# A mean-field Game Of Shipping: Modelling And The Use Of Real Data

Charles-Albert Lehalle * Giulia Livieri † Evan Chow ‡ § ¶ ‖

17 August 2025

**Abstract**

## Contents



* Visting Processor, Imperial College London and Fellow of Institut Louis Bachelier, Paris. c.lehalle@imperial.ac.uk
† London School of Economics and Political Science. Email address:
‡ Abu Dhabi Investment Authority, ADIA. Evan.Chow@adia.ae
§ Affiliation: . Email address:
¶ Affiliation: . Email address:
‖ Affiliation: . Email address:

# 1 Introduction

The mean-field game approach for a continuous state and time symmetric differential games was introduced by Huang et al. [2006] and Lasry and Lions [2007], almost concurrently. It arises in situations, also called games, involving a large number of indistinguishable rational agents where the impact of each individual player on the system is infinitesimal. The idea is that a large-population game of this type should behave similarly to its mean-field counterpart, which may be thought of as an infinite-player version of the game. By exploiting the underlying symmetry while passing to the limit with the number of agents tending to infinity, the sequence of games converges in some sense to a problem in which the game structure is preserved in a simpler form for a "representative player", that responds optimally to the average behaviour of the population. Once the limit problem is solved, its solution can typically be implemented in the large-population game and provides an approximate Nash equilibrium with vanishing error as the number of agents tends to infinity. The solution to the representative agent problem is characterized by a pair of backward Hamilton-Jacobi-Bellman (HJB) and forward Fokker-Planck-Kolmogorov (FPK) equations. The former equation guarantees player-by-player optimality, while the FPK equation guarantees time consistency of the solution. We refer the reader to the lecture notes by Cardaliaguet [2010] and the two-volume monograph by Carmona et al. [2018] for comprehensive presentation of mean-field game theory and its applications from analytic and probabilistic perspective, respectively.

The discretization of the problems introduced by Huang et al. [2006] and Lasry and Lions [2007] has led to numerical analysis questions, as well as motivated the study of mean-field models in discrete-time with a finite number of states; see, e.g., Gomes et al. [2010]. In this case, the very large number of identical agents can be in a finite number of states, and each agent behaves individually and rationally, moving from state to state according to certain optimal criteria. Decisions are based solely on the following information, which is know by every agent: the current state, and the fraction of agents in each state. As in non-cooperative games, there may be interactions between the players in different states. In addition, because the number of agents is very large, the mean-field hypothesis assumes that only the fraction of players in each state at a given time is the relevant information for the global evolution.

Mean-field models in discrete-time with a finate number of states is a framework that lends itself well to the investigation of the connection between the transportation research in general and the recent progress in the mean-field game theory; see Sheffi [1985], for an old reference in this direction. More recently, Salhab et al. [2018] analyze a discrete-time discrete-state mean-field route choice games. In Chevalier et al. [2015], instead, the authors model the interaction between drivers on a straight road as a non-cooperative game and characterize its mean-field equilibrium. Bauso et al. [2016] consider a continuous-time Markov chain to model the aggregated behavior of drivers on a traffic network. In Baillon and Cominetti [2008], instead, the authors introduce a Markov framework for traffic assignment problems, which is similar to the problem formulation adopted in Tanaka et al. [2020]. Here, we also cite the work of Guéant [2015], which provides existence and uniqueness result for mean-field equations on graphs, a setting that allows to take into account the so-called "congestion effects". Finally, a connection between large-population Markov decision processes and mean-field games has been discussed in Yu et al. [2022].

In the present paper, we adopt a formulation similar to that in Tanaka et al. [2020] to analyze seaports competition, which is an accepted and important phenomenon, and a key driver of performance improvement, in the shipping industry; see Subsection 1.1 for an overview of the recent game theory applications on seaport cooperation and competition. In particular, maritime transportation is a vital piece of global trade, with approximately 80% of commerce by volume, and 70% by value is transported by sea and processed at ports worldwide (unc [2023]). In principle, seaports competition can be analyzed from two different perspective. The first is microscopic. It considers an individual ship and tries to understand how it makes choice based on both some personal attributes, such that the type of good it carries, and network related attributes, such as the travel time between the port where it is located and a destination port. The second perspective is macroscopic, and it is concerned with the macroscopic behavior of ships. The model we propose in the present paper lies somewhere between the microscopic and macroscopic approach. More precisely, the analysis at the level of one ship is not realistic since shipping company often own a lot of ships[1]. Therefore, in practice, we should have a few numbers of large companies. However, for the sake of simplicity, we adopt the viewpoint of modelling one representative company at each seaport, that can be considered as the aggregation of all the decisions taken by all companies having ships at this port.

Many factors affect shipping companies decisions on the selection of ports, e.g., availability of hinterland connections, port tariffs, immediacy of consumers, feeder connectivity, environmental issues and the total portfolio of the port; see, e.g., Wiegmans et al. [2008]. In the present paper, we account for two types of cost, transportation cost and congestion costs at destination, in addition to the so-called commercial margin, i.e., the difference between the value of a certain good at the arrival port and its value at the departure port. These costs represent the main drivers of the decisions. In particular, the long delays that ports congestion may have caused in delivering goods to consumers and firms have been gathering increasing attention, especially after the COVID-19 crisis; see, e.g., Komaromi et al. [2022]. On the other hand, the mean field effect is represented by the $K$-dimensional distribution of ships at ports. G.L.: previous paraghraph should be written better.

The main contributions of this paper are as follows G.L.: to be written at the end.:

1)

2)

### 1.1 Review of the literature on game theory applications on seaport cooperation and competition.

This subsection (partially) overviews the recent game theory applications on seaport cooperation and competition; notice that mean-field games theory is a branch of game theory.

We start with the work by Luo et al. [2012] who studies, via the techniques of Bertrand competition and Nash equilibrium, a two-stage game where ports make capacity investment decisions between ports of Hong Kong and Shenzhen when the market demand increases with different service levels, in addition to pricing decisions. They find that both ports would expand with the increasing market demand and that the new port with a smaller capacity will be more likely to expand due to lower investment costs and higher price sensitivity. The same objective and ports are examined in Do et al. [2015] via a two-person game model with uncertain demand and payoff Nash equilibrium; here, ports decide to invest under consideration that demand is uncertain or payoff is uncertain. In this case, Shenzhen was found to be the dominant port in a long-term strategy, whereas Hong Kong should make capacity investments only when Shenzhen does. Ports capacity investment decisions, by increasing their capacity, between ports of Busan, Korea, and Shanghai are analysed in Anderson et al. [2008] via Bertrand competition. They find that investments should only be undertaken in East Asia. In addition, governments should be aware of any current or future competitor developments that may have a chance to gain a more significant share of the market. Nguyen et al. [2016], instead, study a two-stage game via Nash equilibrium where ports make pricing decisions to maximize profit by identifying network links between ports in a network and strategic interaction. In particular, they identify the effects of strategic pricing on ports in their networks in three Australian regions, namely Queensland, South Australia and Victoria, and Western Australia, and find that not all ports set prices through strategic interaction between other ports; some set prices independent of each other. Moreover, the pricing strategy of competing ports may differ from each other. Ishii et al. [2013] analysed a two-person game model with stochastic demand via Nash equilibrium in which ports make pricing decisions in the time of capacity investment. More precisely, they analyze strategic port pricing in a setting of inter-port competition and at the time when ports make capacity investment decisions. They find that when both the demand elasticity and port capacity development activities are high, prices should be set low. Zhuang et al. [2014] investigate service differentiation for ports that manage containerized cargo and dry-bulk cargo via a Stackelberg game where the leader port decides output volumes for both container and bulk cargo operations and the follower port decides output volumes in container and bulk cargo

[1]See, e.g., `https://panjiva.com/sitemap/scac_profiles_directory`, where one can find a list of 7,327 shipping companies with detailed information about the suppliers, buyers, vessels and ports associated with that company. For instance, the shipping company Aa Freight Inc uses 820 vessels.

operations. They find that: (a) port infrastructure investments should be coordinated adequately with other port infrastructure investments and potential demand; (b) government intervention may be required, as it may lead to over-investment and excessive competition; (c) leading ports benefit from making first moves that result in greater profit and larger traffic volume. A two-stage game where ports make terminal award decisions and terminals set port charges competing in quantity is analyzed in Yip et al. [2014], where the authors examine inter- and the intra-port competition on terminal concession awarding. They find that terminal operators prefer to govern more terminals in the region, and that port authorities with considerable market dominance prefer to introduce inter- and intra-port competition. Finally, we mention the work of Kaselimi et al. [2011], where the authors study a two-stage game where terminal operators compete for quantities by taking consideration of their capacity via a Cournot-type competition; terminals compete in both prices and output. In particular, they examine the effects of the transition from a multi-user terminal to a fully dedicated terminal on inter- and intraport competition between the multiuser terminals. They find that the introduction of dedicated terminals will result in less profit to the port authorities and also to the users of multiuser terminals, while multiuser terminals were unaffected by the introduction of dedicated terminals.

# 2 A Framework For Shipping Of Multiple Goods Over Multiple Ports

**Main concepts and notations.** This paper considers the routes taken by ships conveying goods from one port to other ports from the viewpoint of a coordinator standing at each port. At the scale of the globe, we consider they are $K$ ports, and there exist $N$ goods. In each port $i$, there is a set of ships waiting to be loaded at time $t$, companies owning the ships decide which fraction of their fleet has to be sent from port $i$ to port $j$. We make the classical Mean Field Game assumption that companies have no idiosyncrasy: the Mediterranean Shipping Company will take the same decision as the CMA CGM would take if the later would have exactly the same ships at the same positions. The instructions of all the companies can thus be aggregated at the level of each port to form a vector of desired transitions for the whole fleet present at $t$ in port $i$, to each potential destination $j$. It forms the desired flow $\Phi^n_{i,j}(t^+)$ of ships to be sent from $i$ to $j$ loaded with good $n$. The notation $t^+$ stands to underline that this decision is take after the capacity of the fleet at $t$ (i.e. the number of ships) is known.

The companies are taking multiple factors into account:

- The margin $M^n_{i,j}(t)$ they can make by sending a unit of good $n$ from port $i$ to port $j$. This margin will typically include tariffs.
- The transportation costs $c_n g(T_{i,j})$ to send a unit of good $n$ from port $i$ to port $j$, where $T_{i,j}$ is the transportation time, or transportation length. Libraries like searoutes (Eurostat [2022]) are available to get such distances.

