Studies on the Ship Basic Planning Support System using Maritime Logistics Big Data

（海洋物流ビッグデータを利用した船舶基本計画支援システムに関する研究）

August 2021

Mohammad Danil Arifin

D145238

TABLE OF CONTENT

LIST OF TABLE

LIST OF FIGURE

Chapter 1

Introduction

1.1 Present Situation of Marine Logistics Big Data

Mobility and maritime transport have always played an important role in economic, ecological, and social development [1]. With the development of international trade and globalization, traffic and cargo volumes keep increasing. Therefore, numerous stakeholders (such as maritime logistics companies, forwarders, agents, etc.) are compelled to accept changes in the maritime transport sector and turn to more effective practices by introducing technologies that can gather and process massive amounts of information (in a cost-effective way) [2].

A large amount of data is generated from different sources and in different formats in maritime transport on a daily basis. This includes traffic data, cargo data, weather data, and machinery data [3]. Due to the size of the maritime transport network that includes the aforementioned stakeholders, there exist large-scale planning problems at the strategic, tactical, and operational levels [4]. Nowadays, Big Data analytics is applied to many industries (among which the maritime transport industry) to promote a better quality of decision-making processes [5].

According to Acharjya and Ahmed [6], “Big Data means more than just dealing with a large quantity of data. In general, it refers to the collection of large and complex datasets which are difficult to process and analyze using traditional database management tools or data processing applications”. When handling Big Data in large quantities, advanced data-processing techniques and tools are required to effectively analyze and utilize these data [5].

Various stakeholders involved in cargo transports (from origin to destination) are constrained to rely on the information provided by other parties involved. The more exact and detailed such information, the smoother running of shipping operations. The level of cooperation and coordination that is essential for the success of these businesses and competitiveness is based on information systems and information technology [7]. Decisions based on the information extracted from raw data can lead to several advantages: increased safety and resource utilization, as well as a higher degree of efficiency, sustainability, and environmental protection [8].

The concept of Big Data comes with a set of related components that enable organizations to put the data to practical use and solve several business problems. These include the IT infrastructure needed to support Big Data, the analytics applied to the data, technologies needed for Big Data projects, related skill sets, and the actual use cases that make sense for Big Data [9]. According to Nita and Mihailescu, the sources of data could be communication, transport, social media, climate, search engines, GPS signals, online shopping, mobile devices, and many others [10]. The sources of data can be the Internet of Things (IoT) as well, creating large data sets on a daily basis. IoT consists of various devices, which are provided with digital sensors and connected for interacting. The number of these devices, such as mobile phones or Radio Frequency Identification tags (RFID) is rising rapidly [11], [12].

Even though there is no official definition of Big Data, the concept of Big Data goes beyond the literal meaning of the data with great volume. Taylor-Sakry [13] defines Big Data as “large sets of complex data, both structured and unstructured which traditional processing techniques and/ or algorithms are unable to operate on.” This definition also points out the importance of specific technology or methods, such as algorithms, to handle Big Data

As the volume of marine big data has increased dramatically, there is significant potential and high value hidden in the huge volumes of data that are widely used in various fields, one of the main concerns is how to fully exploit the value of such data in the development of marine logistics industries. There are several key uses of big data in the shipping and logistics industry. Thanks to the use of big data engineering, the shipping industry has grown even stronger over the past few years.

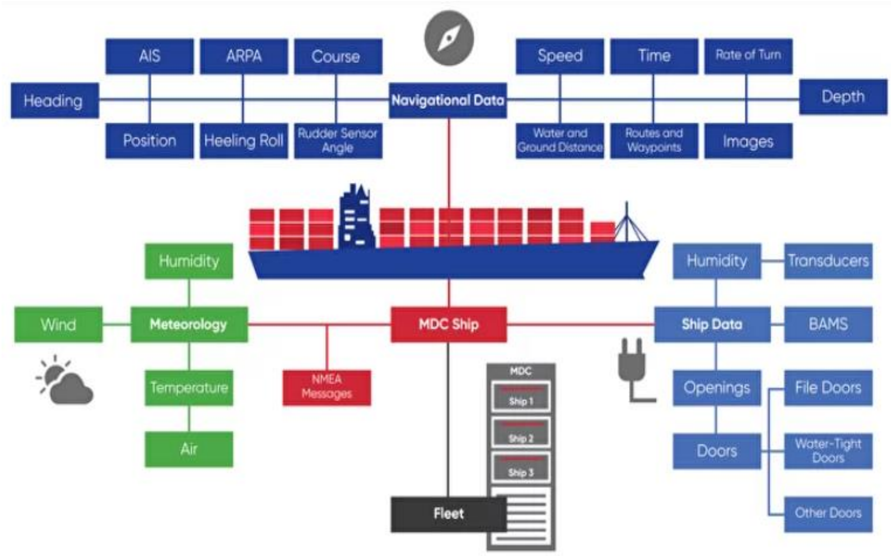


Fig 1.1 Marine Digital, [14]

Big data is used to manage ship sensors and for predictive analysis, which is needed to prevent delays and improve the overall operational efficiency of the industry. In the shipping industry, proper cargo tracking is essential to ensure the necessary safety and confidentiality. Through data obtained through proper tracking of shipments over several years, information on the causes of vessel losses at sea, losses of containers inside or outside terminals or warehouses, and other problems related to dispatch (for example, the reasons for damage to the goods) may be received. This big data for the shipping industry can be used to make decisions in the future to predict and avoid costly problems, and to create more reliable cargo delivery options.

Everyone agrees that big data can play an important role in ship design. Basically, this will be possible by analyzing the results obtained from the sensors of previously used vessels. Data collected and analyzed over the life of the vessel will be useful for future improvements in ship design. Previous datasets could help in testing the proposed ship design without physically developing it. This is a very big advantage for the shipbuilding industry.

As we move towards a more globalized economy, the demand for transportation of goods and related logistical support will continue to grow exponentially. Over time, this growth will increase the need for maximizing time and profitability to have the most profitable delivery processes. Through the use of advanced data processing techniques, the delivery of goods will become more efficient. Improved transport services will increase overall international trade.

Big Data which is effective for forecasting demand and planning process is the oil of the information economy [14]. The future competitiveness of marine industries will be affected by how rapidly we take advantage of it. By identifying the real information extracted from Big Data, we can take the advantage of its full value to help organizations to be more efficient and profitable.

The other research found that Big Data has a big potential to improve operational efficiency in shipping [15]. Furthermore, Big Data in marine industries is providing information that can make a maritime operation more efficient [16]. Moreover, it is widely believed that Big Data can aid in improving forecasts provided that we can analyze and discover hidden patterns [17].

The global marine logistics industry has changed significantly because of the influence of the global movement of goods [18]. Hence, it is important to develop ships that meet specific needs and market requirements by developing the allocation model.

Simultaneously, marine logistics big data can be acquired more easily than ever before (e.g., port, ship, route, international trade, and Automatic Identification System (AIS) data) [19-23]. If these data are effectively utilized, great innovation might be achieved.

1.2 Ship Allocation Model Application

Maritime transport is characterized by a large number of stakeholders such as seaport operators, shipbuilders and ship owners, agents, brokers, shipping and insurance companies or classification societies, etc. A large variety of stakeholders means a large variety of business procedures and interests, among which there are various interests in data types. Thus, a clear definition of the term "Maritime Big Data" does not exist. In dependence on the target, Maritime Big Data includes details of ships' performance, freight rates, weather data, labor costs, oil or even metal prices [24]. In addition, the number of digital sensors in maritime transport is also increasing. Consequently, the generation of Big Data in the maritime context is increasing, adding to the quantity of data, and the data are provided by numerous different sources [25]. Besides weather forecasts and historical data, the most important data resources in the case of voyage data are provided by the bridge equipment to be recorded by the Voyage Data Recorder (VDR), and by external monitoring such as the Automatic Identification System (AIS) [24], [25].

The main purpose of VDR is to provide data for analyses in case of accidents. Therefore, VDR stores all recorded data of a voyage [25]. As the International Maritime Organization (IMO) determines (Regulation 20, Chapter 5of the International Convention for the Safety of Life at Sea – SOLAS, 1974), a VDR needs to be installed on all passenger ships and other ships of 3000 gross tonnages or more on international voyages [26]. Information collected by VDR is date and time, ship's position, speed, heading, bridge audio, communication audio, wind speed and direction, main alarms on the bridge and rudder or engine orders and responses, and more [27]. According to Resolution

MSC.333(90) (adopted on 22 May 2012), the time for which all stored data items are retained should be at least 30 days/720 hours on the long-term recording medium and at least 48 hours on fixed and float-free recording media. Data items older than this may be overwritten with new data. A standard international voyage can last longer than 30 days and data collected during this journey will be overwritten and not extracted and saved on an external device. On short voyages, the data will also be deleted after the voyage or during the upcoming voyage, making them inaccessible for a long-term analysis [24]. VDR collects a massive volume of data, providing deep insights into the voyage. As the VDR data are overwritten after a specific period and replaced with new data, the data should be sent ashore or be saved and exported manually or automatically. Mirović et al. [25] claim that such information is still analyzed only in cases of accidents, and is otherwise deleted without consideration.

The AIS tracking system was originally developed as the collision avoidance tool that enables commercial vessels to 'see' each other more clearly in any conditions and to improve helmsman's information about the surrounding environment. The AIS does this by continuously transmitting the vessel's position, identity, speed, and course, along with other relevant information, to all other AIS-equipped vessels within the range [28].

In this respect, the AIS enables safer and more efficient navigation by tracking all ships within the range (enabling the exchange of ship data among ships and the shore) [29], [25]. AIS regulations are defined in the SOLAS as well and require AIS transponders to be installed on all passenger ships and all other ships of 300 gross tonnages and more [30]. The AIS provides other ships and coastal authorities with static data (IMO number, ship's length, and beam, etc.), voyage related data (destination and ETA, etc), dynamic data (position, course, speed, etc), and VTS data (short content information related to various safety warnings and information on areas with warnings about navigational and other dangers) [31], [25], [32]. Thus, it helps collision avoidance and decision-making onboard in real-time [29]. In this respect, the AIS provides data as a basis for BDA in maritime transport. The frequency of the AIS data refreshing rate depends on the vessel movement and navigational status, meaning that the data quantity will increase as the vessel moves faster and/or alters the course.

However, the growing trend of the AIS has been such that in some most congested waters the system is already overloaded as of today. Given the danger that this overload can represent for the main mission of the AIS, that is collision avoidance, the International Association of Marine Aids to Navigation and Lighthouse Authorities (IALA) and several national maritime authorities have started the work on the VHF Data Exchange System (VDES). Rather than an evolution of AIS, VDES is a communications system encompassing different subsystems of communications, one of them being the AIS [33], [34]. New techniques providing higher data rates than those used for the AIS are a core element of VDES. The VDES supports e-Navigation [33] and could have a significant positive impact on the provision of maritime information services, such as maritime safety information, general data communications at high data rates, locating, vessel traffic management, satellite communications, etc. [35].

Maritime transport offers favorable conditions for Big Data Processes, by being provided with a wide range of data from various sources and by being obliged to record such data. As mentioned before, Big Data Analyses provide deeper insight into processes and events and allow for well-informed decision making. In the same manner, as in other industries, the maritime industry might benefit from Big Data Analyses. In maritime transport (as part of the maritime industry), innovations in BDA will bring advantages to efficient routing, operational planning, and safety improvements, as shown in the following paragraphs.

There exist several current applications of Big Data in maritime transport and one of them is "operations". In this respect, ship owners can determine the optimum speed for fuel consumption, taking into consideration factors such as bunker cost, freight rates, and schedules. Speed reduction will result in fuel consumption optimization and will also significantly contribute to decreasing the emission of harmful gases, thus getting in consonance with the development of ecological legislation [36].

Other applications of Big Data are voyage operations (vessels can be tracked using dashboards instead of relying on notes, emails, or phone calls), tracking, etc. [37]. For example, Freight Metrics provide the location visibility of a ship or vessels located around a particular port. Freight Metrics has plotted locations of major shipping ports around Australia on Google Maps. It enables a simplified connection to the port website for shipping schedules [38].

In Singapore and Malaysia, ports utilize Big Data techniques to create advanced inspection systems to assess the history and cargo type of importers. The Hamburg Port in Germany uses a cloud-based analytics tool called SmartPortLogistics. The tool pulls different types of data, such as vessel positions, height, and breadth of bridges, and planned sailing routes. These data can be viewed in real-time on mobile applications [37].

Other studies also have applied Big Data to the maritime industry. For construction applications, Hiekata et al proposed a high-accuracy block component measurement method that uses point cloud data from a 3D laser scanner [39]. Aoyama et al proposed new methods of extracting and utilizing monitoring data by introducing two additional monitoring technologies and considering the reliability of each for advanced shipbuilding construction management [40].

In the operations field, Perera et al analyzed large ship performance datasets to propose a model for evaluating ship performance under various seagoing conditions [41]. Ando et aland Yoshida et al proposed a data collection platform called the Ship Information Management System and utilized the data collected for many purposes (e.g., energy efficiency determinations, ship performance monitoring, and engine monitoring)

[42] [43]. Note that many of these studies have employed big data to improve ship construction, operation, and performance; few have examined the use of big data for ship demand prediction [44]. However, detailed predictions of the routes on which demand will increase are currently difficult to execute. Therefore, a ship allocation model by harnessing marine logistic big data is conducted.

1.3 The objective of this Study

Based on the above considerations, we have been developing ship allocation models using marine logistics data and its application to bulk carrier demand forecasting and basic planning support

Following is the objective of this study:

- The support system of ship basic planning for bulk carrier

To realize the objective, there are three important points, Marine Logistics Database (MLDB), ship allocation model, and simulation. In the MLDB, integration of all of the data i.e. AIS, ship, port, route, and trade data into a relational database is required. Furthermore, in the ship allocation model, the development of three distinct models—the shipper, shipowner, and operator models is required. And then, in the simulation to develop new ship allocation and examine which specification of the ship is effective, several simulations are required.

The basic concepts of proposed methods and details of information regarding the support system of ship basic planning are described in this study. Moreover, the effectiveness of the developed models is discussed.

1.4 Organization of this Dissertation

The structure of this thesis is presented in Fig.1.1. The summary of each chapter is shown in the following.

(Chapter 1) At the beginning of this thesis, a short overview of the present situation of marine logistics big data is described. In this chapter, the support system of ship basic planning and ship allocation model concept is introduced. Furthermore, the objective and scope of the study are clarified.

(Chapter 2) In this chapter, the related studies of maritime big data applications are described. Some references about maritime big data for several purposes are described. Furthermore, the characteristics of this study are clarified. In addition, the differences of this study compare to the related studies are described.

(Chapter 3) The basic concept to develop a support system of ship basic planning is introduced and discussed. This chapter consists of three important points, such as Marine Logistics Database (MLDB), ship allocation model, and simulation

(Chapter 4) The development of the Marine Logistics Database is defined and illustrated. Detailed information regarding the development of MLDB—input, data structure, error cleaning, etc, are discussed. Furthermore, the evaluation of cargo estimation is described.

(Chapter 5) The development of the ship allocation model starting with the explanation of the development of the shipper model. The results of the shipper model are shown, and the evaluation of the model is discussed.

(Chapter 6). The development of the shipowner model is described. The results of the shipowner model are shown, and the evaluation of the model is discussed.

(Chapter 7). The development of the operator model is described in this section. Results of the operator model are shown and the reproducibility of the proposed model is discussed.

(Chapter 8). Several cases were simulated and performed. The cases; examination of supply-demand balance, examination of effective ship size, and influence of fuel efficiency on demand are specifically discussed.

(Chapter 9) The last chapter presents the conclusions of the study and future tasks.

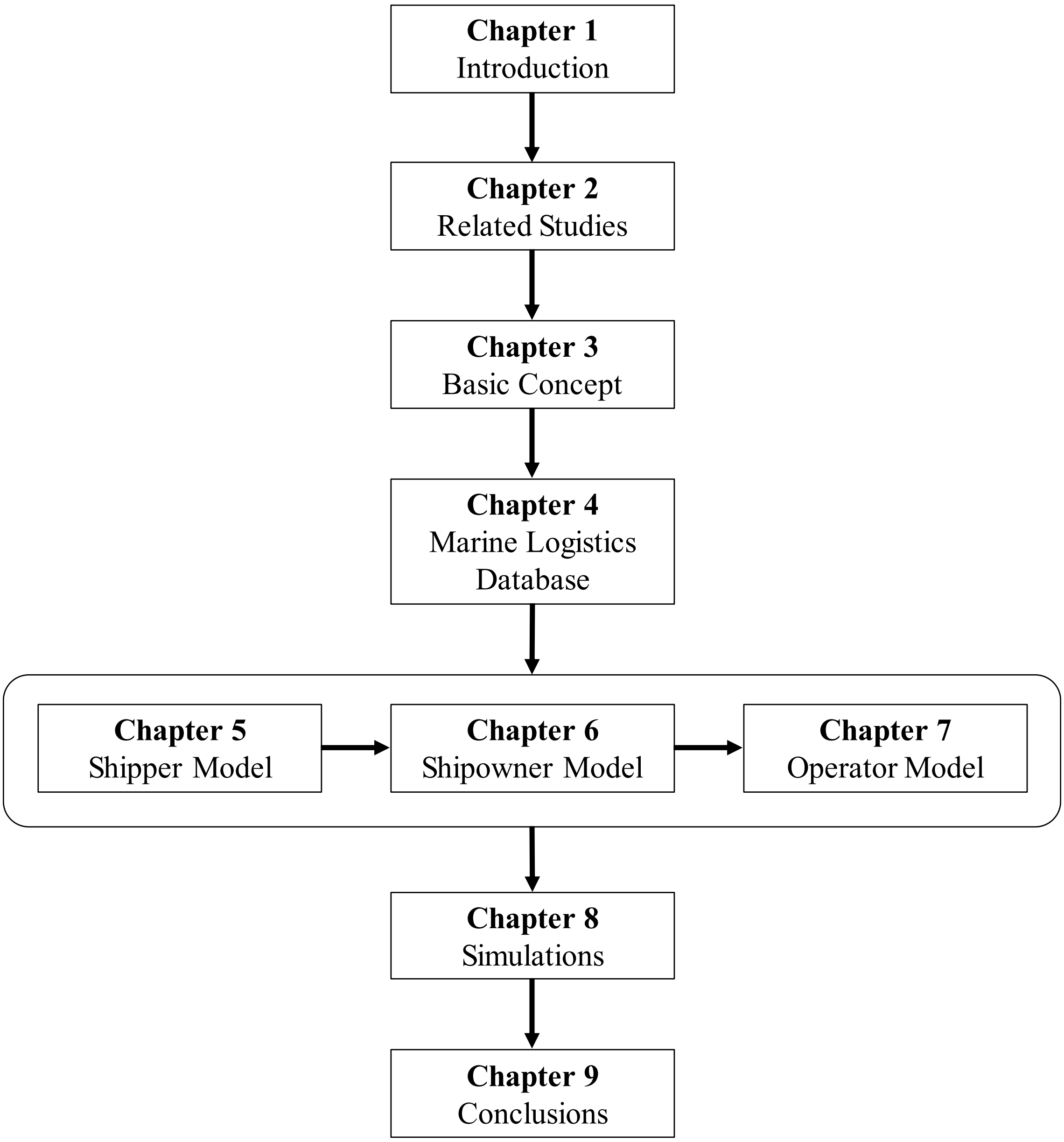


Fig. 1.2 Structure of the dissertation

Chapter 2

Related Studies

2.1 Definition of Maritime Big Data

2.1.1 Big Data in General Sense

The term “big data” was created in the field of space engineering 18 years ago. According to Press [45] and Friedman [46], the first use of the term “big data” appeared in 1997 in the work of Michael Cox and David Ellsworth who were scientists at NASA. They defined big data to describe certain challenges which they were facing in the 1990s. This “challenge” was massive information that was generated by supercomputers and could not be processed and visualized by the technology of that time. Since then, thanks to the rapid development of computing and information technology, the notion of “big data” has spread to various industries.

Apart from this initial definition, in recent years, there have been many perceptions of big data, especially in the information industry. One of the simplest and most familiar expressions of big data may be one that was introduced by IBM, which is one of the leading companies in the information industry [47]. It states “Data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. This data is big data” (IBM, n.d.).

This seems to imply two very features of big data. First, big data is regarded as a kind of huge electronic data that is explosively increasing. Although it may be a matter, of course, that big data is electronic, it is important to recognize and put an emphasis on this feature to precisely define big data. Second, big data is being derived through sensors that have been developed rapidly. According to IBM's definition, the term “sensor” means not only ones such as speed meters and gyroscopes but also such as smartphones and digital cameras, which capture information and interpret it into electronic data. Thanks to such sensors, big data can be generated or detected.

Taking into account the definitions introduced so far, it seems possible to say that there are three fundamental features of big data.

- Electronic form
- Derived through various sensors (which have been rapidly developed)
- The difficulty of capturing, storing, managing, and analyzing

As a general consideration of big data, three important features are identified. Big data takes the form of electronic information, it is derived through various sensors, and there are difficulties in capturing, storing, managing, and analyzing it. Concerning difficulties, it is commonly considered that there are four main aspects as in Fig. 2.1, which are volume, velocity, variety, and veracity as in [47]. These findings can be summarized as in Fig. 2.2

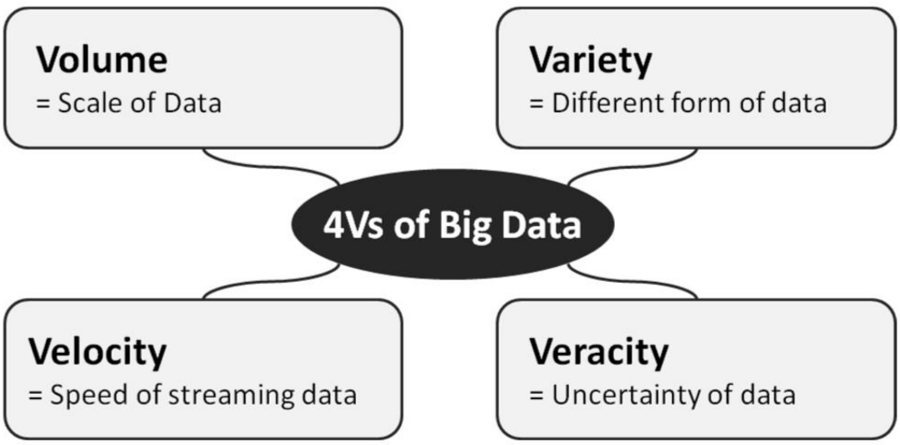


Fig. 2.1 Essence of 4Vs as main difficulties of treating big data, [47]

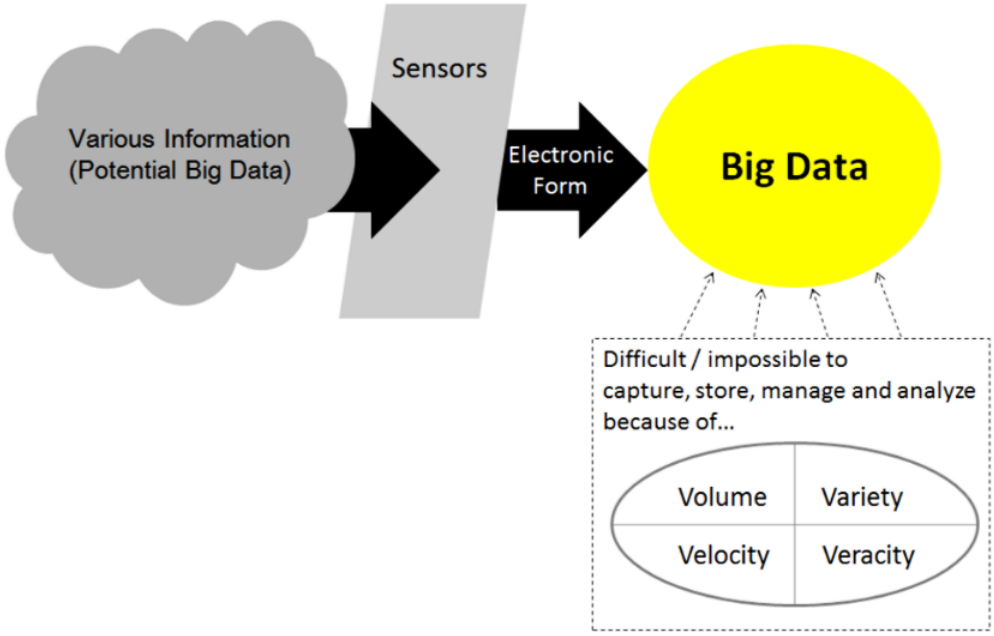


Fig. 2.2 Overview of the important feature of big data, [47]

2.1.2 Big Data in Logistic and Transportation

Big Data analytics brings many benefits to the logistics and transportation industry. The data are collected from a very large network of sensors and devices. Big data analytics tools are efficient in storing the data and processing them in real-time to monitor traffic and make predictions that improve service quality and companies' revenues [48].

Data generated by traffic sensors can be used to identify issues in real-time, which means that road users can make informed decisions to save time while road authorities may control traffic and intervene quickly when needed [49], [50]. Los Angeles, for example, uses the data to control traffic lights, which has reduced traffic congestion by an estimated 16 percent [51]. In Dublin, the data collected from traffic sensors, bus GPS devices, and other sources are used to build a real-time digital map of the city transportation network. It helps identify traffic problems and make decisions regarding the bus transportation network. As a result, the traffic flow in the city has been improved [52].

The private sector may gain a competitive advantage and increase productivity using Big Data [53]. Tracking vehicles' locations using satellite navigation and sensors enables customers to know exactly where their shipment is and when it will be delivered. The vehicles' routes can be optimized by taking into account delivery addresses and data regarding road conditions, traffic jams, weather conditions, locations of gas stations, etc. Routing optimization saves a lot of fuel, which reduces financial costs and environmental impact.

Fuel consumption can also be reduced by optimizing fuel input based on data from sensors that monitor the engine. Sensors can also monitor the state of the equipment and vehicle performance in real-time. This helps predict errors and detect when maintenance is needed. Safety can also be improved with sensors that monitor vehicle speed, whether the driver is complying with the traffic laws and if the driver has been behind the wheel for too long [54].

US Xpress, an American logistics company, is an example of a company taking advantage of Big Data [54]. It has installed almost 1,000 sensors in each truck to monitor their locations, driving speed, petrol use, how often they break, if they are on idle for too long, when maintenance is required and even the drivers' capabilities. Hundreds of billions of data records generated help the company save over $6 million a year. It is presented in [55] how data regarding the fuel, speed, acceleration, etc. are collected from vehicles' sensors and GPS devices.

They are then analyzed, which enables monitoring the driving behaviors to improve productivity, detecting negative driving patterns, determining which trucks are idling and wasting fuel, which trucks have the worst gas mileage, and which drivers have the highest risk factor. The data flow is shown in Fig. 2.3

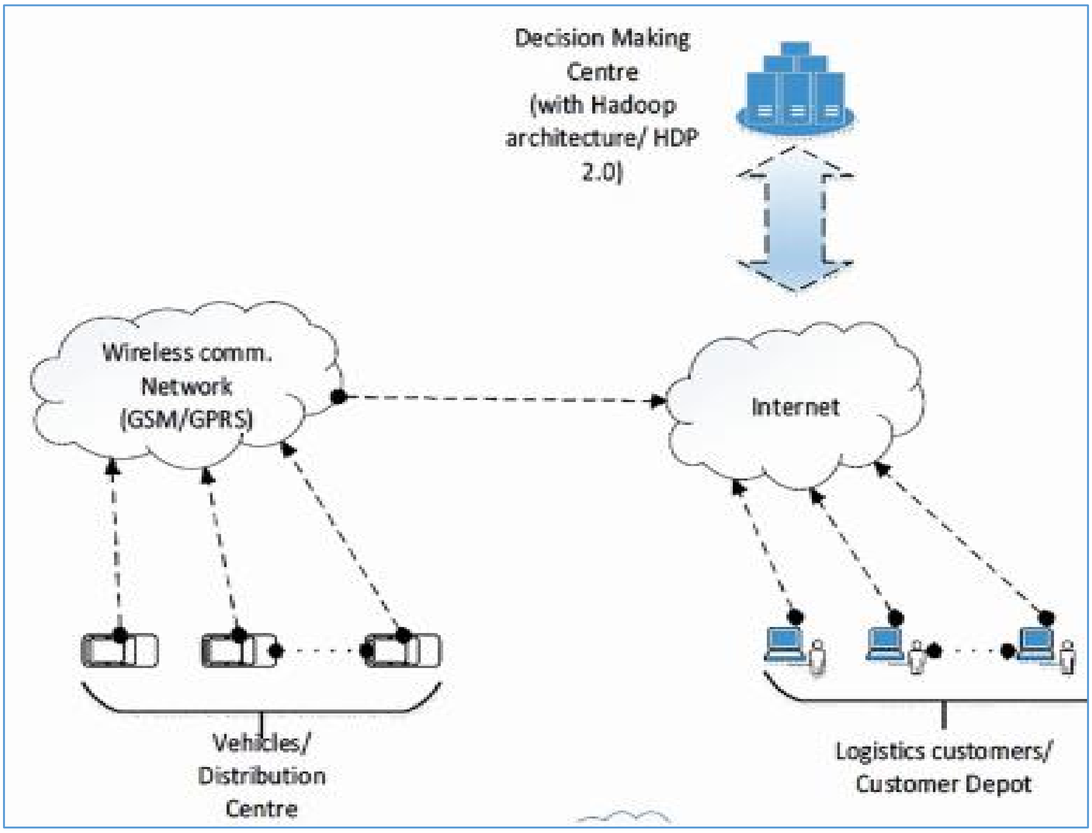


Fig. 2.3 System data flow, [55]

2.1.3 Maritime Big Data in General

Recently, a huge amount of big data is potentially available within the maritime industries, considering the number of ships carrying vast amounts of goods to and from the numerous ports worldwide. For example, Big Data such as voyage data, machinery data, weather data, business data, trouble data, AIS data, port data, ships data, route data, and trade data are available in the maritime industries. To date, only a fraction of this information is used [56].

