Reducing CO2 Emissions through Freight Traffic Analytics

Neal Ashinsky
Georgia Institute of Technology

Hayden Cornwell
Georgia Institute of Technology

Tiong Tjin Saw
Georgia Institute of Technology

Fen Zhong Wu
Georgia Institute of Technology

Numaer Zaker
Georgia Institute of Technology

ABSTRACT

Freight transportation is a significant contributor to CO2 emissions which have caused an increase in global temperatures with adverse environmental effects (Al-Ghussain, 2018). Using a route optimisation algorithm to cluster similar ports into hubs-and-spokes we can improve the CO2 efficiency of shipping traffic in the specific example of the US east coast.

KEYWORDS

Global Warming | CO2 | Freight Traffic | Optimization

PROBLEM DEFINITION

Gu specifically discusses maritime freight transport as a significant component of CO2 emissions (Gu, 2019). Tatar supports Gu's point with geospatial visualizations and statistics showing an increasing trend in CO2 (Tatar, 2018). Furthermore, according to the Guardian (Laville, 2019), in France, Germany, UK, Spain, Sweden and Finland shipping emissions in 2018 were larger than the emissions from all the passenger cars registered in the largest cities in each of those countries. Moreover, this has been a topical issue with the UN IMO setting a target of shipping emissions reduction of at least 50% by 2050 (UN Secretary General, 2019), as international transport contributes 2.5% of the global emissions (UN Environment Programme 2019). Various agency resources are currently being devoted to this issue with an estimated $1 trillion needed to de-carbonize shipping between 2030 and 2050 (University Maritime Advisory Services, 2019). Fleet management and voyage plan optimization is estimated to reduce CO2 shipping emissions shipping by 5% (Energy Transitions Commission, 2018).

SURVEY

Bauer attempted to solve the CO2 emissions problem on railroad freights, which she describes as a "service network design problem" (Bauer, 2015) by formulating a mixed-integer linear program. By leveraging this approach, Team 42 uses an optimization model and derives a model that generalizes the size of a freight ship to determine appropriate CO2 emissions.

Currently, CO2 emissions in freight traffic are managed based on global policies known as the MBM, where freight traffic abide by freight rates and pay for ticket costs (Amann, 1993). Team 42 plans to improve on Amann's paper by incorporating a reduction of CO2 emissions into these policies, based on the concept introduce in Shapiro's paper in assessing the environmental costs of CO2 emissions (Shapiro 2016). There are existing methods of route optimization to reduce freight costs (Andersson, 2015), but not CO2 emissions. Team 42 leverages ideas from Andersson's paper to optimize traffic routes and will improve those ideas by solving an optimization problem to minimize CO2 emissions (rather than freight efficiency).

In order to optimize routes to minimize CO2 emissions, Team 42 proposes the following clustering model: ports of close proximity (based on latitude/longitude), will be clustered together as a "single" port cluster. This "hub and spoke" network will also provide the model assessment framework. O'Kelly noted in 1986 that hub-and-spoke models are a way to work with transportation needs and to perform a type of network optimization model between two points (Bryan O'Kelly, 1999).

INTUITION

Leveraging the hub-and-spoke model, as well as the clustering algorithm, Team 42 plans to create an optimization model that minimizes the amount of CO2 emissions in maritime freight traffic, while also optimizing the available freight traffic around certain clustered areas. By focusing on the hub-and-spoke model, CO2 emissions in the shipping industry can be reduced by simply transporting the same freight in trucks versus shipping. This is consistent with Kim and Wee's approach, noting that performing transportation utilizing only one model (whether trucking or maritime only) was incredibly inefficient from both a cost and a CO2 emissions perspective. Instead, Kim and Wee notes that if inter-modal transportation routes were derived (that is, trucks brought their shipments from one centralized port to their local distributors), efficiencies both from a cost and emissions perspective could be realized (Kim and Wee 2014). Team 42 plans to leverage both O'Kelly and Kim and Wee's papers in order to develop a model that leverages this hub-and-spoke functionality, to improve on maritime traffic efficiencies both from a cost and emissions perspective.