  Following the literature we will soon assume the transportation cost is a power law in the number of units to be sent (cf. Micco and Pérez [2002]); we will focus on quadratic costs to get simpler results.
- The congestion costs that is a function of the flow arriving at destination ports $j$. This is more subtle to handle than the transportation costs, because the flow $\Phi^n_{i,j}(t^+)$ will reach port $i$ at $t + T_{i,j}$, the expected crowd at $j$ at arrival is hence

$$\sum_{\ell \neq i} \Phi^n_{\ell,j}(t + T_{i,j} - T_{\ell,j}).$$

Our framework aims to *solve the stationary version of the shipping problem.* That for we will need to make some approximations that could be relaxed in future researches. First we will consider that the total transportation capacity $F^n$ for a given good does not change: companies are not introducing ships into or removing ships from the ecosystem we consider. It will allow us to renormalise all the quantities and flow we observe by these $F^n$. For instance, we will denote by $F^n P_i^n(t)$ the shipping capacity for good $n$ waiting in port $i$ at time $t$.

Once the problem is stationarized, the flows are following the same law whatever the considered date. It implies that, at the correct time scale, quantities like $\sum_{\ell \neq i} \Phi_{\ell,j}^n$ are random variables, and not stochastic processes. That simplifies for instance the writing of the congestion costs.

## 2.1 The Details Of The Framework

**Notations.** The control is a transition matrix $Q_{i,j}^n(t^+)$; it corresponds to the fraction of the ships waiting to be loaded with good $n$ at port $i$ that is to be sent to port $j$. As the consequence, the expected flow of good $n$ from port $i$ to port $j$ is

$$\Phi_{i,j}^{n,Q}(t^+) = F^n P_i^n(t)\, Q_{i,j}^n(t^+).$$

Note that there is no notion of randomness in this relation, even $Q_{i,j}^n(t^+)$ can be read as the desired probability to send from port $i$ to port $j$ ships waiting to transport good $n$. The reader can immediately think about a stochastic version where the ships that are effectively sent are one realization drawn according to this probability. But for this paper, and since we are modeling decisions for large ports, with an intense traffic of ships, we will consider that *the expected flow is realized*. It means that the (random) variations around the concentration provided by the central limit theorem is small enough to be neglected. It is beyond the scope of this paper to study these variations. Extending our results to a stochastic version of this model is possible but would need to introduce convexity terms around cost terms that are non linear functions of the flows of ships. For ports having a high flow of ships, this term should be small enough to not drastically change the obtained results that have the virtue to be remarkably simple.

The costs are made of two main components:

(*i*) The expected *Transportation costs* perceived from port $i$ for the good $n$ at the decision time reads

$$C_{i,\cdot}^{n,Q}(t^+) = c_n \sum_{j \neq i} \left| \Phi_{i,j}^{n,Q}(t^+) \right|^{\gamma} \cdot g(T_{i,j}). \tag{1}$$

Considering quadratic costs, i.e. $\gamma = 2$, will lead to an easier mathematical framework.

(*ii*) To write the *Congestion* (or Crowding) *costs*, it is needed to understand that the decision to send goods from port $i$ to port $j$ is a function of all the goods that will arrive simultaneously at $j$. With the notation $T_{i,j}$ for the transportation time from $i$ to $j$, it means that the goods sent from $i$ to $j$ at $t^+$ will arrive at $\tau = t^+ + T_{i,j}$. Similarly goods sent from another port $\ell$ to the same destination $j$ and arriving at the same moment $\tau$ will have to be sent at $\tau - T_{\ell,j}$, i.e. at $t^+ + T_{i,j} - T_{\ell,j}$. It boils down to the following notation for the cost related to the decision $Q$ taken at port $i$:

$$R_i^Q(t^+) = \mathbb{E}_{t^+} \left\{ \sum_j r_j \left| \sum_n \sum_{\ell \neq i} \Phi_{\ell,j}^{n,Q}(t^+ + T_{i,j} - T_{\ell,j}) + \sum_n \Phi_{i,j}^{n,Q}(t^+) \right|^{\rho} \right\}. \tag{2}$$

The conditional expectation term $\mathbb{E}_{t^+}(\cdot)$ means that the decision is taken at time $t^+$, but the effective cost can depend of some randomness of the transportation time. Again considering $\rho = 2$, i.e. quadratic costs, leads to simpler mathematical conclusions.

The expression is split between a part $\sum_n \sum_{\ell \neq i} \Phi_{\ell,j}^{n,Q}(t^+ + T_{i,j} - T_{\ell,j})$ that is not controlled at the level of this port (that is clearly a *mean-field term*) and another part $\sum_n \Phi_{i,j}^{n,Q}(t^+)$ that is related to the decisions that is currently taken.

**Assumption 1** (Stationarity of the Costs)**.** *When the game is stationarized, the crowdedness costs* $R_i^Q(t^+)$ *reads*

$$R_i^Q = \sum_j r_j \left| \sum_n \sum_{\ell \neq i} \Phi_{\ell,j}^{n,Q} + \sum_n \Phi_{i,j}^{n,Q} \right|^{\rho}. \tag{3}$$

This assumption combines a true stationarity property and an assumption that, as the transition probabilities $Q^n_{i,j}$, the traveling times (or traveling distances) are not random (or that their randomness can be reduced to a change in the term $r_j$).

**Gain.** The gain is expressed as the difference between the value $v^n_j$ of the good $n$ at the arrival port $j$ and its value $v^n_i$ at the departure port $i$:

$$\mathbf{M}^{n,Q}_i = \sum_j \Phi^{n,Q}_{i,j}(t^+) \underbrace{[v^n_j - v^n_i]}_{M^n_{i,j}}. \tag{4}$$

We use the notation $M^n_{i,j}$ for the commercial margin $v^n_j - v^n_i$.

**Optimization.** All this together reads the maximization program to be solved at each port $i$ for each good $n$:

$$\max_{Q^n_{i,\cdot}} \left\{ \sum_j \Phi^{n,Q}_{i,j} M^n_{i,j} - \sum_j r_j \left| \sum_{m\neq n} \sum_\ell \Phi^{m,Q}_{\ell,j} + \sum_{\ell\neq i} \Phi^{n,Q}_{\ell,j} + \Phi^{n,Q}_{i,j} \right|^\rho - c_n \sum_{j\neq i} \left| \Phi^{n,Q}_{i,j} \right|^\gamma \cdot g(T_{i,j}) \right\}. \tag{5}$$

The split of the crowdedness cost in three terms show the three sources of such costs: the one $\Phi^{n,Q}_{i,j}$ that is controlled at the level of this specific port $i$ and the good $n$, the ones $\sum_{\ell\neq i} \Phi^{n,Q}_{\ell,j}$ that comes from the same good $n$ but coming from other departure ports, and all the rest (i.e. other goods arriving from other ports).

**The Mean Field.** For us, the mean-field is a matrix $[\varphi]_{n,i} = \varphi^n_i$. Its component $n, i$ contains the amount of good $n$ available at port $i$. The controlled flow $\Phi^{n,Q}_{i,j}$ reads as a simple multiplication of the mean-field (what is at port $i$) by the transition (what is sent from this port to the other ports):

$$\Phi^{n,Q}_{i,j} = \varphi^n_i \cdot Q^n_{i,j}. \tag{6}$$

## 2.2 Solving the quadratic version of the MFG

In the quadratic case, i.e. $\rho = \gamma = 2$, the maximization program becomes:

$$\max_{Q^n_{i,\cdot}} \left\{ \sum_j \varphi^n_i Q^n_{i,j} M^n_{i,j} - \sum_j r_j \left[ \sum_{m\neq n} \sum_\ell \varphi^m_\ell Q^m_{\ell,j} + \sum_{\ell\neq i} \varphi^n_\ell Q^n_{\ell,j} + \varphi^n_i Q^n_{i,j} \right]^2 - c_n \sum_{j\neq i} \left( \varphi^n_i Q^n_{i,j} \right)^2 \cdot g(T_{i,j}) \right\}. \tag{7}$$

That is coupled with the static (mass conservation) equation

$$\forall n : \left\{ \forall j : \varphi^n_j = \sum_i \varphi^n_i Q^n_{i,j} \right\} \tag{8}$$

that is saying $\varphi^n$ is the stationary distribution for $Q^n$.

**Proposition 2.1** (mean-field Approximation on Ports of Origin)**.** *When the number of ports is large enough, each agent focuses on the optimization at the level of one port of origin and one good, and perceives the other flows as one independent effect:*

$$\sum_{\ell\neq i} \varphi^n_\ell Q^n_{\ell,j} + \varphi^n_i Q^n_{i,j} \simeq \sum_\ell \varphi^n_\ell Q^n_{\ell,j} + \varphi^n_i Q^n_{i,j} = \varphi^n_j + \varphi^n_i Q^n_{i,j}. \tag{9}$$

*As a consequence the maximization (7) reads*

$$\max_{Q^n_{i,\cdot}} \left\{ \varphi^n_i \sum_j Q^n_{i,j} M^n_{i,j} - \sum_j r_j \left[ \left( \sum_{m\neq n} \varphi^m_j \right) + \varphi^n_j + \varphi^n_i Q^n_{i,j} \right]^2 - c_n \sum_{j\neq i} \left( \varphi^n_i Q^n_{i,j} \right)^2 \cdot g(T_{i,j}) \right\}. \tag{10}$$

*Proof.* This approximation is typically the simplification mean-field Games provide: replacing multiple interactions by one interaction with the mean-field plus one "isolated control".

In this specific case it expresses that when an agent focuses on the control $Q^n_{i,j}$, she or he is not directly sensitive to the isolated decisions taken from other ports of origin. From a continuum of ports perspective, it can be read as

$$\int_{\ell \neq i} \varphi^n_\ell Q^n_{\ell,j}\, dm(\ell) = \varphi^n_j.$$

Once the approximation is accepted, $\sum_\ell \varphi^n_\ell Q^n_{\ell,j} = \varphi^n_j$ is nothing more than (8), that is expressing the stationarity of the mean-field. □

As expected the control equations (10) are coupled thanks to the $N \times J$ dimensional mean-field that is updated via equation (8).

**Solving the control problem.** (10) is not very difficult to solve. Seen from each port $i$ and for each good $n$, the control is a vector $(q_j)_{1\leq j\leq J}$ in the simplex corresponding the fraction of what is available at port $i$ for good $n$ that will be sent to port $j$ (the configuration $j = i$ is possible; it is the amount of good that stays at $i$).

