InPlain

Port Statistics: The Rise of a New Era for Open Data?

Dong Yang, Xiwen Bai, Venus Lun

Practitioners and scholars in the maritime sector have suffered from data availability for a long time. Conventionally, in the public domain, port statistics refers to official data released by port authorities or shipping companies; they are usually delayed, heterogeneous in terms of reporting methodology or reporting format, and most importantly, not easily accessible. With the Automatic Identification System (AIS), it becomes possible to produce port statistics at higher frequencies and on a global level in an alternative way. In the previous two years, we have developed various techniques to build a global-scale port monitoring platform covering three sets of key indicators on port performance: throughput, congestion, and connectivity. This platform generates indicators for global ports and at high frequencies. It does not only require fewer inputs and save running time but can be generalized to accommodate ports of different geographical and economic characteristics. Of course, accuracy should be enforced.

We first developed an iterative multi-attribute clustering algorithm to automatically identify berth and anchorage areas, and it can be applied globally. Tracking a ship within the layout of a port thus becomes possible, i.e., berthing, mooring, and movement. Based on this, we estimate high-frequency throughput, congestion level, and global connectivity index. In the following, we briefly introduce the construction of the platform.

Automatic berth and anchorage area identification

Ports around the world have vastly different shapes of anchorages and berths. Portraying port zones is crucial for shipping traffic analysis in ports. Existing research rely on nautical charts, while the nautical chart suffers from some drawbacks, such as static and seldomly updated, hard to collect globally and sometimes cannot reflect the dynamics of ship movement, to name but a few. We propose a creative, spatial clustering algorithm leveraging both AIS data and domain knowledge (e.g., the sequential of ship mooring and the heading of ships during mooring) based on the difference in the density of mooring points. This technique consists of two layers of clustering, effectively eliminating noises; more specifically, iteration in the second layer functions as an innovative approach to parameter setting. In the first layer, we derive the trajectories of ships at a particular port from AIS data. The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is used to remove noises and cluster all the mooring points of each ship that has been filtered by speed. The first layer DBSCAN can identify mooring area but is insufficient to distinguish between berthing and anchorage areas. As ship density in the berthing area is much higher than at anchorage, in the second layer, we employ DBSCAN again to distinguish the two. The points identified in the clusters are considered as the berthing area while noises represent the anchorage area. Parameter setting in a clustering algorithm may lead to huge

variations in accuracy. The result of DBSCAN strongly depends on initial parameters, especially, the radius of a neighborhood concerning given points and the minimum number of points to form a dense region. Setting them for every port is a formidable amount of work. It is necessary to find a generalized setting for the two parameters. Here, we run iterations of DBSCAN to minimize the gradient until the algorithm converges to a predetermined threshold.

Furthermore, the non-spatial values (e.g., headings, timestamps) in AIS provide additional information that is useful in distinguishing between different berths. In this study, we further add a rule-based algorithm based on domain knowledge, such the heading of ship in berth and anchorage are different, ships will never overlap in one berth to further improve the identification. Figure 1 shows the algorithm being applied to Ningbo-Zhoushan Port, China, chosen for its complicated layout of terminal and anchorage areas. Our algorithm is proved to be accurate in distinguishing the berthing and anchorage areas.

(a) The terminals and anchorages

(b) Berths in Yuandong port zone

Figure 1. Berth and anchorage identification of Ningbo-Zhoushan Port

Port throughput estimation

Container throughput refers to the quantity of both loaded and unloaded containers at a port complex in a given period. With available ship data, we calculate it by multiplying the berthing duration of a ship with the handling efficiency (turnover rate) at its berthing terminal; daily estimation thus becomes possible. Berthing duration is calculated by analyzing the ship trajectory in the berthing area. The berthing areas of global port are extracted from our previous berth identification algorithm. A tough challenge is to determine the turnover rate of a given terminal. A few studies proposed a quay cranes (QC)-operation-based method to calculate the turnover time, in which the number, handling speed, and working time of cranes need to be known (Chen et al., 2015). This method does not apply to many ports because such data is not easily accessible. In our study, we assume ports are in full utilization and calculate the turnover rate for every ship size in a given port with input berthing time and its previous throughput.

With the real-time berthing time derived from AIS, we can estimate the throughput of each port at a high-frequency level. Our estimated throughputs of some top container ports in 2019 can achieve roughly 95% accuracy (Figure 2). Notably, the handling efficiency is determined by multiple factors, for example, port facility, weather, working hours of a port, and time effect.