An additional crucial point is that both O'Kelly and Kim and Wee was verbose in their papers. Team 42 notes that the use of visualizations will help in describing the model to a larger audience due to human perception limitations when interacting with data (Olshannikova, 2016); visualizations will be used to represent the complex relationships between the different ports (leveraging Olshannikova's paper as reference for good visuals). By leveraging recent technological developments in big data (AWS, Postgres), and visualization platforms (Dash, Plotly, D3, React), Team42 will improve on the visualizations referenced in the aforementioned papers.

An additional consideration may be the seasonality of the data and whether weather and climate impact on the traffic. Additional assessments may be performed with the power of big data and the availability of geospatial and climate data (Retchless, 2018). The team may use Retchless' book as a reference for interactive geospatial visualizations.

DESCRIPTION OF APPROACH

Data Sources

This analysis was performed using Automatic Identification System (AIS) vessel traffic data collected by the U.S. Coast Guard through an onboard navigation safety device that transmits the location and characteristics of large vessels. The entire dataset includes records of ships in U.S. coastal waters from 2009 to 2017, with a sampling rate of every minute. The original data is located: https://marinecadastre.gov/ais/.

The U.S. is split up into 20 zones, as seen in figure (1). Since there will be little traffic from one coast to another, only the eastern zones, 16 through 20, will be used for December 2017. This represents the most recently available monthly cross-section of data.

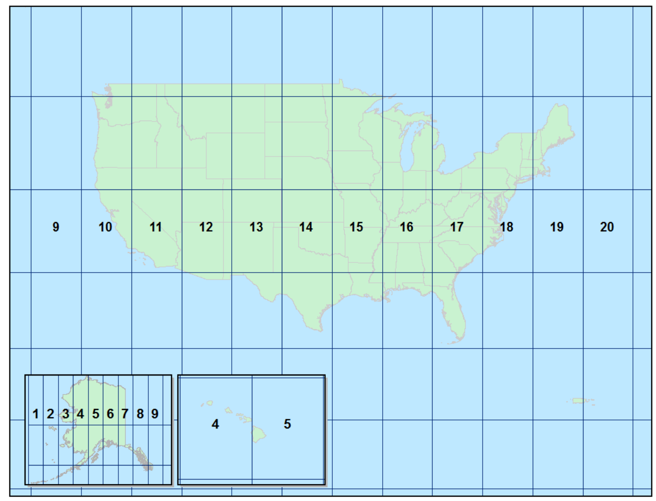

Figure 1: Map of respective zones split up by latitude and longitude

The raw AIS data includes the following attributes: MMSI (Maritime Mobile Service Identity), date/time, latitude, longitude, speed over ground (SOG), course over ground (COG), heading, vessel name, international maritime organization ID (IMO), call sign, vessel type, status, length, width, draft, and cargo.

Cleaning the Data

Jupyter Notebooks with Pandas was the primary software used to clean and manipulate the AIS data for the analysis. Once loaded into a Jupyter notebook, the initial cleaning was as follows:

- Remove columns: SOG, COG, Heading, VesselName, IMO, CallSign, Draft, and Cargo.
- Remove all rows that include NaN.
- Filter out rows with an ambiguous status.
- Keep vessel types that are either Cargo or Tanker.

With this partially cleaned dataset, the location of each ship was still recorded for every minute of the month. This data needed to further parsed to reference only the single data point for a ship arriving or leaving a port. The steps to implement are as follows:

- Sort data by VesselName and BaseDateTime.
- Copy the status column and shift one row down.
- Filter out to remove rows where the two status rows are the same.
- Filter out to remove rows where the status is not "moored" or "at anchor".

The original dataset had around 95 million rows of data. After the initial filtering, it is reduced to around 11 million rows, and after the last cleaning steps down to 3325 rows.

Creating the Hub-and-Spoke Model

The next step is to create a "graph" of the entire shipping network. Although the AIS data does not assign ships to a port, we can assign ships to a port using a clustering model. The clustering algorithm used was Density-Based Spatial Clustering of Applications with Noise (DBSCAN) imported from the python library "scikit-learn". The size of the clusters is controlled by the hyperparameter "epsilon".

