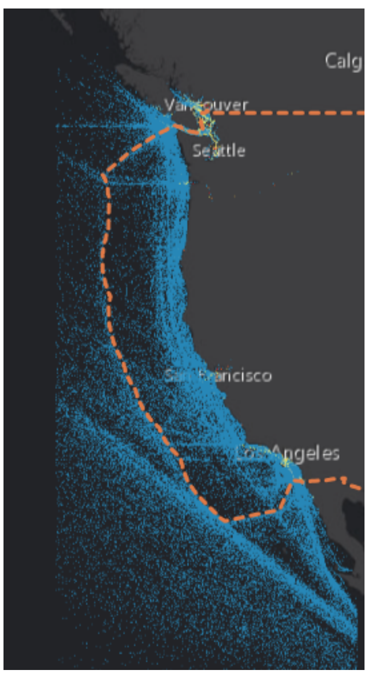

Fig. 1. ScreenShot of vessel analysis tool [4]

A. Data

The concentration of the data for this analysis was picked from the Marine Cadastre Vessel Traffic AIS dataset. This website tracks and logs all the vessels going through U.S. and international waters sourced by U.S. Coast Guards [1]. The data is obtained through an onboard safety device. This dataset contains a plethora of data for each vessel.

- Maritime Mobile Service Identity(MMSI)
- Date
- Vessel Type
- Vessel Name
- Call Sign
- Longitude and Latitude
- Length, Width, Draft
- Speed over Ground(SOG)
- Course over Ground(COG)
- Heading

The majority of the information received from the dataset isn't needed to complete the analysis of how major events impact the vessel traffic over time between the two years specified earlier. The type of data that we'll be focusing on is the MMSI code to be able to identify and distinguish between each vessel. This code is unique to each vessel, so using this code is the key to organizing and compress the data down to a more manageable size. The data that will be used to make our analysis will be the longitude and latitude, speed, cargo type, and time.

B. Filtering data

In our first step of data analysis, we want to be able to filter out vessels that don't give us any information on the West Coast ports. We ideally would be able to filter out any vessels that have longitudes and latitudes that are below the following: Lat ¡ 23.267256179549886 and Lon ¿ -106.36553418675241; These coordinates are some arbitrary cutoff somewhere on the coast of Mexico. We do this to filter vessels out that are not on the West Coast.

C. Analysis

In our data analysis, we will focus on the West Coast of the United States by monitoring activity in and out of major ports throughout 2019-2020. We have not yet decided on how we want to represent the results that we obtain, but the results should show a difference between vessel activity from 2019-2020 due to the COVID-19 pandemic, as well as any trade wars between the United States and Russia. We should be able to see this through various graphs and number analyses.

V. Timeline

For this project, the timeline is broken down into roughly 4 different tasks. The first task that needs to be completed is cleaning the data and compressing it all down into a more manageable size. Next, focus on the two US coasts and track the information of the vessels. Comparing the information that has been stored between the years 2019 and 2020 is what will get done next. Finally, the project will conclude with condensing the results and reporting them.

VI. Expected Outcomes

This project will produce results that show that COVID-19 had a negative effect on the traffic of vessels entering the US. The main highlights of the results of this project will be on how the West Coast was greatly affected because of the vast majority of the vessel traffic from Asia is imported there. Another main outcome that will be expected is which specific ports along the East and West Coasts are more greatly affected when compared to others. The data gained from this analysis will provide a broader understanding of how vessel traffic was impacted by this major event.

REFERENCES

[1] "Vessel Traffic - Marine Cadastre," Marinecadastre.gov, 2016. https://hub.marinecadastre.gov/pages/vesseltraffic (accessed Nov. 21, 2024).

[2] M. Chambers and M. Liu, "Maritime Trade and Transportation by the Numbers — Bureau of Transportation Statistics," Bts.gov, 2017. https://www.bts.gov/archive/publications/by_the_numbers/maritime_trade_and_transportation/index

[3] U.S. International Trade Commission, "The Impact of the COVID-19 Pandemic on Freight Transportation Services and U.S. Merchandise Imports — USITC," www.usitc.gov, 2020. https://www.usitc.gov/research_and_analysis/tradeshifts/2020/special_topic.html

[4] Marine Cadastre, "Automatic Identification System (AIS) Data," Available at: https://marinecadastre.gov/accessais/, Accessed: November 28, 2024.