That huge amount of Big Data is potentially important within the marine industries and becomes a useful knowledge bank if handled carefully. Hence, big companies are largely investing in research and harnessing Big Data [57].

2.2 The Studies of Maritime Big Data Application Area

Many studies have applied big data to the maritime industry. The followings studies are conducted by utilizing big data.

2.2.1 Maritime Big Data and E-Navigation

With the development of navigational systems, sensors, and tracking systems following recent advances in technology, the maritime industry is opening up to the benefits of the digital era [25]. In this respect, the International Maritime Organization (IMO) has developed an e-Navigation concept to improve shipping through better organization of data on ships and onshore, and better data exchange and communication between ships and between the ship and the shore [58]. The e-navigation Strategy Implementation Plan (SIP) introduces a vision of e-navigation which is embedded in general expectations for onboard, onshore, and communications elements. The main objective of the present SIP is to implement the five prioritized e-navigation solutions [59]:

- Improved, harmonized, and user-friendly bridge design;
- Means for standardized and automated reporting;
- Improved reliability, resilience, and integrity of bridge equipment and navigation information;
- Integration and presentation of the available information in graphical displays received via communication equipment; and
- Improved Communication of Vessel Traffic Service (VTS) Service Portfolio (not limited to VTS stations).

Consideration should be given to the IMO Formal Safety Assessment (FSA) from which several required tasks have been identified. These tasks, when completed in the period 2015–2019, should provide the industry with harmonized information, to start designing products and services to meet the e-navigation solutions [59]. The Maritime

Connectivity Platform (MCP), cloud-based technology, and an open-source digital maritime domain enable common Internet standards to maritime navigation and transport systems [60]. It is a communication framework enabling efficient, secure, reliable, and seamless electronic information exchange between all authorized maritime stakeholders across available communication systems [61]. As navigation systems become more advanced, the amount of ship performance and navigation data thus generated is becoming ever more significant. Big Data analytics tools make it possible for these large quantities of data to be analyzed to gain the insight that supports the decision-making [25].

2.2.2 Maritime Big Data Innovation in the field of Weather Routing

Weather conditions may cause several issues in maritime transport. Due to the direction or strength of winds and waves, the engine power needs to be increased to maintain the speed. By BDA of historical weather data, eventual harsher weather conditions on certain routes can be detected and data could be used to adapt the future strategic route planning [62]. Furthermore, data about the ship's performance and navigation information, such as the location, average draught, main engine power, speed, and fuel consumption combined with wind speed and direction, can support navigation strategies, as well as intelligent decisions, in real-time as well [63]. With the provision of real-time weather data (wind speed and direction, wave heights, storm forecasts) and ship data (regarding the trim and rolling), dangerous situations can be identified immediately. Furthermore, alternative (less hazardous) routes outside critical areas can be calculated faster. The data regarding the ships that are already within a hazardous area can help other ships to avoid the same difficulties, thus improving the safety of the crew and other ships. The risk of sinking or cargo loss due to severe weather conditions may be minimized.

For example, the port of Singapore uses a platform that enables information on ship positions and weather data to be collected. The platform helps avoid accidents by inferring the most likely path ships would take in a given situation [37].

FastSeas [64], as another example, is a cloud-based weather routing and passage planning tool that enables calculating the fastest route (from point A to point B) according to the current NOAA GFS weather forecast (a weather forecast model produced by the National Centers for Environmental Prediction), current oceanic currents, vessel performance, and "comfort criteria". FastSeas has to be "taught" the vessel performance. The performance settings must be set only once. FastSeas will save these settings and use them next time. After the calculation is complete, the route will be displayed on the map.

To achieve an efficient maritime transport service, shipping companies can also implement specialized software for performance and route analyses [65]. Weather-routing software can help users determine the optimal departure time and routing based on user-specified parameters. These parameters can include maximum acceptable wind speeds or wave heights, as well as the percentage of time that the vessel will be on a particular point of sail [66].

Weather-routing can also enable fuel savings, which will not only bring the benefit such as the cost reduction but will also support the achievement of emission standards and support the Ship Energy Efficiency Management Plan (SEEMP) [25], [63]. There exist several examples of such software such as the Transas Navi-Planner, Octopus, and BonVoyage. Transas Navi-Planner utilizes one of the world's largest navigational databases, as well as artificial intelligence to auto-create a route that is safe to sail. It calculates weather optimization, supports hazard identification, creates a voyage plan, and provides at all times the latest charts and data automatically [67]. The OCTOPUS software assists ship officers and engineers in making real-time decisions enabling them to be proactive in safety and efficiency actions, resulting in more immediate benefits than just a traditional post voyage analysis [68].

The BonVoyage System is a combination of software and data. BonVoyage is a user-friendly icon-driven graphical marine weather briefing system. By using real-time weather data to generate color-coded graphics, the software lets the user see into the heart of severe weather systems. This provides a much more complete picture of the storm than traditional text-based weather bulletins or radio facsimiles because it captures the detailed shape of each storm system and visually displays dangerous wave generating areas [69].

2.2.3 Maritime Big Data Innovation in the field of Monitoring/Tracking

To avoid collisions of ships in seaport areas, forecasts of vessel arrival times with real-time information about their current speed are made to illustrate the estimated traffic [58]. Through this technology, collision issues are identified and vessels are notified before accidents could happen. The encounter of two vessels in narrow areas can be avoided. For example, the Vessel Traffic Service is designed to improve maritime safety, support security activities in the maritime sector, promote smooth and efficient maritime traffic and prevent accidents and the potential environmental harm they may cause. The system displays a graphical environment with movements of vessels in the approach areas, putting each of these overlapping vessels to a digital nautical chart, in its real geodesic position and informs the identification of each vessel [70].

Risky shore areas (such as bunkering facilities for fuel or liquid cargo) require a special control in order, for example, to detect unauthorized ship movements as early as possible by monitoring the information about all vessel and vehicle movements [71]. Collisions outside the seaport areas can be avoided by collecting data about the standard routes of vessels on a long-term basis. These data reveal if a vessel deviates from the declared path [25].

Monitoring of technical equipment enables higher safety for the crew and environment and increased the longevity of ship components. During the use of the equipment and engines, data are collected and sent to manufacturers, who can detect irregularities and increase their service quality [24]. The Satellite AIS (S-AIS) is a vessel identification system that tracks the location of vessels in the most remote areas of the world, especially over open oceans and beyond the reach of terrestrial-only AIS systems. S-AIS is used for collision avoidance, identification, and location information as well as for maritime domain awareness, search and rescue, environmental monitoring, and maritime intelligence applications [72]. Analyses of oceangoing ship routes are also important in the function of environmental protection. The analyses of ship routes in the mentioned oceanic areas covered by S-AIS could be very important to estimate air pollution emissions from ships, seeking to minimize their negative impact [73].

2.2.4 Maritime Big Data Application in Ports and Terminals

Similarly, to previous analyses, Big Data Analysis (BDA) can also be used for predictions of arrival times and calculations of the required speed. In general, shipping companies are under the pressure of meeting the agreed arrival times at the seaport, while being faced with the problem of the named uncertainties of weather conditions. Consequently, managers often decide to deviate from the efficient low steaming to prevent delays [62]. On the other hand, seaport operations are sometimes delayed due to different reasons, such as arrival times of prior ships, loading or unloading processes, etc. Both circumstances will lead to waiting times on available slots for the arriving ship, being anchored outside of the port.

To prevent this situation, the performance and weather data can be analyzed and considered with tracking data, as provided by the AIS. The planned (optimum) route matched with real-time data on outside circumstances, like the weather, and the tracked position can thus be used to forecast arrival times [25]. However, it is necessary to divide the standard AIS transceiver from AIS Application Specific Messages (referred here). An AIS transceiver transmits and receives static data (Maritime Mobile Service Identity, ship name, and callsign), as well as dynamic data (position data, speed, and course over ground) via VHF on two channels specially reserved for AIS [74]. AIS Application Specific Messages (ASMs) are a quick way to get important information through the VDL in small, predefined packages. A ship, for example, can use the ASMs to transmit important details such as dangerous cargo information, number of persons on board, extended ship static and voyage-related data, and route information [75].

Data Science Campus (DSC) [76] has explored the likelihood that a ship would be delayed and would arrive at its intended destination sometime after its estimated time of arrival. Ship delays were predicted using supervised machine learning and these predictions can be used to predict port loading at a point in time and can support any subsequent operational port planning. Both the Automatic Identification System (AIS) and the Consolidated European Reporting System (CERS) datasets have been used and a means by which they could be joined has been developed. The common identifier between the AIS and CERS data is the Maritime Mobile Service Identity (MMSI, the unique identifier of each ship) [76]. To merge the two datasets they were first inner joined on the MMSI, the datasets were then filtered by restricting each timestamp within the AIS to its closest estimated time of arrival (ETA) or estimated time of departure (ETD) (both ETA and ETD were considered since the AIS data contained both the inbound and outbound portions of the journey). Records were retained where the ETA or ETD were falling within 24 hours of the timestamp. As more than one ETA or ETD can meet these criteria, the ETA or ETD closest to the timestamp was selected. After the merging, the complete dataset included 727 voyages relating to 235 unique ships. An in-depth explanation of the research can be found in the report [76].

Big Data enables optimized usage of resources and infrastructures as well. For example, a typical crane operator works only one-quarter of the time, remaining idle for three-quarters of the time, waiting to get a container ready to load or for an empty truck to load a container. Increasing the number of trucks may not be a viable solution owing to the congestion it would cause. Rather, Big Data analysis could synchronize movements, so that the crane operator can work more time. For instance, signals related to the crane position, status, and GPS position signals could sync the movement of trucks and containers, to reduce idling time [77].

The Bahri maritime company has rolled out a series of initiatives aimed at harnessing the power of "Data Innovation" to enhance productivity, unlock growth opportunities, and transform the existing operations model in the shipping industry [78]. In this respect, the company has created BahriData, a Big Data platform, to improve the operating performance and to unlock growth opportunities. The company has developed various data models in its BahriData platform to cover various key business areas such as chartering, voyage management, fleet operations, maintenance, and customer services [37].

The following are examples of big data technology implementation in ports:

- **Ports in Singapore and Malaysia**

  In Singapore and Malaysia, ports utilize big data techniques to create advanced inspection systems to assess the history and cargo type of importers.

  Purpose

  The purpose is to segregate importers that most require inspection and allow other importers to operate quickly, without impacting the port's security objectives.

- **Maritime and Port Authority, Singapore**

  In August 2015, Singapore's Maritime and Port Authority (MPA) signed a Memorandum of Understanding (MOU) with IBM to establish a two-year agreement for big data. The agreement involved developing a platform using IBM's Traffic Prediction tool to forecast vessel arrival times and estimate potential traffic congestion using fusion analytics. The platform also relies on data mining and anomaly detection by using IBM Incident Detection Module and IBM System G. According to Goh Kwong Heng, CIO of MPA, the port authority plans to invest in big data to improve port operations and activities. The MPA aims to use data analytics platforms to complement its port management systems in detecting anomalies and supporting both operations and planning processes. The authority also plans to invest in technologies such as drones and mobile apps to manage marine accidents and improve the efficiency of its port workers.

  Purpose

  The port uses the platform to improve productivity and marine safety at the major trans-shipment hub. For example, by collecting information on ship positions and weather data, the platform helps avoid accidents by inferring the most likely path ships would take in a given situation. This also helps prevent illegal bunkering by detecting unusual movement patterns. According to Andrew Tan, the CEO of MPA, the port authority's collaboration with IBM enabled a mix of research expertise, software technologies, and maritime domain knowledge to create new capabilities for the maritime industry in Singapore.

- **Hamburg Port, Germany**

  The Hamburg Port in Germany uses a cloud-based analytics tool called SmartPort Logistics. The tool pulls different types of data, such as vessel positions, height, and width of bridges, and planned driving routes. This data can be viewed in real-time on mobile applications.

  Purpose

  The tool aims to streamline the flow of goods. It allows port workers to know precisely when ships are expected to dock, and truck drivers to know when cargo is expected to be offloaded.

- **Port of Cartagena, Columbia**

  The port uses Cisco and IBM solutions for IoT analytics. The solutions help forecast equipment failures and keep our machines running effectively by ensuring the equipment degradation with needed maintenance.

  Purpose

  The solutions ensure proper and timely maintenance of port machinery equipment and ensure the fleet runs even more efficiently, keeping vessels and cargo moving smoothly in and out of the port.

- **Port of Rotterdam, The Netherlands**

  The Port of Rotterdam has recently implemented big data analytics. CargoSmart Limited, a global shipment management software provider, recently announced that the Port of Rotterdam Authority is using its sailing schedule data to provide shippers with greater visibility to the available route options through its gateway. The Port of Rotterdam launched Navigate on May 10, 2017.

The online tool incorporates the deep sea and short sea schedules of 550 ports worldwide as well as the rail and barge connections between Rotterdam and over 150 European inland terminals. Navigate also contains a business directory with over 1,500 companies that are active in and around the Port of Rotterdam, as well as an empty depot tool. Users can find and compare the best connections through Rotterdam based on modes and expected transit times to make more informed decisions.

Incorporating CargoSmart's ocean schedules with the port's intermodal schedules allows the Port of Rotterdam to may provide a complete picture of its logistics coverage for shippers. CargoSmart may also save the port time from obtaining and continuously updating the schedule data from ocean carriers.

Purpose

Sustainable supply chain analysis for Reefer containers.

2.2.5 Studies of Maritime Big Data for Safety Improvement

The security of ship operations can be affected by several different reasons. To improve the general security of ship operations and to eliminate causes such as acts of piracy, various data about ship details and positions need to be monitored. The use of Big Data can not only prevent accidents caused by weather conditions, but also those caused by other participants in maritime traffic. In some seas, acts of piracy still pose a great risk to shipping. Information regarding the most recent pirate activities can help to divert other ships that are currently on the same route. One of the systems in the field of security is the Long Range Tracking and Identification (LRIT). It is an international tracking and identification system incorporated by the IMO under the SOLAS convention to ensure a thorough tracking system for ships across the world [79]. The system aims to enhance security for government authorities. LRIT provides ship identity and current location information insufficient time for a government to evaluate the security risk posed by a ship off its coast and to respond to reduce the risk if necessary. The system complements existing modes of tracking ships using coast-based AIS stations. Furthermore, the combined use of LRIT and Satellite-AIS data can increase the tracking quality and coverage of ships registered in the EU CDC [80].

Safety at sea can also benefit from implementing anomaly detection in marine operations. An overview of several machine learning techniques that can be used to detect anomalies from data gathered on vessel movement is presented in [81]. In [82], an application of sensor-based anomaly detection in maritime transport is demonstrated. Sensor data, which include environmental and ship system information, are streamed from a ship to shore where they are analyzed through a Big Data analytics platform. Auto Associative Kernel Regression and the Sequential Probability Ratio Test technique are used to detect anomalies and trigger alarms so that appropriate action can be taken as soon as possible. Critical points identified along a vessel trajectory are shown in Fig. 2.4.

Fig. 2.4 Critical points identified along a vessel trajectory, [83]

A solution for real-time monitoring of sensor data in a seaport infrastructure implemented in the Puerto de La Luz seaport in the Canary Islands is described in [84]. The system integrates data from AIS, various sensors, and external sources, and provides a 3D map showing the ingoing and outgoing vessels, as well as the environmental conditions. It is equipped with an alert system which means that the port can easily identify issues and notify the vessels to prevent collisions, help the vessels avoid high waves, etc. Another example of Big Data used for anomaly detection is CMAXS, a maintenance system developed by ClassNK [85]. It uses real-time data collected from flow, pressure, and temperature sensors on all engines and pumps.

The data are analyzed for anomalies to detect potential damage as well as generate recommendations that help minimize downtime and reduce repairs. In [86], the arrangement of Precaution Areas, whose purpose is minimizing the possibility of collisions, is optimized using AIS data. On the other hand, [87] shows how Big Data can be used to gain a better understanding of maritime activities, which is especially useful in remote areas such as the Arctic where shipping activity needs to be monitored to ensure sustainability and the information is otherwise difficult to access. It also discusses anomaly detection, such as detecting when a vessel deviates from the declared path, falsifies its AIS reports, or turns off its AIS transponder to potentially engage in illegal activities.

2.2.6 Maritime Big Data Application by NYK for Operational Efficiency, Safety, Maintenance.

In shipping, potential technical revolution areas are shifting from the Eco Ship to Smart Ship and from the hardware to software field. A key element of the revolution in Information Communication Technology (ICT). Throughout the maritime industries, researchers and developments are taking place that analyzes and utilizes any types of data taken from the ship to improve operational safety and efficiency. ICT toolkits include a variety of fields such as IoT, satellite communications, data analysis, Apps, system integrations, automation, and robotics.

In the industrial machinery fields that have taken the lead in this, for example in the wind turbine industry, the products are being developed with sensors and intelligent systems. It can prevent unexpected downtime, reduce maintenance costs and improve energy efficiency through their life cycle with condition monitoring, big data analysis, and supports from service engineers. IoT data collected from the real wind turbine are fed into virtual wind turbines on computers. The actual conditions of each wind turbine can be reproduced on computers using IoT data and data analysis and simulation technologies. It represents the current status of each wind turbine and allows forecasting under various estimated scenarios.

The digitalization activities related to harnessing maritime Big Data from NYK (Nippon Yusen Kabushiki Kaisha) is shown as follows:

- **Implementing Ship Information Management System [SIMS]**

  Through the implementation of SIMS (Ship Information Management System) in 2008, the NYK Group has been able to share data among workplaces on land and sea in real-time, including detailed hourly updates on shipping operations and data related to fuel consumption. Optimized economic vessel operations and energy-saving operations are realized by visualization of information and close information-sharing among crew members, shipowners, ship operators, and ship managers. SIMS has been installed on 190 of the operating vessels of NYK (as of the end of March 2019).

- **Making Onboard Work More Efficient [UMS Check System]**

  An unmanned machinery space (UMS) check system enables the operation status of onboard machinery to be checked and recorded when conducting the unmanned operation of an engine room at night. We introduced digital tablet entry for the approximately 2,000 items that were previously checked on paper, and this reduced crew work for identifying anomalies and allowed us to easily share data with other vessels. NYK also promoting the development of a system to enable detected anomalies to be notified to shoreside personnel and thereby identify machinery trouble at an early stage.

- **Optimizing navigation plans, making route planning more efficient, and enabling swift support of emergencies [J-Marine NeCST]**

NYK, MTI, and Japan Radio Co. Ltd. recently teamed up to develop J-Marine NeCST, an operational support tool that lets users manage and share electronic charts and other voyage information on large-screen displays. By fusing its years of experience in vessel operation management with its partners' technological capabilities, the Group has successfully developed a one-of-a-kind tool that not only lets users write information by hand on electronic charts but also boasts incredible operability and maneuverability. The tool even has a feature that integrates prepared route information with the Electronic Chart Display and Information Systems (ECDIS) that all large cargo vessels making international voyages must now have. The tool also lets users superimpose meteorological and sea conditions information, thereby streamlining and optimizing the process of drafting voyage plans. By digitizing vessel-specific information, as well, the tool enables information-sharing among ships and land-based sites to make onboard and onshore work processes more efficient.

- **Launching Shipmanagement Platform [NiBiKi]**

The NYK Group developed the ship management platform for data sharing called "NiBiKi" and launched services in December 2018. By the Safety Management System (SMS) manual, the onboard crew is required to file reports about safety management to the ship-management company. However, the conventional operation flow had some inefficiencies. Crews would have to draft various reports and applications, and then send emails with attachments to the ship management company to obtain approval. Afterward, crews would need to file printouts of the reports to retain on board. Moreover, the information reported by the ship was not being used efficiently to analyze this data, etc., because each vessel and ship management company would file the data independently.

Given this issue, the NYK Group developed the NiBiKi system, which digitalizes SMS manual documents and the application and approval workflow, thereby making it possible for crews to report and request approval accurately in a short period of time by simply completing the prescribed forms following the instructions. Furthermore, the data accumulated in the NiBiKi system will be shared with ship operators and ship management companies for high-quality big data analyses, enabling linkages with conventional safety and crew health activities. Going forward, we plan to build a more comprehensive system and further utilize data for crew training and drills.

- **Supporting Safe Vessel Handling [i.Master]**

Docking and undocking are some of the tensest times in vessel operations. To assist in reducing the risk of colliding with the quay, the Group has introduced i.Master software for handheld digital charts. The software gives crew members a bird's eye view of the vessel's movement and the surrounding situation. Via a tablet computer, the software shows crew members the course of a vessel and its docking or undocking speed and automatically identifies other vessels. Furthermore, the system allows crew members to monitor docking and undocking constantly without being on the bridge.

Collisions with the quay could force vessels to lay up for long periods. As well as inconveniencing customers, such delays would lead to a loss of trust. By using i.Master effectively, it can reduce the risk inherent in docking and undocking.

i.Master is just one example of the innovative efforts to develop useful technologies for a range of operational situations and thereby build systems that ensure stable, safe, and efficient vessel operations.

- **Taking Part in "i-Shipping," a Japanese Government R&D Project for Advanced Safety Technology**

  IoT and other information technology are showing enormous potential to transform the shipping industry. Japan's Ministry of Land, Infrastructure, Transport, and Tourism (MLIT) is promoting research and development projects in this new technology area for a "productivity revolution in the maritime business" (i-Shipping). NYK, MTI, and other industry partners are now cooperating and pursuing technological development initiatives on four i-Shipping projects, including "vessel machinery plant trouble-prevention" and "collision avoidance and autonomous operation."

Other researchers also harnessing maritime big data for energy efficiency improvement. Perrera et al. [88] propose the data flow chart as presented in Fig. 2.5. The data are collected from various onboard sensors and data acquisition systems, preprocessed and transmitted to shore-based data centers where they are stored and analyzed. The result of the analysis is information that supports decision-making, for example, to improve energy efficiency and system reliability.

Monitoring fuel consumption, various emissions, the use of lighting, heating, and similar processes can result in insights that support decision-making. In [53], data such as wind speed and direction, average draft, trim, main engine power, shaft speed, and fuel consumption are analyzed and several higher fuel consumption trends under these parameter variations are noted. The optimal trim configuration is identified concerning the fuel consumption rates. Applying strategies based on this information enables ships to meet energy efficiency and emission control standards. Along with environmental benefits, this is also significant for cost reduction.

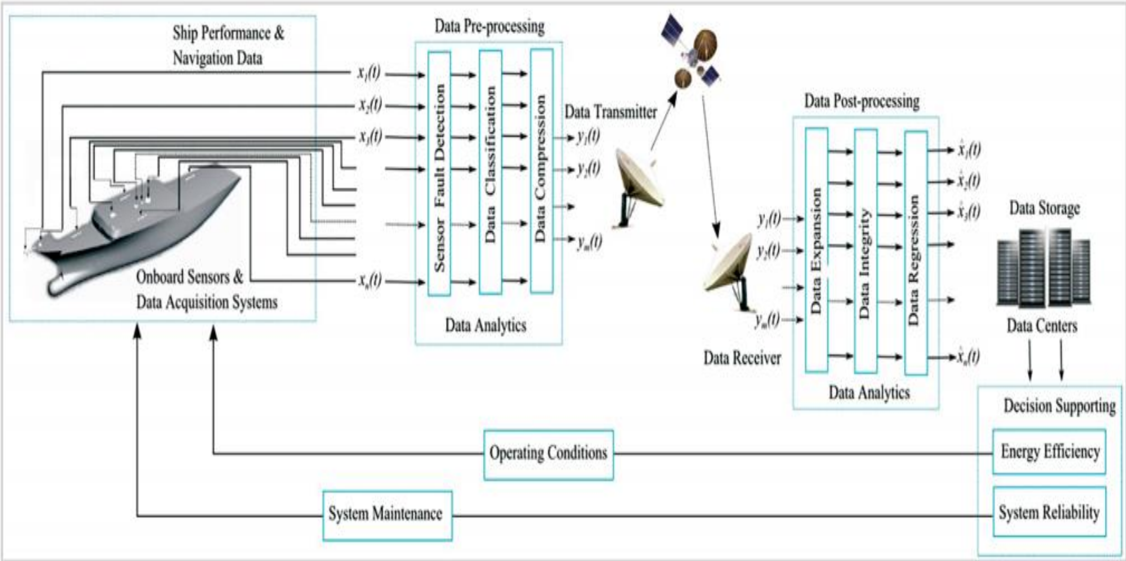


Fig. 2.5 Data flow chart in ship performance and navigation information, [53]

2.2.7 Maritime Big Data Application for Other Purposes

The following table provides a snapshot of the application and the key plan application areas for big data in the maritime industry [89].

Table 2.1. Application area for big data in the maritime industry

Role	Function	Example of Big Data Application
Ship Operator	Operator	Energy-saving operation Save operation Schedule management
	Fleet Planning	Fleet allocation Service planning Chartering
Shipowner	Technical management	Safe operation Condition monitoring and maintenance Environmental regulation compliance Hull and propeller cleaning Retrofit and modification
	New building	Design optimization

- **Chartering**

  A key function of charterers is to find the right ship for cargo at the most economical price. The task is highly dependent on information provided to them by known brokers and ship owners. However, as this information is limited, it may or may not be most efficient. Big data analytics can provide charters with readily available, accurate, and actionable information to improve decision-making.

  Charters can integrate Automatic Identification System (AIS) information, position reports, estimated times of arrival, vessel particulars (such as size), and market information into an exchange portal to find all available alternatives as well as the freight forecast. This can give charterers and ship owners access to more options thus improving transparency and competitiveness. Bahri, the national shipping company of Saudi Arabia, has developed various data models in its new data platform, BahriData, to cover various key business areas such as chartering, voyage management, fleet operations, maintenance, and customer services.

- **Operations**

  Speed: Ships, like automobiles, have optimum speeds, and various tests are conducted at the time of vessel delivery to determine the optimum speed for fuel consumption. However, operating a vessel at its optimum speed is difficult as it changes over time due to a variety of factors such as engine wear and maintenance. Big data analytics can help shipowners determine the optimum speed for fuel consumption, taking into consideration factors such as bunker cost, freight rates, and schedules.

  Maintenance: Decisions regarding vessel maintenance, including hull cleaning and propeller polishing, are taken based on intuition or a schedule rather than on actual vessel performance. Fuel consumption data can also be used for cost-benefit analysis of vessel maintenance. Data analytics can make it easier for operators to decide the timing and the benefits of performing maintenance.

- **Voyage Operations**

  Terminal operators, voyage managers, or port agents need estimated time of arrival (ETA) and cargo information. Vessels can be tracked using dashboards instead of relying on notes, emails, or phone calls. This helps in making more effective decisions about terminal and berth allocation, cargo handling, and route tracking. Dashboards can also provide information about any deviations from optimum performance. The ideal route, the weather service-provided route, and the actual route can be tracked in real-time. Any changes to speed, ETA, and other factors can be tracked and managed in real-time, thus ensuring that the voyage goes as planned and remains profitable.

  * ClassNK-NAPA GREEN offers a real-time big data analysis performance monitoring and optimization solution. The solution passes data collected from both on-board and shore-side sources through advanced and predictive algorithms to deliver information on current operations and on potential operational changes to allow vessels to reduce fuel consumption.