Figure 2 demonstrates how epsilon changes the hub and spoke model. For each segment, it is assumed that the ship starts at its original GPS coordinate, the starting "spoke", and travels to its assigned "hub" determined by the clustering results. The ship then travels to the ending GPS locations "hub", then finally traveling to the ending GPS location, the ending "spoke". This various paths travelled are to determine the total CO2 output of each ship.

Calculating CO2 Output

Previous research to gauge the efficiency of different shipping methods reports the CO2 efficiency through the metric grams of CO2 per tonne per kilometer, seen in figure 4. Maritime shipping is the most efficient method and becomes more efficient as the size of the ship increases. CO2 efficiency of different ship sizes has been documented in prior work (Responsible Care 2011) and can be used to fit a curve. Although there are not many data points, the CO2 to TEU relation can be loosely fitted into a quadradic curve as seen in figure 5. Ship size is defined by twenty-foot equivalent units (TEUs), which is the carrying capacity of a ship for twenty-foot containers.

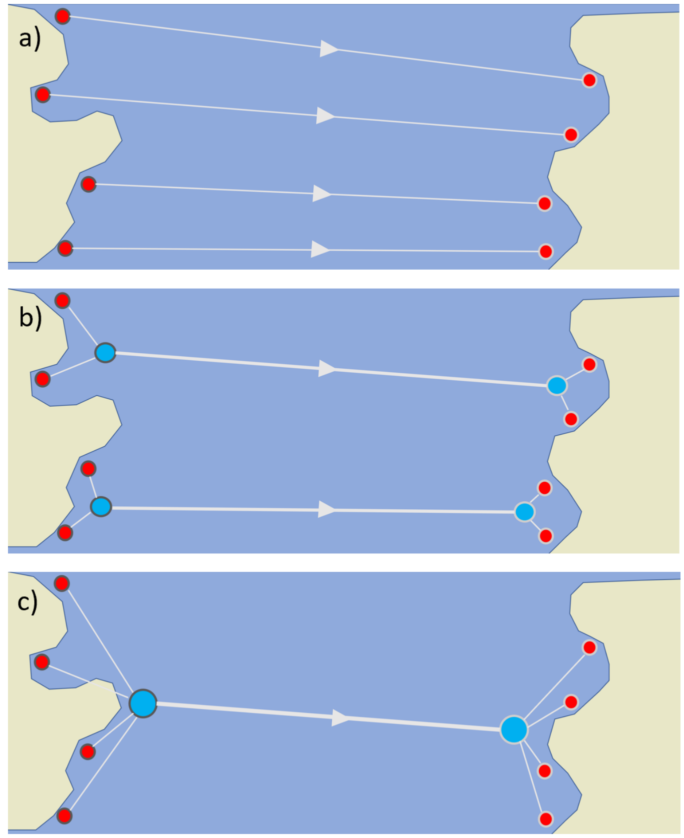

Figure 2: a) is an example of an "Unoptimized" model where no hubs are assigned. The other models are hub and spoke models with b) using a lower value of epsilon and c) using a larger value of epsilon.

Length and width of the ship are used calculate the TEU for each ship. For a random sample of 28 ships of varying lengths and widths in the dataset, the MMSI of the ship was researched online to find its associated TEU. As seen in figure 5, the ship area (the product of length and width) correlates very well with TEU, and thus can approximate TEU.

The equations for both regression curves are listed below:

$$TEU = (3.17688908e - 01Area) + (2.06756025e - 05Area^2) - 219.2003311 \tag{1}$$

$$CO2 = (-4.32978571e - 03TEU) + (1.39479467e - 07TEU^2) + 36.15242554 \tag{2}$$

With CO2 efficiency, to calculate the total CO2 output the distance traveled and amount of cargo need to be

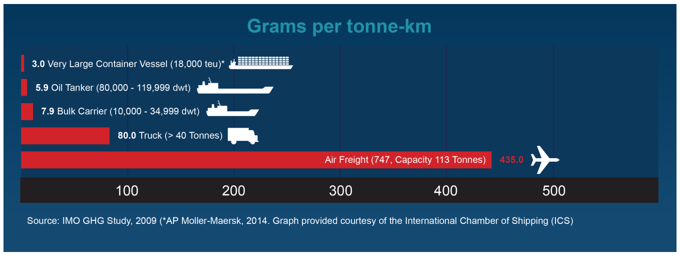

Figure 3: Source: IMO GHG Study, 2009 ('AP Moller-Maersk, 2014').