In the future, we plan to develop image recognition and machine learning algorithms to collect information and study the association relationship between outputs, namely historical efficiency (can be derived from our platform), and inputs (port facility, economic development level, weather, etc.) so as to further improve accuracy.

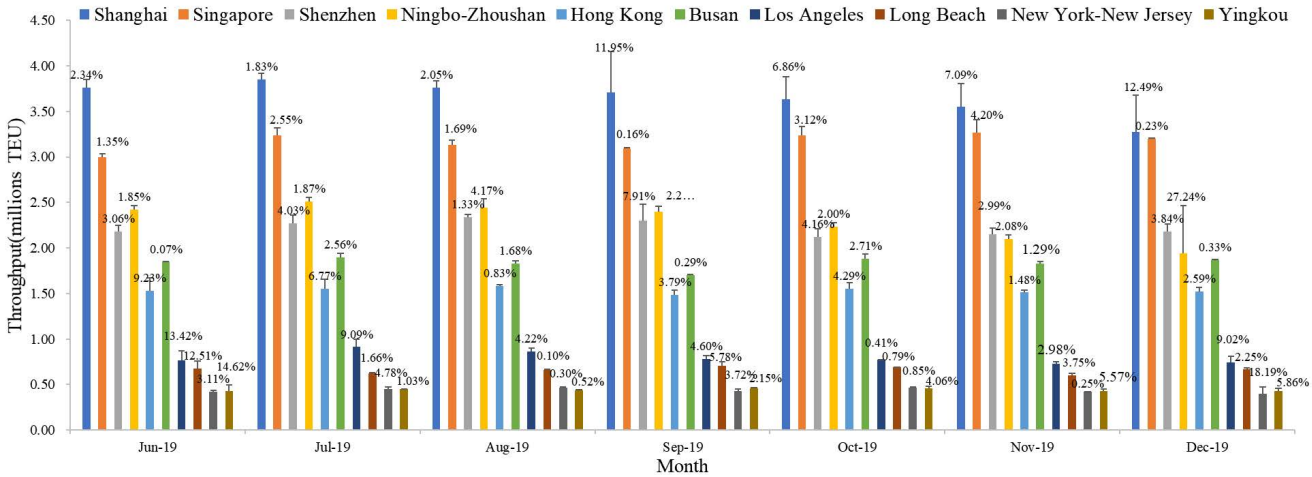

Figure 2 The estimation of monthly throughputs for 7 major container ports in 2019

Port congestion threatens

The effectiveness and sustainability of the global supply chain. It does not only stagnate cargo flows, but also triggers ripple effects across in the transport network. The congestion level of a port is one of the key indicators that affects a port's competitiveness and attracts shipping companies to call. Port congestion happens when ships arrive at a port, they cannot load or unload immediately, instead, having to queue up at an anchorage area and waiting for their turn to berth at the terminal. Waiting time at anchorage is broadly accepted as a congestion measure. In our study, we define two congestion indicators: (1) the ratio of ships which moor at anchorage before berthing at the terminal over the total ship visits at a port, and (2) the average waiting time of a ship at the anchorage before it berths at the terminal. With the location information of anchorage and berthing areas of each port, the two indicators can be calculated based on the ship trajectories within the anchorage and berthing areas identified. Figure 3 illustrates our estimation of average waiting for the world's 20 major container ports in 2020.