- **Vetting**

  Vessel owners and operators try to ensure that their fleets are acceptable for use by charterers. Instead of improving the vessel quality, they focus on meeting or passing the acceptance criteria. The process of vetting includes getting feedback from various entities such as inspectors, terminals, and port state authorities, as well as operator self-assessment. Data analytics can help charterers and vetting organizations analyze the different sources of information and select the right vessel with the least amount of risk involved in pollution preparedness, safety management, and navigation

2.2.8 Maritime Big Data in this Studies

Recently, the era of shipping records big data has been started along with mandatory digitalized ship movement with AIS (Automatic Identification System). Substantial data amount is generated by AIS, such as the ship's unique identification of international maritime organization number (IMO number), position, course, speed, and destination [89]. AIS is required to be installed for the ship on international voyages of over 300 gross tonnage and all passenger ships.

An AIS transmitter exchanges data with other nearby ships, AIS land-based systems, and satellites with the purpose of collision avoidance [89]. Moreover, maritime logistics big data, such as ship and port specification data, route data, international trade data, and data provided by AIS, are currently available and can be used. Based on that, some studies have been conducted on big data utilization to improve ship construction, ship operations, and ship maintenance [90-92]. However, studies regarding the utilization of such maritime logistics big data for ship basic planning or ship design are limited and have been identified as important future tasks [89].

Big data use in this study is consists of AIS data, ship operational data, ship data, port data, route data, and trade data.

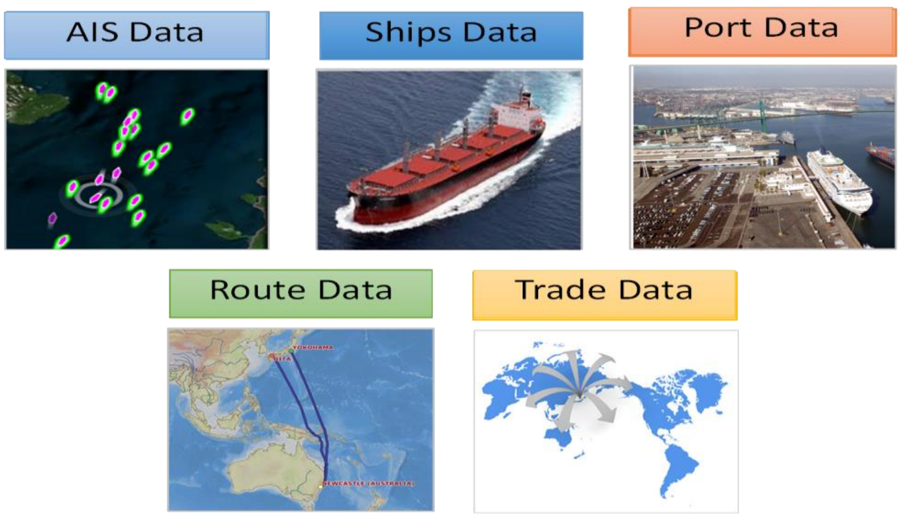


Fig. 2.6 Big data in this study

2.3 Characteristics of this Study

The characteristics of this study are represented based on the following comparison with the previous studies as follows:

Study object	Main analysis	Data Sources	Literature
E-Navigation	Ship navigation, ship communication	VTS, AIS, Sensors & Tracking System	[25][58][59][60][61]
Weather Routing	Route decision, planning, collision avoidance	Data Acquisition Module, AIS	[25][37][62][63][64] [65][66][67][68][69]
Ports & Terminals	Ship congestion, delays, safety and maintenance	AIS, ETA & ETD	[25][62][74][75] [76][77]
Safety Improvement	Ship security and safety at sea	Long Range Tracking and Identification (LRIT)	[79][80][81][82] [83][84][85][86][87]
Operational Efficiency	Ship management, FOC monitoring	J-Marine NeCST NiBiKi, i-Shipping	[53][63][88][90][92]
Demand Forecasting	System dynamic, demand forecasting	Statistical Data, Big Data	[14][17][39]
Marine Logistics	Ship basic planning Cargo assessment Ship allocation	MLDB (Marine Logistics Database)	This Study

2.3.1 Marine Logistic Databases (MLDB)

In this study, the authors developed the MLDB using AIS and statistical data. The MLDB consists of the latest marine logistics data, i.e., operation information from AIS, ship, port, route, and international trading information, as shown in Fig. 2.7. The data are managed, integrated, and structured to derive valuable insights from information buried in marine logistics data.

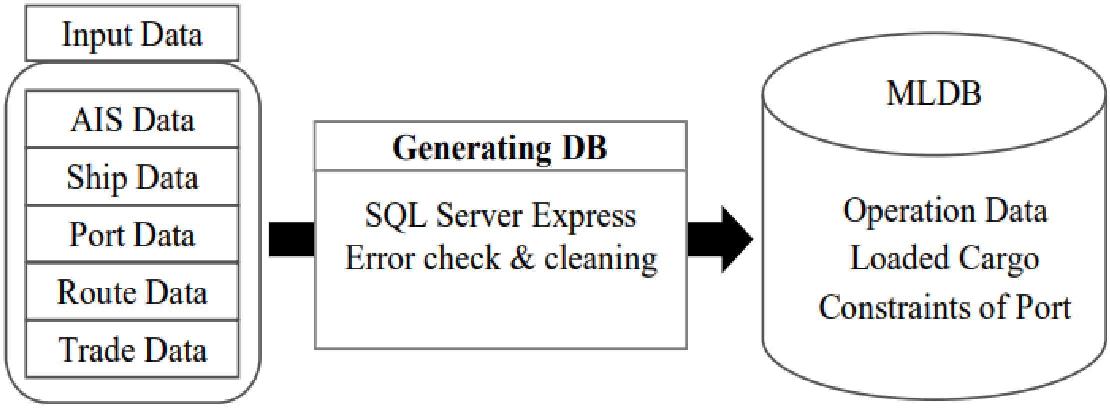


Fig. 2.7 MLDB

2.3.2 Ship Allocation Model

The ship allocation model constructed aims to replicate the actual ship allocation. The actual shipping market has the essential affection in the actual ship allocated. By utilizing the MLDB, the scheme of ship allocation model used in this study is shown in Fig. 2.8. With the ship allocation model scheme, the ship allocated in chartered contract scenarios, and long-term contracts are to be concluded on an annual cycle where a single shipowner manages certain freight requests continuously in one year.

The ship allocation model consists of three specified models, shipper model, shipowner model, and operator model. The shipper model creates the clusters of exporter and importer ports, and its cargo demand is to be transported in the selected year. Later, the demand estimated will be used in the operator model as a freight transport request to the shipowner model. Following the request, the shipowner bid the cost and transportable volume on the route where the shipowner ship operates. To finalize the charter contract between operator and shipowner, the operator will choose the served bid in economical consideration and create ship allocation.

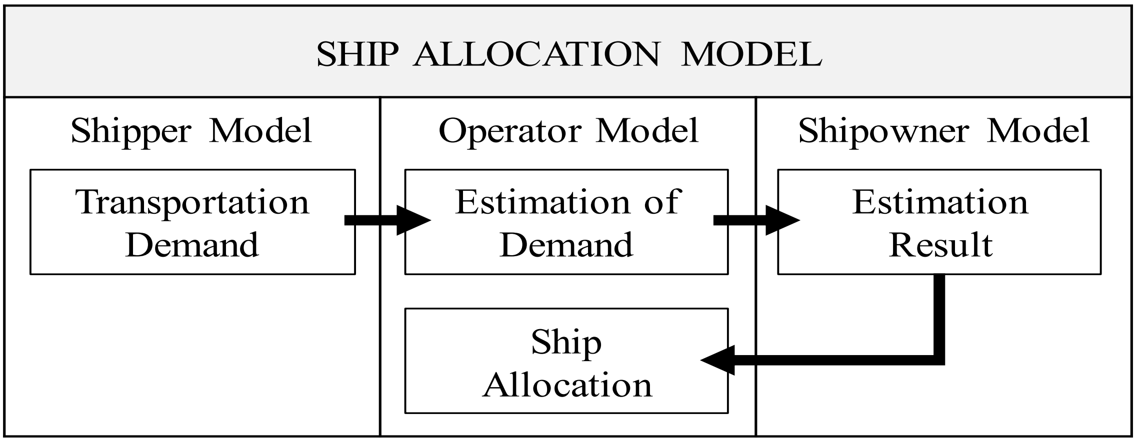
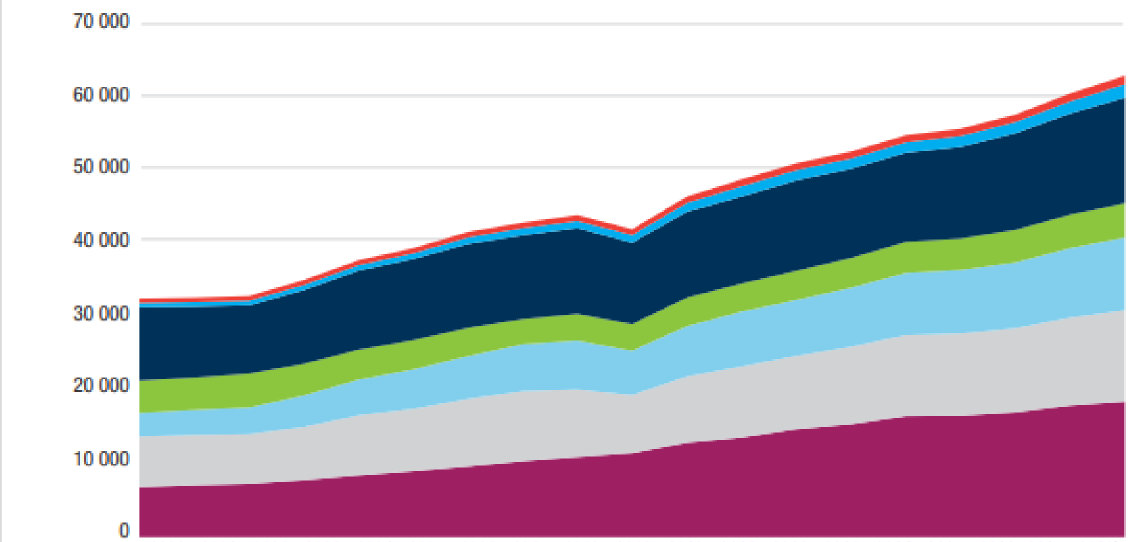


Fig. 2.8 Ship allocation model

2.3.3 Importance of Bulk Carrier

2.3.3.1 World Seaborn Trade

Seaborne trade was measured in ton-miles to reflect distances traveled and the employment of ship capacity increased by 5 percent in 2017, up from 3.41 percent in 2016. Overall ton-miles generated by seaborne trade in 2017 amounted to an estimated 58,098 billion tons (Fig. 2.9). Much of the growth was driven by crude oil and coal shipments, which have greatly benefited the shipping industry, given the growth in volumes and distances. Crude oil trade contributed 17.5 percent to ton-mile growth while major dry bulks contributed nearly one-third. Together, minor bulks and other dry cargo accounted for 17.7 percent of ton-mile growth, while containerized shipments contributed 17.4 percent. The contributions of gas and petroleum products were much smaller [93].



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017[a]	2018[b]
Chemicals	580	589	620	632	650	676	713	747	759	783	848	888	903	925	920	961	993	1 058	1 111
Gas	576	591	611	662	717	735	833	913	956	958	1 148	1 344	1 333	1 337	1 381	1 421	1 462	1 595	1 766
Oil	9 614	9 303	8 938	9 665	10 348	10 654	10 984	10 981	11 211	10 679	11 255	11 420	11 831	11 657	11 659	11 993	12 657	13 216	13 809
Other dry cargo	4 233	4 245	4 414	4 150	3 920	3 818	3 712	3 257	3 517	3 481	3 723	3 645	3 795	3 923	4 065	4 139	4 242	4 384	4 497
Containers	3 111	3 279	3 512	4 124	4 687	5 158	5 601	6 178	6 431	5 815	6 588	7 206	7 352	7 712	8 157	8 290	8 635	9 117	9 535
Minor dry bulks	6 638	6 573	6 538	6 965	7 876	8 170	8 852	9 160	8 817	7 586	8 705	9 312	9 624	10 172	10 617	10 775	11 018	11 510	11 967
Main Bulks	6 509	6 793	6 937	7 448	8 061	8 626	9 245	9 941	10 476	11 006	12 336	13 019	14 099	14 764	15 828	15 897	16 314	17 217	17 729

Source: UNCTAD secretariat calculations, based on data from Clarksons Research, 2018a.

Fig. 2.9 World seaborne trade in cargo ton-miles, 2000–2018 (BTM), [93]

2.3.3.2 Dry-cargo trades: The mainstay of seaborne trade in 2017

Following a limited expansion in 2015–2016, global dry bulk trade1 grew by about 4 percent in 2017, bringing total volumes to 5.1 billion tons (Table 2). A sharp increase in iron ore imports to China, a rebound in the global coal trade, and improved growth in minor bulk trades supported the expansion. Overall, strong import demand in China remained the main factor behind growth in global dry bulk trade. An overview of global players in the dry bulk commodities trade sector is presented in Table 3-6 [93].

Iron ore imports to China increased by 5 percent in 2017, bringing total volumes to nearly 1.1 billion tons. With a market share of more than 70 percent, China remains the main source of global iron ore demand. A rise in steel production and the closure of more than 100 million tons per annum of outdated steelmaking capacity in 2016–2017 boosted the country’s demand for imports. Further, the increased use of higher-grade imported iron ore displaced domestic supplies. The leading iron ore exporters were Australia, Brazil, and South Africa; Australia and Brazil supplied over 85 percent of the demand for imports in China. Nevertheless, Australia is by far the largest exporter, supplying nearly two-thirds of iron ore requirements in China. The country imports 21 percent of its iron ore requirements from Brazil, which benefits the dry bulk shipping industry through long distances. South Africa generates 4 percent of all iron ore imports to China. Other suppliers, such as India, the Islamic Republic of Iran, and Sierra Leone, have also increased their exports to China.

Table 2.2 Dry bulk trade 2016–2017 (MT and % annual change)

Dry Bulk	2016	2017	Percentage change 2016-2017
Main Bulks of which	3040.9	3196.3	5.1
Iron ore	1418.1	1472.7	3.9
Coal	1141.9	1208.5	5.8
Grain	480.9	515.1	7.1
Minor Bulks of which	1874.6	1916.5	2.2
Steel products	406.0	390.0	-3.9
Forest Products	354.6	363.6	2.5
Total Dry Bulks	4915.5	5112.8	4.0

Global grain trade, including wheat, coarse grains, and soybeans, reached 515.1 million tons in 2017, a 7.1 percent increase over 2016. Exports are dominated by a few countries, notably the United States; importers tend to be regionally diverse. As in other dry bulk trades, Asia was a driving force of growth, albeit not the only one. In 2017, grain trade was underpinned by a 14.7 percent increase in soybean imports to China and growing exports from Brazil and the United States.

Table 2.3. Major dry bulks and steel: Producers and Users 2017

Steel Producers	%	Steel Users	%
China	49	China	46
Japan	6	United States	6
India	6	India	5
United States	5	Japan	4
Russian Federation	4	Republic of Korea	4
Republic of Korea	4	Germany	3
Germany	3	Russian Federation	3
Turkey	2	Turkey	2
Brazil	2	Mexico	2
Others	19	Others	25

Table 2.4. Major dry bulks and Iron Ore: Exporters and Importers 2017

Iron Ore Exporters	%	Iron Ore Importers	%
Australia	56	China	72
Brazil	26	Japan	9
South Africa	4	Europe	8
Canada	3	Republic of Korea	5
India	2	Others	6
Others	9	-	-

Table 2.5. Major dry bulks and Coal: Exporters and Importers 2017

Coal Exporters	%	Coal Importers	%
Indonesia	32	China	18
Australia	30	India	17
Colombia	7	Japan	15
United States	7	European Union	13
South Africa	7	Republic of Korea	12
Canada	2	Taiwan Province of China	6
Others	15	Malaysia	3
-	-	Others	16

Table 2.6. Major dry bulks and Grain: Exporters and Importers 2017

Grain Exporters	%	Grain Importers	%
United States	25	East and South Asia	34
Russian Federation	23	Africa	21
Ukraine	15	Developing America	20
Argentina	11	Western Asia	16
European Union	9	European Union	7
Australia	8	Transition Economies	2
Canada	7	-	-
Other	2	-	-

China dominates the soybean trade and accounted for nearly two-thirds of the global soybean import demand in 2017. Outside Asia and the European Union, some lesser consuming regions, such as Africa and Western Asia, also contributed to such growth. Tariffs by the United States on certain goods imported from China, including steel and aluminum, and retaliation by China, may lead to restricting soybean import from the United States.

China is the world's largest consumer and importer of uncrushed soybeans. However, it may decide to replace imports from the United States and source its soybean requirements from alternative suppliers such as Brazil. While trade restrictions generally portend ominous consequences for shipping, a shift in suppliers and routes in this context may have an unintended positive effect on ton-miles generated.

Growing manufacturing activity and construction demand supported a 2.2 percent increase in minor bulks commodity trade. Rising demand for commodities such as bauxite, scrap, and nickel ore pushed volumes to 1.9 billion tons. However, the large drop (less 30.8 percent) in exports of steel products from China due to reforms in the country's steel sector undermined the expansion to some extent. Bauxite shipments expanded by 19.5 percent, accounting for 13 percent of minor dry bulks commodities trade in 2017.

The continued rise in Chinese aluminum production and the availability of bauxite ore, following years of export disruptions, led to an expansion in the bauxite trade. While China dominates the import side with a market share of more than two-thirds, key players on the supply side are more varied and include Australia, Brazil, Guinea, and India. Nickel ore trade rose by 7.6 percent, highlighted in particular by increased growth in nickel ore shipments from Indonesia, following its decision to relax its export ban on unprocessed ores.

Based on the information above, the global bulk seaborne trade is expected to be growing higher in the next years. Based on this condition, Bulk Carrier transportation is important. However, the research of ship basic planning support system for Bulk Carrier has not been developed. Based on the information appointed above, this study is focused on the decision of ship specifications considering ship operations for the latest market conditions. To realize that, the system which can generate ship allocation that matches a real ship allocation is developed based on the data extracted from MLDB. Therefore, thorough executing the simulation by changing the ship specifications or future scenarios, such as fuel price and cargo demand between ports, the competitive ship principal particulars and performance can be examined.

Chapter 3

Basic Concept

3.1 Marine Logistic Database (MLDB)

Marine Logistic Database denoted as MLDB is defined as a collection of Marine Logistic Big Data sets that are integrated into a relational database. MLDB is developed by integrating maritime logistics big data, such as AIS data, ship operation data, ship and port data, route data, and trade data. The basic concept of MLDB is described below.

3.1.1 MLDB Input Data

To develop the MLDB, we employed the following data as input.

- AIS data: indicated speed, indicated draft, ship position, timing arrival, and departure dates, and arrival and departure port collected from the Market Intelligence Network
- Port data: port name, longitude, latitude, port dimension, and cargo handling collected from Sea-web Port
- Ship data: ship name, deadweight, International Maritime Organization number, classification, ship dimension, operator, shipbuilder, ship status, and build year collected from Sea-web Ship
- Route data: departure port, arrival port, route choices, and distances collected from Sea-web Port and IHS-Fairplay
- Trade data: commodity trade, the period of trade between countries, commodity code, trade value, trade quantity, reporter, and partner collected from UN Comtrade.

3.1.2 MLDB Structure

To more easily extract valuable information, the structure for the MLDB should be defined and modified unstructured data into a relational database. For example, by integrating ship and port data with operation data, some information related to a ship's operational state can be analyzed (e.g., berthing, anchoring, or sailing).

3.1.3 Error Cleaning

Massive amounts of data are available for the organization which will influence their business decision. Data collected from the various resources are dirty and this will affect the accuracy of prediction results. Data cleansing offers a better data quality which will be a great help for the organization to make sure their data is ready for the analyzing phase. However, the amount of data collected by the organizations has been increasing every year, which is making most of the existing methods no longer suitable for big data. The data cleansing process mainly consists of identifying the errors, detecting the errors, and corrects them. Despite the data need to be analyzed quickly, the data cleansing process is complex and time-consuming to make sure the cleansed data have a better quality of data [94].

Incomplete information will generate uncertainties during data analysis and this must be managed in the data cleansing stage. Errors or missing values in the dataset will produce a different result and may affect the business decision. The data must be accurate to avoid losses, problems, and additional costs due to the poor quality of data. For example, according to the "Price Waterhouse Coopers" survey conducted in 2001, 75% of 599 companies have suffered losses due to data quality issues [95]. Since these businesses rely on data like customer relationship management and supply chain management, therefore they need to have excellent quality data to achieve a more precise and useful result. Quality data only can be produced by data cleansing as the data collected from the various sources might be dirty [96].

Data quality can be defined as the fitness of data to fulfill the business requirement. It is achieved through people, technology, and processes. It ensures compliance and consistency particularly when data from different databases are combined. Without proper data quality management, even a minor error might cause revenue loss, process inefficiency, and failure to comply with the industry and government regulations [97]. Thus, data quality and data cleansing are always linked together as ensuring data quality is critical and necessary before the sharpening of analytic focus can occur [98].

Investigations into the problems related to data quality can be traced back to as early as the late 1960s when a mathematical theory for considering the duplicate problem in statistical data sets was proposed by Fellegi and Sunter [99]. However, it is only in the 1990s that the data quality problem has been considered in computer science with the data stored in databases and data warehouse systems. More and more people have become aware that poor data quality is one of the main reasons for the failure of a database project. Though a variety of definitions for data quality have been given [100-102], studies show that still, no formal definition for data quality exists [101].

From the literature, data quality can be defined as "fitness for use", i.e., the ability of data to meet the user's requirement. The nature of this definition directly implies that the concept of data quality is relative. Orr states "the problem of data quality is fundamentally intertwined in how our system fits into the real world; in other words, with how users use the data in the system" [101]. This has two interpretations: one is that if a data set is available and is as good as it can be, there are no other options than to use it. The other one is that what is considered quality data in one case may not be sufficient in other cases.

For example, an analysis of the financial position of a company may require data in units of thousands of pounds while an auditor requires precision to the pence, i.e. in real life, it is the business policy or business rules that determine whether or not the data is of quality. Generally speaking, data quality can be measured or assessed with a set of characteristics or quality properties called data quality dimensions [103]. Some commonly used data quality dimensions include accuracy, completeness, timeliness, and consistency, which can be defined as:

- Accuracy – conformity of the recorded value with the actual value;
- Timeliness – the recorded value is not out of date;
- Completeness – all values for a certain variable are recorded;
- Consistency – the representation of data is uniform in all cases.

Therefore, data quality can be considered a multi-dimensional concept. These data-quality dimensions measure data quality from different angles. Within each of these dimensions, a set of data quality rules generated by real business policies can be used to assess the data quality reflected by each dimension [104]. For example, a data quality rule defined as „the value of date must follow the pattern of DD/MM/YYYY" can be used for the consistency dimension.

Data cleansing is an operation that is performed on the existing data to remove anomalies and obtain the data collection which is an accurate and unique representation of the mini world [105]. It involves eliminating errors, resolving inconsistencies, and transforming the data into a uniform format [106]. With the vast amount of data collected, manual data cleansing is almost impossible as it is time-consuming and prone to errors. The data cleansing process is complex and consists of several stages which include specifying the quality rules, detecting data error, and repairing the error [107].

There is no commonly agreed formal definition of data cleaning. Depending on the particular area in which data cleaning has been applied, various definitions have been given. The major areas that include data cleaning as part of their defining processes are data warehousing, knowledge discovery in databases (KDD), and total data/information quality management (TDQM).

Within the data warehousing field, data cleaning is typically employed when several databases are merged. Records referring to the same entity are often represented in different formats in different data sets. Thus, duplicate records will appear in the merged database. The issue is to identify and eliminate these duplicates. The problem is known as the merge/purge problem [108]. Other instances of this problem are also referred to as record linkage, semantic integration, instance identification or the object identify the problem in the literature [109]. There are a variety of methods proposed to address this issue: knowledge bases [110], regular expression matches and user-defined constraints [111], filtering [112], and others [113-115].

In the KDD process, data cleaning is regarded as a first step or a pre-processing step. However, no precise definition and perspective over the data cleaning process is given and data cleaning activities are performed in a very domain-specific fashion. For example, Simoudis et al [116] defined data cleaning as the process that implements computerized methods of examining databases, detecting missing and incorrect data, and correcting errors. In data mining, data cleaning is emphasized concerning the garbage in garbage out principle and its techniques such as outlier detection where the goal is to find exceptions. For example, the problem of outlier detection where the goal is to find exceptions [117, 118] can be used in data cleaning. Total data quality management is an area of interest both within the research and business communities. From the literature, the data quality issue and its integration in the business process are tackled from various points of view [103, 117, 101, 119-122]. It is also referred to as the enterprise data quality management problem. However, none of the literature refers to the data cleaning problem explicitly. Most of this work deals with the process management issues from the data quality perspective, others with the definition of data quality.

Of particular interest in this area, the definition of data quality can help to define the data cleaning process to some extent. For example, within the model of data life cycles proposed by Levitin and Redman [120], data acquisition and data usage cycles contain the following series of activities: assessment, analysis, adjustment, and discarding of data. This series of activities proposed in Levitin and Redman"s model define the data cleaning process from the perspective of data quality. Fox et al [119] proposed four data quality dimensions of the data, i.e., accuracy, currentness, completeness, and consistency. The correctness of data is defined in terms of these dimensions. Thus, the data cleaning process within Fox et all, data quality framework can be defined as the process that assesses the correctness of data and improves its quality.

With the above in mind and related literature [123, 124], data cleaning must be viewed as a process that is tied directly to data acquisition and definition or is applied to improving data quality in an existing system. For example, in Müller and Freytag"s work, comprehensive data cleaning is defined as the entirety of operations performed on existing data to remove anomalies and receive a data collection being an accurate and unique representation of the mini-world [123]. According to Müller and Freytag"s work, the three major steps within the data cleaning process are (i) define and determine error types, (ii) search and identify error instances, and (iii) correct the uncovered errors. Müller and Freytag include four major steps within the process of data cleaning: (i) auditing data to identify the types of anomalies reducing the data quality, (ii) choosing appropriate methods to automatically detect and remove them (specification of data cleaning), (iii) applying the methods to the tuples in the data collection (execution of data cleaning), and (iv) the post-processing or control step where the results are checked and the exception handling for tuples not corrected within the actual processing are handled.

In this study, to ensure and the reliability and quality of the data used to construct the MLDB, the following error cleaning methods should be performed.

- Keeping the first data recorded in AIS based on the arrival date and time, and deleting duplicate data points.
- Deleting unrealistic voyage data by checking the average voyage speed, which is calculated by considering the navigation days and distance between two ports. If the average voyage speed exceeds the service speed, it is defined as an error and the data are deleted.
- Deleting inappropriate zero values, such as 0-m drafts, null data, and unavailable data.

3.2 Ship Allocation Model

Allocation is the process of assigning product items from the inventory to shipping orders and then fulfilling the shipping orders from appropriate fulfillment sites such as drop-ship vendors, virtual sites, warehouses, or a retail store. Goods are carried by sea under a contract of carriage between the shipper and the shipowner. The shipper may employ a forwarding agent to arrange the transport, while the Shipowner may employ a loading broker to control the allocation of space and advertise the service, and to make the loading arrangements and prepare documents on the shipowner's behalf.

When a shipper wants to send a particular cargo with a particular ship on a scheduled service, a "shipping note" for the consignment is completed by the shipper and forwarded to the shipowner or his agent. This note will have to contain a brief description of the commodity. The loading broker then compiles a list of the consignments intended for shipment, the booking list. This is sent to the ship to enable the Master to plan the stow and to the stevedore to arrange the loading. The shipper may receive a "booking note", which specifies that the carrier reserves space for a specified volume and kind of cargo in a named vessel between named ports. The broker may also issue a "calling forward notice" to the shipper, advising him of the time and place at which he is to deliver the goods.

When the cargo is delivered to the warehouse or the ship, a receipt for that cargo must be obtained by the shipper. When the cargo is placed on board, this is called a "mate's receipt". This receipt acknowledges that the goods have been loaded and have been properly and carefully handled, loaded, and stowed. If there are any damages to the goods before loading, this will be recorded on the receipt, and it is no longer "clean".

In some trades, it is customary for the shippers to have a "boat note" following the cargo. When the "boat note" is signed by the cargo officer aboard the ship, it becomes a "mate's receipt". With many shipping companies, it is the practice to give an official "mate's receipt" irrespective of the fact that a boat note may be provided by the shipper.

Modern practice is to present a copy of the shipping note as the boat note, which when endorsed, becomes the "mate's receipt".