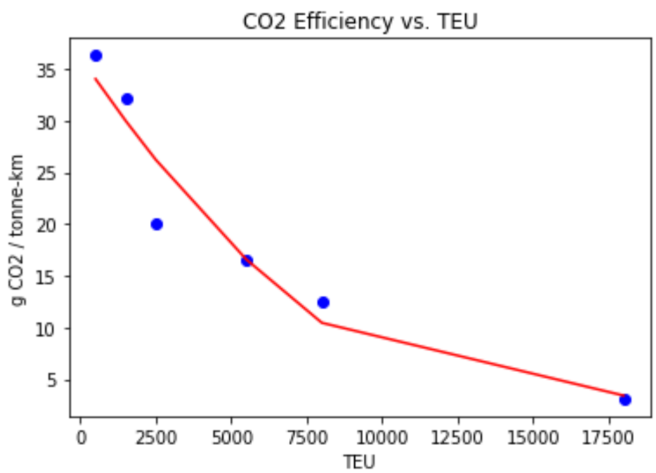

Figure 4: Overview of CO2 Efficiency versus TEU, from Team 42's model.

calculated. The Haversine distance formula, equation 3, was used to calculate distance between any two GPS coordinates. Using this formula calculates the "as-the-crow-flies" distance between two points, assuming the ship travels a straight line between two ports, potentially crossing over land. This over-land assumption is reasonable as the model is a generalized heuristic. The multi-modal approach accepts that it is sometimes more efficient to have overlapping land and sea transport.

$$d = 2r\arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2-\varphi_1}{2}\right)+\cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\lambda_2-\lambda_1}{2}\right)}\right) \tag{3}$$

The final portion of the CO2 calculation is to determine the weight of the ship in tonnes, as this was not included in the data. The calculation is based on the average filled weight of a single twenty-foot container can be estimated as 15 tonnes of dead-weight (Pike 2011). The

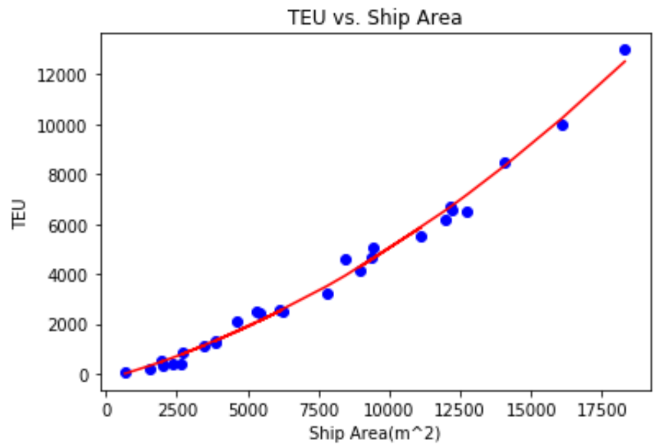

Figure 5: Overview of TEU versus Ship Area, from Team 42's model.

final assumption for the CO2 calculation is that all of the ships with the same hub-to-hub segment will combine their cargo to be transported by a single larger ship. As a result, there will be two TEUs to factor when calculating CO2. When the ship is traveling from the spoke to hub, the TEU to calculate the CO2 efficiency and the cargo weight will be the same. However, when a ship travels from hub to hub, the TEU to calculate the CO2 efficiency will change to the sum of all of the ship's TEU that are traveling the same segment. With this higher TEU, there will be a relatively lower CO2 output for the hub-to-hub segment. This assumes that it is possible to load up the larger hub-to-hub segment ships with a lot more additional cargo and suggests the need for larger capacity ships to drive down CO2 emissions.