**Theorem 2.2** (Optimal Control for the quadratic case)**.** *The solution of the quadratic maximisation (10) is, for each good $n$ and for an origin $i$ and destination $j$:*

$$Q^n_{i,j} = \frac{w^n_{i,j}}{\varphi^n_i}\left\{\widetilde{M}^n_{i,j} - \sum_\ell \underline{w}^n_{i,\ell}\,\widetilde{M}^n_{i,\ell} - \left[r_j\varphi^\bullet_j - \sum_\ell \underline{w}^n_{i,\ell}\, r_\ell \varphi^\bullet_\ell\right]\right\} + \underline{w}^n_{i,j}, \tag{11}$$

*with the notations:*

$$\widetilde{M}^n_{i,j} := \frac{M^n_{i,j}}{2}, \quad \underline{w}^n_{i,j} := \frac{w^n_{i,j}}{\sum_\ell w^n_{i,\ell}}, \quad w^n_{i,j} = \frac{1}{r_j + c_n\, g(T_{i,j})\mathbf{1}_{j\neq i}}. \tag{12}$$

*Proof.* Using $\lambda$ as Lagrange multiplier to keep $q$ in the simplex, the optimal $q$ has to cancel the augmented gradient (with the notation $\varphi^\bullet_j = \sum_m \varphi^m_j$):[2]

$$\forall n, \forall i: \; \frac{\lambda}{\varphi^n_i} =: \lambda^n_i = M^n_{i,j} - 2r_j(\varphi^\bullet_j + \varphi^n_i q_j) - 2c_n g(T_{i,j})\varphi^n_i q_j \mathbf{1}_{j\neq i}.$$

It is then enough to express $q_j$ (that is $Q^n_{i,j}$) as a function of the Lagrange multiplier:

$$Q^n_{i,j} = q_j = \frac{w^n_{i,j}}{2\varphi^n_i}\left(M^n_{i,j} - 2r_j\varphi^\bullet_j - \lambda^n_i\right), \tag{13}$$

where the Lagrange multiplier is obtained via saturating the constraint[3] to find the value of $\lambda^n_i$, that is

$$\lambda^n_i = \frac{1}{\sum_\ell w^n_{i,\ell}}\left(\sum_\ell w^n_{i,\ell}(M^n_{i,\ell} - 2r_\ell\varphi^\bullet_\ell)\right) - 2\frac{1}{\sum_\ell w^n_{i,\ell}}\varphi^n_i.$$

Injecting $\lambda^n_j$ in the upper expression of $Q^n_{i,j}$ leads to (11). □

The interpretation of the optimal destinations starting from the port $i$ is the following:

[2]It comes from cancelling the partial derivative with respect to $q_j$ of:

$$-\lambda\sum_j q_j + \varphi^n_i \sum_j q_j M^n_{i,j} - \sum_j r_j\left[\left(\sum_{m\neq n}\varphi^m_j\right) + \varphi^n_j + \varphi^n_i q_j\right]^2 - c_n\sum_{j\neq i}(\varphi^n_i q_j)^2 \cdot g(T_{i,j}).$$

[3]it means to write

$$\sum_j Q^n_{i,j} = 1 = \sum_j \frac{w^n_{i,j}}{2\varphi^n_i}\left(M^n_{i,j} - 2r_j\varphi^\bullet_j - \lambda^n_i\right).$$

- the higher the expected margins $\widetilde{M}^n_{i,j}$ (relative to $\sum_\ell \underline{w}^n_{i,\ell}\, \widetilde{M}^n_{i,\ell}$ the barycenter of the margins at destination ports), the more quantity send to $j$;

- the more crowded expected at destination (once again relative to the barycenter of the crowd at destination ports), the less quantity send there;

- the overall costs (especially transportation costs) are penalizing the sent quantity too;

- the averaging scheme weights less the places where crowded costs more (without respect of what is sent, because of the term $r_j$), and the places that suffers from high transportation costs (because of the term $c_n g(T_{i,j})$, except for the considered port $i$, since the transportation cost is zero).

**The positivity constraint.** We do not take care of the positivity constraint in the optimization problem. It is usually a tricky constraints because it leads to sparsity in the solutions. It is nevertheless obvious that in realistic cases a non empty domain for the parameters of the problem (i.e. costs and commercial margins) should exist that will not push transitions to be negative.

The reader may be convinced that it should not happen for realistic problems, since if it is more optimal to import goods from port $j$ than to export goods to $j$, in this framework there is no reason that an export transition matrix has a negative term since it is enough that the export matrix having its origin at port $j$ has a positive term going to port $i$.

**Closing the loop to get the mean-field.** Solving this quadratic problem in the simplex gives the $i$th row of the transition matrix for good $n$. It is enough to stack them to get $[Q^n_{i,j}]_{i,j}$ that is the transition matrix (i.e. the control) for good $n$. Its first eigenvector is $\varphi^n_i$. This can be done for all goods and we can go back to the optimal control problem until we get to a fixed point.

**Units and scaling of costs.** The terms $Q$ and $\underline{w}$ are unit-less, $w$ is in one over cost, and $M$ over costs should have the same unit as $\varphi$ so that the formula (11) makes sense.

## 2.3 Convergence results

**Special case with only one good.** In such a case one can write that $\varphi$ is the row eigenvector of $Q$ associated with the eigenvalue 1 (i.e. $\varphi = \varphi\, Q$). The good point is that not only (10) is quadratic in $Q$, but it is such a way that $Q_{i,j}$ is always multiplied by $\varphi_j$; because of that (13), the obtained expression for $Q_{i,j}$, has a prefactor of $1/\varphi_i$, that will cancel when writing $\sum_i \varphi_i Q_{i,j}$. As a consequence:

$$
\begin{aligned}
\sum_i \varphi_i Q_{i,j} &= \sum_i w_{i,j}\left[\widetilde{M}_{i,j} - \sum_\ell \underline{w}_{i,\ell}\widetilde{M}_{i,\ell} - \left(r_j\varphi_j - \sum_\ell \underline{w}_{i,\ell}\, r_\ell\, \varphi_\ell\right)\right] + \sum_i \underline{w}_{i,j}\varphi_i \\
\Rightarrow \left[1 + r_j \sum_i w_{i,j}\right]\cdot \varphi_j &= \underbrace{\sum_i w_{i,j}\left[\widetilde{M}_{i,j} - \sum_\ell \underline{w}_{i,\ell}\widetilde{M}_{i,\ell}\right]}_{\left(\sum_i w_{i,j}\right)\overline{M}_j} + \sum_\ell\left(\sum_i w_{i,j}\underline{w}_{i,\ell}\right) r_\ell\, \varphi_\ell + \sum_i \varphi_i \underline{w}_{i,j} \\
\Rightarrow \left[1 + r_j \sum_i w_{i,j}\right]\cdot \varphi_j &= \left(\sum_i w_{i,j}\right)\overline{M}_j + \sum_\ell\left(\sum_i w_{i,j}\underline{w}_{i,\ell}\right) r_\ell\, \varphi_\ell + \sum_\ell \varphi_\ell\, \underline{w}_{\ell,j},
\end{aligned}
$$

Thanks to the notation $\overline{M}_j$ for the *overall average relative margin*:

$$
\overline{M}_j = \sum_i \frac{w_{i,j}}{\sum_{i'} w_{i',j}}\left[\widetilde{M}_{i,j} - \sum_\ell \underline{w}_{i,\ell}\widetilde{M}_{i,\ell}\right].
$$

It allows to state the following result:

**Theorem 2.3** (Existence of a mean-field Equilibrium for one good)**.** *The mean-field problem defined by*

$$Q_{i,j} = \frac{w_{i,j}}{\varphi_i}\left\{\widetilde{M}_{i,j} - \sum_{\ell} \underline{w}_{i,\ell}\,\widetilde{M}_{i,\ell} - \left[r_j\varphi_j - \sum_{\ell} \underline{w}_{i,\ell}\, r_\ell\varphi_\ell\right]\right\} + \underline{w}_{i,j},$$

*for the control $Q_{i,j}$ of the probability to send a good from port $i$ to $j$; and by*

$$\forall j:\ \varphi_j = \sum_i \varphi_i Q_{i,j},$$

*for the stationarity of the mean-field, has a unique solution if and only if*

$$\det\left(\left[\left(1 + r_j \sum_i w_{i,j}\right)\mathbf{1}_{j=\ell} - \left\{\left(\sum_i w_{i,j}\underline{w}_{i,\ell}\right) r_\ell + \underline{w}_{\ell,j}\right\}\right]_{j,\ell}\right) \neq 0. \tag{14}$$

The reader can keep in mind that the expression for the optimal control corresponds to (11) with only one good, and it is is the solution of the maximization program (5) in the quadratic case.

*Proof.* The proof is the continuation of the upper observation, since using the notations:

$$R_j := \frac{1}{r_j} + \sum_i w_{i,j}, \quad m_j := \frac{\sum_i w_{i,j}}{r_j}\,\overline{M}_j, \quad C_{j,\ell} := \left(\sum_i w_{i,j}\underline{w}_{i,\ell}\right)\frac{r_\ell}{r_j} + \frac{\underline{w}_{\ell,j}}{r_j},$$

it reads $R_j\varphi_j = m_j + \sum_\ell C_{j,\ell}\,\varphi_\ell$. This gives a *linear expression* for $\varphi$:

$$\sum_\ell [R_j\mathbf{1}_{j=\ell} - C_{j,\ell}]\cdot\varphi_\ell = m_j. \tag{15}$$

In algebraic notations with $m$ being the vector of $m_j$ components and $\Omega$ a matrix with components $\Omega_{j,\ell} := R_j\mathbf{1}_{j=\ell} - C_{\ell,j}$, it reads

$$\Omega\cdot\phi = m. \tag{16}$$

In other terms: *the fixed point exist if and only if $\Omega$ is an invertible matrix.*

A way to write this condition is that the upper determinant is not null. □

A first remark is that, if the solution is a function of the margins, *the existence of an equilibrium seems to be a condition on the costs only.*

CAL: just write it for 2 ports and look at the equation.

**With $N$ goods.** The linear equality over $\varphi$ reads, for any $j$ and $n$:

$$\varphi_j^n + \left(\sum_i w_{i,j}^n\right) r_j \left(\sum_k \varphi_j^k\right) = \left(\sum_i w_{i,j}^n\right)\overline{M}_j^n + \sum_\ell\left[\left(\sum_{i,j}^{n} \underline{w}_{i,\ell}^n\right) r_\ell \left(\sum_k \varphi_\ell^k\right) + \underline{w}_{\ell,j}^n\varphi_\ell^n\right]. \tag{17}$$

# 3 Using data to infer the parameters of the game

## 3.1 The Panjiva dataset

EC to provide a short description of the dataset and some statistics

It is estimated that 80% of all goods transported worldwide travel by sea at some point (Statista). To study the global flow of cargo between seaports, we use S&P Global Panjiva, a large import-export dataset of customs records corresponding to goods entering/exiting ports. These bills of lading correspond to individual shipments ("box on a ship") and capture information such as: the type of product (HS6 code[4]), quantity and weight, the origin and destination, importing/exporting company information, vessel information and more, depending on the country. Table 2 and Table 3 give examples of Panjiva export and import records for the USA, respectively, and Figure 1 shows an example bill of lading form for the USA.