Special tally companies are engaged by the shipowner to check or keep a record of all cargo loaded into and discharged from a vessel. This is an essential part of cargo work to prevent claims upon the ship for so-called "short" discharge, i.e. when some of the cargo is missing. It is sometimes customary for the shipper or consignee to provide his tally clerks, particularly with cargoes of a straight nature, such as bags, bales, etc.

A copy of the "mate's receipt" will be returned to the shipowner, so that a "bill of lading" can be issued to the shipper. The "bill of lading" acknowledges that the goods have been "shipped in apparent good order and condition" if the "mate's receipt" is clean. Otherwise, comments are transferred to the "bill of lading". This document is issued under all forms of shipping, scheduled or not. The complete list of cargo loaded, as compiled from the "bills of lading" form the "manifest" of the ship. Customs regulations at most ports require at least one copy of the manifest and copies are also required for stevedores at discharging ports.

While cargoes are in transit, they may be sold so that the goods change ownership. Such a sale will be represented by the "bill of lading" changing hands. At the port of discharge, the consignment will be handed over to the party presenting the original "bill of lading".

The basic concept of the ship allocation model is shown in Fig. 3.1 which highlights the two important steps for realizing the objectives: first, developing a ship allocation model, and second, carrying out simulations using the ship allocation model. A ship allocation model can reproduce actual ship allocation. Input and output data of the ship allocation model are shown in Table 3.1.

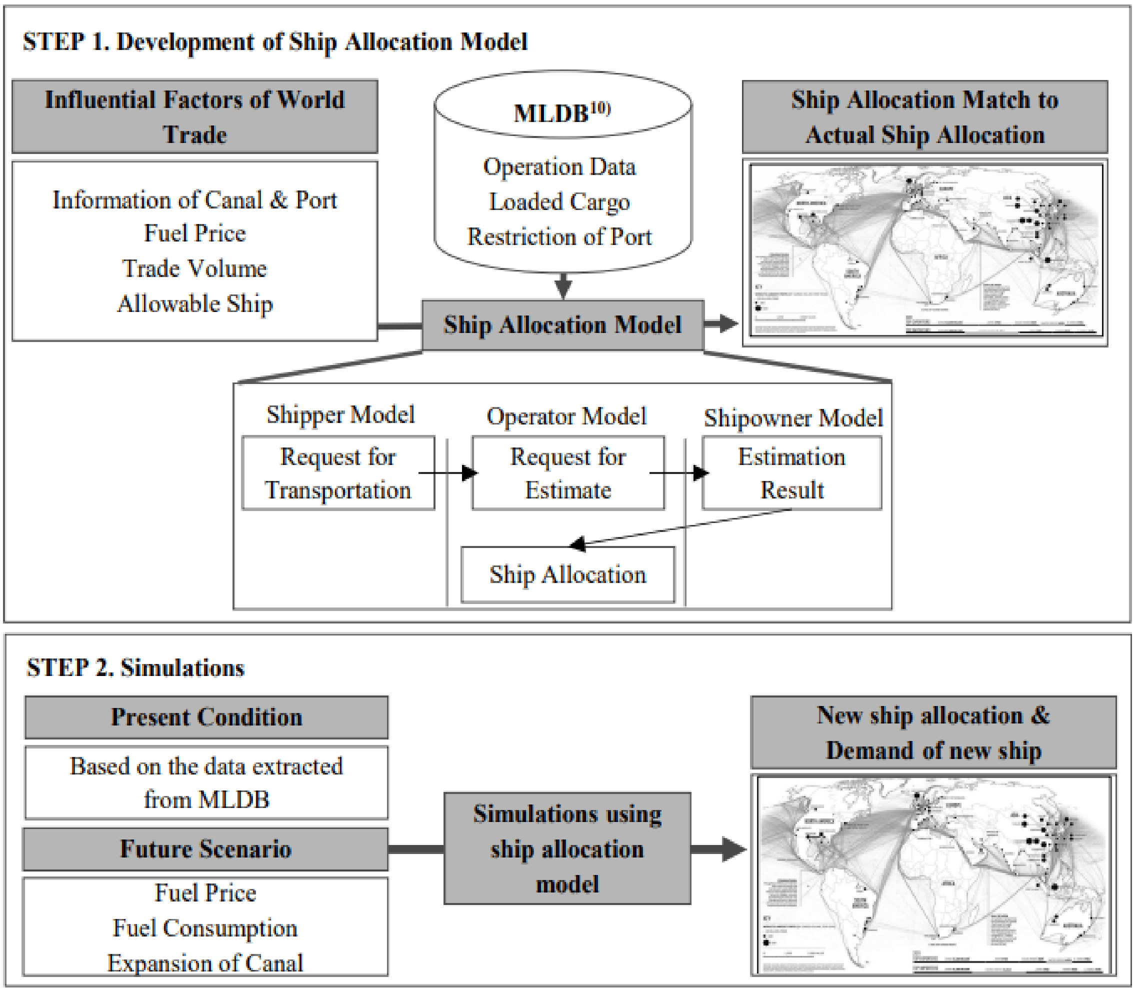


Fig. 3.1 Basic concept of ship allocation

Table 3.1. Input and output of ship allocation model

Input	Trade conditions: fuel price, trade volume between the ports.
	Allowable ships specifications: DWT, LOA (m), B (m), D (m), d (m), service speed (knot), horsepower
	Constraints of the canal and ports: Max DWT, Max LOA (m), Max B (m), and Max D (m), etc.
	Number of allowable ships
Output	Ship allocation of all ships

3.2.1 Configuration of the ship allocation model

To realize actual ship allocation conditions, we develop three distinct models—the shipper, shipowner, and operator models.

- The shipper model issues a request for cargo transportation between two or more ports. The shipper model is defined using cluster analysis. •
- The shipowner model estimates the shipment days, amount of cargo, and operating cost in response to the cargo transportation requests. The shipowner model is defined using deep learning analysis.
- The operator model requests all shipowner models to estimate shipment costs, cargo volume, and transport time based on shipper requests; then, based on the answer from the shipowner model, the operator model decides on a ship for cargo transport. A detailed explanation of ship allocation and confirmation of the proposed models is shown in Section 5.

3.3 Simulation Using Ship Allocation Model

In the ship allocation model, a competitive ship is allocated to a prioritized route. Moreover, the following data can be set freely in executing the simulation:

- Future scenario (fuel price, trade volume between ports)
- New ship specifications
- Number of ships (freely selectable by the operator)
- Constraints of ports and canals

With these characteristics, we can execute the following simulations:

(1) Examination of the supply-demand balance of various ships

In our system, supply is defined as a ship allocation only using existing ships, and demand is defined as a ship allocation in which all the ships can be used freely for cargo transportation. Therefore, by changing the number of ships that can be used in the simulation, we can estimate the supply-demand balance.

(2) Examination of effective ship specifications

As discussed in the previous section, we can change the specifications freely. Then, by changing ship specifications and simulating ship allocation, we can understand the demand for various kinds of ships. Therefore, by executing this simulation, we can examine effective ship specifications and the kinds of ships that are attractive for operation on the intended routes.

(3) Influence of economic situations on demand.

In our system, port constraints can be changed. Moreover, fuel prices and trade volume between the ports can be freely changed. By forecasting such a future situation using a ship allocation model, we can understand logistics and demand results. Moreover, we can understand the kinds of ships that will operate effectively on intended routes in the future. In this paper, we take bulk carriers that operate between Australia and Japan as an example. Detailed simulations are shown in Section 6.

Chapter 4

Development of Marine Logistics Database

As mentions in the previous section, the objective of this study is to develop a support system of ship basic planning using Marine Logistics Big Data. There are 3 important points to realize the objectives; Marine Logistics Database (MLDB), Ship Allocation Model, and Simulations. Below is the detailed description of the Marine Logistics Database (MLDB)

4.1 Definition of Marine Logistic Database (MLDB)

The marine logistic database is defined as an integration of maritime Big Data that is structured and managed into a relational database and denoted as MLDB. The MLDB consists of the latest marine logistics data, i.e., operation information from AIS, ship, port, route, and international trading information, as shown in Fig. 4.1. The data are managed, integrated, and structured to derive valuable insights from information buried in marine logistics data.

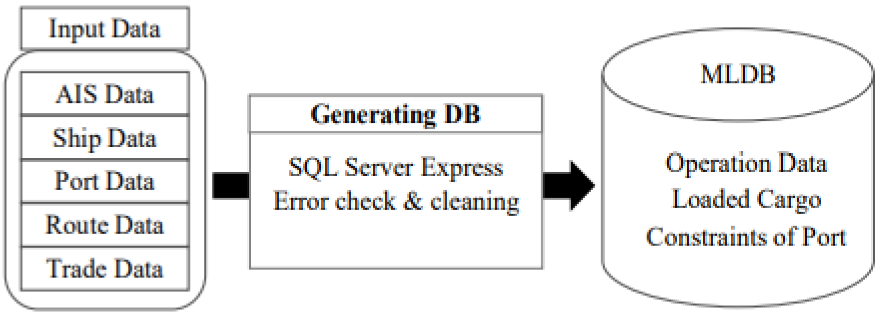


Fig. 4.1 Basic concept of MLDB

4.2 Development of MLDB

Marine logistic database (MLDB) is developed by integrating the following big data i.e. operation information from AIS, ship, port, route, and international trading information as an input. The development of MLDB is described as follows:

4.2.1 Input of MLDB

A. Automatic Identification System (AIS) Data

The Automatic Identification System (AIS) is a worldwide automatic positioning system based on fitting small transponders to vessels that continuously transmits a signal. This alerts other vessels and shore stations with AIS receivers to the presence of that vessel. The position information is supplemented with additional information about the vessel. The signals and accompanying information can then be received by any vessel, land station, or satellite fitted with an AIS receiver and is then typically displayed on a screen using interactive chart-plotting software [125].

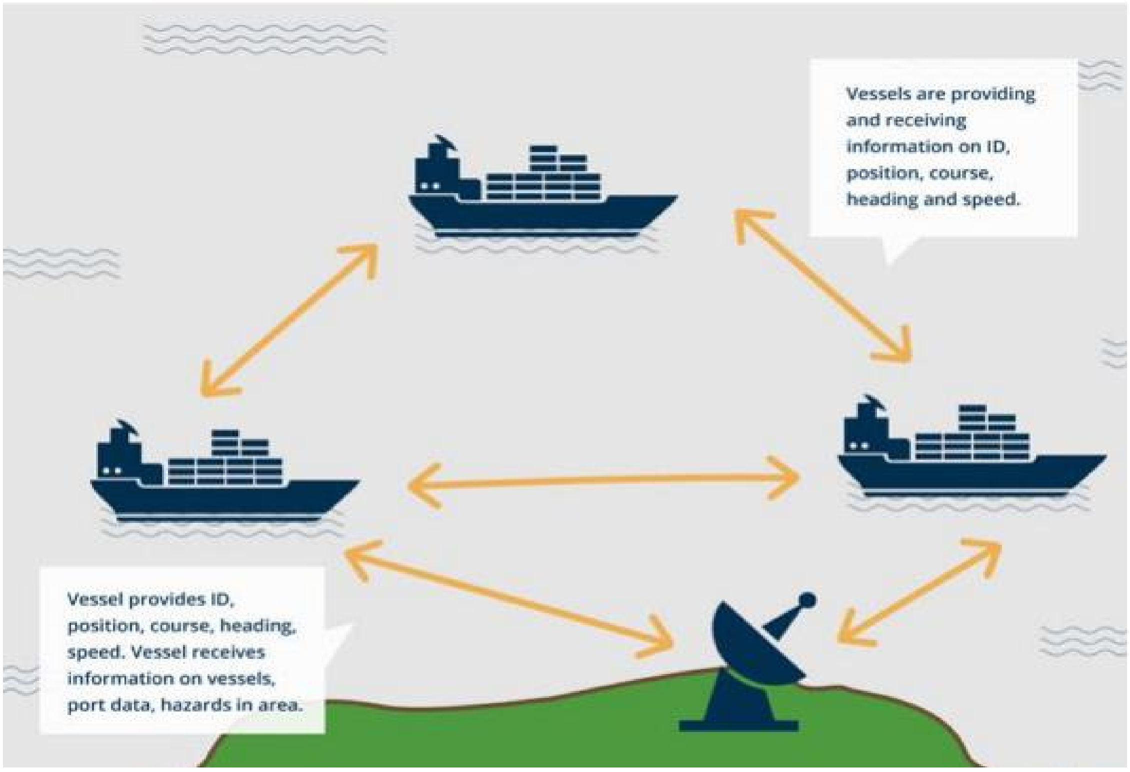


Fig. 4.2 AIS Data Transponder, [125]

International maritime law requires AIS transponders to be fitted aboard international voyaging ships with a gross tonnage of 300 tonnes or more, and on all passenger ships regardless of their size. Given its visibility and safety advantages many smaller vessels also voluntarily install AIS units. In many countries, no license is required to purchase and operate either transponders or receivers. The result is that AIS is used almost universally in the worldwide commercial maritime industry and increasingly so in the leisure marine sector. Not all vessels can be tracked by AIS. Naval and security ships generally prefer not to be tracked when on active duty, and cases are regularly reported of commercial vessels underway with their transponders turned off for unspecified reasons.

- **A brief history of AIS**

  It's not entirely clear who it was that came up with the first AIS vessel tracking and identification system, but like so much other technology that we take for granted these days, it came out of the introduction of GPS for civilian purposes, which achieved global coverage in the early 1990s. It was then the integration of GPS time and position data with long-standing VHF radio technology that enabled the development of AIS.

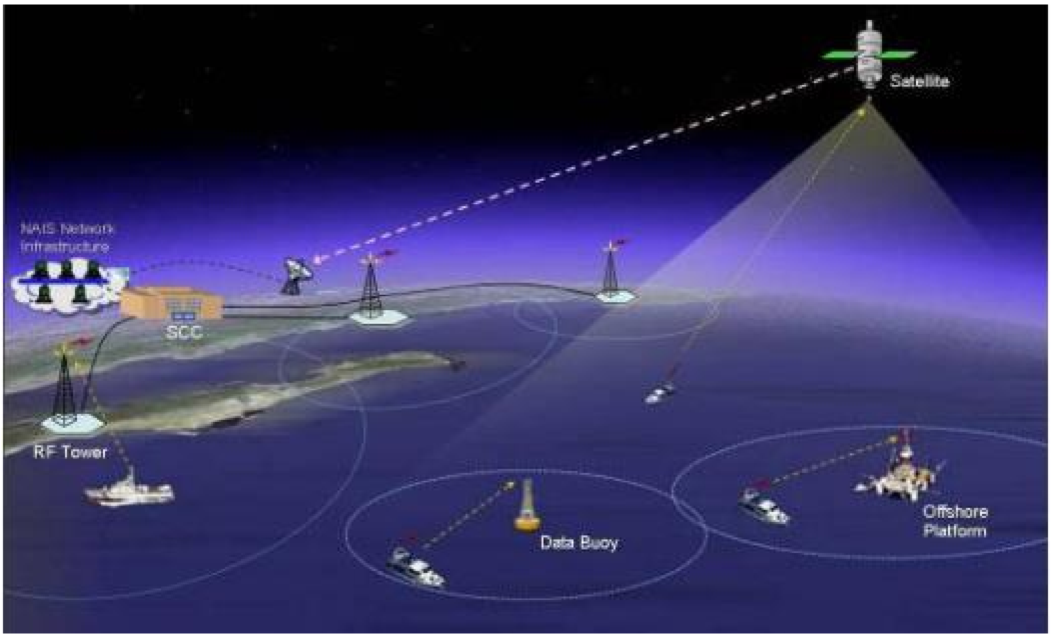


Fig. 4.3 AIS Conceptual Operation View, [125]

In its early years, the primary use of AIS was as a ship-to-ship anti-collision system for use in poor visibility and at night, in support of radar and conventional watch keeping. Over time the amount of information that could be transmitted in the VHF signal grew and its usefulness increased. In 2002 it finally went global when the IMO in its landmark SOLAS convention mandated that all passenger ships and other commercial vessels over 300 GT should carry Class-A AIS transceivers. At the time this affected around 100,000 ships, but since then use has expanded as the cost of transceivers has fallen and both compulsory and voluntary adoption has increased.

Originally developed as a short-range identification and vessel tracking system, at the start of the 21st century it was discovered that AIS transmissions could be received at ranges of up to 400km above the surface of the earth, whereas on the surface the maximum effective distance is around one-tenth of that. This revolutionized AIS, taking it from a coastal and ship-to-ship tracking application to a vessel management system with global coverage. However, the challenge for satellite operators looking to develop this opportunity now is managing the enormous volumes of data that this creates for individual satellites each monitoring thousands of square kilometers of ocean.

The AIS format uses TDMA radio access that allows for just 4,500-time slots per minute. A one-time slot equates to a single vessel transmission. Any more than that and the individual signals start to interfere with each other, corrupting the data held within. The terrestrial AIS infrastructure with its short-range and higher density does not have the same capacity problems. The satellite developers are, however, working on ways of receiving and processing incoming data at faster rates and rapid advances are being made.

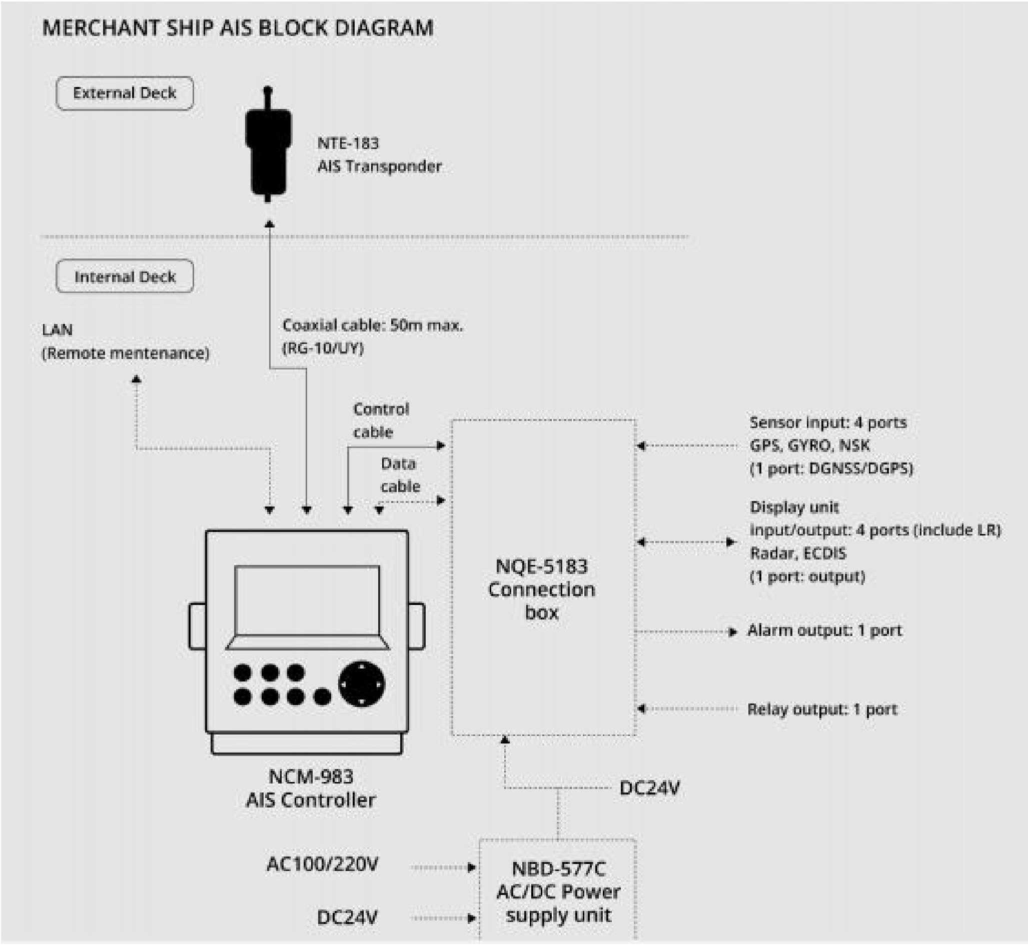


Fig. 4.4 Merchan Ship AIS Diagram Block, [125]

There is no perfect vessel tracking system, but AIS is becoming increasingly effective as accuracy and refresh rates get even better. Its ability to interface with other detection sources makes it an important component of integrated navigation and warning systems, and the addition of supplementary environmental and situational data makes it yet more versatile. Without a doubt, AIS is now one of the most valuable information sources available for anyone involved in the maritime sector.

- **AIS in Practice**

There are two classes of AIS transponders: A and B. Broadly speaking, the higher specification class A is mandated for commercial vessels, while the lower specification class B is intended for smaller, mostly leisure, vessels. Capabilities and prices vary greatly between the two classes. Information from Class A units is also prioritized over that from Class B equipment.

Class A: Class A transponders are mandated under international SOLAS regulations for all international voyaging ships with a gross tonnage of 300 tonnes or more, and on all passenger ships regardless of size. They transmit continuously at 12.5 Watts and have a horizontal range of up to 40nm. They use SOTDMA (Self-Organized TDMA) technology so that each automatically adjusts its transmissions to avoid interfering with others within range. In areas with high-density shipping, the system also shrinks the area of coverage when necessary to ensure that the system isn't overloaded.

Fig. 4.5 AIS Class A, [125]

Class B: Class B transponders were developed to give smaller vessels voluntary access to the benefits of the AIS system. They transmit every 30 seconds at 2 Watts and the horizontal range is line-of-sight. They use CSTDMA (Carrier Sense TDMA) technology that checks for Class A transmissions before sending its signal. Class B information is therefore only broadcast when there is sufficient space on the AIS channel. Class B AIS capability is increasingly being included in low-cost chart plotters and multifunction displays to overlay the information on electronic charts.

Fig. 4.6 AIS Class B, [125]

- **AIS Data Type**

  The availability of the data provided by the AIS depends on the AIS type. The characteristics of data recorded by the AIS is described as follows:

  Class A: Provides three types of information:

  1. Fixed, or static information, which is entered into the AIS on installation and needs only be changed if the ship changes its name or undergoes a major conversion from one ship type to another. Includes data such as:
     - MMSI (Maritime Mobile Service Identity)
     - Call sign and name of the vessel
     - IMO Number
     - Length and beam
     - Type of ship
     - Location of Position-fixing antenna

2. Dynamic information, which, apart from navigational status information, is automatically updated from the ship sensors connected to AIS. Includes:
   - Ship's position with accuracy indication and integrity status
   - Position Timestamp in UTC
   - Course over ground (COG)
   - Speed over ground (SOG)
   - Heading
   - Navigational status (e.g. underway by engines, at anchor, engaged in fishing, etc)
   - Rate of turn (ROT)

3. Voyage-related information, which might need to be manually entered and updated. Such as:
   - Ship's draught
   - Hazardous cargo (type) (e.g. DG (Dangerous goods), HS (Harmful substances)
   - or MP (Marine pollutants)
   - Destination and ETA
   - Route plan (waypoints) (at the discretion of the master)

Class B: Class B transponders transmit only static information every six minutes. This should include:
- MMSI (Maritime Mobile Service Identity)
- Call sign and name of the vessel
- Length and beam
- Type of vessel

The original purpose of AIS was to reduce the risk of vessels colliding with each other in poor visibility. However, the combination of satellite AIS allowing near-global coverage and the ability to access AIS data online has now made it a valuable resource for anyone wishing to monitor individual or groups of vessels, different classes of vessels, or volumes of total traffic in certain areas or the whole world.

As well as ships' officers, AIS data feeds with mapping overlays are therefore used by vessel owners and operators to monitor and manage their fleets, as well as port managers and service providers, shippers, maritime security providers, insurers, other maritime professionals, marine intelligence analysts, government agencies, economists, academics, and family and friends of crew members.

Users vary greatly in how they use the information. Many vessels owners use it simply to see where their vessels are at any given time while sitting at their desks or indeed on the beach or at home. Port managers can use the information to view incoming vessels and their latest expected times of arrival. Yachts off cruising or racing can be tracked by family and friends seeking peace of mind that all is well. On a bigger scale, academics and economists can analyze AIS data on a global scale to identify patterns and changes in trade flows and volumes, while environmentalists can use it to identify areas that might be at risk from excess marine traffic.

AIS today provides anyone who wants to see it with the most complete view available of the activity that takes place 24/7/365 on the world's oceans, seas, and inland waterways, together with a treasure trove of information on the size, type, and often cargoes of the vessels themselves. Its evolution over the years from a simple collision-avoidance system is truly remarkable.

In the last few years, new utilization of AIS such as logistics and transport economy have emerged. AIS data can be used in the various mapping of marine traffic to be used in the various simulation. Further, AIS has a wide interpretation, as it can be used solely, or can be combined to gain deeper insights. Therefore, AIS-based studies can benefit parties, from official authorities to private parties such as cargo owners or shipping companies [126].

In this study, AIS data was acquired from IHS Markit. AIS data include the ship name, movement timestamp, indicated sailing speed, indicated draught, and latitude/longitude coordinates, etc. The illustration of AIS data is shown below.

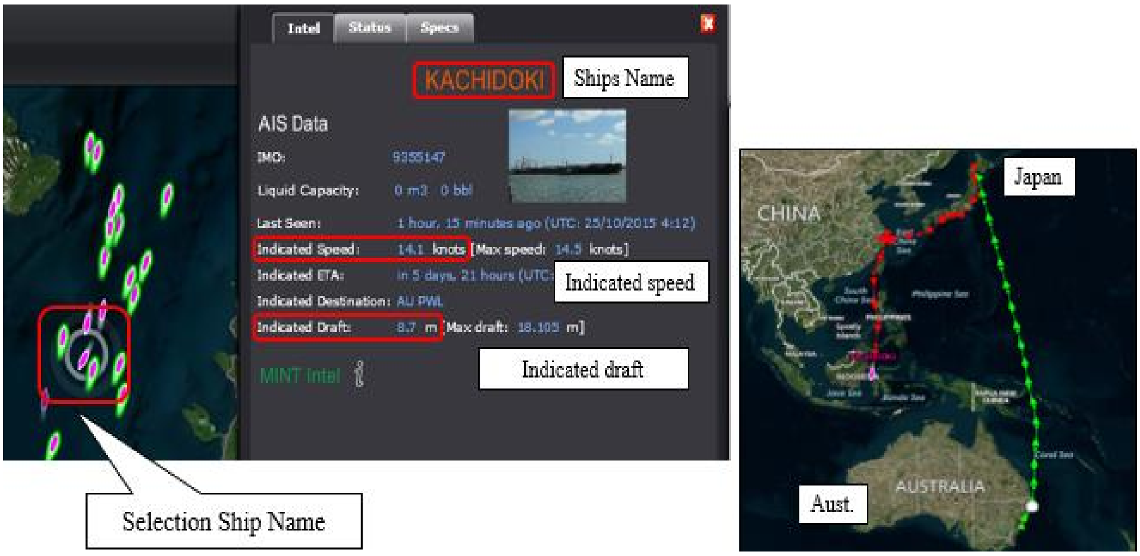


Fig. 4.7 AIS Data Illustration, [20]

B. Ship Operation

Besides AIS data, ship operation data was acquired from Market Intelligence Network (MINT) managed by IHS Markit. It is possible to obtain the current position of each ship, the ship name, IMO number, indicated draught, arrival/departure dates, and arrival/departure ports for 2014 until 2017. As shown in **Table 2.2**, typical ship operation data is comprehension from AIS data. It consists of the longer-range timestamp in the manner of ship position or ship movement defined by its sailing speed. In the same size ratio of data, ship operation data enclosed more information rather than AIS data.

Table 4.1 Typical Data of Ship Operation Data, [20]

Arrival Port	Vessel Name	Vessel Type	Vessel IMO	Arrival Date	Arrival Draught	Max Draught	Previous Port Departure Dtae
Kashima	SHIP A	BC	9355147	2015-12-02 01:55:46	18	18,105	2015-11-17 05:26:02
Kashima	SHIP A	BC	9355147	2014-03-07 23:10:20	17,2	18,105	2014-02-18 02:10:20
Kashima	SHIP A	BC	9355147	2013-01-31 07:20:31	18	18,105	2013-01-17 05:42:29
Kawasaki	SHIP A	BC	9355147	2017-08-04 05:41:07	18,1	18,105	2017-07-22 13:23:11
Kisarazu	SHIP A	BC	9355147	2016-01-12 07:56:03	18	18,105	2015-12-30 02:57:05
Kisarazu	SHIP A	BC	9355147	2015-04-19 03:10:48	15,4	18,105	2015-04-04 11:40:29
Kisarazu	SHIP A	BC	9355147	2014-10-25 01:40:22	15,4	18,105	2014-10-06 20:40:33
Kisarazu	SHIP A	BC	9355147	2014-09-01 14:55:47	17,7	18,105	2014-08-19 20:18:09
Muroran	SHIP A	BC	9355147	2015-09-26 10:25:35	15,1	18,105	2015-09-09 07:25:48
Muroran	SHIP A	BC	9355147	2013-08-09 05:45:29	15,9	18,105	2013-07-23 22:30:27
Oita	SHIP A	BC	9355147	2015-08-10 21:11:01	16,7	18,105	2015-07-22 01:25:33
Oita	SHIP A	BC	9355147	2015-03-01 09:10:36	17,6	18,105	2015-02-02 20:25:30
Oita	SHIP A	BC	9355147	2014-01-19 02:55:20	17,6	18,105	2014-01-04 15:25:18
Oita	SHIP A	BC	9355147	2013-12-08 20:55:18	17,5	18,105	2013-11-25 06:25:19
Oita	SHIP A	BC	9355147	2013-10-19 06:55:23	16,5	18,105	2013-10-01 22:00:33
Oita	SHIP A	BC	9355147	2013-05-25 09:45:25	15,2	18,105	2013-05-05 07:20:21

C. Ship Data

The ship data used in this study was collected from Sea-web Ships. Ship data consists of ship name, IMO number, built year, flag, and principal dimensions, such as DWT, length, draught, and breadth. Figure 4.8 shown the illustration of ship data, and typical ship data collected in MLDB is shown in Table 4.2.