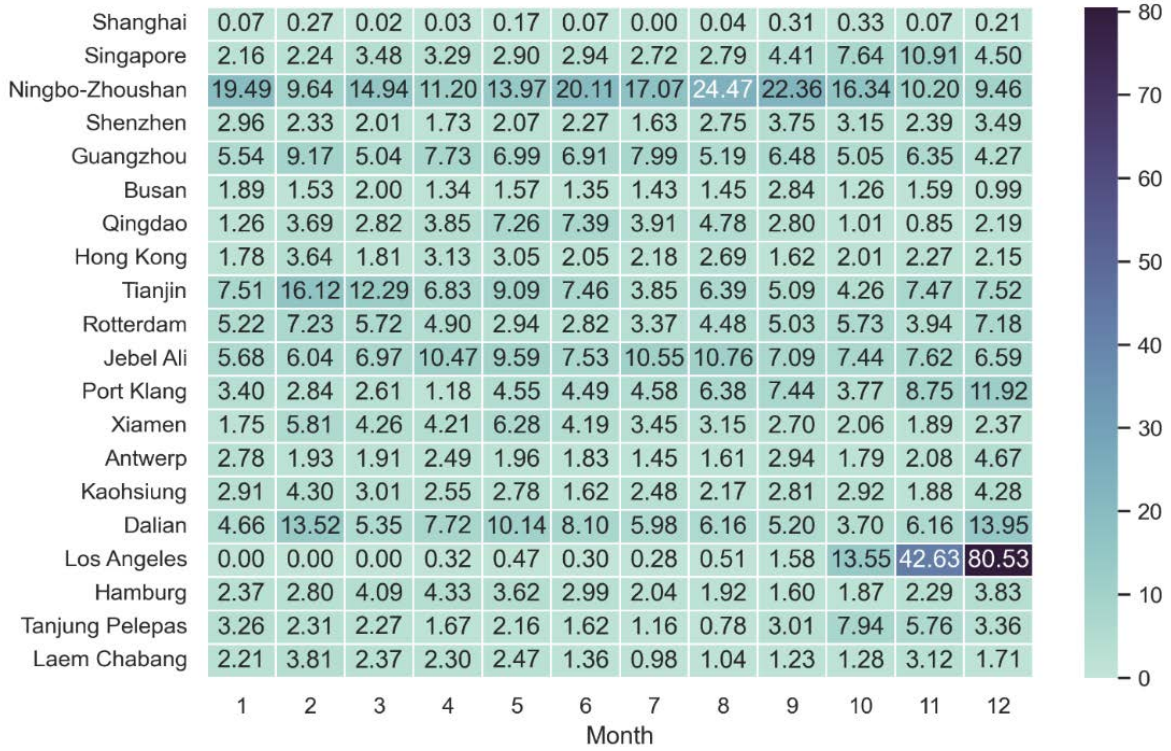

Figure 3. The average waiting time of 20 major container ports in 2020

Port connectivity estimation

The container port connectivity reflects one region's access to the world markets. Both local authorities and shipping practitioners attach great importance to the measurement of container port connectivity. Many inter-governmental organizations (e.g., UNCTAD) and enterprises (e.g., Drewry Shipping Consultants) publish their report on port connectivity regularly. Building upon their work, we create a new connectivity index which can reflect not only the ports' traffic volume and route diversity but also its network properties.

AIS data enables us to calculate the required network factors. We take four steps to calculate the connectivity indicators. First, we extract the shipping trajectory and trip length of each ship from AIS; previously generated port throughput is also an input. Second, based on trajectories and trip lengths, we calculate the number of ship visits to a port, the number of connected countries, and the strategic importance of a port in the global shipping network. Strategic importance is represented by the possibility of a port to attract long-haul (e.g., inter-continental voyage) ships with the highest load factor. Third, we construct the global liner shipping network basing on all the trajectories and use it to calculate three network indicators: degree centrality (number of connected peers of a port); closeness centrality (average number of midway ports on direct liner services); betweenness centrality (frequency that

a port lies on the direct liner services between any other two ports). Fourth, we standardize all the indicators and calculate the total score. Principle Component Analysis, which reflects the relationship of many indicators, can be an option. Figure 4 shows the port connectivity of illustrative ports in 2020. We can see that, although some port, such as Hong Kong (China), Antwerp (Belgium), Los Angeles (US), and Tanjung Pelepas (Malaysia) have relatively fewer throughputs, they rank high in our system because of better network characteristics.

We believe this project can improve the port datatransparency and standardization, and hence generates good research value in improving port performance. High-frequency port statistics are highly valued by the industry but very costly to access. For example, it took Lam et al. (2011) two years just to construct a database to study the dynamics of port connectivity and inter-port relationship, while Tovar et al. (2015) stated that obtaining the required data was the most challenging part of their research on port connectivity. These problems can now be addressed by our platform that makes high-frequency indicators available for a wide range of users. It will save substantial financial and labor resources and encourage more academic and practical outputs.