The Panjiva dataset is sourced from government records, port authorities and other data partners in a set of 16 countries (at the time of this research). However, it captures an estimated 40% of global trade flow (S&P Global Market Intelligence [2020a]) due to the hub-and-spoke structure of the global supply chain (Ganapati et al. [2021]). It has been used for a variety of economic applications from tracking medical supply chains (Wu et al. [2022]) to predicting stock performance from corporate shipping activity (Jain and Wu [2023]).

A large stream of literature uses the Panjiva dataset to study the granular flows of specific goods across geographies. For instance, Cho et al. [2022] examine the flow of natural rubber from Sri Lanka to the USA, in the context of measuring deforestation, while VanderWilde et al. [2023] does a similar case study of palm oil transported from Guatemala. Datta et al. [2022] categorize illegal timber imports to the USA. Cho [2020] studies the environmental impacts of avocado trade between Mexico and the USA. Other studies have examined coffee and chocolate (Ramchandani et al. [2020]), cotton (Murphy [2021]), cigarettes (Krylova [2023]), automobiles (Hsu et al. [2022]) and more.

At the same time, other research uses Panjiva to study specific ports of interest and their shipping activity. Flaaen et al. [2023] estimate congestion at the Port of Long Beach and find, during busy periods, ships were rerouted to other US ports. Moreover, S&P Global itself tracks port activity and congestion via Panjiva (S&P Global Market Intelligence [2020b]) (S&P Global Market Intelligence [2021]).

Lastly, in comparison to other maritime shipping datasets, our choice is to use Panjiva in order to focus on the flow of goods across ports. UN Comtrade captures import-exports of specific goods, but only monthly and at an aggregated country-to-country level (Chen et al. [2022]). AIS data tracks the movements of ships, but does not directly include cargo information as granular as customs records. Other datasets in the literature address aspects of the shipping industry and cargo, but do not necessarily focus on the physical flow of granular shipments through ports.

### 3.2 Structure and curation of the data

We limit our analysis of the dataset to the end of 2022. At the time of of our research, the Panjiva dataset consists of 16 countries,[5] with counts of shipments, histories, and number of product types by HS code. By yearly shipment count, the largest exporters are India, Mexico and Vietnam, while the largest importers are Mexico, India and Indonesia. Across countries, there are are about 100+ HS2 codes, 1000+ HS4, and 6000+ HS6. The earliest countries in the dataset are Uruguay (2003), Colombia (2007) and the United States (2007), and all countries have data until the end of 2022 except for Indonesia which was removed in 2021. However, customs records are often handwritten, misclassified, or improperly parsed by computer vision. To curate the classification of goods for inaccuracies, we only use HS codes that are officially listed in any nomenclature release (UNSTATS [2024]). For our Panjiva dataset, this yields a product universe of 97 HS2 codes, 1,253 HS4, and 6,234 HS6 codes.

Similarly, to obtain a curated list of ports, we needed to geocode English port names in Panjiva to geographical coordinates and compute distances between them. To do this, we queried each port name with the Google Maps Geocoding API to obtain geographic coordinates. For those that returned no results, we fuzzy-matched port names against UNLOCODE, a UN-published list of transit locations (UNLOCODE [2024]). To filter out inland river ports, airports and customs offices, we only select ports that are within 100 geodesic kilometers of the ocean, as defined by the NOAA-maintained global shoreline data of (Wessel and Smith [1996]). We also filter out small ports with less than 1000 observations over all time periods. Lastly, we cluster together seaports that are within a 10 kilometer radius, and assign the cluster centroid to the largest port by shipment count. After this filtering and

[4] The WCO's Harmonized System is an international nomenclature for classifying traded products at a granular level, consisting of about 5,000 hierarchical codes. (WCO [2024])

[5] Some countries have been added and removed by the vendor due to data availability (e.g., China).

curation process we obtain a port universe of 2,156 individual ports and 1,140 port clusters connected by 6,425 routes.

Using these curated goods and ports, we visualize their international trade flows in Figure 2 to motivate our analysis. Distances are computed via EUROSTAT's Searoute, which uses historical AIS data (Eurostat [2024]). The figure shows global coverage of port flows with some of the largest flows through the Suez Canal, from South America to Asia.

## 3.3 Constructing high-activity baskets of goods and ports

To motivate our large-sample analysis, we focus on the transport of plastics, which were the 6th-most traded product globally in 2022 (OEC [2024]). Their supply chains are generally difficult to track even with official statistics (Wang et al. [2021]). To track the international trade flow of plastics across ports, we narrow down the curated data above by constructing several measures of trade activity, then use those to construct stable baskets of goods and ports. We focus on HS2 code 39: *Plastics and Articles Thereof*, which yields a product universe of 32 HS4 and 149 HS6, and the associated maritime subnetwork of 601 port clusters connected by 2,620 routes. Figure 3 shows this flow of plastic goods specifically, which is strongest between Americas, Europe, and Asia.

### 3.3.1 Selection of plastic goods

Students to explain to EC why they choose the goods and ports we will focus on, EC to write the associated section.

## 3.4 Recipes to get parameters from the data Inference of parameters of the model

Relation (13) is the one linking the outgoing flows to the other parameters. For inference purposes, it can be rewritten as

$$\varphi_i^n Q_{i,j}^n = w_{i,j}^n \left\{ \widetilde{M}_{i,j}^n - \sum_{\ell} \underline{w}_{i,\ell}^n \, \widetilde{M}_{i,\ell}^n - \left[ r_j^n \sum_m \varphi_j^m - \sum_{\ell} \underline{w}_{i,\ell}^n \, r_\ell \sum_m \varphi_\ell^m \right] \right\} + \varphi_i^n \underline{w}_{i,j}^n.$$

To use data to infer the parameters of the model, one can try to take profit of the small variations of the flows that are observed around the equilibrium. For one observed week $t$ in a database, we can expect this relation to hold

$$\varphi_i^n(t) Q_{i,j}^n(t) \simeq A_{i,j}^n - b_{i,j}^n \sum_m \varphi_j^m(t') + \bar{b}_{i,j}^n \sum_{m,\ell \neq j} \varphi_\ell^m(t') + \underline{w}_{i,j} \varphi_i^n(t) + \epsilon_{i,j}^n(t).$$

$i$ is what you observe. One in principle can put a big bracket around $\varphi_i^n(t) Q_{i,j}^n(t)$ because this observation is observed simultaneously, that it is what it sent at time $t$ from $i$ to $j$. What it sent at time $t$ from $i$ to $j$ is made of what is available $\varphi_i^n(t)$ with the positive coefficient $\underline{w}_{i,j}$, and what will be the crowdedness term $\sum_m \varphi_j^m(t')$ that is in a close future at destination; $m$ indicates all the other goods, including $n$, in the same basket. Instead, $\sum_{m,\ell \neq j} \varphi_\ell^m(t')$ is the sum over all the other goods and the other ports except $j$. Probably the latter term will be very small. There is multi-col linearity because the two of them are quite correlated. The second term is linked to the constraint that the sum is equal to one. We expect that $A_{i,j}$ will be proportional to the margin, and we are assuming that the margin is the same whatever the time. If we observe the variation of what it send corrected by what was available and the crowdedness factor, what remains is a mix of the margin (so why you want to send). The margin is independent of what you have. You can even short sell There is no other reason to think than the margin to one port more than another port except of the margin. $\phi_i$ to capture a little bit of variation that explain what he sent. The term with $-b_{i,j}$ is important is a penalization term that comes from what will be available. The correlation between what he send $\varphi_i$ and what is available (r.h. $\varphi_i$) is expected just because of how you compute that at the same $t$. You observe month after month all that it is send from port $i$ to any other port, and for all the $j$'s whatever they are you

should have a port that it is proportional to what was available but this proportion is different for each destination port. The term $w_{ij}$ is a mixture of all the cost, is an arbitrage between the cost. If you have no margins and no crowdness term then $w_{ij}$ is something as one over the time to travel, i.e., 1 over the transportation cost. It will simply say that we expect that going from $i$ to $j$ is simply one over the transportation cost. You need to do linear regression for each $i$ and $j$. You fix $i$ and $j$ and you simply regress what it send by what it was available, you have different coefficients for each terms and this should reflect the transportation cost. it should be positive. Then we have the crowdeness effect, what will be at the destination port in a close future $t^{'}$. If we sum we expect a negative effect, the more we will have a destination the less we will send. It will capture like a crowdedenss cost. We expect the term to be positive. It may happen that the margins are so large that the previous term can become negative. Find ports where different goods of the same portfolio are sent to help the statistical significance of the crowdeness term; there no reason to have a square matrix.
Here what we do is to estimate the coefficient, it is not bad to do not select data point that we know that add noise to the estimation of the coefficients. We don't really want to do a linear regression to be able to do a prediction. We want to capture data points that are as less noise as possible to estimate the parameter. We do not want to mate a prediction, how much the variation that we see from one month to another is able to say something about the parameter of the cost, and indeed the margins, which are maybe captured from the intercept.

In this expression, $t$ stands for when the decision is taken and $t'$ the week of arrival of the goods at destination. It is not exactly an equality because the term $\bar{b}^n_{i,j} \sum_{m,\ell\neq j} \varphi^m_\ell(t')$ should rather be $\sum_\ell \underline{w}^n_{i,\ell}\, r_\ell \sum_m \varphi^m_\ell$. But it will add a lot of parameters, and be detrimental to the robustness of the inference. To write this equality, we also made the approximation that the crowding cost coefficient $r_j$ are all the same, independently from the considered port, i.e. $\forall j:\ r_j \simeq r$.

The parameters $A^n_{i,j}, b^n_{i,j}, \bar{b}^n_{i,j}, \bar{\bar{b}}^n_{i,j}$ and $\underline{w}_{i,j}$ can be obtained by a linear regression. Since the sign of these coefficients are expected to be non negative, this constraints can be injected in the regression.

Last but not least, since the values $v_i$, and hence the margins $\widetilde{M}^n_{i,j}$ are not known at a multiplicative coefficient, and hence one can replace $wM^n_{i,j}$ by $rwM^n_{i,j}$. Hence in a second stage, one would expect the following equalities to stand:

$$w^n_{i,j} r \left\{ \widetilde{M}_{i,j} - \sum_\ell \underline{w}_{i,\ell}\, \widetilde{M}_{i,\ell} \right\} = A^n_{i,j}, \quad \frac{1}{1 + c_n g(T_{i,j})/r_j} = w^n_{i,j} r_j = b^n_{i,j}. \tag{18}$$

students to perform the inference and explain it; look at the `_old.tex` version of the file for suggestions.