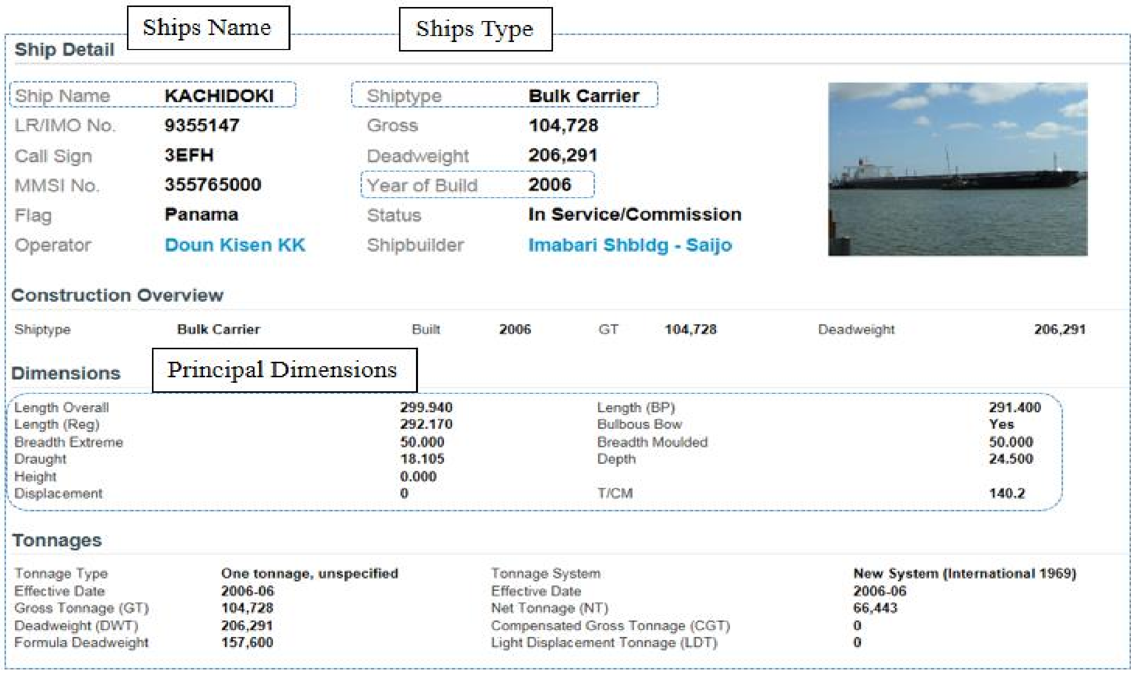


Fig. 4.8 Illustration of Ship Data, [20]

Table 4.2 Typical Data of Ship Data

Vessel IMO	Name of Ship	MMSI	Deadweight	Year	Class	Shipbuilder	Service Speed	Length	Breadth	Draught
9592446	AANYA	373434000	179628	2012	LR	HHIC-Phil Inc	14,5	292	45	18,2
9583897	AARGAU	269108000	32790	2010	BV	Universe Shipbuilding Yangzhou	13,7	179,9	28,4	10,15
9087738	AARTI PREM	636014966	69087	1994	NK	Imabari Shbldg - Marugame	14,5	224,98	32,2	13,295
9592458	AASHNA	373916000	179523	2012	LR	HHIC-Phil Inc	14,5	292	45	18,2
9019535	AAZAM	422535000	1852	1990	AS	Sungkwang Shipbuilding Co Ltd	12	70,49	12,51	4,312
9571040	ABDALA	370533000	34938	2011	LR	Shanghai Shipyard Co Ltd	13,7	179,9	28,4	10,8
7116781	ABDUL M	671359000	6370	1972	DR	Viana Do Castelo	11	101,48	15,97	7,217
9132923	ABDULLAH	405000132	45653	1996	NK	Tsuneishi Shbldg - Fkym - earl	14,9	185,74	30,4	11,62
8902929	ABDULLATIF	312793000	13790	1992	NK	Szczecinska Stocznia SA	12,8	143,7	20,6	8,43
9334882	ABIGAIL N	636014327	297430	2009	NV	Universal Shbldg - Tsu	14,3	327	55	21,4
9213399	ABILA	422034700	75249	2001	X	Samho Heavy Industries Co Ltd	14,8	225	32,25	14,15
8912261	ABK TRADER	312338000	28452	1991	NK	Imabari Shbldg - Imabari	13,7	169,03	27,2	9,745
9006643	ABM DISCOVERY	353527000	39110	1992	NK	IHI - Tokyo	13	180,8	30,5	10,93
8400971	ABM NAVIGATOR	312320000	42183	1987	NK	Sasebo Sasebo	14	185,91	30,41	11,469
9481702	ABML EVA	229068000	106659	2012	GL	STX Dalian Shipbuilding Co Ltd	14,5	253,5	43	13,6
9224738	ABML GRACE	229567000	172319	2002	BV	Daewoo Shipbuilding & Marine	14,5	289	45	17,82

D. Port Data

The port data consists of port name, port specification/dimension, latitude/longitude information, and handled cargo types. Port data was used in the previous studies and this study was collected from Sea-web Ports. Fig 4.9 shown the illustration of port data, and typical port data collected in MLDB is shown in Table 4.3.

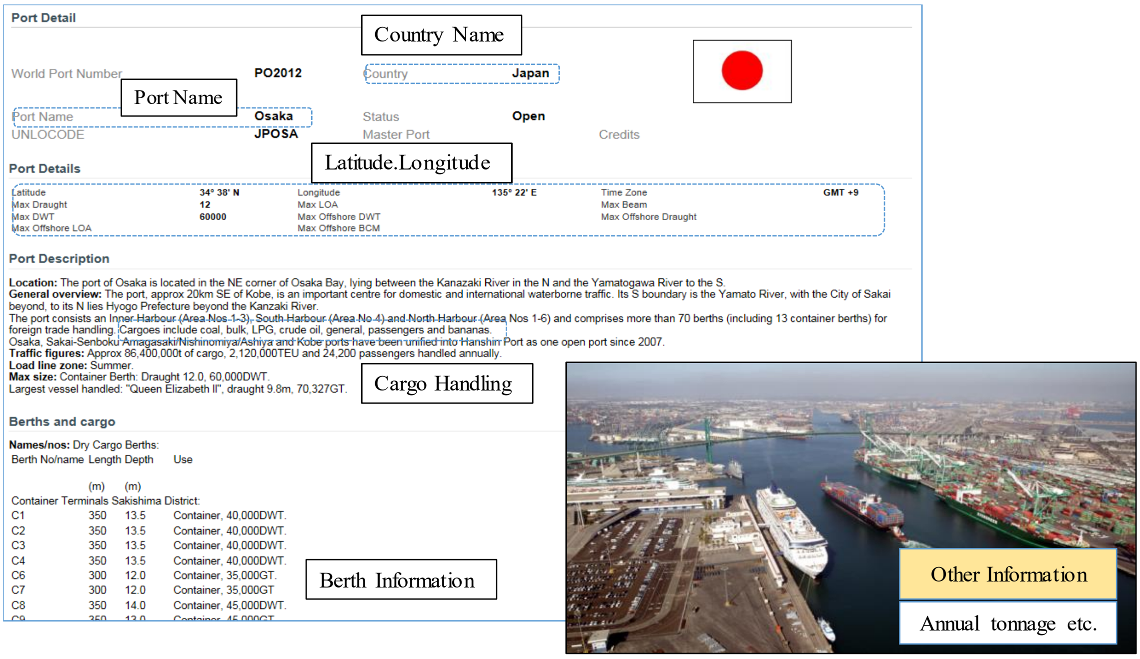


Fig. 4.9 Illustration of Port Data, [19]

Table 4.3 Typical Data of Port Data

Port Name	Country	Latitude	Longitude	Max DWT	Max Draft	Max LOA	Max Breadth	Annual Tonnnage	Port Number	UN Locode	Cargo Type
Abbot Point	Australia	-19,85	148,0833	200000	17,6	300	50	14443487	PO3978	AUABP	Coal
Albany (Australia)	Australia	-35,0333	117,8833	67000	11,5	225	33	3501077	PO1046	AUALH	Grain, Others
Brisbane	Australia	-27,3833	153,1666	148160	14,2	294	46	31877104	PO1052	AUBNE	Iron ore, Grain, Coal, Others
Bunbury	Australia	-33,3166	115,6333	80000	11,6	234	32,2	13866969	PO1054	AUBUY	Others
Dampier	Australia	-20,6666	116,7	250000	19,5	340	55	165025204	PO1060	AUDAM	Iron ore, Others
Esperance	Australia	-33,8666	121,8833	203000	18	300	50	10000000	PO3709	AUEPR	Iron Ore, Grain
Geraldton	Australia	-28,7666	114,6	80000	12,8	225	32,5	9005508	PO1068	AUGET	Iron Ore, Grain
Gladstone	Australia	-23,8166	151,3	231850	17,9	315	55	55602406	PO1069	AUGLT	Grain, Coal, Others
Hay Point	Australia	-21,2666	149,3166	231851	17,5	315	56	99500000	PO1071	AUHPT	Coal
Lucinda	Australia	-18,5333	146,3333	70000	12,3	230	32,2	575497	PO1075	AULUC	Others
Milner Bay	Australia	-13,8833	136,45	50000	12,2	200	32,6	3893821	PO1079	AUMIB	Iron Ore, Others
Newcastle	Australia	-32,9166	151,8	232000	16,2	300	55	114575744	PO1080	AUNTL	Grain, Coal, Others
Newcastle (Australia)	Australia	-32,9166	151,8	232000	16,2	300	55	114575744	PO1080	AUNTL	Grain, Coal
Port Hedland	Australia	-20,3	118,5833	260000	19,2	330	55	199002079	PO1083	AUPHE	Iron ore
Port Kembla	Australia	-34,4666	150,9	206306	16	300	50	24036000	PO1084	AUPKL	Coal
Port Latta	Australia	-40,85	145,3833	130000	15,3	245	38,5	2200000	PO1085	AUPLA	Iron Ore
Port Walcott	Australia	-20,6166	117,1833	340000	19,5	335	60	54626968	PO1089	AUPWL	Iron Ore
Weipa	Australia	-12,6666	141,85	85000	11,7	256	35,4	22111499	PO1100	AUWEI	Others

E. Route Data

The route data comprise arbitrary route data. Nautical mile distance for each voyage route obtained from IHS Fairplay. Figure 4.10 shown the illustration of route data and typical route data collected in MLDB shown in Table 4.4.

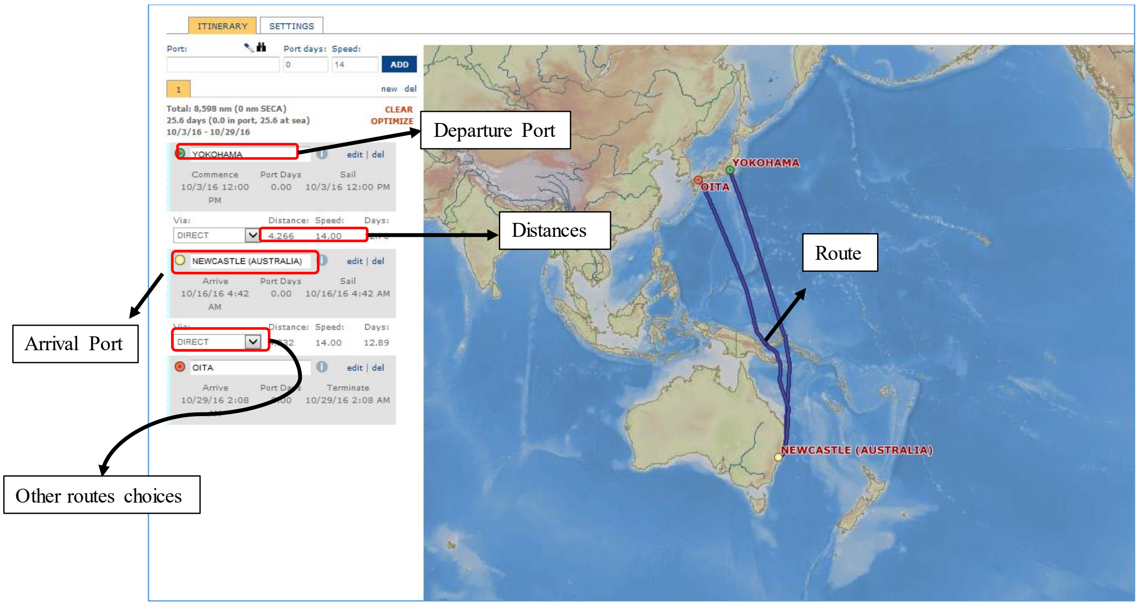


Fig. 4.10 Illustration of Route Data, [21]

Table 4.4 Typical Data of Route Data

Route Number	Arrival Port Name	Departure Port Name	Route	Distance	Speed	Days	Via
RN0053	Chiba	Gladstone	ChibaGladstone	3793	14	11,29	DIRECT
RN0125	Chiba	Abbot Point	ChibaAbbot Point	3598	14	10,71	DIRECT
RN0133	Chiba	Dampier	ChibaDampier	3665	14	10,91	DIRECT
RN0204	Fukuyama	Gladstone	FukuyamaGladstone	3856	14	11,48	DIRECT
RN0258	Fukuyama	Abbot Point	FukuyamaAbbot Point	3644	14	10,85	DIRECT
RN0335	Fukuyama	Dampier	FukuyamaDampier	3585	14	10,67	DIRECT
RN0538	Hibikinada	Gladstone	HibikinadaGladstone	3970	14	11,82	DIRECT
RN0564	Hibikinada	Abbot Point	HibikinadaAbbot Point	3756	14	11,18	DIRECT
RN0615	Higashi-Harima	Gladstone	Higashi-HarimaGladstone	3832	14	11,4	DIRECT
RN0637	Higashi-Harima	Abbot Point	Higashi-HarimaAbbot Point	3620	14	10,77	DIRECT
RN0680	Higashi-Harima	Dampier	Higashi-HarimaDampier	3568	14	10,62	DIRECT
RN1075	Kanda	Abbot Point	KandaAbbot Point	3653	14	10,87	DIRECT
RN1168	Kashima	Dampier	KashimaDampier	3714	14	11,05	DIRECT
RN1190	Kashima	Gladstone	KashimaGladstone	3791	14	11,28	DIRECT
RN1330	Kawasaki	Gladstone	KawasakiGladstone	3778	14	11,24	DIRECT
RN1338	Kawasaki	Dampier	KawasakiDampier	3650	14	10,86	DIRECT
RN1340	Kawasaki	Abbot Point	KawasakiAbbot Point	3582	14	10,66	DIRECT
RN1434	Kinuura	Abbot Point	KinuuraAbbot Point	3587	14	10,68	DIRECT

F. Trade Data

The trade data comprise the transported cargo amount between countries. Trade data include the exporter/importer countries name, trade period, trade flow, code, trade value, net weight (kg), etc. The illustration of UN Comtrade data selection is shown in Fig. 4.11, and the typical trade data collected in MLDB is shown in Table 4.5.

Fig. 4.11 Illustration of Trade Data, [22]

Table 4.5 Typical Data Trade from UN Comtrade

Iron Ore

Period	Trade Flow	Reporter	Partner	Code	Trade Value	NetWeight (kg)	Quantity Unit	Trade Quantity	(Ton)
2014	Import	Japan	Australia	2601	$9.001.321.700	82.884.716.192	8	82.884.716.192	82.884.716

Coal

Period	Trade Flow	Reporter	Partner	Code	Trade Value	NetWeight (kg)	Quantity Unit	Trade Quantity	(Ton)
2014	Import	Japan	Australia	2701	$12.660.422.478	119.139.855.456	8	119.139.855.456	119.139.855

Grain

Period	Trade Flow	Reporter	Partner	Code	Trade Value	NetWeight (kg)	Quantity Unit	Trade Quantity	(Ton)
2014	Import	Japan	Australia	1001	$316.168.354	928.869.984	8	928.869.984	928869,984
2014	Import	Japan	Australia	1003	$190.139.529	646.754.016	8	646.754.016	646754,016
2014	Import	Japan	Australia	1004	$5.944.581	13.918.000	8	13.918.000	13918
2014	Import	Japan	Australia	1005	$577.148	911.000	8	911.000	911
2014	Import	Japan	Australia	1006	$27.925.641	38.401.000	8	38.401.000	38401
2014	Import	Japan	Australia	1007	$293.263	301.000	8	301.000	301
2014	Import	Japan	Australia	1008	$407.231	313.000	8	313.000	313
						1.629.468.000			1.629.468

4.3 Data Structure of MLDB

To more easily extract valuable information, we defined a structure for the MLDB and modified unstructured data into a relational database. For example, by integrating ship and port data with operation data, some information related to a ship's operational state can be analyzed (e.g., berthing, anchoring, or sailing). All available marine logistics data are integrated into a one-to-many relationship. The data structure of MLDB is shown in Fig. 4.12.

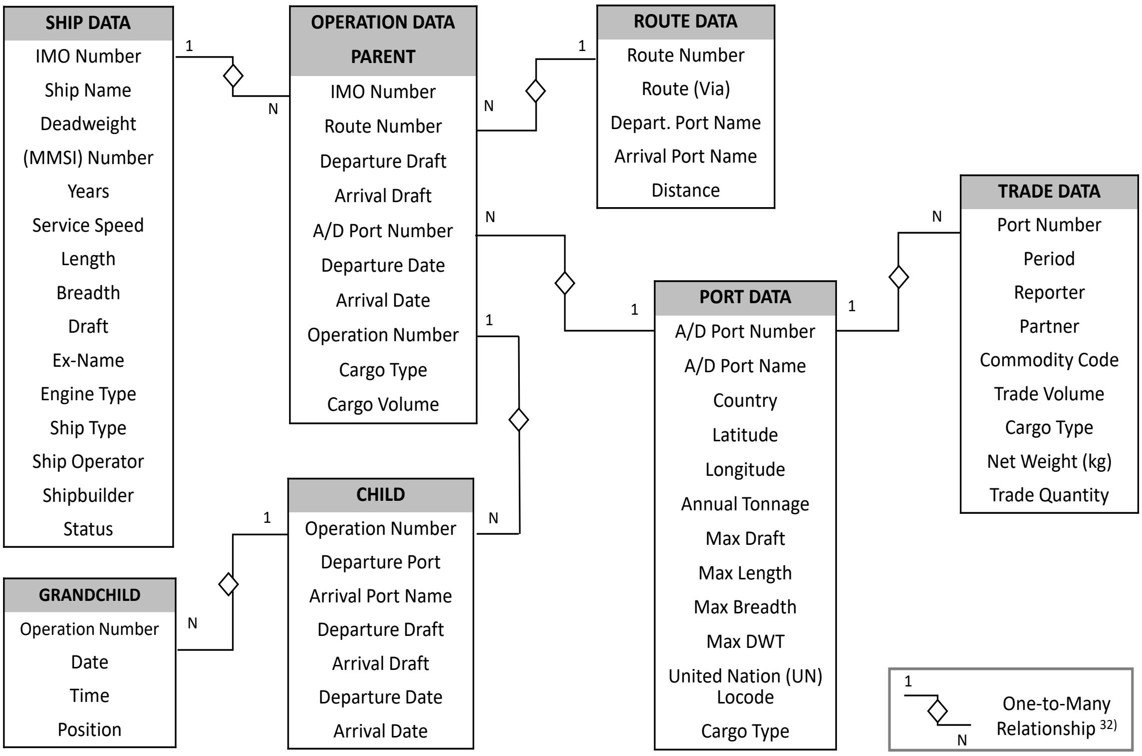


Fig. 4.12 Data Structure of MLDB

4.4 Data Cleaning

Essentially, big data utilization encounter issues of data quality. Some data may be flawed caused by sensor faults or mistakes made during manual entry. Extensive validity check considered to proper routine [127]. To ensure and the reliability and quality of the data used to construct the MLDB, the following error cleaning methods were performed.

4.4.1 Duplicated Data

Duplication in recorded data of arrival port and/or departure port. The data is defined as duplicated data managed by taking the earliest arrival date and the latest departure date, the data in between is deleted. The example of duplicated data is shown in Table 4.6.

Table 4.6 Example of Duplicated Data at Oita Port

Port	Country	Arrival Date	Days Ago	Arrival Draught	Departure Draught
Oita	Japan	08/04/2015	344	7,4	7,4
Cristobal	Panama	07/04/2015	345	7,4	7,4
Oita	Japan	07/04/2015	345	7,4	7,4
Cristobal	Panama	07/04/2015	345	7,4	7,4
Oita	Japan	07/04/2015	345	7,4	7,4
Cristobal	Panama	07/04/2015	345	7,4	7,4
Oita	Japan	07/04/2015	345	7,4	7,4
Cristobal	Panama	07/04/2015	345	7,4	7,4
Oita	Japan	07/04/2015	345	7,4	7,4
Cristobal	Panama	07/04/2015	345	7,4	7,4
Oita	Japan	07/04/2015	345	7,4	7,4
Cristobal	Panama	06/04/2015	346	7,4	7,4
Oita Quarantine Anchorage	Japan	06/04/2015	346	7,4	7,4
Tonda	Japan	04/04/2015	348	6,8	6,8
Cristobal	Panama	02/04/2015	350	10,8	6,8

4.4.2 Unrealistic Ship Movement Data

Too fast sailing speed of the ship in the manner of departure and arrival date, and distance between ports. Thus, misreported location or arrival date will result from unrealistic voyage data. The solution is deleting unrealistic voyage data by checking the average voyage speed, which is calculated by considering the navigation days and distance between two ports. If the average voyage speed exceeds the service speed, it is defined as an error and the data are deleted.

4.4.3 Unrealistic Ship Draught Data

Recorded data contains zero (null) draught data, or draught data is not changed during operation considered to be unrealistic ship draught data. For this case of error, the draught was corrected with average draught data from the same ship size category.

4.5 Generating Cargo Information

Cargo information on an operating ship are important for demand forecasting and understanding the ship's use. However, such information does not exist in AIS data. Therefore, we estimated the cargo type and volume of each operation. In the case of a bulk carrier, the cargo type is selected from three types: iron ore, coal, and grain and minor bulk (MB). The estimation methods used in this study are described as follows.

4.5.1 Checking Data Reliability

Confirmation of data's reliability is required for a good cargo volume estimation. In our study, data reliability was evaluated by checking the draft rate di by using Eq. (1)

$$d_i = \frac{d_{sail\,i}(m)}{d_{max\,i}(m)} \tag{1}$$

Where d_{aily} *(m)* is the sailing draught, $_{maxi}$ *(m)* is the ship maximum draught, and d_i is the draught rate.

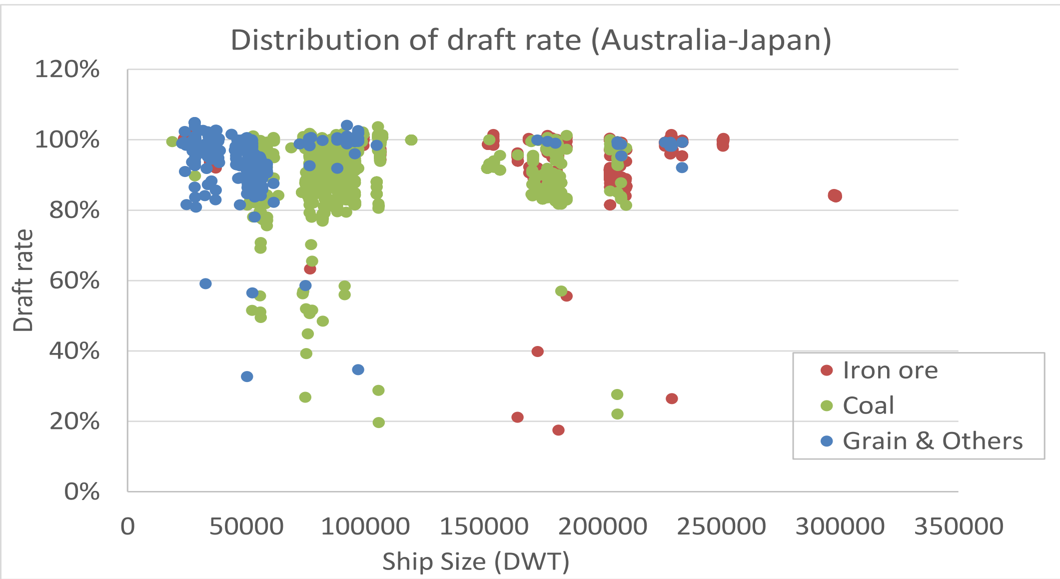


Fig. 4.13 Bulk Carrier Draught Rate in Australia-Japan 2014

Table 4.7 Bulk Carrier Status based on its Draught Rate

Draft Rate [%]	Arrival>Departure	Arrival=Departure	Arrival<Departure
100+	Unknown	Unknown	Unknown
65~100	Loading	Unknown	Unknown
45~65	Unknown	Unknown	Ballast
0~45	Unknown	Unknown	Unknown

4.5.2 Estimating Cargo Type Using Port Data

By identifying the cargo type from port data, the cargo of each operation could be estimated. As shown in Table 4.8, cargo type estimation was conducted by checking the combination of cargo from the arrival and departure ports. In the case of operation from Port A to Port D, the only common cargo is coal. Therefore, the cargo type was estimated to be coal. In contrast, in the case of operation from Port B to Port D, there are two common cargos: coal and iron ore. In this case, cargo type was defined as multi-cargo and decided using the ship size.