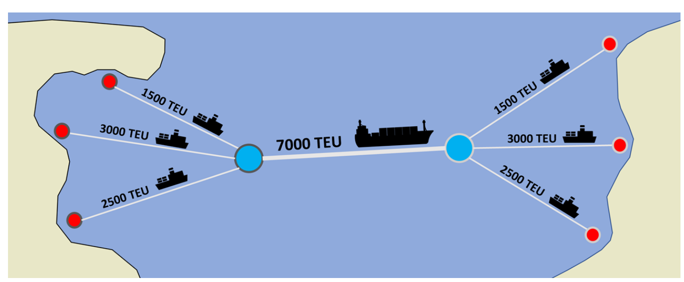

Figure 6: Illustration of how cargo is combined at the hubs.

By summing all of the CO2 values for each ship segment, we can have the total CO2 output for each hub-and-spoke model.

User Interface and Backend

Dash Framework.
We utilize the open-source library Dash to build our user-interface (UI) because it allows us to create highly custom and reactive web-base visualizations using simple native python.

Dash is reactive which makes it easy to create web-interface elements which interact with each other. It abstracts away web technologies such as Plotly.js (build on top of D3.js) and React.js and allows you to interact with the HTML and CSS elements directly using the Python abstraction. This also mean that every element of the application can be customized (the CSS, layout etc). Dash further allows you encapsulate D3.js charts directly if necessary. The frontend is rendered with React.js which works in tandem with the declarative python libraries provided as part of the toolset.

Dash also has a backend Flask web-server baked into it that allows us to query the results produced above. The front-end will interact with the web-servers using JSON packages over HTTP request. Since the application state is stored in the front-end , it also allows Dash apps to be multi-tenant with multiple users having independent sessions.

Visualizations.
Our application follows a model-view-controller design pattern that let us separate our visuals from the processed data. We utilize the application logic to control the flow of reprocessed data through the interface.

The main user control “Filters” are on the top right of the UI. The dashboard dynamically updates to match the control criteria and re-renders only the relevant visuals on the page with this logic flow. We have filters of "model epsilon”, “individual hub”, and “vessel type”.

The visualizations in the dashboard are based on the key calculations produced above:

- The top high-level key metrics show shows the total trips, model hubs, original actual CO2 emissions vs the optimized CO2 to show the impact of the model.
- The “Hub and Spoke Optimized Traffic Network” map below allows us to see the difference in the network routes based on different model epsilon values optimizing the paths. It defaults to displays all of the Hubs for a particular epsilon selection. You can select Hub values to visualize the more granular hub-and-spoke lines for a particular hub. For example, changing epsilon from 0.1 to 0.35 would divert more of the ports to a single hub and we can visualise the impact of the hub-and-spoke improvements.
- The “Cluster Size vs CO2 emissions” interactively updates based on the mouse-over location of a spoke on the hub-and-spoke map. It gives a finer grain understanding of how cluster size affects the CO2 emissions. You can directly visualize how CO2 emissions change as the epsilon value varies.
- The “Original Freight Traffic Network” map allows the user to see the differences in network routes once the model optimizes the paths. It will color the original network nodes based on which hub they belong to. It gives a high-level overview to how the paths are re-categorized based on differing epsilon values.
- The 3D scatter-plot visual for “Vessel Dimensions vs CO2 Efficiency” color-coded by vessel type to demonstrate how vessel dimensions affect vessel CO2 out. You can rotate this 3D plot to visualize the pattern of CO2 efficiencies based on vessel length/width.
- The “Overall Optimal Epsilon vs CO2 Emissions” indicates what the overall best clustering size of epsilon is for the network.

DESCRIPTION OF YOUR TESTBED

The list of questions that the experiment is designed to answer includes:

- Can CO2 emissions be reduced based on route optimisation techniques? If so by how much?
- What is the optimal route clustering size for a hub-and-spoke optimisation model?
- Can a heuristic be modeled, based off of length and width of a ship and distance traveled, in order to approximate the carbon emissions?