We would like to test

$$\begin{aligned} A^n_{i,j} &= v_j - v_i && (19)\\ A^n_{i,j} &= v^n_j - v^n_i && (20)\\ A^n_{i,j} &= a_n(v_j - v_i) && (21) \end{aligned}$$

There is another version replacing $A^n_{i,j}$ by $A^n_{i,j}/b^n_{i,j}$.

Any of these linear equations boils down to something like

$$Sv = A. \tag{22}$$

where $S$ is a selection matrix. That can be solved using linear regression. The last one is a little more complex since it is non linear. It can either be solved in two steps (first solve (20) and then explain the differences with $a_n$.

Another way to solve (21) is to ask to `scipy.minimize`.

# 4 Some Monte-Carlo simulations What if scenarios? when causality goes further than naive inference

CAL to write what it means, we will see who will perform and explain the MC simulations.

Considering the destinations of goods leaving one port (typically $(\Phi^{n,Q}_{i,j})_{n,j}(t)$ is the matrix of weights of goods $n \in \{1, \dots, N\}$ leaving port $i$ to ports $j \in \{1, \dots, P\}$) as a quantity of interest, statistics under standard stationarity assumptions will capture the law such that each matrix $(\Phi^{n,Q}_{i,j})_{n,j}(t)$ is drawn from, conditionally to variables of context (the day of the week, the month of the year, and may be the offer and demand in the US if $i$ is a located in the US).

With the mean-field Game formulation proposed in this paper, we capture statistically the parameters of the process leading users and owners of the ships to take decisions. The stationarity assumptions are on the costs parameters $r_j$, $c_n$ and $g(T_{i,j})$. That is far more generic even if to really being able to capture them, the considered history of observations has to cover enough diversity of context such that the differences between the costs really influenced the decisions.

### 4.1 Monte-Carlo simulation

(MB) I have added the dataframe containing the useful regressions and the results of all three equations, using the plastics basket, to the 'plastics basket results' folder. We can now start with the simulations.

Note that the $v$ vector has some negative values, which doesn't make sense for monetary values. However, the entire vector can be adjusted by adding any constant $c$, since the results are obtained from equations involving differences any solution for $v$ can be adjusted accordingly.

I have no idea what amount of $c$ to add though...

We are now going to do the same procedure on the other three baskets.

### 4.2 On real data

to be done...

## 5 Future work

we should at least write the convexity terms...

Some notes/statements from some zoom meetings are all written in blue.

- We have good of different baskets. Do we need select always the same ports or can we create a different basket of ports for each basket of goods? Probably the best subset of ports and even the way in which we aggregate ports is different from one good to the other. For instance, for petrol or things like that there are ports that cannot just host any ship with this kind of material. check the Python function seaborn clustermap.

## References


Review of maritime transport 2023. towards a green and just transition. *United Nations publication*, E.23.II.D.23, 2023.

C. M. Anderson, Y.-A. Park, Y.-T. Chang, C.-H. Yang, T.-W. Lee, and M. Luo. A game-theoretic analysis of competition among container port hubs: the case of busan and shanghai. *Maritime Policy & Management*, 35(1):5–26, 2008.

J.-B. Baillon and R. Cominetti. Markovian traffic equilibrium. *Mathematical Programming*, 111(1): 33–56, 2008.

D. Bauso, X. Zhang, and A. Papachristodoulou. Density flow in dynamical networks via mean-field games. *IEEE Transactions on Automatic Control*, 62(3):1342–1355, 2016.

R. J. Boucherie. Aggregation of markov chains. *Stochastic processes and their applications*, 45(1): 95–114, 1993.

P. Cardaliaguet. Notes on mean field games. Technical report, Technical report, 2010.

R. Carmona, F. Delarue, et al. *Probabilistic theory of mean field games with applications I-II*. Springer, 2018.

C. Chen, Z. Jiang, N. Li, H. Wang, P. Wang, Z. Zhang, C. Zhang, F. Ma, Y. Huang, X. Lu, et al. Advancing un comtrade for physical trade flow analysis: review of data quality issues and solutions. *Resources, Conservation and Recycling*, 186:106526, 2022.

G. Chevalier, J. Le Ny, and R. Malhamé. A micro-macro traffic model based on mean-field games. In *2015 American Control Conference (ACC)*, pages 1983–1988. IEEE, 2015.

K. Cho. *Environmental impacts of the US-Mexico avocado supply chain*. PhD thesis, 2020.

K. Cho, B. Goldstein, D. Gounaridis, and J. P. Newell. Hidden risks of deforestation in global supply chains: A study of natural rubber flows from sri lanka to the united states. *Journal of Cleaner Production*, 349:131275, 2022.

D. Datta, S. Muthiah, J. Simeone, A. Meadows, and N. Ramakrishnan. Scrutinizing shipment records to thwart illegal timber trade. *arXiv preprint arXiv:2208.00493*, 2022.

T. M. H. Do, G.-K. Park, K. Choi, K. Kang, and O. Baik. Application of game theory and uncertainty theory in port competition between hong kong port and shenzhen port. *International Journal of e-Navigation and Maritime Economy*, 2:12–23, 2015.

Eurostat. Searoute: Maritime route computation library. GitHub repository, 2022. URL `https://github.com/eurostat/searoute`. Latest commit: Jan 10, 2022.

Eurostat. searoute: A python library for calculating sea routes. `https://github.com/eurostat/searoute`, 2024. Accessed: 2024-08-18.

A. Flaaen, F. Haberkorn, L. Lewis, A. Monken, J. Pierce, R. Rhodes, and M. Yi. Bill of lading data in international trade research with an application to the covid-19 pandemic. *Review of International Economics*, 31(3):1146–1172, 2023.

S. Ganapati, W. F. Wong, and O. Ziv. Entrepot: Hubs, scale, and trade costs. Technical report, National Bureau of Economic Research, 2021.

D. A. Gomes, J. Mohr, and R. R. Souza. Discrete time, finite state space mean field games. *Journal de mathématiques pures et appliquées*, 93(3):308–328, 2010.

O. Guéant. Existence and uniqueness result for mean field games with congestion effect on graphs. *Applied Mathematics & Optimization*, 72(2):291–303, 2015.

J. Hsu, Z. Li, and J. Wu. Supply chain restructuring: Nearshoring in response to regional value content requirements. *Available at SSRN 4246225*, 2022.

M. Huang, R. P. Malhamé, and P. E. Caines. Large population stochastic dynamic games: closed-loop mckean-vlasov systems and the nash certainty equivalence principle. 2006.

M. Ishii, P. T.-W. Lee, K. Tezuka, and Y.-T. Chang. A game theoretical analysis of port competition. *Transportation Research Part E: Logistics and Transportation Review*, 49(1):92–106, 2013.

N. Jain and D. Wu. Can global sourcing strategy predict stock returns? *Manufacturing & Service Operations Management*, 25(4):1357–1375, 2023.

E. N. Kaselimi, T. E. Notteboom, and B. De Borger. A game theoretical approach to competition between multi-user terminals: the impact of dedicated terminals. *Maritime Policy & Management*, 38(4):395–414, 2011.

A. Komaromi, D. A. Cerdeiro, M. D. A. Cerdeiro, and Y. Liu. *Supply chains and port congestion around the world*. International Monetary Fund, 2022.

Y. Krylova. Smugglers' paradises in the global economy: Growing threats of hubs of illicit trade to security & sustainable development. 2023.

J.-M. Lasry and P.-L. Lions. Mean field games. *Japanese journal of mathematics*, 2(1):229–260, 2007.

M. Luo, L. Liu, and F. Gao. Post-entry container port capacity expansion. *Transportation Research Part B: Methodological*, 46(1):120–138, 2012.

A. Micco and N. Pérez. Determinants of maritime transport costs. 2002.

L. Murphy. Laundering cotton: How xinjiang cotton is obscured in international supply chains. 2021.

H.-O. Nguyen, A. Chin, J. Tongzon, and M. Bandara. Analysis of strategic pricing in the port sector: The network approach. *Maritime Economics & Logistics*, 18:264–281, 2016.

OEC. Plastics & articles thereof, 2024. URL `https://oec.world/en/profile/hs/plastics-articles-thereof`. Accessed: 2024-08-23.

P. Ramchandani, H. Bastani, and K. Moon. Responsible sourcing: the first step is the hardest. *Available at SSRN 3656497*, 2020.

R. Salhab, J. Le Ny, and R. P. Malhamé. A mean field route choice game model. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 1005–1010. IEEE, 2018.

Y. Sheffi. *Urban transportation networks*, volume 6. Prentice-Hall, Englewood Cliffs, NJ, 1985.

S&P Global Market Intelligence. Ship to shore: Mapping the global supply chain with panjiva shipping data in xpressfeed. Technical report, S&P Global Market Intelligence, 2020a. URL `https://www.spglobal.com/marketintelligence/en/documents/ship-to-shore-mapping-the-global-supply-chain-with-panjiva-shipping-data-in-xpressfeed.pdf`.

S&P Global Market Intelligence. Tracking vessel activity at key ports throughout covid-19, 2020b. URL `https://www.spglobal.com/marketintelligence/en/mi/research-analysis/tracking-vessel-activity-at-key-ports-throughout-covid19.html`. Accessed: 2024-08-12.

S&P Global Market Intelligence. Dsv panalpina diversifies further as systemic congestion continues, 2021. URL `https://panjiva.com/research/dsv-panalpina-diversifies-further-as-systemic-congestion-continues/40707`. Accessed: 2024-08-12.

Statista. Ocean shipping. `https://www.statista.com/topics/1728/ocean-shipping/`. Accessed: 2024-08-12.

T. Tanaka, E. Nekouei, A. R. Pedram, and K. H. Johansson. Linearly solvable mean-field traffic routing games. *IEEE Transactions on Automatic Control*, 66(2):880–887, 2020.

UNLOCODE. UN/LOCODE Code List by Country and Territory. `https://unece.org/trade/cefact/unlocode-code-list-country-and-territory`, 2024. Accessed: 2024-08-18.

UNSTATS. United nations economic classifications. `https://unstats.un.org/unsd/classifications/Econ`, 2024. Accessed: 2024-08-18.

USA GSA. U.s. government bill of lading (gsa forms library). `https://www.gsa.gov/reference/forms/u-s-government-bill-of-lading`, 2024. Accessed: 2024-08-18.

C. P. VanderWilde, J. P. Newell, D. Gounaridis, and B. P. Goldstein. Deforestation, certification, and transnational palm oil supply chains: Linking guatemala to global consumer markets. *Journal of Environmental Management*, 344:118505, 2023.