Table 4.8 Estimation Cargo Type Using Port Data

PORT	Port D (Coal, Iron Ore)	Port E (Coal, Iron Ore, Grain & MB)	Port F (Coal, Iron Ore)
Port A (Coal, Grain & MB)	Coal	Coal, Grain & MB	Coal
Port B (Coal, Iron Ore)	Coal, Iron Ore	Coal, Iron Ore	Coal, Iron Ore
Port C (Iron Ore, Grain & MB)	Iron Ore	Iron Ore, Grain & MB	Iron Ore

4.5.3 Estimating Cargo Type Using Ship Size

If two or more common cargo types exist in port data, the cargo types were estimated using ship size. Since ship size and cargo type are closely related, the remaining operation could be estimated. The flowchart of estimation steps using port data and ship size is shown as follows:

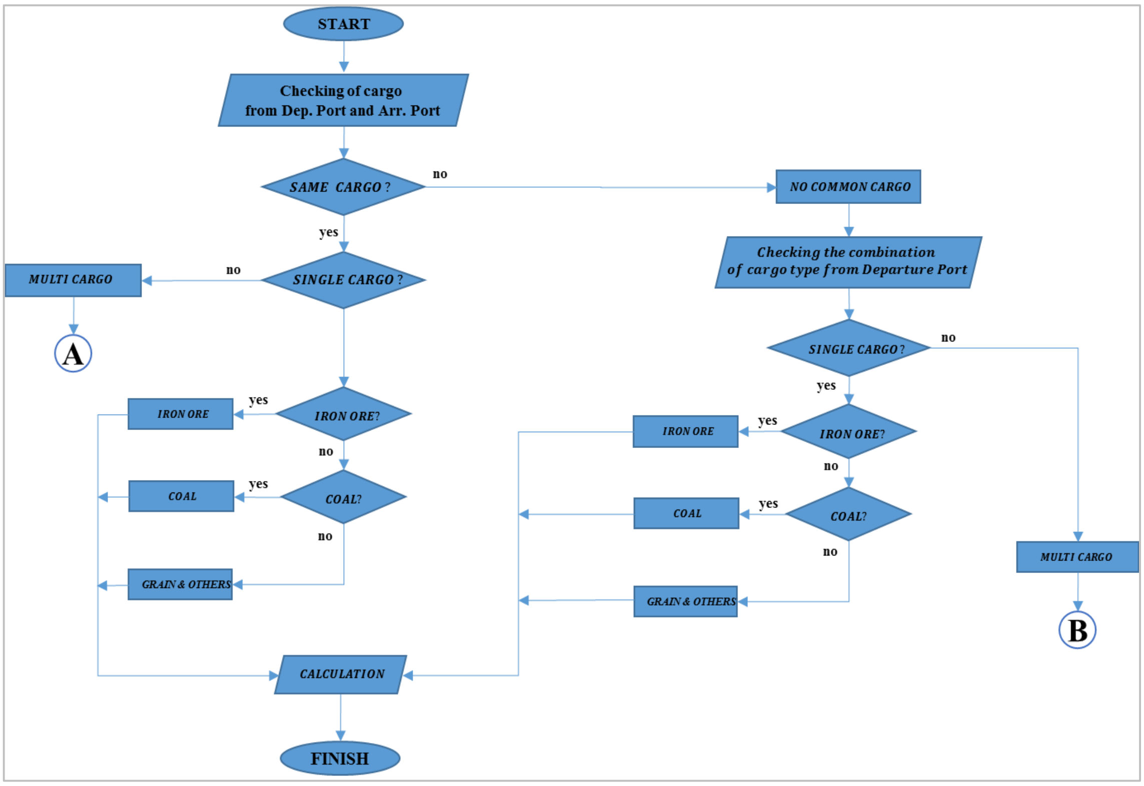


Fig. 4.14 Estimating Cargo Using Port Data (Single Cargo)

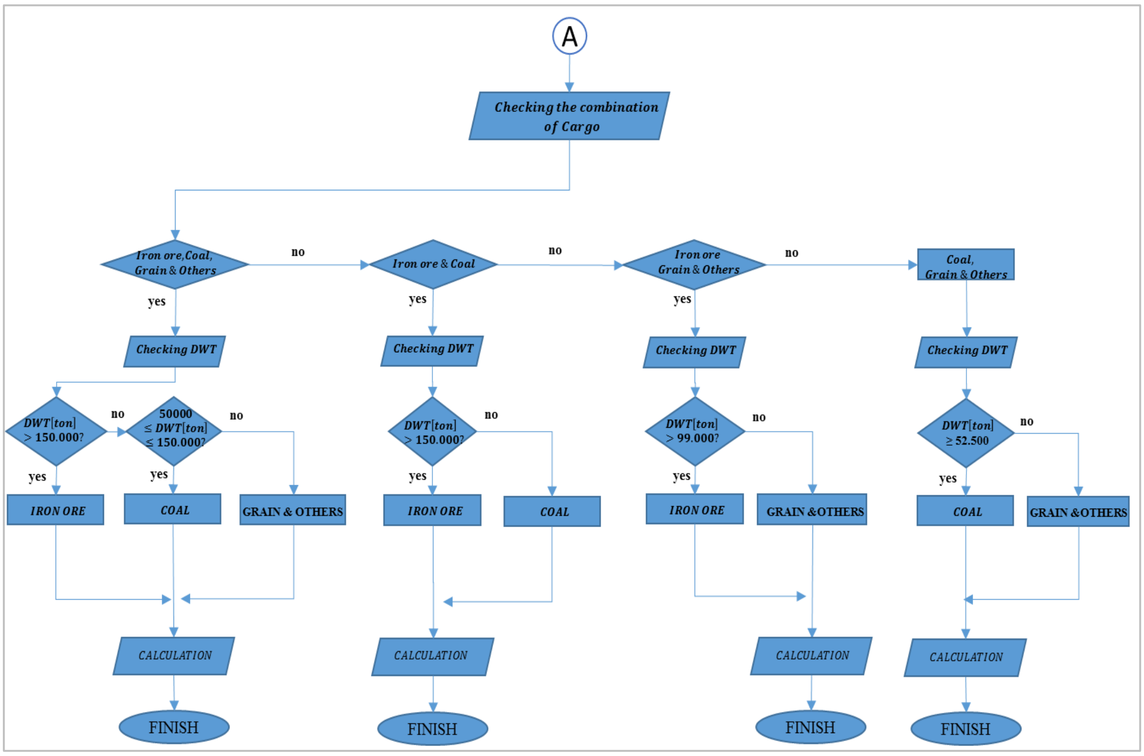


Fig. 4.15 Estimating Cargo Using Port Size (Multi Cargo)

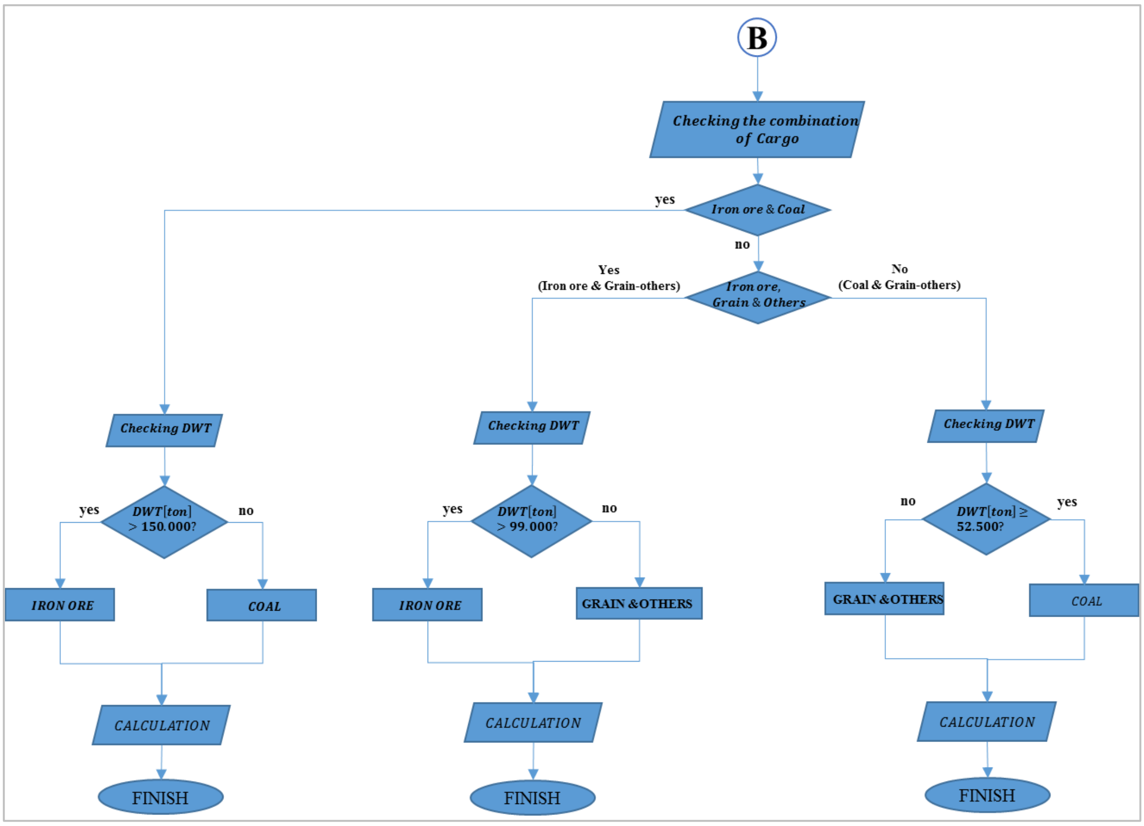


Fig. 4.16 Estimating Cargo Using Port Size (No Common Cargo)

4.5.4 Estimating Cargo Volume

Ship data has information on the deadweight and maximum draft of the target ship, while AIS data has information about the sailing draft. The cargo volume was estimated using Eq. (2).

$$V_i[ton] = DWT_i \times \frac{(d_i - 0.2)}{(1 - 0.2)} \tag{2}$$

where V_i (ton) is cargo volume, DWT_i is deadweight, and d_i is the draft rate. As explained in the previous section, the operation conditions of ships were defined as loading, ballast, and unknown. In the unknown condition, cargo volume was estimated by considering the average draft of ships of the same size operating on the same route.

4.6 Confirmation of Cargo Estimation of MLDB

To verify the cargo estimation in Section 4.6, we compared our results with actual trade value from UN Comtrade data, using bulk carriers operating from Australia and Brazil to Japan, Australia and Brazil to Korea, and Australia and Brazil to China. The confirmation results of cargo estimation for each route are shown as follows:

4.6.1 Cargo Estimation from Australia-Brazil to Japan

The cargo estimation exported from Australia to Japan, and Brazil to Japan is shown in the following table. The results are shown based on the several scenarios; (1) All operation (including error data), (2) All operation + error cleaning, (3) All operation + error cleaning + draft cleaning. The results are shown as follows:

Table 4.9 Result of Estimation Cargo 2014 (Australia~Japan)

Scenario 1

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	15465577	262	12351717	125%
Iron Ore	85150411	448	82884716	103%
Coal	115836800	1145	119749649	97%
TOTAL	216452788	1855	214986082	101%

Scenario 2

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	12191371	243	12351717	99%
Iron Ore	83958690	440	82884716	101%
Coal	116641588	1135	119749649	97%
TOTAL	212791649	1818	214986082	99%

Scenario 3

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	11624683	243	12351717	94%
Iron Ore	78742212	440	82884716	95%
Coal	108976830	1135	119749649	91%
TOTAL	199343724	1818	214986082	93%

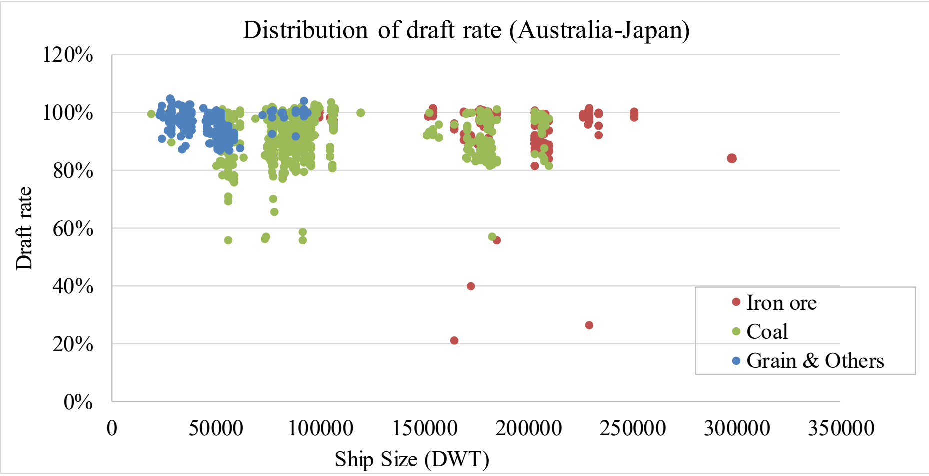


Fig. 4.17 Distribution of Draught Rate (Australia~Japan)

Table 4.10 Result of Estimation Cargo 2014 (Brazil~Japan)

Scenario 1

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	3067319	52	2698556	114%
Iron Ore	36467312	203	36993022	99%
Coal	0	0	0	0%
TOTAL	39534631	255	39691578	100%

Scenario 2

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	2727349	45	2698556	101%
Iron Ore	35244640	198	36993022	95%
Coal	0	0	0	0%
TOTAL	37971989	243	39691578	96%

Scenario 3

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	2504870	45	2698556	93%
Iron Ore	34122409	198	36993022	92%
Coal	0	0	0	0%
TOTAL	36627280	243	39691578	92%

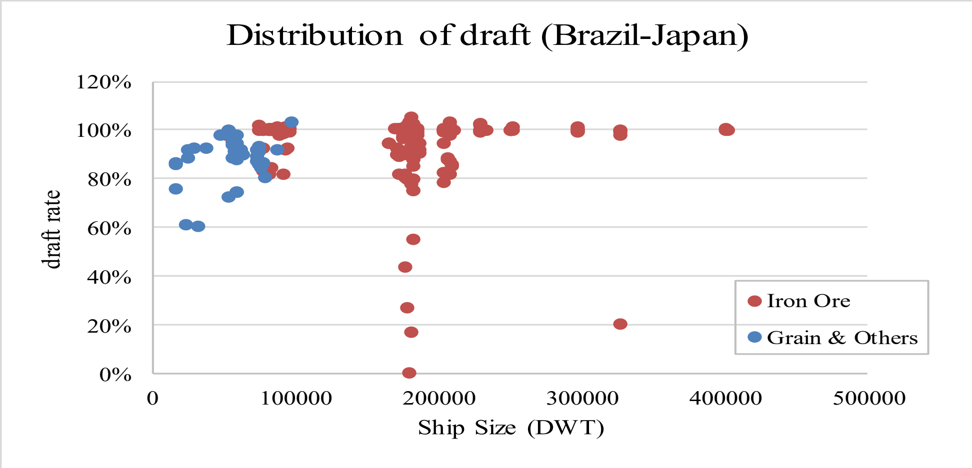


Fig. 4.18 Distribution of Draught Rate (Brazil~Japan)

4.6.2 Cargo Estimation from Australia-Brazil to Korea

The cargo estimation exported from Australia to Korea, and Brazil to Korea is shown in the following table. The results are shown based on the several scenarios; (1) All operation (including error data), (2) All operation + error cleaning, (3) All operation + error cleaning + draft cleaning. The results are shown as follows:

Table 4.11 Result of Estimation Cargo 2014 (Australia~Korea)

Scenario 1

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	7132344	130	7624620	94%
Iron Ore	50317540	248	50949931	99%
Coal	50411964	359	54996867	92%
TOTAL	107861848	737	113571418	95%

Scenario 2

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	6873369	125	7624620	90%
Iron Ore	49902348	246	50949931	98%
Coal	49894285	356	54996867	91%
TOTAL	106670002	727	113571418	94%

Scenario 3

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	6558950	125	7624620	86%
Iron Ore	48872168	246	50949931	96%
Coal	47065859	356	54996867	86%
TOTAL	102496977	727	113571418	90%

Table 4.12 Result of Estimation Cargo 2014 (Brazil~Korea)

Scenario 1

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	4244416	51	4261497	100%
Iron Ore	15899091	71	15849483	100%
Coal	0	0	0	0%
TOTAL	20143507	122	20110980	100%

Scenario 2

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	4244416	51	4261497	100%
Iron Ore	15711227	70	15849483	99%
Coal	0	0	0	0%
TOTAL	19955643	121	20110980	99%

Scenario 3

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	3838960	51	4261497	90%
Iron Ore	15206546	70	15849483	96%
Coal	0	0	0	0%
TOTAL	19045506	121	20110980	95%

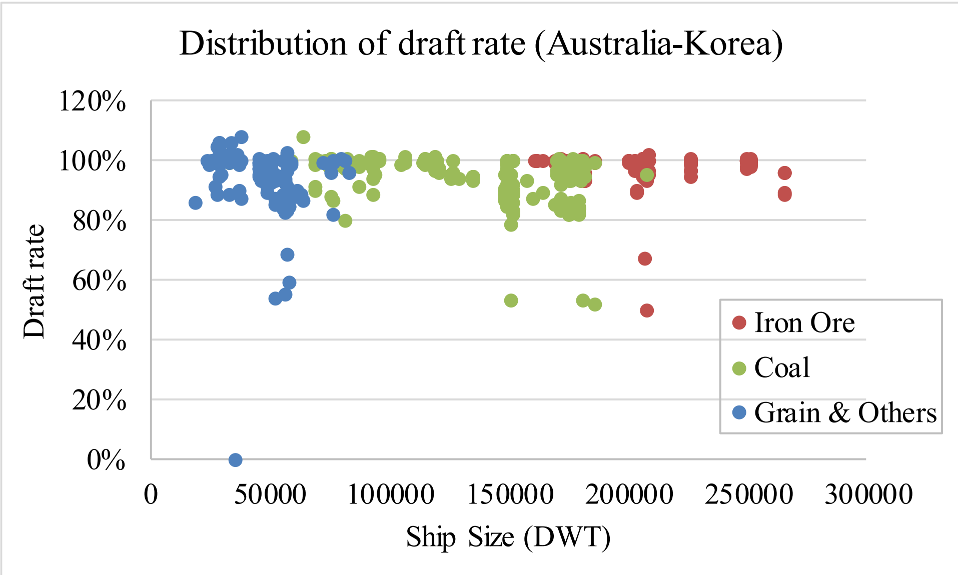


Fig. 4.19 Distribution of Draught Rate (Australia~Korea)

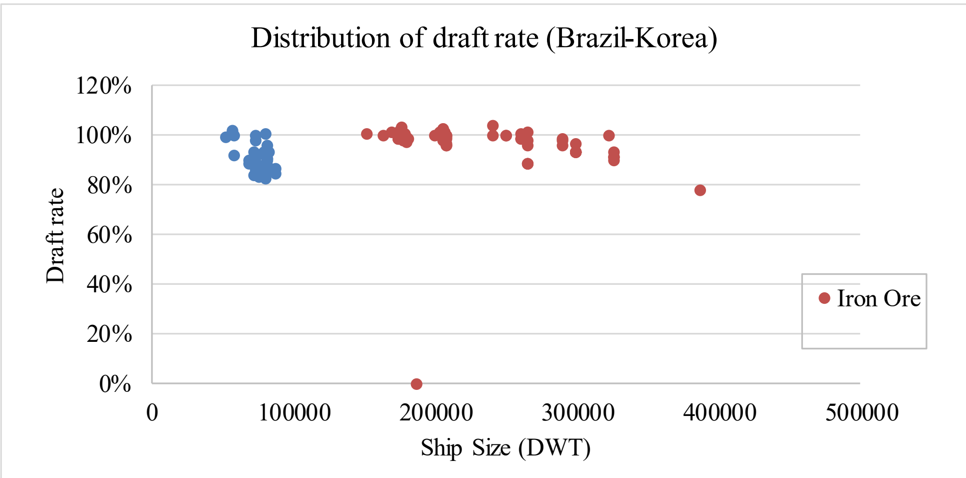


Fig. 4.20 Distribution of Draught Rate (Brazil~Korea)

4.6.3 Cargo Estimation from Australia-Brazil to China

The cargo estimation exported from Australia to China, and Brazil to China is shown in the following table. The results are shown based on the several scenarios; (1) All operation (including error data), (2) All operation + error cleaning, (3) All operation + error cleaning + draft cleaning. The results are shown as follows:

Table 4.13 Result of Estimation Cargo 2014 (Australia~China)

Scenario 1

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	50463648	603	46754508	108%
Iron Ore	508335211	2760	548354530	93%
Coal	79836626	732	79595444	100%
TOTAL	638635485	4095	674704481	95%

Scenario 2

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	43075896	543	46754508	92%
Iron Ore	491943119	2671	548354530	90%
Coal	78141995	718	79595444	98%
TOTAL	613161010	3932	674704481	91%

Scenario 3

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	39706038	543	46754508	85%
Iron Ore	475288948	2671	548354530	87%
Coal	74804274	718	79595444	94%
TOTAL	589799260	3932	674704481	87%

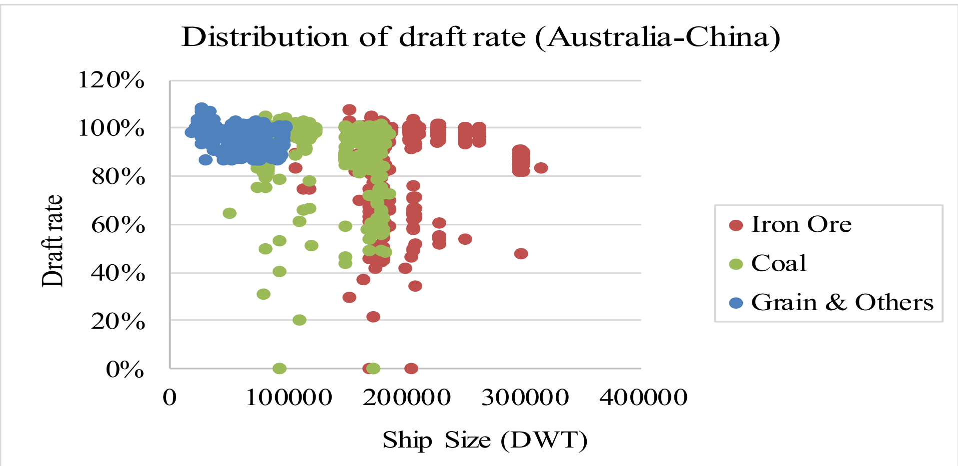


Fig. 4.21 Distribution of Draught Rate (Australia~China)

Table 4.14 Result of Estimation Cargo 2014 (Brazil~China)

Scenario 1

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	46708026	573	48840289	96%
Iron Ore	138113282	681	134450605	103%
Coal	0	0	0	0%
TOTAL	184821308	1254	183290894	101%

Scenario 2

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	45552261	559	48840289	93%
Iron Ore	133886109	662	134450605	100%
Coal	0	0	0	0%
TOTAL	179438370	1221	183290894	98%

Scenario 3

Cargo Type	Cargo Volume	Amount of Ship	UN (T)	Ratio
Grain & Others	41883160.75	559	48840289	86%
Iron Ore	130320675.8	662	134450605	97%
Coal	0	0	0	0%
TOTAL	172203837	1221	183290894	94%

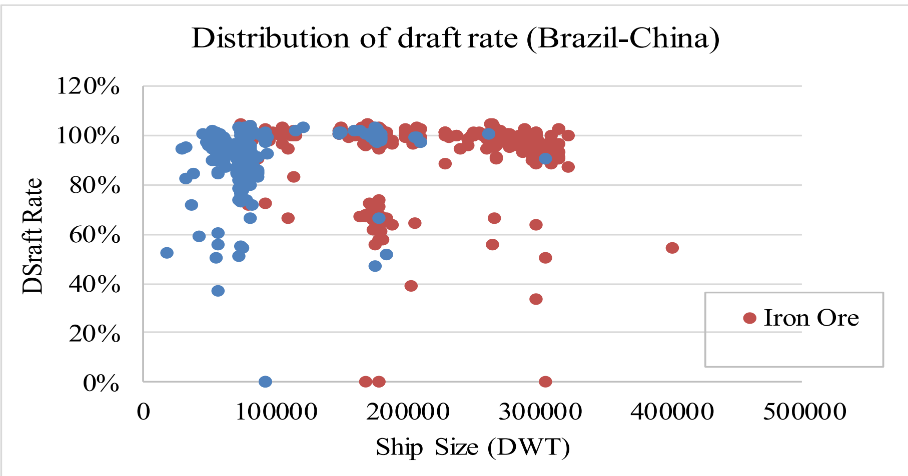


Fig. 4.22 Distribution of Draught Rate (Brazil~China)

4.7 Data Findings of MLDB

Based on the operational data extracted from MLDB, it can be seen that the availability of ship operational data especially (draft) is important for cargo volume estimation results. Lack of ship operational data might have caused the estimated volume different from the trade data volume from UN Comtrade. The detailed explanation is described as follows:

4.7.1 Findings of Data Availability

To check the availability of ship operational data, the selected country is divided into two categories: (1) Developed Country i.e. Japan, Australia, China, Korea, and America; (2) Developing Country i.e. Malaysia, India, Vietnam, Ghana, and Indonesia. The availability of each category is shown as follows:

(1) Developed Country

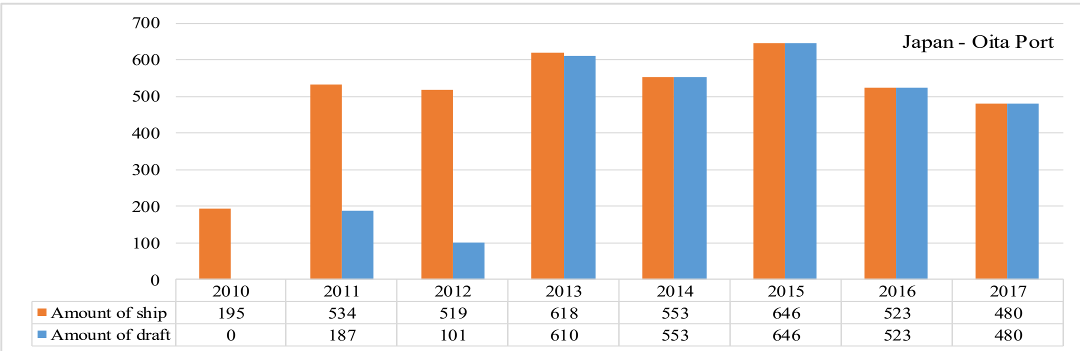


Fig. 4.23 Data Availability of Oita Port

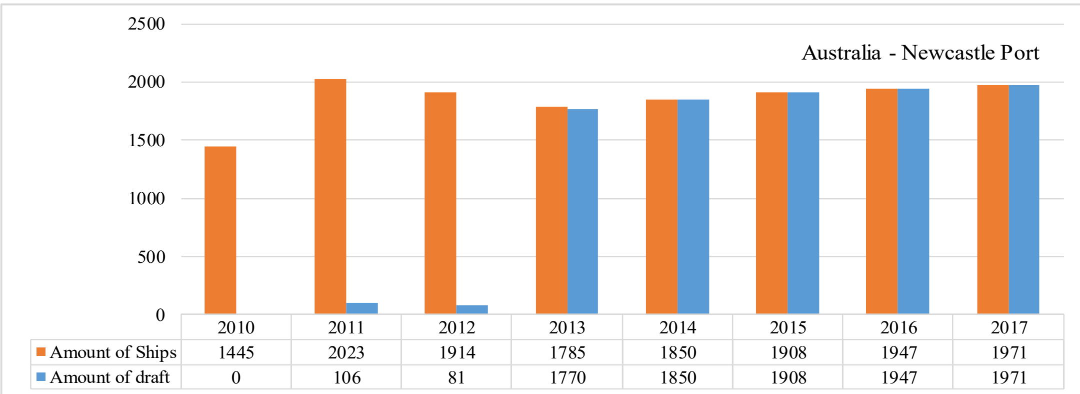
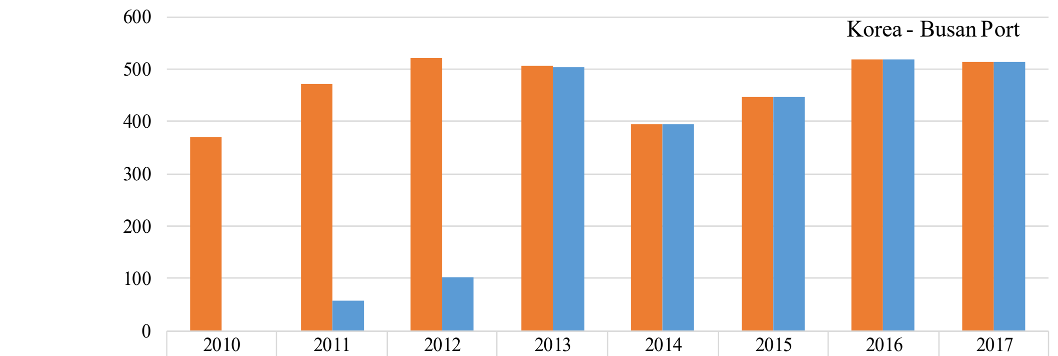
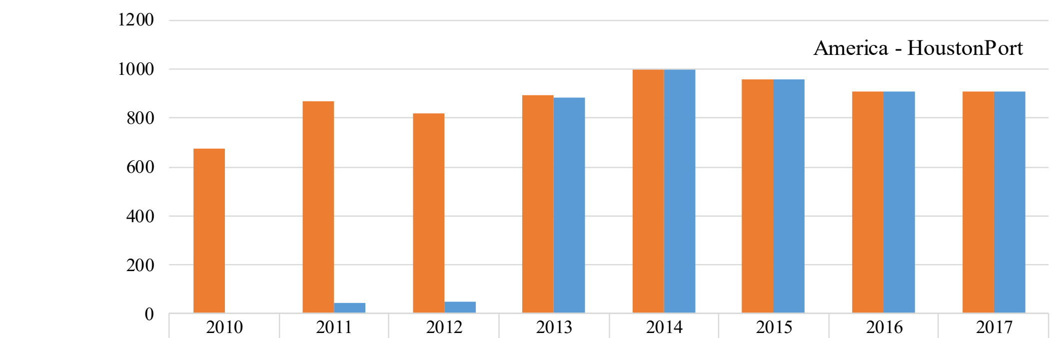