DETAILS OF THE EXPERIMENTS

The main variable of interest for the hub-and-spoke models is the cluster size of the hub. By varying epsilon from 0.00001 to 2.0, many hub-and-spoke models were created. Epsilon is in units of GPS coordinates, so the max value of 2.0 is relatively large and will create clusters that will start to look unrealistic; the corollary is also true in that a small epsilon will result in many

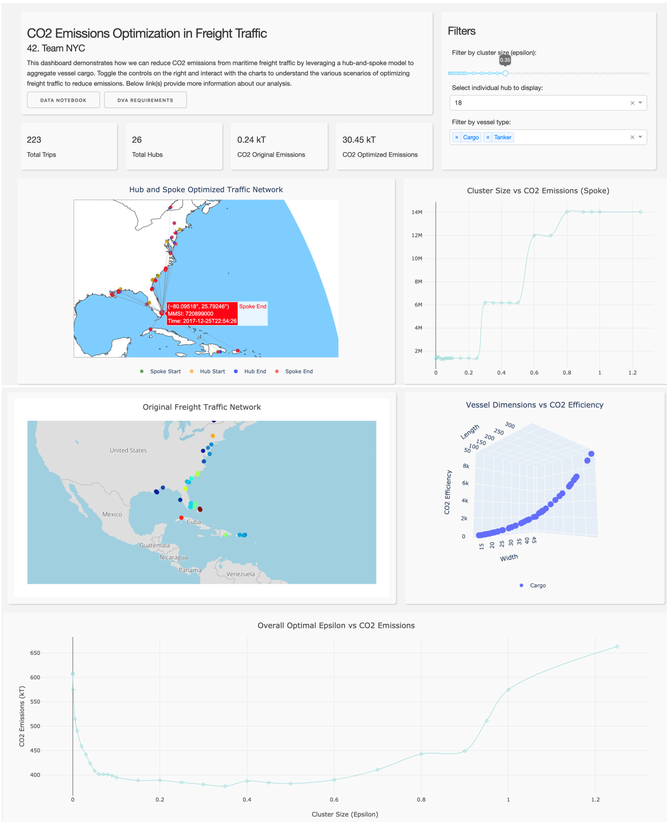

Figure 7: Team Visualization Overview

clusters creating a network where no hubs exist (the assumed prediction for actual CO2 emission is an epsilon value of 0.0001).

Another assumption of this model is that hub-to-hub travel breaks down when epsilon values exceed a particular threshold. As the port cluster becomes larger, the frequency of ships starting and ending in the same cluster increases, resulting in additional distance for each trip to travel. Although the hub-to-hub traveling distance will be 0km in these cases since there is only one hub, the more optimal route would be for the ship to go directly to the end spoke, instead of visiting the intermediate out-of-the-way hub first.

Figure 8 plots the total CO2 output versus epsilon. The unoptimized model reports a CO2 output of 608 kilotonnes (kt). With only a small increase in epsilon, the CO2 output decreases rather quickly to 400kt. This is the point where the hub-to-hub optimizations give a positive impact. In the plot, the minimum CO2 output is for an epsilon value of 0.35. As epsilon increases from this minimum, the hub and spoke model starts to see diminishing returns with the CO2 output beginning to increase at an epsilon value of 0.5. This occurs because you have more inefficient spoke-hub-spoke diversions rather than the more efficient spoke-hub-hub-spoke optimisations.

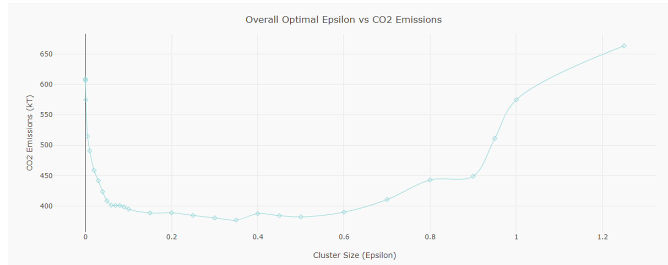

Figure 8: Model Optimization

TEAM ASSESSMENT

All team members have contributed similar amount of effort

CONCLUSIONS AND DISCUSSIONS

From our analysis, an optimal hub-and-spoke network was found by clustering the ships destinations using an epsilon value of 0.35. This reduces the predicted CO2 output from 608kt to 377kt. Although this is the optimal value our analysis, it is not certain if this value of epsilon should be used universally to optimize other shipping networks. Our results show that even a very low epsilon value of 0.05 gives a CO2 output of 400kt, which is very close to the optimal value. As epsilon is increased the model complexity is increased, which leads us to recommend evaluation all values from 0.05 to 0.35.