C. Wang, Y. Liu, W.-Q. Chen, B. Zhu, S. Qu, and M. Xu. Critical review of global plastics stock and flow data. *Journal of Industrial Ecology*, 25(5):1300–1317, 2021.

WCO. WCO Harmonized System Nomenclature. https://www.wcoomd.org/en/topics/nomenclature/overview.aspx, 2024. Accessed: 2024-08-18.

P. Wessel and W. H. F. Smith. A global, self-consistent, hierarchical, high-resolution shoreline database, 1996. URL https://www.ngdc.noaa.gov/mgg/shorelines/. Accessed: 2024-08-18.

B. W. Wiegmans, A. V. D. Hoest, and T. E. Notteboom. Port and terminal selection by deep-sea container operators. *Maritime Policy & Management*, 35(6):517–534, 2008.

J. Wu, S. Agca, and J. R. Birge. The impact of the covid-19 pandemic on global sourcing of medical supplies. *European Society of Medicine*, 2022.

T. L. Yip, J. J. Liu, X. Fu, and J. Feng. Modeling the effects of competition on seaport terminal awarding. *Transport Policy*, 35:341–349, 2014.

Y. Yu, D. Calderone, S. H. Li, L. J. Ratliff, and B. Açıkmeşe. Variable demand and multi-commodity flow in markovian network equilibrium. *Automatica*, 140:110224, 2022.

W. Zhuang, M. Luo, and X. Fu. A game theory analysis of port specialization—implications to the chinese port industry. *Maritime Policy & Management*, 41(3):268–287, 2014.

## A Tables and figures

| field | data type | description |
|---|---|---|
| panjiva_record_id | numeric | The Panjiva-assigned ID to a shipment. |
| bill_of_lading_number | string | The bill of lading of the shipment. |
| shpmt_date | date | The date of processing by customs authorities. |
| shp_name | string | Shipper name. |
| shp_full_address | string | Shipper's full address. |
| shp_route | string | Shipper's address. |
| shp_city | string | Shipper's city. |
| shp_state_region | string | Shipper's region. |
| shp_postal_code | string | Shipper's postal code. |
| shp_country | string | Shipper's country. |
| shp_panjiva_id | numeric | Shipper's Panjiva-assigned ID. |
| shp_original_format | string | Shipper's info in original format. |
| carrier | string | |
| shpmt_destination | string | Destination of the shipment. |
| port_of_lading | string | Port of lading. |
| port_of_lading_region | string | Port of lading, region. |
| port_of_lading_country | string | Port of lading, country. |
| port_of_unlading | string | Port of unlading. |
| port_of_unlading_region | string | Port of unlading, region. |
| port_of_unlading_country | string | Port of unlading, country. |
| place_of_receipt | string | Place the cargo was received. |
| vessel | string | Vessel name. |
| vessel_country | string | Vessel's country. |
| is_containerized | boolean | Whether transported in a container. |
| volume_teu | numeric | Cargo in TEUs. |
| item_quantity | string | Cargo quantity, if applicable. |
| weight_kg | numeric | Cargo weight in kilograms. |
| weight_t | numeric | Cargo weight in tons. |
| weight_original_format | string | Original text for the cargo weight. |
| value_of_goods_usd | numeric | Dollar value of the cargo. May be estimated. |
| equipment_type | string | Handling info. |
| equipment_dimensions | string | Handling info. |
| divided_lcl | string | Less than container load. |
| hs_code | string | HS code. May be parsed/classified (e.g. USA). |

*Table 1:* Selected shipment fields in Panjiva, USA exports. Fields vary based on country and trade direction.

## B Old Sections

For the sake of the reader, we collect and define here the main notations we will use in the rest of the paper; as said, our problem formulation is similar to that of Tanaka et al. [2020].

1. $N \in \mathbb{N}_+$, with $\mathbb{N}_+$ the set of natural numbers strictly greater than zero, denotes the number of players in the game. ambiguity: player, good, ship $\rightarrow$ it is a game of decisions in ports

2. $T$: it denotes the number of stages of the (dynamic) game; $\mathcal{T} := \{1, \ldots, T\}$ denotes the time steps of the game.

3. $\mathcal{G} = (\mathcal{V}, \mathcal{E})$: it denotes a directed graph, where $\mathcal{V} = \{1, 2, \ldots, V\}$ is the set of nodes (intersections) and $\mathcal{E} = \{1, 2, \ldots, E\}$ is the set of directed edges.

4. Given a directed graph, for each $i \in \mathcal{V}$, $\mathcal{V}(i) \subseteq \mathcal{V}$ denotes the set of intersections to which there is a directed link from the intersection $i$.

5. Given a directed graph, $i_{n,t} \in \mathcal{V}$ denotes the node at which the $n^{th}$ player is located at time step $t$.

6. $\Delta^J$, $J \in \mathbb{N}_+$: it denotes the $J$-dimensional probability simplex.

| field | value |
|---|---|
| panjiva_record_id | 51975201 |
| bill_of_lading_number | MEDUU0207917 |
| shpmt_date | 2021-07-04 00:00:00 |
| shp_name | Sims Global Commodities Pte Lt |
| shp_full_address | 80, ROBINSON ROAD, UNIT #26-00 068898 JERSEY CITY NEW JERSEY 07305 |
| shp_route | 200 Morris Pesin Drive |
| shp_city | Jersey City |
| shp_state_region | New Jersey |
| shp_postal_code | 07305 |
| shp_country | United States |
| shp_panjiva_id | 61186754.000000 |
| shp_original_format | Sims Global Commodities Pte Ltd |
| carrier | MSCU - Msc Mediterranean Shipping Company S A |
| shpmt_destination | Thailand |
| port_of_lading | Port of Oakland, Oakland, California |
| port_of_lading_region | Pacific Region |
| port_of_lading_country | United States |
| port_of_unlading | Laem Chabang, Thailand |
| port_of_unlading_region | South-Eastern Asia |
| port_of_unlading_country | Thailand |
| place_of_receipt | NaN |
| vessel | Cap San Juan |
| vessel_country | NaN |
| transport_flight_code | NaN |
| is_containerized | True |
| volume_teu | 1.000000 |
| item_quantity | See 'Container Quantities' column |
| weight_kg | 21527.000000 |
| weight_t | 21.527000 |
| weight_original_format | 21527.0 kg |
| value_of_goods_usd | 107000.000000 |
| equipment_type | NaN |
| equipment_dimensions | NaN |
| divided_lcl | NaN |
| file_date | 2021-07-18 00:00:00 |
| hs_code | Parsed: 7404.00 |

*Table 2:* USA exports, example shipment in Panjiva. The product (HS 7404.00, 2017 nomenclature) corresponds to copper waste and scrap.

7. Given a directed graph, $Q^n_{i,j}(t) = \{Q^n_{i,j}(t)\}_{j \in \mathcal{V}(i)} \in \Delta^{|\mathcal{V}(i)|-1}$ denotes the probability distribution according to which player $n$ at intersection $i$ selects the next destination $j \in \mathcal{V}(i)$.
8. $Q_{n,t} = \{Q^n_i(t)\}_{i\in\mathcal{V}}$: it denotes the collection of probability distributions in 7., which is the policy of player $n$ at time $t$; $\mathcal{Q} = \{\{Q_i\}_{i\in\mathcal{V}} : Q_i \in \Delta^{|\mathcal{V}(i)|-1}\, \forall i \in \mathcal{V}\}$ is the space of admissible policies.
9. $P^n(0) = P_0 \in \Delta^{|\mathcal{V}|-1}$: it denotes the probability distribution of the initial location of the players.
10. $P^n(t) = \{P^n_i(t)\}_{i\in\mathcal{V}}$: it denotes the probability distribution of the location of player $n$ at time $t$; it is computed recursively by $P^n_j(t+1) = \sum_i P^n_j(t+1) Q^n_{i,j}(t)$, $\forall t \in \mathcal{T}$, $j \in \mathcal{V}$.

## B.1 Organization of the paper.

The rest of the paper is organized as follows. G.L.: to be written at the end.

## B.2 Intuition towards a Mean Field model of shipping From a set of ships to a fleet

Let explore what happens if we look at the most atomic level:

- We consider only one good for simplicity of the notation, hence we will not mention the index $n$;
- The fleet is made of $K$ ships, and we assume the capacity of every ship is the same, of one standardized unit;
- The position of ship $k \in [[K]]$ is stored in a variable $S_k(t)$ at time $t$, i.e. $S_k(t) = i$ when the ship number $k$ is at port $i$ at time $t$;
- We assume a random policy by the owners of the ship, like in Tanaka et al. [2020]: each policy is a transition matrix $q_k(t)$, hence the position of the ship $k$ is a random vector $p_k(t) = [\mathbb{P}(S_k(t) = i)]_{1 \le i \le J}$ It belongs to the $J-1$-dimensional probability simplex.
- As a consequence, the $K$-dimensional vector $F(t)$ of the capacity at ports at time $t$ is (measured in standardized unit)

$$F(t) = \sum_{k=1}^{K} p_k(t), \tag{23}$$

and at time $t+1$ it is:

$$F(t+1) = \sum_{k=1}^{K} p_k(t)\, q_k(t). \tag{24}$$

At this stage, I am not sure there is a equivalent Markov chain on $F/\bar{F}$, where $\bar{F}$ is the total capacity of the fleet (note that in this model $\bar{F}$ does not change over time). In all generality, it may be complicated, see Boucherie [1993].