Fig. 4.24 Data Availability of Newcastle Port



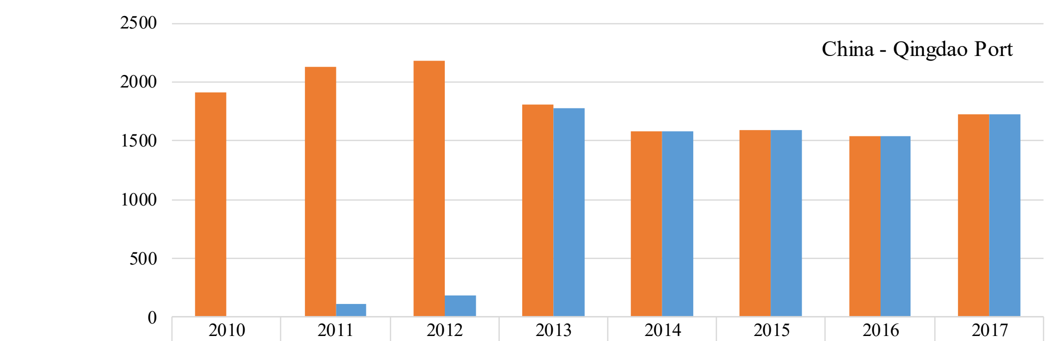
	2010	2011	2012	2013	2014	2015	2016	2017
Amount of ships	369	471	522	506	396	448	518	513
Amount of draft	0	57	103	503	396	448	518	513

Fig. 4.25 Data Availability of Busan Port



	2010	2011	2012	2013	2014	2015	2016	2017
Amount of ships	674	869	821	893	999	959	909	908
Amount of draft	0	46	51	885	999	959	909	908

Fig. 4.26 Data Availability of Houston Port



	2010	2011	2012	2013	2014	2015	2016	2017
Amount of Ships	1908	2125	2180	1809	1578	1588	1538	1728
Amount of draft	0	110	181	1774	1578	1588	1538	1728

Fig. 4.27 Data Availability of Qingdao Port

Table 4.15 Availability Data of Developed Country

Period	Japan		Australia		China		Korea		America	
	Oita		Newcastle		Qingdao		Busan		Houston	
	∑Ships	∑draft	∑Ships	∑draft	∑Ships	∑draft	∑Ships	∑draft	∑Ships	∑draft
2010	195	0	1445	0	1908	0	369	0	674	0
2011	534	187	2023	106	2125	110	471	57	869	46
2012	519	101	1914	81	2180	181	522	103	821	51
2013	618	610	1785	1770	1809	1774	506	503	893	885
2014	553	553	1850	1850	1578	1578	396	396	999	999
2015	646	646	1908	1908	1588	1588	448	448	959	959
2016	523	523	1947	1947	1538	1538	518	518	909	909
2017	480	480	1971	1971	1728	1728	513	513	908	908

(2) Developing Country

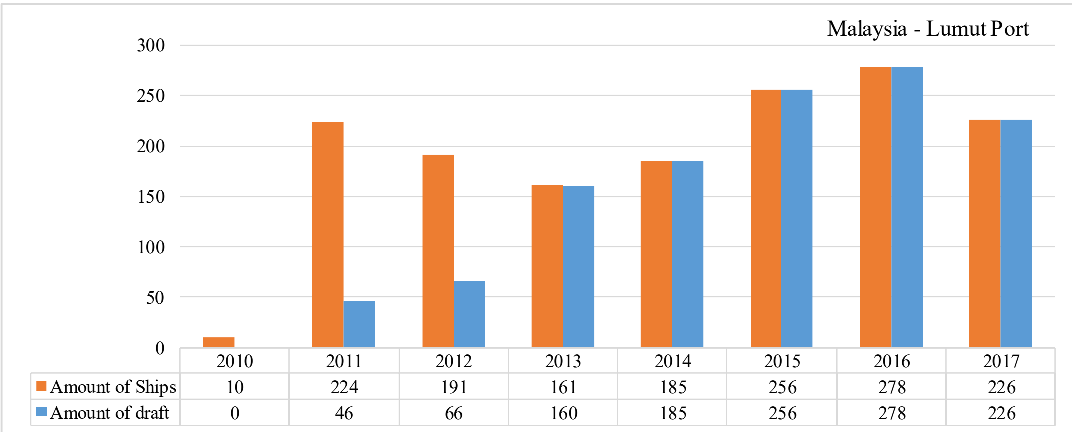


Fig. 4.28 Data Availability of Lumut Port

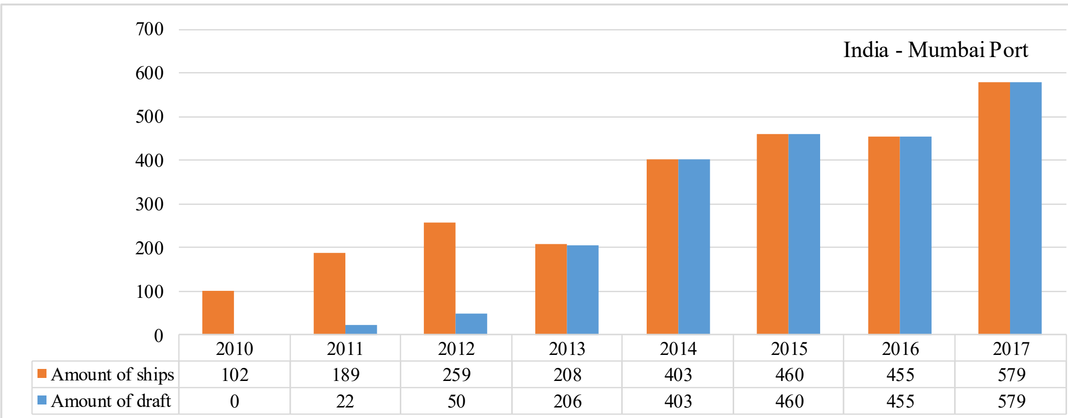


Fig. 4.29 Data Availability of Mumbai Port

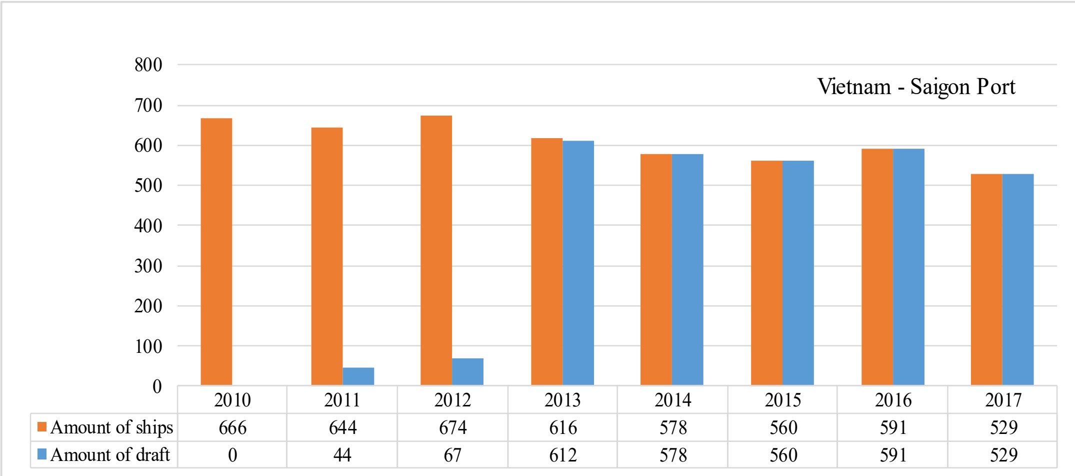


Fig. 4.30 Data Availability of Saigon Port

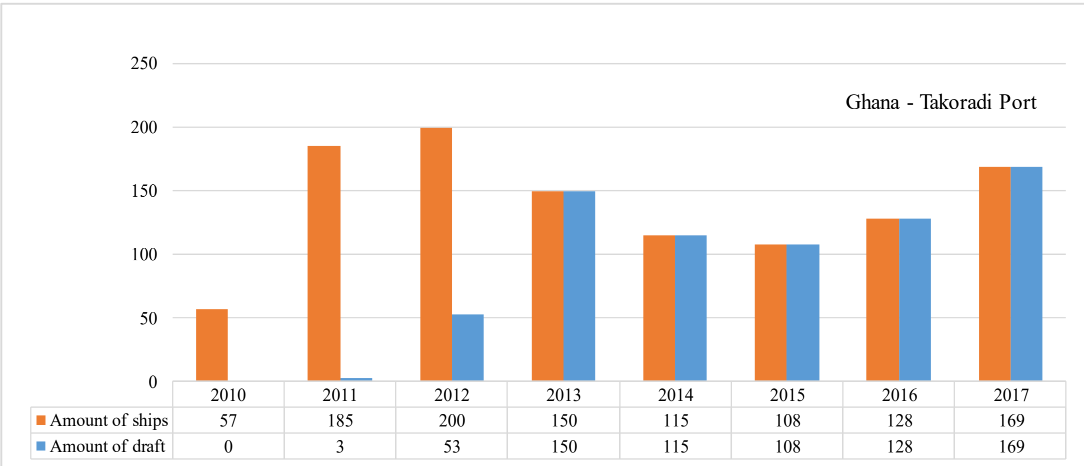


Fig. 4.31 Data Availability of Takoradi Port

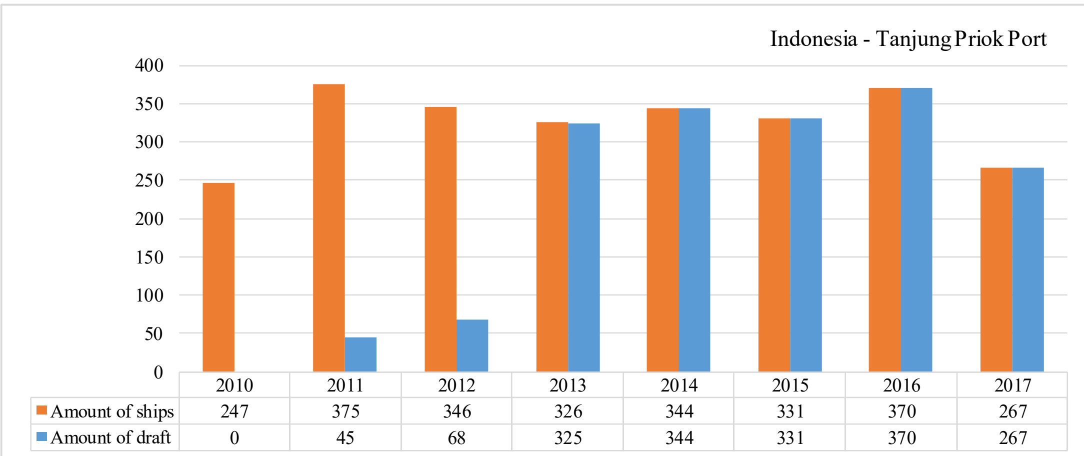


Fig. 4.32 Data Availability of Tanjung Priok Port

Table 4.16 Availability Data of Developing Country

Period	Malaysia		India		Vietnam		Ghana		Indonesia	
	Lumut		Mumbai		Saigon		Takoradi		Tanjung Priok	
	∑Ships	∑draft	∑Ships	∑draft	∑Ships	∑draft	∑Ships	∑draft	∑Ships	∑draft
2010	10	0	102	0	666	0	57	0	247	0
2011	224	46	189	22	644	44	185	3	375	45
2012	191	66	259	50	674	67	200	53	346	68
2013	161	160	208	206	616	612	150	150	326	325
2014	185	185	403	403	578	578	115	115	344	344
2015	256	256	460	460	560	560	108	108	331	331
2016	278	278	455	455	591	591	128	128	370	370
2017	226	226	579	579	529	529	169	169	267	267

Based on the results of the finding, we can conclude that based on the number of arrival ships, the data are reliable from 2013 onward both developed countries and developing countries. However, based on the availability of draft, the data is reliable from 2014 onward, which is the number of ships and the amount of draft are the same for both developed countries and developing countries.

4.7.2 Findings of Cargo Estimation Results

Below are the estimation cargo results of a developed country and developing country shown as follows:

(1) Developed Country

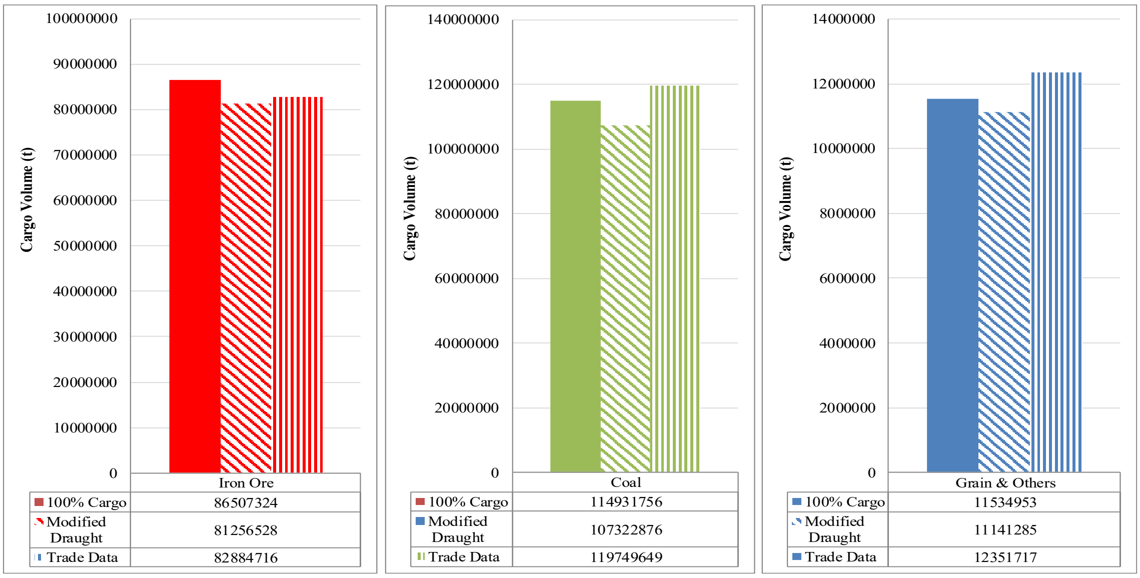


Fig. 4.33 Cargo Estimation Results (Australia~Japan)

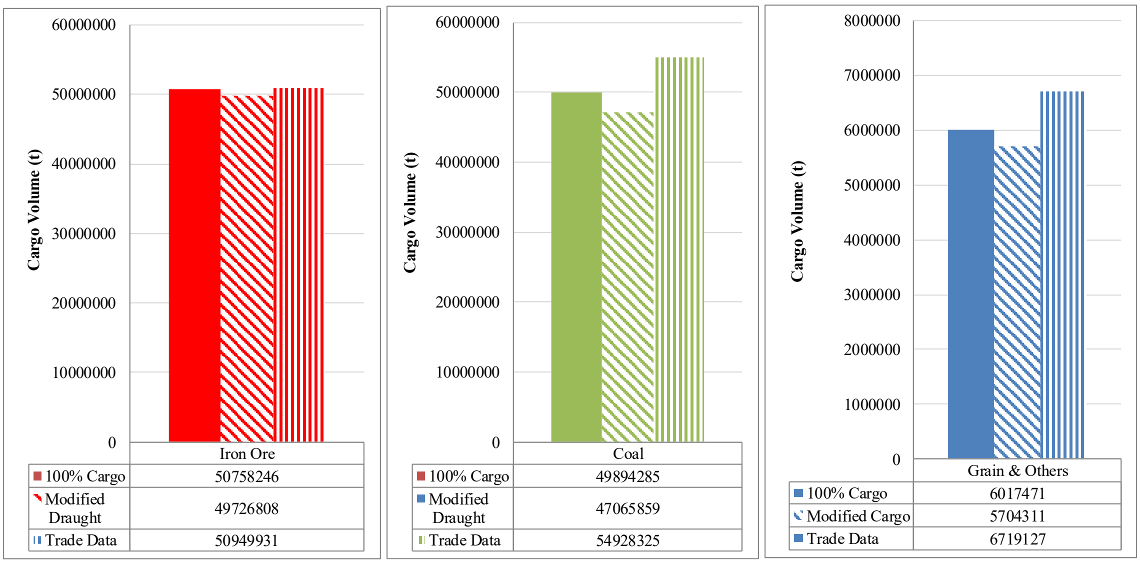


Fig. 4.34 Cargo Estimation Results (Australia~Korea)

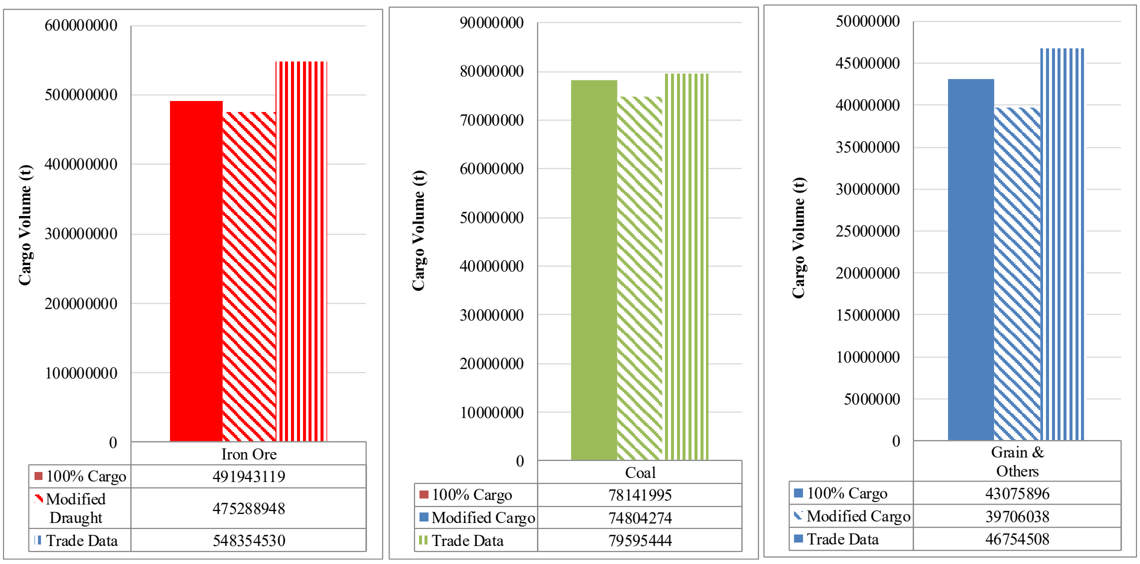


Fig. 4.35 Cargo Estimation Results (Australia~China)

The findings of country trade coverage are shown in Fig. 4.33 ~ Fig. 4.35. In the case of Australia to Japan, Australia to Korea, and Australia to China, the cargo estimation is fulfilling the same amount as statistics trade data. Most of them reaching more than 90% coverage of cargo.

(2) Developing Country

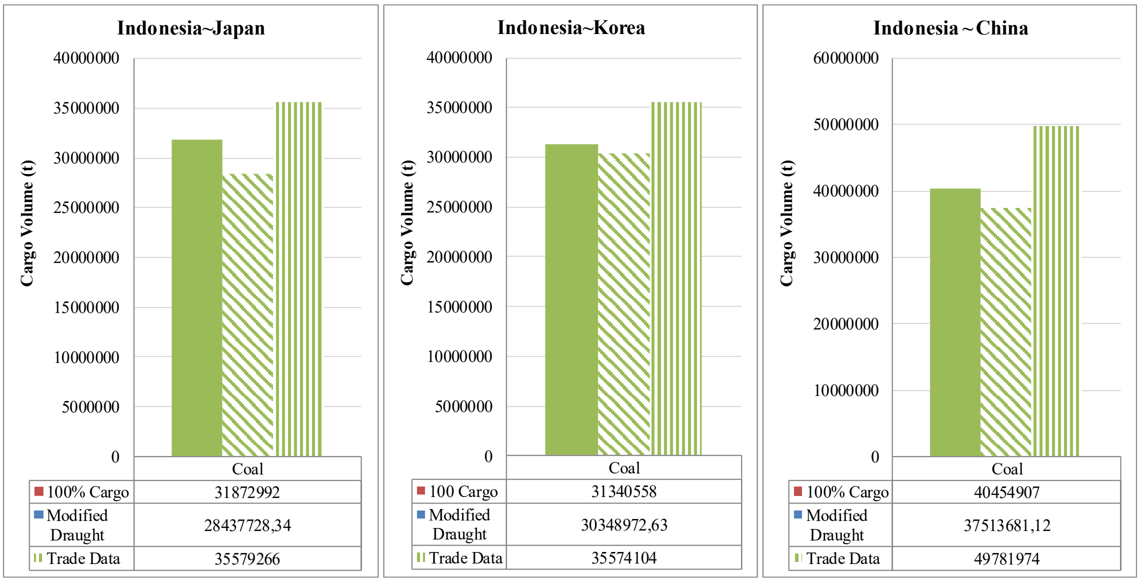


Fig. 4.36 Cargo Estimation Results (Indonesia: Japan, Korea, China)

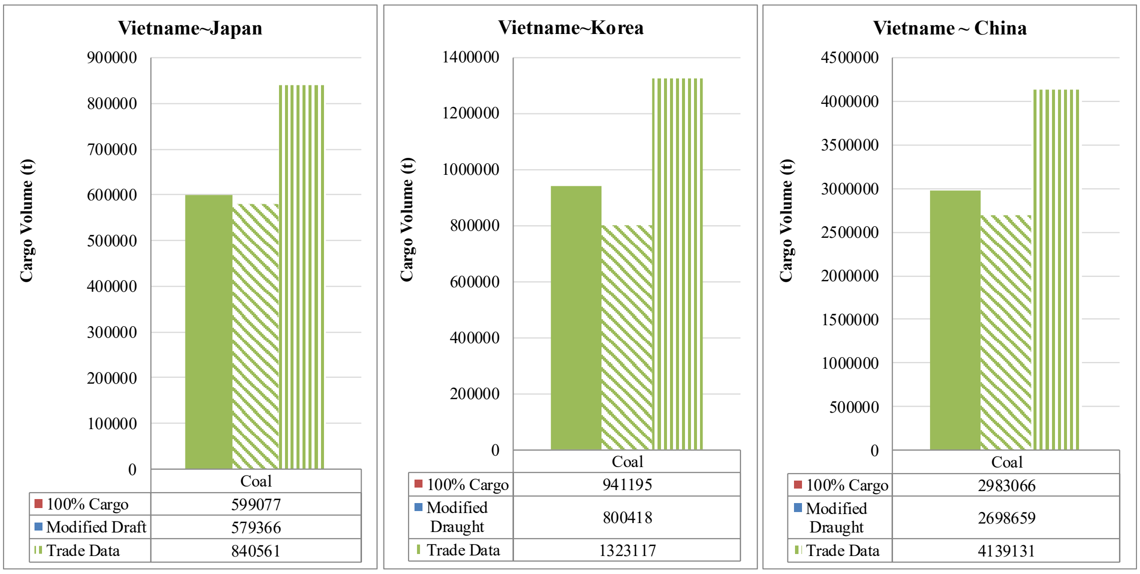


Fig. 4.37 Cargo Estimation Results (Vietnam: Japan, Korea, China)

The findings of country trade coverage are shown in Fig. 4.36 ~ Fig. 4.37. In the case of Indonesia to Japan-Korea-China, and Vietnam to Japan-Korea-China, the cargo estimation is not fulfilling the same amount as statistics trade data. Most of them are below 90 % coverage of cargo. It indicates AIS data and operation data may be missing in some countries, especially from developing countries like Indonesia and Vietnam.

Chapter 5

Shipper Model

As discussed in Section 3.2, the ship allocation model is composed of shipper, shipowner, and operator models. The data extracted from the MLDB were used to develop these models. Details of the shipper model are discussed in this section, and the other two sections will be discussed in the next section. The target of the development ship allocation model in this study are:

- Data used : All of the operation data in 2014
- Country : Australia to Japan
- Ship Type : Bulk Carrier
- Cargo Type : Iron Ore

5.1 Development of Shipper Model

5.1.1 Overview of Shipper Model

The shipper model issues a request for cargo transportation between two or more ports from Australia to Japan. Herein, the shipper model was generated using cluster analysis, which is a method of defining similarities in data, grouping similar items, and classifying them into clusters. Using hierarchical cluster analysis, we clustered shippers between Japan and Australia.

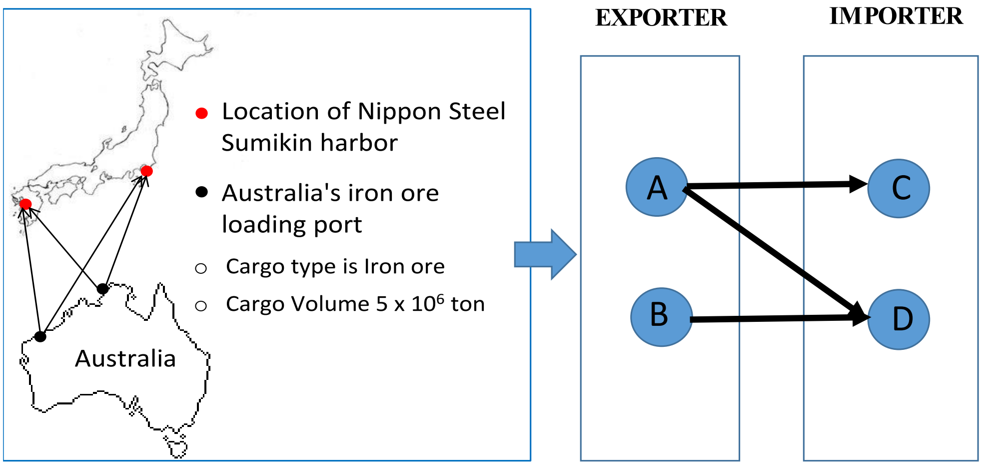


Fig. 5.1 Illustration of Shipper Model

5.1.2 Method of Shipper Model

In this study, the shipper model is generated by using the following steps:

A. Extracting Data from MLDB

Operation data from 2014 from Australia to Japan were extracted from the MLDB. The information extracted from the MLDB included operation, port (origin and destination), and ship (name, principal particulars, etc.) data. By utilizing these data, we easily analyzed the number of port callings from Australia to Japan.

Table 5.1 Example Data of Ship Operation from MLDB

Arrival Port	Vessel Name	Vessel Type	Vessel IMO	Arrival Date	Arrival Draught	Max Draught	Previous Port Departure Dtae
Kashima	SHIP B	BC	9355147	2015-12-02 01:55:46	18	18,105	2015-11-17 05:26:02
Kashima	SHIP B	BC	9355147	2014-03-07 23:10:20	17,2	18,105	2014-02-18 02:10:20
Kashima	SHIP B	BC	9355147	2013-01-31 07:20:31	18	18,105	2013-01-17 05:42:29
Kawasaki	SHIP B	BC	9355147	2017-08-04 05:41:07	18,1	18,105	2017-07-22 13:23:11
Kisarazu	SHIP B	BC	9355147	2016-01-12 07:56:03	18	18,105	2015-12-30 02:57:05

B. Define a Shipper Using Cluster Analysis

Data clustering is the process of identifying natural groupings or clusters within multidimensional data based on some similarity measure (e.g. Euclidean distance) [128-130]. It is an important process in pattern recognition and machine learning [131] [132]. Furthermore, data clustering is a central process in Artificial Intelligence (AI) [133]. Clustering algorithms are used in many applications, such as image segmentation [134-136], vector and color image quantization [137-139], data mining [140], compression [141], machine learning [142], etc. A cluster is usually identified by a cluster center (or centroid) [143]. Data clustering is a difficult problem in unsupervised pattern recognition as the clusters in data may have different shapes and sizes [145].

Clustering analysis divides data into groups (clusters) that are meaningful, useful, pr both. If meaningful groups are the goal, then the clusters should capture the natural structure of the data. In some cases, however, cluster analysis is only a useful starting point for other purposes such as data summarization. Whether for the understanding of utility, cluster analysis has long played an important role in a wide variety of fields: psychology and other social sciences, biology, statistics, pattern recognition, information retrieval, machine learning, and data mining.

There are three important techniques in cluster analysis i.e. K-means, Agglomerative Hierarchical Clustering. And DBSCAN. In this study, the Agglomerative Hierarchical Clustering concept is used. The approach of this concept refers to a collection of closely related clustering techniques that produce a hierarchical clustering by starting with each point as a singleton cluster and then repeatedly merging the two closest clusters until a single, all-encompassing cluster remains. Some of these techniques have a natural interpretation in terms of graph-based clustering, while others have an interpretation in terms of a prototype-based approach.

Agglomeration hierarchical clustering techniques are by far the most common and often displayed graphically using a three-like diagram called a dendrogram, which displays both the cluster-subcluster relationship and the order in which the clusters were merged (agglomerative view) or split (divisive view). Below is sample data to illustrate the behavior of the various hierarchical clustering algorithms i.e. sample data that consists of 6 two-dimensional points, which are shown in Fig. 5.2. The x and y coordinates of the points and the Euclidean distances between them are shown in Tables 5.2 and 5.3, respectively.