REFERENCES

[1] Amann, M. (1992). Estimating Emission Control Costs: A Comparision of the Approaches Implemented in the EC-EFOM-ENV and the IIASA-RAINS Models.

[2] Andersson, H., Fagerholt, K., Hobbesland, K. (2015). Integrated maritime fleet deployment and speed optimization: Case study from RoRo shipping. Computers and Operations Research, 55, 233-240. https://doi.org/10.1016/j.cor.2014.03.017

[3] Bauer, J., Bektaş, T., Crainic, T. G. (2010). Minimizing greenhouse gas emissions in intermodal freight transport: an application to rail service design. Journal of the Operational Research Society: Special Issue, 61(3), 530-542. https://doi.org/10.1057/jors.2009.102

[4] Bryan, D. L., O'Kelly, M. E. (1999). Hub-and-spoke networks in air transportation: an analytical review. Journal of Regional Science, 39(2), 275.

[5] Energy Transitions Commission (2018). Mission Possible: Reaching Net-Zero Carbon Emissions from Harder-to-Abate Sectors by Mid-Century, 10. http://energy-transitions.org/sites/default/files/ETC_MissionPossible_ReportSummary_English.pdf

[6] Gu, Y., Wallace, S., Wang, X. (2018). Can an Emission Trading Scheme really reduce CO2 emissions in the short term? Evidence from a maritime fleet composition and deployment model. IDEAS Working Paper Series from RePEc.

[7] Laville, S. (2019, December 9). European shipping emissions undermining international climate targets. Retrieved from https://www.theguardian.com/environment/2019/dec/09/european-shipping-emissions-in-way-of-nations-meeting-paris-climate-targets

[8] Olshannikova, E., Ometov, A., Koucheryavy, Y., Olsson, T. (2016). Visualizing big data. https://doi.org/10.1007/978-3-319-44550-2_4

[9] Pike, J. (2011, July 7). Container Ship Types. Retrieved from https://www.globalsecurity.org/military/systems/ship/container-types.htm

[10] Responsible Care, ECTA, cefic. (2011). Guidelines for Measuring and Managing CO2 Emission from Freight Transport Operations (1st ed.). Retrieved from https://www.ecta.com/resources/Documents/Best%20Practices%20Guidelines/guideline_for_measuring_and_managing_co2.pdf

[11] Retchless, D. (2018). Bringing the big data of climate change down to human scale: Citizen sensors and personalized visualizations in climate communication.

[12] Shapiro, J. (2016). Trade Costs, CO2, and the Environment. American Economic Journal. Economic Policy, 8(4), 220-254. https://doi.org/10.1257/pol.20150168

[13] TATAR, V , ÖZER, M . (2018). THE IMPACTS OF CO2 EMISSIONS FROM MARITIME TRANSPORT ON THE ENVIRONMENT AND CLIMATE CHANGE. International Journal of Environmental Trends (IJENT) , 2 (1) , 5-24 . Retrieved from https://dergipark.org.tr/en/pub/ijent/issue/36545/405416

[14] UN Environment Programme (2019). Emissions Gap Report, 5. https://wedocs.unep.org/bitstream/handle/20.500.11822/30797/EGR2019.pdf

[15] UN Secretary General (2019). Report of the Secretary-General on the 2019 Climate Action Summit – The Way Forward in 2020, 5. https://www.un.org/en/climatechange/assets/pdf/cas_report_11_dec.pdf

[16] University Maritime Advisory Services, Energy Transitions Commission (2019). Getting to Zero Coalition Insight Series: The Scale of Investment Needed to Decarbonize International Shipping, 1. https://u-mas.co.uk/LinkClick.aspx?fileticket=03IebWyJns8%3d&portalid=0