Nevertheless, we know that if each random policy $q_k$ has a stationary state $\mu_k$, then $F$ is stationary, simply by letting time go to infinity in (24), and

$$\mu_F = \sum_k \mu_k. \tag{25}$$

**A mean-field of ships.** On the other hand, we could assume a mean-field of ships, meaning that at each time $t$, the ships are randomly (uniformly) redistributed to the owners. In such a case, each owner "receives"

$$\bar{p}(t) = \frac{1}{K} \sum_k p_k(t),$$

random ships, and as a consequence applies her random policy $q_k$ to it. In such a framework

$$F(t+1) = \sum_k \left( \frac{1}{K} \sum_{k'=1}^{K} p_\ell(t) \right) q_k(t) = F(t) \frac{1}{K} \sum_k q_k(t). \tag{26}$$

And then, with the notation

$$Q(t) = \frac{1}{K} \sum_k q_k(t), \tag{27}$$

we get a natural Markov chain on $P = F/\bar{F}$:

$$\frac{F(t+1)}{\bar{F}} = \frac{F(t)}{\bar{F}}\, Q(t) \Leftrightarrow F(t+1) = F(t)\, Q(t). \tag{28}$$

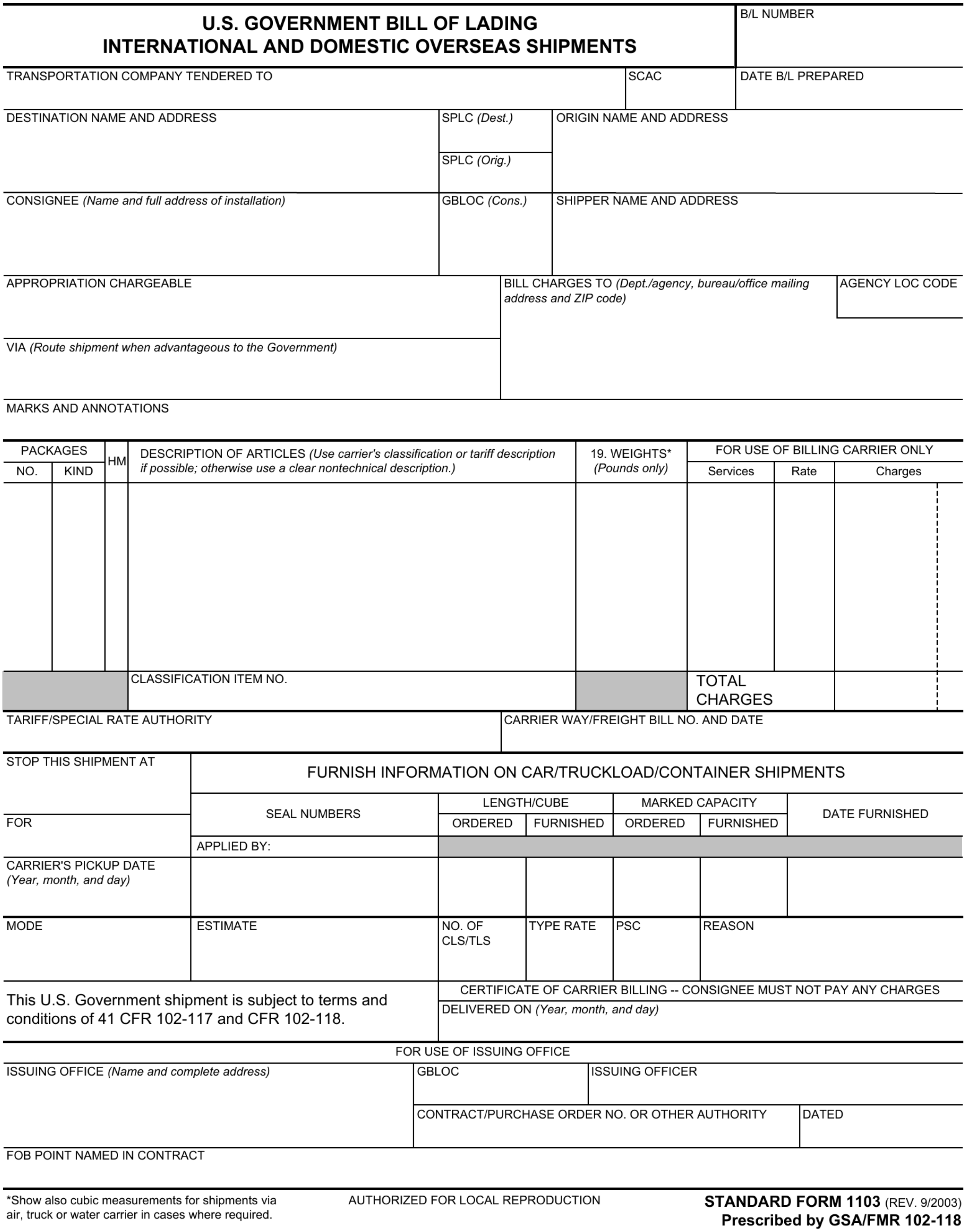
Note that it corresponds to the interchangeability of the owners of the ships: if they all have the same preference, it is obvious. Otherwise, we are back to the assumption of a "representative agent", that is now properly defined: she receives a random ship and applies her desired policy (that takes her individual preferences into account). If we say that each policy $q_k(t)$ is a generic function of the state (that is the collection of the $p_k$, and of the preferences $a_k$ of this specific owner, namely

$$q_k(t) = \mathbf{Q}(p_k, (p_\ell)_{\ell \neq k}, a_k),$$

then the equation (27) tells us that our representative agent applies a policy that is "averaged over the preferences":

$$Q(t) = \frac{1}{K} \sum_k \mathbf{Q}(\bar{p}(t), (\bar{p}(t))_{1 \le u \le K-1}, a_k).$$

The second argument is indeed no more useful since, after the uniform reshuffling, all the $p_\ell$ are the same; they are all equal to $\bar{p}$.



**U.S. GOVERNMENT BILL OF LADING**
**INTERNATIONAL AND DOMESTIC OVERSEAS SHIPMENTS**

B/L NUMBER

TRANSPORTATION COMPANY TENDERED TO | SCAC | DATE B/L PREPARED

DESTINATION NAME AND ADDRESS | SPLC *(Dest.)* | SPLC *(Orig.)* | ORIGIN NAME AND ADDRESS

CONSIGNEE *(Name and full address of installation)* | GBLOC *(Cons.)* | SHIPPER NAME AND ADDRESS

APPROPRIATION CHARGEABLE | BILL CHARGES TO *(Dept./agency, bureau/office mailing address and ZIP code)* | AGENCY LOC CODE

VIA *(Route shipment when advantageous to the Government)*

MARKS AND ANNOTATIONS

| PACKAGES | | HM | DESCRIPTION OF ARTICLES *(Use carrier's classification or tariff description if possible; otherwise use a clear nontechnical description.)* | 19. WEIGHTS* *(Pounds only)* | FOR USE OF BILLING CARRIER ONLY | | |
|---|---|---|---|---|---|---|---|
| NO. | KIND | | | | Services | Rate | Charges |
| | | | | | | | |
| | | | CLASSIFICATION ITEM NO. | | TOTAL CHARGES | | |

TARIFF/SPECIAL RATE AUTHORITY | CARRIER WAY/FREIGHT BILL NO. AND DATE

STOP THIS SHIPMENT AT

FOR

CARRIER'S PICKUP DATE *(Year, month, and day)*

FURNISH INFORMATION ON CAR/TRUCKLOAD/CONTAINER SHIPMENTS

| SEAL NUMBERS | LENGTH/CUBE | | MARKED CAPACITY | | DATE FURNISHED |
|---|---|---|---|---|---|
| | ORDERED | FURNISHED | ORDERED | FURNISHED | |
| APPLIED BY: | | | | | |

MODE | ESTIMATE | NO. OF CLS/TLS | TYPE RATE | PSC | REASON

This U.S. Government shipment is subject to terms and conditions of 41 CFR 102-117 and CFR 102-118.

CERTIFICATE OF CARRIER BILLING -- CONSIGNEE MUST NOT PAY ANY CHARGES

DELIVERED ON *(Year, month, and day)*

FOR USE OF ISSUING OFFICE

ISSUING OFFICE *(Name and complete address)* | GBLOC | ISSUING OFFICER

CONTRACT/PURCHASE ORDER NO. OR OTHER AUTHORITY | DATED

FOB POINT NAMED IN CONTRACT

*Show also cubic measurements for shipments via air, truck or water carrier in cases where required.

AUTHORIZED FOR LOCAL REPRODUCTION

**STANDARD FORM 1103** (REV. 9/2003)
**Prescribed by GSA/FMR 102-118**

*Figure 1:* Example bill of lading form, specifically for US government shipments. (USA GSA [2024])

| field | value |
|---|---|
| panjiva_record_id | 108954839 |
| bill_of_lading_number | GVAI15600157 |
| arrival_date | 2015-03-06 00:00:00 |
| con_name | Smi |
| con_full_address | 815 WEST WHITNEY RD FAIRPORT NEW YORK 14450 US |
| con_route | 815 Whitney Road West |
| con_city | Fairport |
| con_state_region | New York |
| con_postal_code | 14450 |
| con_country | United States |
| con_panjiva_id | 27813270 |
| con_original_format | SMI C/O LIDESTRI FOODS INC<br>815 WEST WHITNEY ROAD FAIRPORT NEW YORK 14450 US |
| shp_name | Smi SpA |
| shp_full_address | VIA PIAZZALUNGA, 30 SAN GIOVANNI BIANCO, BERGAMO 24015 |
| shp_route | 30 Via Piazzalunga |
| shp_city | San Giovanni Bianco |
| shp_state_region | Lombardia |
| shp_postal_code | 24015 |
| shp_country | Italy |
| shp_panjiva_id | 44348484.000000 |
| shp_original_format | SMI S.P.A. VIA CARLO CERESA,<br>10 24015 S.GIOVANNI BIANCO BERGAMO IT240015 IT. TE034540111 |
| carrier | GVAI - Gava International Freight Spa |
| notify_party | NaN |
| notify_party_scac | ZIMU - Zim Integrated Shipping Services Ltd |
| bill_of_lading_type | House |
| master_bill_of_lading_number | ZIMUGOA343047 |
| shpmt_origin | Italy |
| shpmt_destination | New York, New York |
| shpmt_destination_region | New York Region |
| port_of_unlading | New York, New York |
| port_of_unlading_region | New York Region |
| port_of_lading | Genoa, Italy |
| port_of_lading_region | Southern Europe |
| port_of_lading_country | Italy |
| place_of_receipt | TV |
| transport_method | Maritime |
| vessel | ZIM CONSTANZA |
| vessel_voyage_id | 033W |
| vessel_imo | NaN |
| is_containerized | True |
| volume_teu | 4.000000 |
| quantity | 6 PCS |
| measurement | NaN |
| weight_kg | 10180.000000 |
| weight_t | 10.180000 |
| weight_original_format | 10180 KG |
| value_of_goods_usd | NaN |
| frob | NaN |
| manifest_number | 461838 |
| inbond_code | NaN |
| number_of_containers | 2 |
| has_lcl | NaN |
| file_date | 2015-03-10 00:00:00 |
| hs_code | Classified: 8422.90; Classified: 8422.90 |

*Table 3:* USA imports, example shipment in Panjiva. The product (HS 8422.90, 2012 nomenclature) corresponds to parts for dishwashers, machines for cleaning/bottling, etc.