Port	Vessel Visits	Connected Countries	Degree Centrality	Betweenness Centrality	Closeness Centrality	Strategic Importance
Shanghai, China	1206	56	0.0612	0.0709	0.3344	4.4502
Singapore	1327	80	0.0633	0.0535	0.3676	4.2433
Shenzhen, China	950	67	0.0496	0.0677	0.2608	4.0067
Ningbo-Zhoushan, China	868	53	0.0426	0.0477	0.3482	4.4332
Busan, South Korea	697	28	0.0362	0.0796	0.1696	4.4233
Hong Kong, S.A.R, China	1073	54	0.056	0.0824	0.2339	4.0629
Qingdao, China	528	33	0.0272	0.0368	0.3271	4.1238
Tianjin, China	281	22	0.0137	0.0218	0.2889	4.2404
Jebel Ali, Dubai, United Arab Emirates	366	55	0.0177	0.0112	0.3018	4.1025
Rotterdam, The Netherlands	550	61	0.0289	0.0399	0.2024	4.2863
Port Klang, Malaysia	628	56	0.0316	0.0325	0.3277	4.2126
Antwerp, Belgium	390	75	0.0197	0.0191	0.3747	4.4863
Kaohsiung, Taiwan, China	704	41	0.0362	0.0583	0.2567	4.4121
Xiamen, China	301	27	0.0161	0.0313	0.2659	4.3469
Dalian, China	213	19	0.0109	0.0227	0.2342	3.9768
Los Angeles, U.S.A	86	19	0.0035	0.0018	0.5919	4.206
Tanjung Pelepas, Malaysia	415	61	0.0209	0.0205	0.3531	4.7835
Hamburg, Germany	286	53	0.0151	0.0222	0.2721	4.4747
Long Beach, U.S.A.	76	26	0.0032	0.0021	0.4858	4.5753
Keihin Ports, Japan	462	31	0.0229	0.0354	0.2855	4.0922
Tanjung Priok, Jakarta, Indonesia	311	26	0.0149	0.0118	0.3868	4.1365
New York-New Jersey, U.S.A.	162	50	0.0076	0.0064	0.5103	4.0552
Colombo, Sri Lanka	317	46	0.0147	0.0056	0.455	4.0416
Ho Chi Minh City, Vietnam	415	20	0.0214	0.0314	0.3044	4.1107
Bremen/Bremerhaven, Germany	216	49	0.0109	0.017	0.2519	4.4909
Hanshin Port, Japan	406	23	0.0214	0.0389	0.2193	4.2438
Manila, Philippines	212	16	0.0105	0.012	0.3366	4.045

Figure 4. Port connectivity of illustrative major ports in 2020

References

Chen, L., Zhang, D., Ma, X., Wang, L., Li, S., Wu, Z. and Pan, G., 2015. Container port performance measurement and comparison leveraging ship GPS traces and maritime open data. IEEE Transactions on Intelligent Transportation Systems, 17(5), 1227-1242.

Lam, J. S. L. and Yap, W. Y., 2011. Dynamics of liner shipping network and port connectivity in supply chain systems: analysis on East Asia. Journal of Transport Geography, 19(6), 1272-1281.

Tovar, B., Hernández, R. and Rodríguez-Déniz, H., 2015. Container port competitiveness and connectivity: The Canary Islands main ports case. Transport Policy, 38, 40–51.

Dong Yang
dong.yang@polyu.edu.hk

Dr. YANG Dong is an associate professor at Department of Logistics and Maritime Studies, Director of International Shipping and Transport Logistics Master Program, Deputy Director of Shipping Research Centre, the Hong Kong Polytechnic University. He serves as associate editor of International Journal of Shipping and Transport Logistics (IJSTL) and Maritime Business Review (MBR). He has published over 50 SSCI/ SCI journal papers.

Xiwen Bai
xiwenbai@tsinghua.edu.cn

Dr. Xiwen Bai is an assistant professor at Department of Industrial Engineering, Tsinghua University. Her main research interests include maritime economics and shipping big data analytics. She received the B.S. degree in Maritime Studies in 2015 and Ph.D. degree in 2019 from Nanyang Technological University. She has published over 20 SCI/SSCI papers. She also serves as the associate editor of Maritime Policy & Management.

Venus Lun
vlun@lscm.hk

IDr Venus Lun (PhD CEng FCILT FIET FIMechE) is Special Project Director of LSCM R&D Centre in Hong Kong. She has published more than 50 articles in SCI/SSCI journals. She has received the following awards: Designated Inventor from European Paten Office, Erasmus Mundus Scholar from European Commission, and two prizes from Inventions Geneva.