| Country | Trade Direction | # Shipments | # Shipments / Yr. | Start | End | # HS2 | # HS4 | # HS6 |
|---|---|---|---|---|---|---|---|---|
| Chile | export | 6.53 | 0.47 | 2009-01-01 | 2022-12-31 | 97 | 1179 | 4696 |
| Colombia | export | 15.31 | 0.96 | 2007-01-01 | 2022-12-01 | 88 | 1363 | 5245 |
| Costa Rica | export | 12.46 | 1.38 | 2014-01-01 | 2022-12-31 | 97 | 1327 | 4774 |
| Ecuador | export | 1.63 | 0.18 | 2014-01-01 | 2022-12-31 | 0 | 0 | 0 |
| India | export | 260.63 | 37.22 | 2016-01-01 | 2022-12-31 | 102 | 1426 | 6325 |
| Indonesia | export | 25.63 | 9.62 | 2019-02-01 | 2021-09-30 | 88 | 1314 | 4997 |
| Mexico | export | 218.18 | 18.17 | 2011-01-01 | 2022-12-31 | 89 | 1409 | 5700 |
| Pakistan | export | 5.87 | 1.17 | 2018-01-01 | 2022-12-31 | 87 | 1192 | 3640 |
| Panama | export | 0.55 | 0.04 | 2009-01-01 | 2022-12-31 | 88 | 1215 | 4049 |
| Paraguay | export | 1.45 | 0.16 | 2014-01-02 | 2022-12-31 | 85 | 962 | 3078 |
| Peru | export | 11.95 | 0.76 | 2007-03-23 | 2022-12-31 | 88 | 1366 | 5153 |
| Sri Lanka | export | 5.91 | 0.84 | 2016-01-01 | 2022-12-31 | 99 | 1440 | 5365 |
| United States | export | 33.64 | 2.44 | 2009-04-01 | 2022-12-31 | 99 | 1402 | 6043 |
| Uruguay | export | 1.33 | 0.07 | 2003-08-11 | 2022-12-28 | 87 | 1227 | 4063 |
| Venezuela | export | 0.15 | 0.02 | 2014-01-01 | 2022-12-01 | 88 | 1031 | 2901 |
| Vietnam | export | 179.6 | 35.92 | 2018-01-01 | 2022-12-31 | 87 | 1356 | 5261 |
| Chile | import | 51.42 | 3.67 | 2009-01-02 | 2022-12-31 | 88 | 1408 | 6033 |
| Colombia | import | 46.43 | 2.9 | 2007-01-01 | 2022-12-31 | 88 | 1383 | 5725 |
| Costa Rica | import | 96.19 | 10.68 | 2014-01-01 | 2022-12-31 | 97 | 1422 | 5743 |
| Ecuador | import | 1.59 | 0.18 | 2014-01-01 | 2022-12-31 | 0 | 0 | 0 |
| India | import | 266.57 | 38.07 | 2016-01-01 | 2022-12-31 | 98 | 1436 | 6096 |
| Indonesia | import | 56.84 | 21.34 | 2019-02-01 | 2021-09-30 | 98 | 1430 | 5570 |
| Mexico | import | 760.83 | 63.37 | 2011-01-01 | 2022-12-31 | 89 | 1415 | 5926 |
| Pakistan | import | 12.44 | 2.07 | 2017-01-01 | 2022-12-31 | 98 | 1414 | 5559 |
| Panama | import | 16.82 | 1.2 | 2008-12-31 | 2022-12-30 | 88 | 1401 | 5757 |
| Paraguay | import | 8.54 | 0.95 | 2014-01-01 | 2022-12-30 | 87 | 1289 | 5009 |
| Peru | import | 112.6 | 8.66 | 2010-01-01 | 2022-12-31 | 88 | 1357 | 5570 |
| Sri Lanka | import | 11.05 | 1.58 | 2016-01-01 | 2022-12-31 | 88 | 1369 | 5390 |
| United States | import | 165.15 | 10.32 | 2007-01-01 | 2022-12-31 | 100 | 1597 | 6195 |
| Uruguay | import | 13.53 | 0.68 | 2003-01-09 | 2022-12-28 | 87 | 1345 | 5433 |
| Venezuela | import | 3.78 | 0.42 | 2014-01-01 | 2022-12-01 | 88 | 1300 | 4803 |
| Vietnam | import | 3.78 | 0.42 | 2014-01-01 | 2022-12-01 | 88 | 1300 | 4803 |

*Table 4:* Raw data from the 16 countries covered by Panjiva, with counts and coverage. Officially, there are nearly seven thousand unique HS6 codes across all periodic nomenclatures. Shipment counts in millions. HS codes not available for Ecuador.

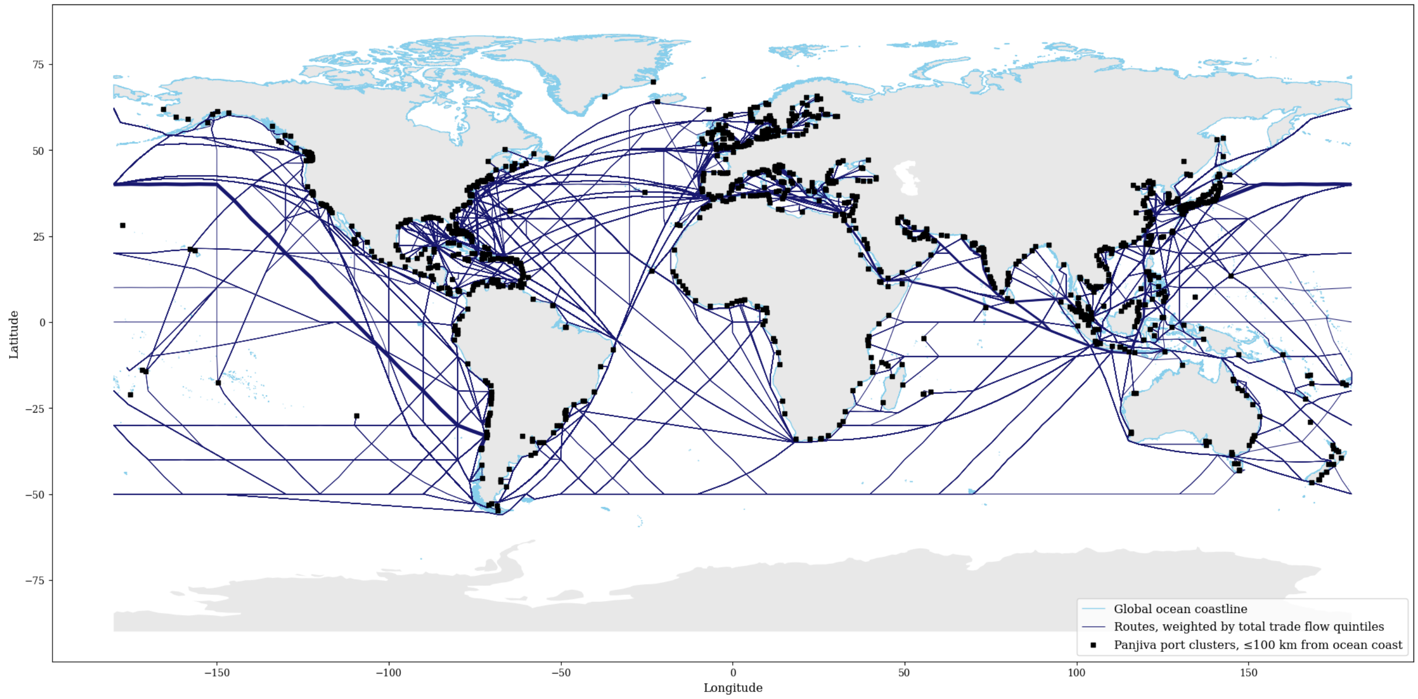



*Figure 2:* The global maritime network in Panjiva, over all times, with computed routes for the curated universe of goods and port clusters.

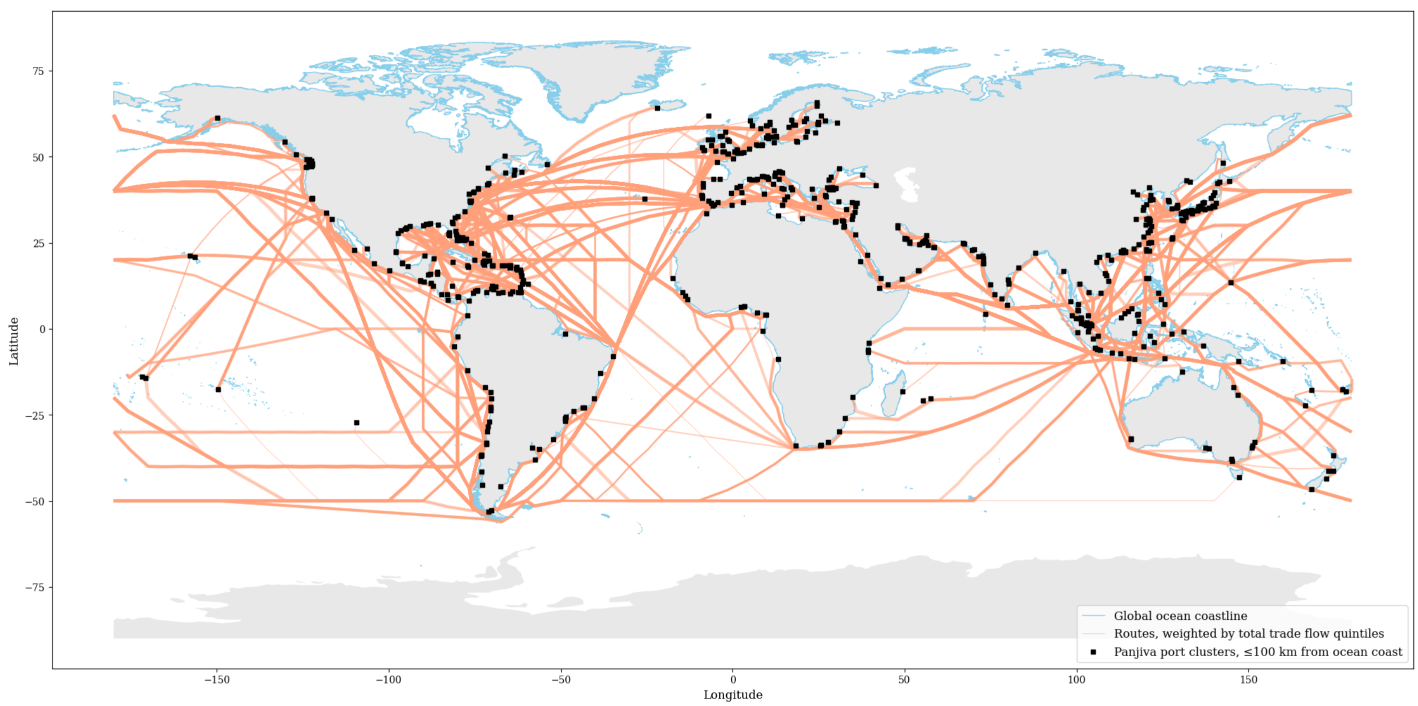



*Figure 3:* Sub-network of Figure (2), for plastics (HS39) specifically.