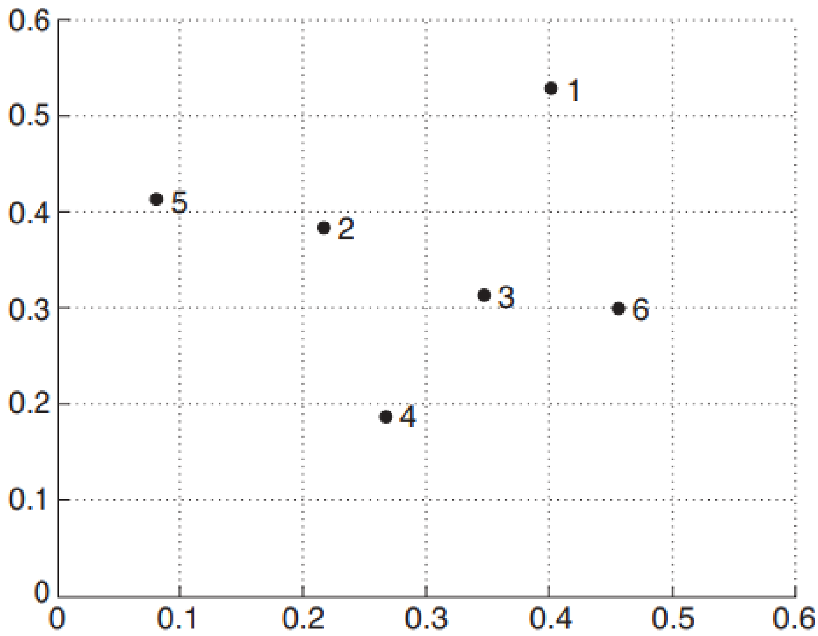


Fig. 5.2 Set of 6 two-dimensional points, [145]

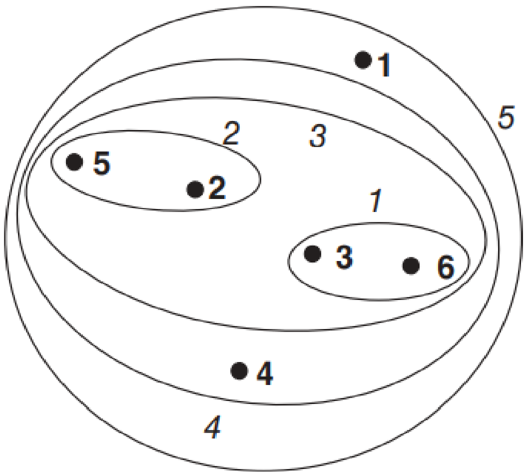
Table. 5.2 *x y* coordinates of 6 points

Point	x Coordinate	y Coordinate
p1	0.40	0.53
p2	0.22	0.38
p3	0.35	0.32
p4	0.26	0.19
p5	0.08	0.41
p6	0.45	0.30

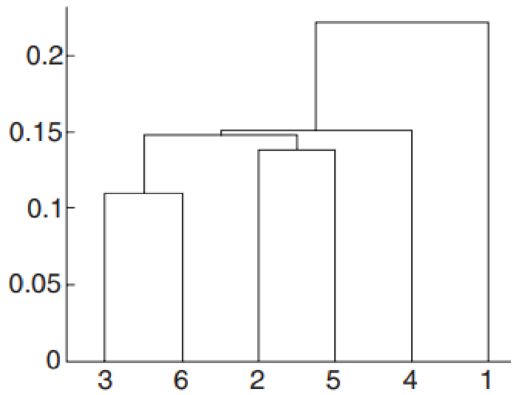
Table. 5.3 Euclidean distance of 6 points

	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00

The hierarchical clustering result of the 6 points using several types i.e. Single linkage, Complete linkage, Group average linkage, and Ward's shown as follows:



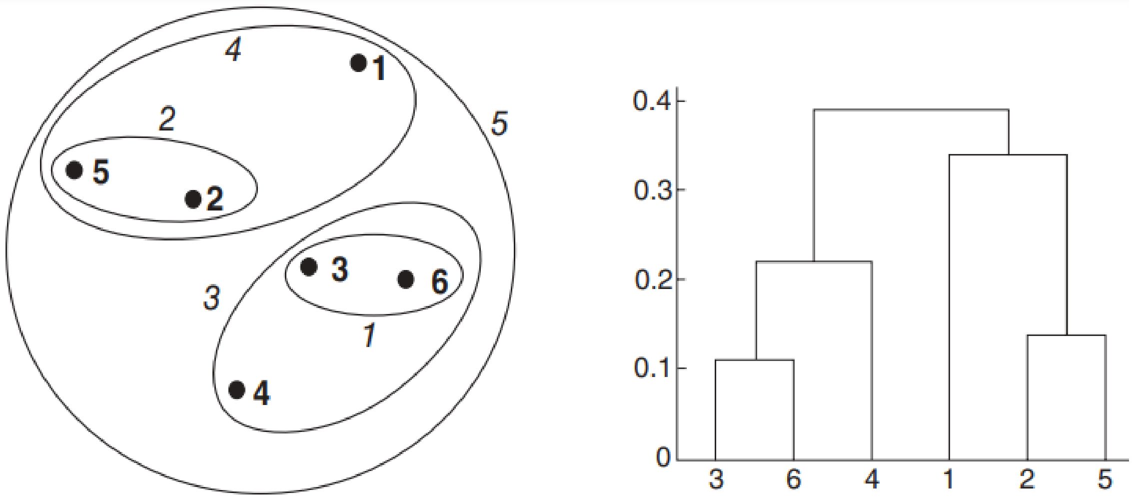
(a) Single link clustering.



(b) Single link dendrogram.

Fig. 5.3 Single link clustering of 6 two-dimensional points, [145]

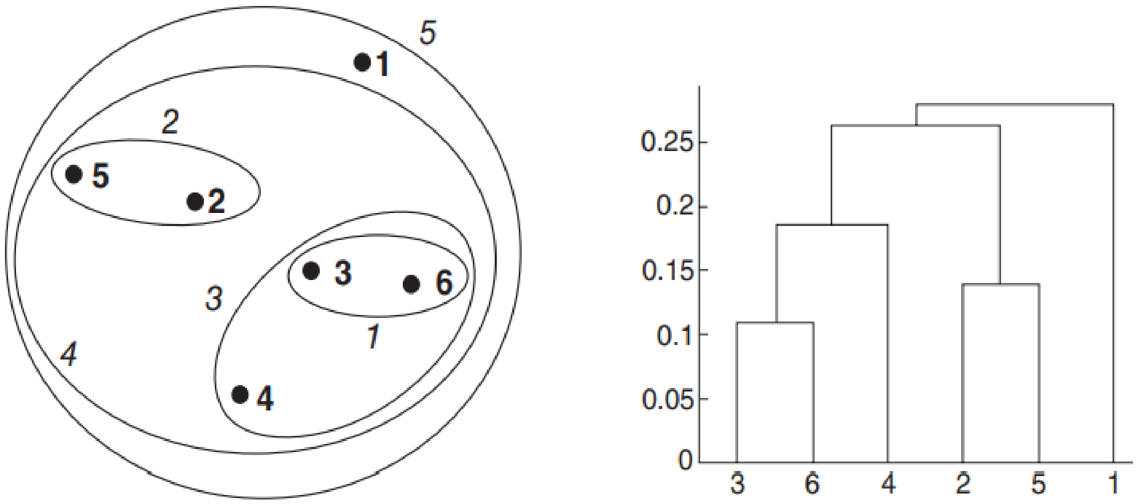
Fig. 5.3 shows the result of applying the single linkage link technique to the example data set of six points above. Fig. 5.3 (a) shows the nested clusters as a sequence of nested ellipses, where the numbers associated with the ellipses indicated the order of the clustering. Fig. 5.3 (b) shows the same information but as a dendrogram. The height at which two clusters are merged in the dendrogram reflects the distances of the two clusters.



(a) Complete link clustering. (b) Complete link dendrogram.

Fig. 5.4 Complete link clustering of 6 two-dimensional points, [145]

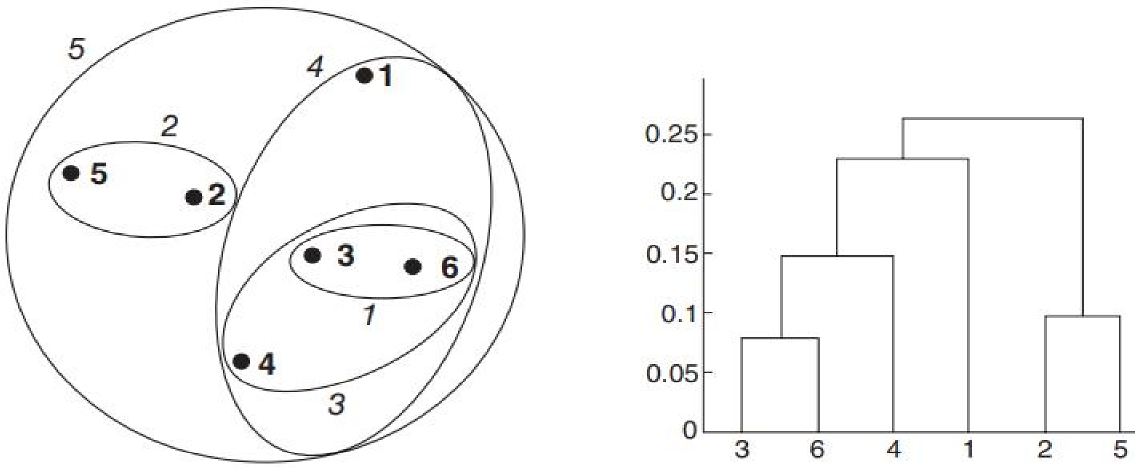
Fig. 5.4 shows the result of applying the complete linkage link technique to the example data set of six points above. Fig. 5.4 (a) shows the nested clusters as a sequence of nested ellipses, where the numbers associated with the ellipses indicated the order of the clustering. Fig. 5.4 (b) shows the same information but as a dendrogram. The height at which two clusters are merged in the dendrogram reflects the distances of the two clusters.



(a) Group average clustering. (b) Group average dendrogram.

Fig. 5.5 Group average clustering of 6 two-dimensional points, [145]

Fig. 5.5 shows the result of applying the group average clustering technique to the example data set of six points above. Fig. 5.5 (a) shows the nested clusters as a sequence of nested ellipses, where the numbers associated with the ellipses indicated the order of the clustering. Fig. 5.5 (b) shows the same information but as a dendrogram. The height at which two clusters are merged in the dendrogram reflects the distances of the two clusters.



(a) Ward's clustering. (b) Ward's dendrogram.

Fig. 5.6 Ward's clustering of 6 two-dimensional points, [145]

In this study, to define a shipper between Australia and Japan, we identified the number of port callings in 2014 using cluster analysis. Where the clustering process can be described as follows:

(1) Calculate the number of port callings

The number of port callings was calculated by identifying data extracted from the MLDB. By using a matrix between the ports (P1, P2, …, Pn) and ships (S1, S2, …, Sn). As shown in Table 5.4 (1), the number of port callings could be calculated.

(2) Measure the Euclidean distance

Euclidean distance is a measure of the true straight-line distance between two points in Euclidean space. In hierarchical clustering, in which the distance measure is Euclidean, data must first be normalized or standardized to prevent the covariant with the highest variance from driving the clustering. The data consist of many calling ships whose weights and numbers of calls differ. Therefore, it was necessary to standardize the differences in each property. Data standardization was performed as shown in Table 5.4 (2) for each port. Then, the Euclidean distance was calculated using Eq. (3), the result of which is shown in Table 5.4 (3). Where x_i and y_i are the numbers of calls after standardizing ship i at ports x and y, respectively.

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \quad (3)$$

Table. 5.4 Cluster analysis process

(1) Port Calling Calculation

Ship / Port	S1	S2	S3	S4	S5	S6
P1	0	0	1	0	2	0
P2	0	0	1	0	3	2
P3	0	0	0	1	0	0
P4	5	1	0	0	0	0
P5	2	0	0	0	0	0

(2) Standardization

Ship / Port	S1	S2	S3	S4	S5	S6
P1	-0.7	-0.7	0.65	-0.7	1.96	-0.7
P2	-0.9	-0.9	0	-0.9	1.73	0.87
P3	-0.4	-0.4	-0.4	2.24	-0.4	-0.4
P4	2.19	0	-0.5	-0.5	-0.5	-0.5
P5	2.24	-0.4	-0.4	-0.4	-0.4	-0.4

(3) Calculation of Euclidean distances

Port / Port	P1	P2	P3	P4	P5
P1		1.71	3.94	4.04	3.94
P2	1.71		4.08	4.21	4.08
P3	3.94	4.08		3.87	3.79
P4	4.04	4.21	3.87		0.49
P5	3.94	4.08	3.79	0.49	

(3) Clustering using hierarchical cluster analysis

First, before any clustering was performed, it was necessary to populate a proximity matrix with the distance between each point using a distance function. Then, the matrix was updated to display the distance between each cluster. In this study, to measure the distance between two clusters, we applied the average linkage method, which is commonly used and represents a natural compromise between linkage measures to provide a more accurate evaluation of the distance between clusters [145]. The distance between two clusters is calculated using Eq. (4).

$$d(C_1, C_2) = \frac{1}{|C_1||C_2|} \sum_{x_1 \in C_1} \sum_{x_2 \in C_2} d(x_1, x_2), \qquad (4)$$

where C_n is a cluster, x_n is a port, and $d(C_1, C_2)$ is the distance between cluster C_1 and C_2.

The goal of this method was to group heterogeneous port data into homogeneous clusters. By doing so, we could identify groups without previous knowledge of group membership or even the number of possible groups. Thus, shippers operating between Australia and Japan could easily be defined. Hierarchical cluster analysis is best illustrated using a dendrogram (a visual display of the clustering process).

5.2 Result of Shipper Model

The shipper model in this study is defined based on the exporter and Consignee's point of view. The results of shippers operating between Australia and Japan (2014) illustrated using a dendrogram shown in Fig. 5.7 (Importer point of view) and Fig. 5.9 (Eksporter point of view). Fig. 5.7 shows the results of shipper clusters based on the Consignee’s point of view. Based on the Consignee’s point of view, the ports were grouped into four clusters (Shippers A–D), defined as follows:

- Shipper A (Kawasaki, Mizushima, Chiba, and Fukuyama)
- Shipper B (Oita, Kashima, and Kisarazu)
- Shipper C (Nagoya, Wakayama, and Tobata)
- Shipper D (Higashi-Harima, Himeji, Kure, Saganoseki, Tomakomai, and Hachinohe)

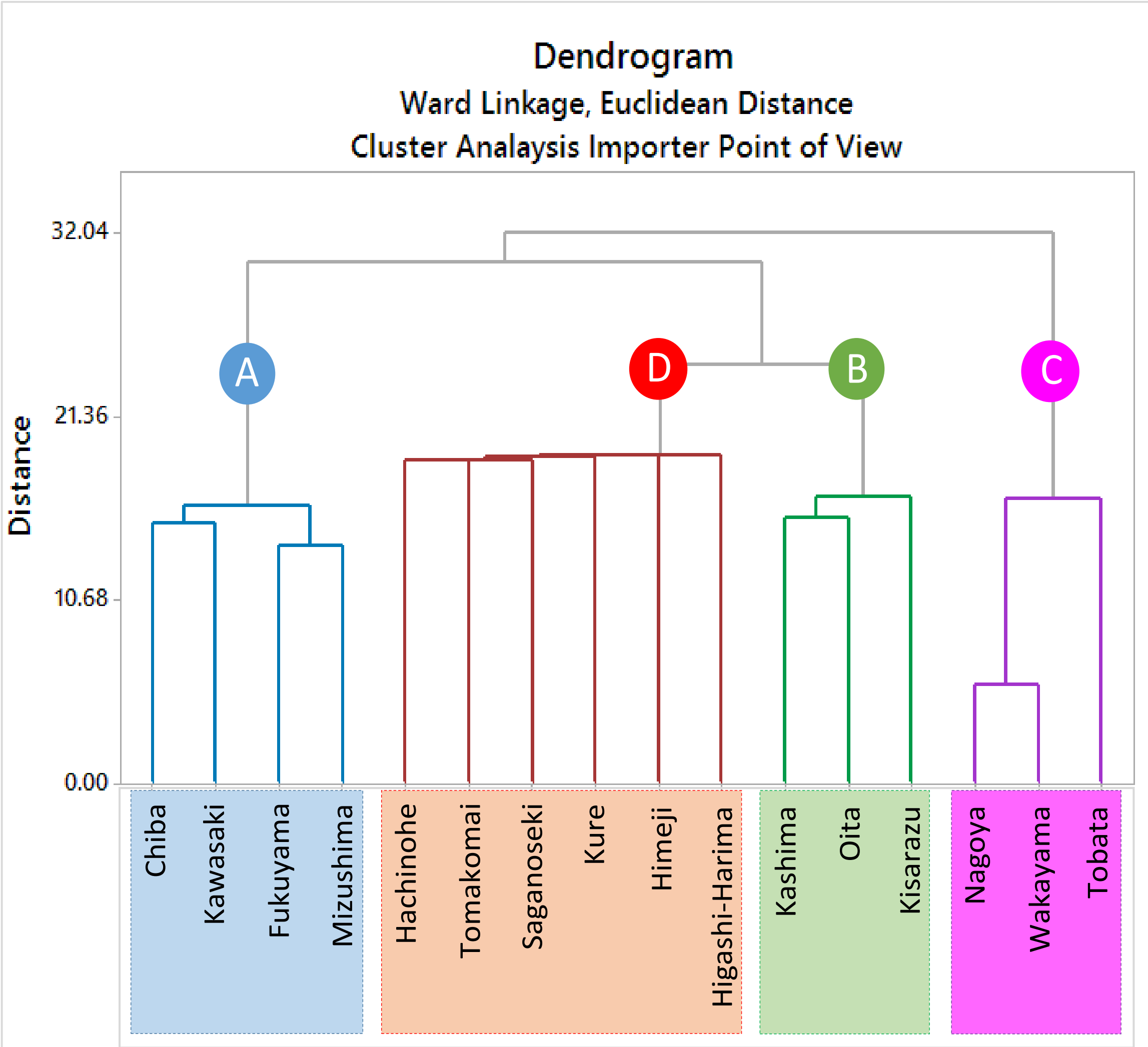


Fig. 5.7 Shippers Cluster Based on Consignee's Point of View

Table. 5.5 Amount of Iron Ore Imported from Australia (2014)

Cluster	Port Name	Amount of Ship	Amount of Cargo	DWT	L (m)	B (m)	d (m)
A	Chiba	38	6,736,992	220,000	300	50	18
	Fukuyama	47	8,843,141	220,000	300	50	18
	Kawasaki	19	3,829,906	260,000	340	50	18
	Mizushima	47	9,387,567	220,000	340	50	18
B	Kashima	56	11,203,358	300,000	340	60	19
	Oita	80	15,847,591	400,000	450	60	25
	Kisarazu	52	11,680,407	300,000	330	60	19
C	Tobata	3	311,937	160,000	327	43	16
	Wakayama	11	1,056,076	160,000	300	43	14
	Nagoya	14	1,439,954	110,000	300	43	16
D	Kure	19	3,124,094	276,000	360	45	18
	Himeji	8	240,329	257,000	335	47	16
	Higashi-Harima	40	6,199,892	180,000	330	47	17

By using the clustering analysis, we can easily calculate the amount of cargo (iron ore) imported from Australia in 2014 as shown in Table 5.5. Moreover, we can easily identify the correlation of port sizes, amount of cargo, and clusters as shown in Fig. 5.8.

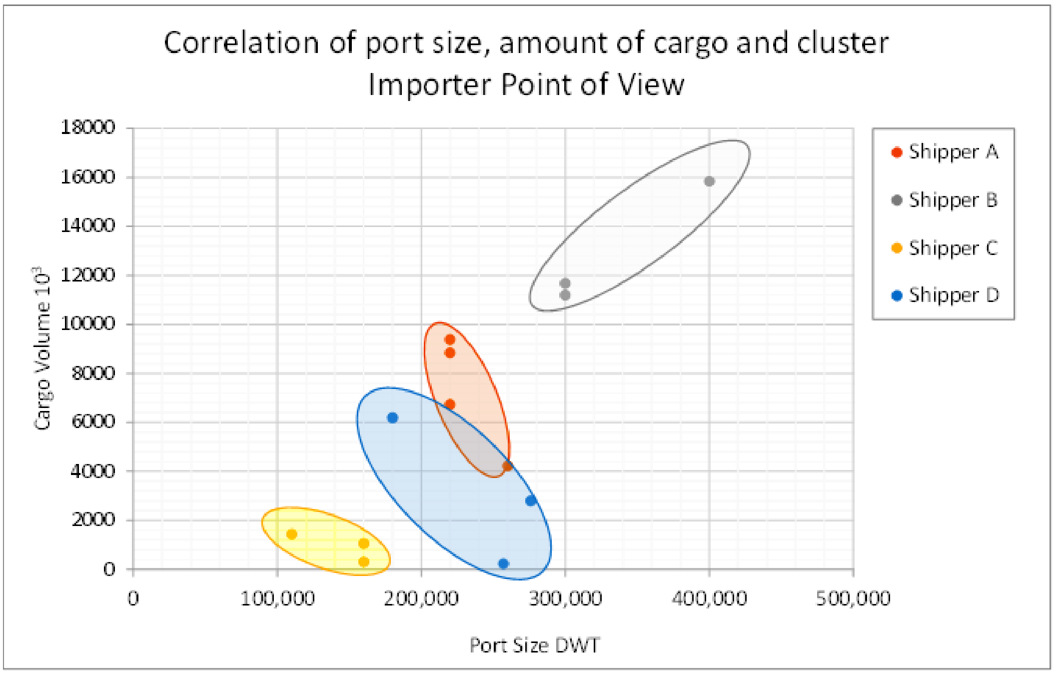


Fig. 5.8 Correlation of port based on the Consignee's point of view

However, the results of shipper clusters based on the Shipper's point of view are shown in Fig. 5.9. Based on the Shippers's point of view, the ports were grouped into three clusters (Shippers A–C), defined as follows:

- Shipper A (Port Hedland)
- Shipper B (Dampier, Port Walcott)
- Shipper C (Experance, and Parker Point)

Not only importer point of view, but by using the clustering analysis we also can easily calculate the amount of cargo (iron ore) exported to Japan from Australia in 2014 as shown in Table 5.6. Moreover, we also can easily identify the correlation of port sizes, amount of cargo, and clusters as shown in Fig. 5.10.

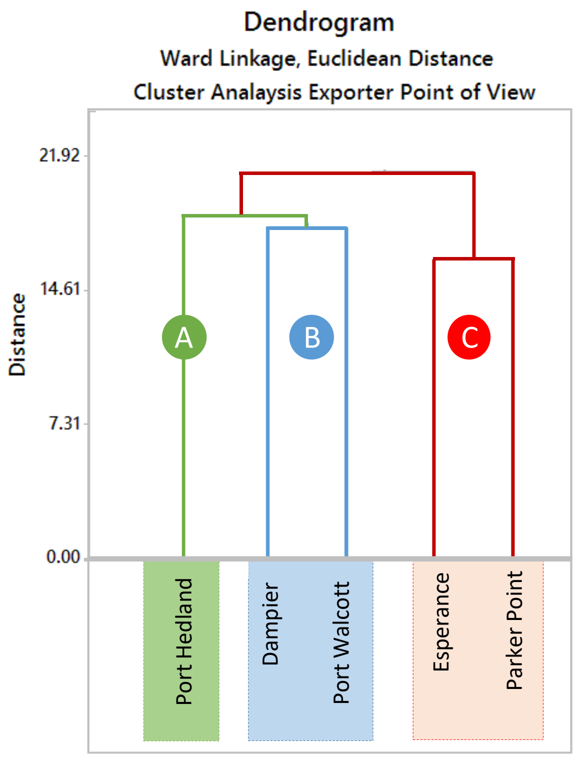


Fig. 5.9 Shippers Cluster Based on Shipper's Point of View

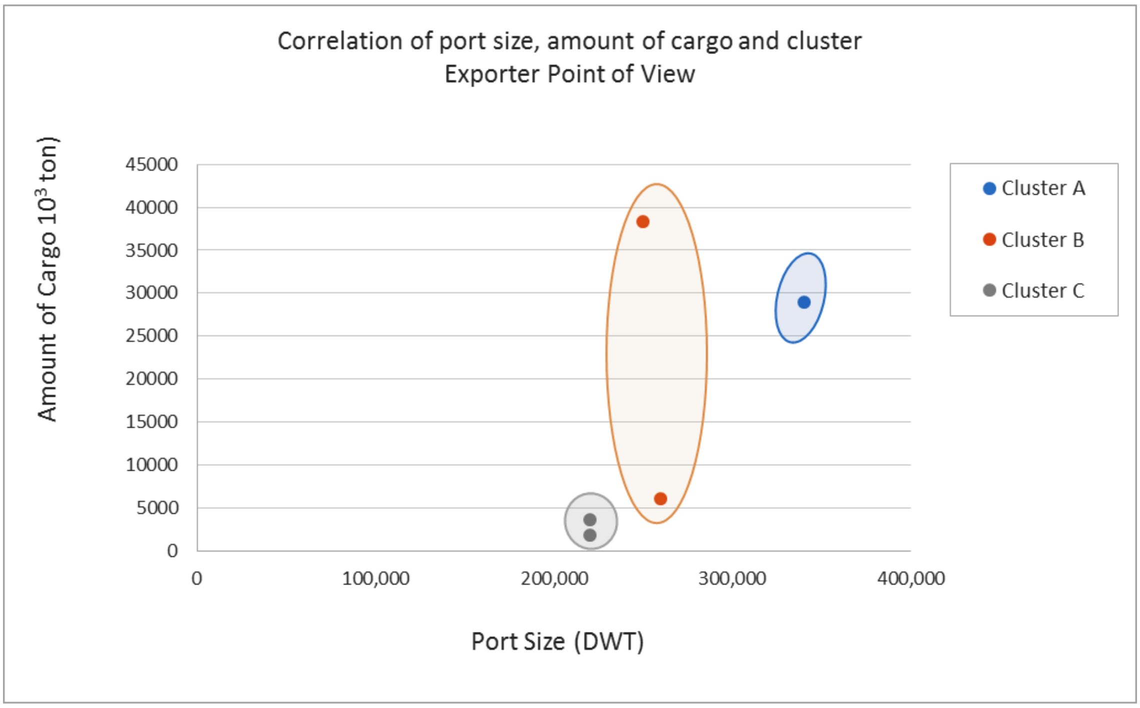


Fig. 5.10 Correlation of port based on the Shipper's point of view

Table. 5.6 Amount of Iron Ore Exported to Japan (2014)

Cluster	Port Name	Amount of Ship	Amount of Cargo	DWT	L (m)	B (m)	d (m)
A	Port Hedland	150	28 MT	340,000	335	60	19.5
B	Dampier	30	6 MT	260,000	330	55	19.2
	Port Walcott	200	38 MT	250,000	340	55	19.5
C	Esperance	13	1.9 MT	220,000	300	50	18.5
	Parker Point	21	3.7 MT	220,000	300	50	18

As shown in Fig. 5.7 and Fig. 5.9, the destination ports (Japan) and the origin ports (Australia) can be identified clearly. As a result, the shippers from Australia to Japan were generated as in Table 5.7.

Table. 5.7 Shippers from Australia to Japan, 2014

Cluster	Origin Port (Consignee)	Destination Port (Shipper)
A	Dampier	Chiba
	Parker Point	Fukuyama
	Port Hedland	Mizushima
	Port Walcott	Kawasaki
	Esperance	
B	Dampier	Kashima
	Parker Point	Kisarazu
	Port Hedland	Oita
	Port Walcott	
	Esperance	
C	Port Walcott	Nagoya
		Tobata
		Wakayama
D	Port Hedland Port Walcott	Higashi-Harima
		Kure
		Himeji

5.3 Evaluation of Shipper Model

As mentioned in the previous section, ports were grouped into 4 clusters based on the Consignee's point of view, and based on Shipper's point of viewports were grouped into 3 clusters. We confirmed the cluster analysis result based on the following points:

5.3.1 Comparison with actual locations

Based on Fig. 5.11, it can be seen that based on the Consignee point of view the Cluster A (Chiba, Kawasaki, Fukuyama, and Mizushima) matched with the location of JFE Steel. Cluster B (Kashima, Oita, and Kisarazu) and Cluster C (Nagoya, Wakayama, and Tobata) are matched with the Nippon Steel & Sumitomo Metal Terminal (NSSMT), and the last Cluster D (Kure, Himeji, and Higashi-Harima) are matched with the location of Kobelco Nisshin. From the result can be concluded that based on the Consignee point of view, the cluster analysis is matched with the current location of the company.

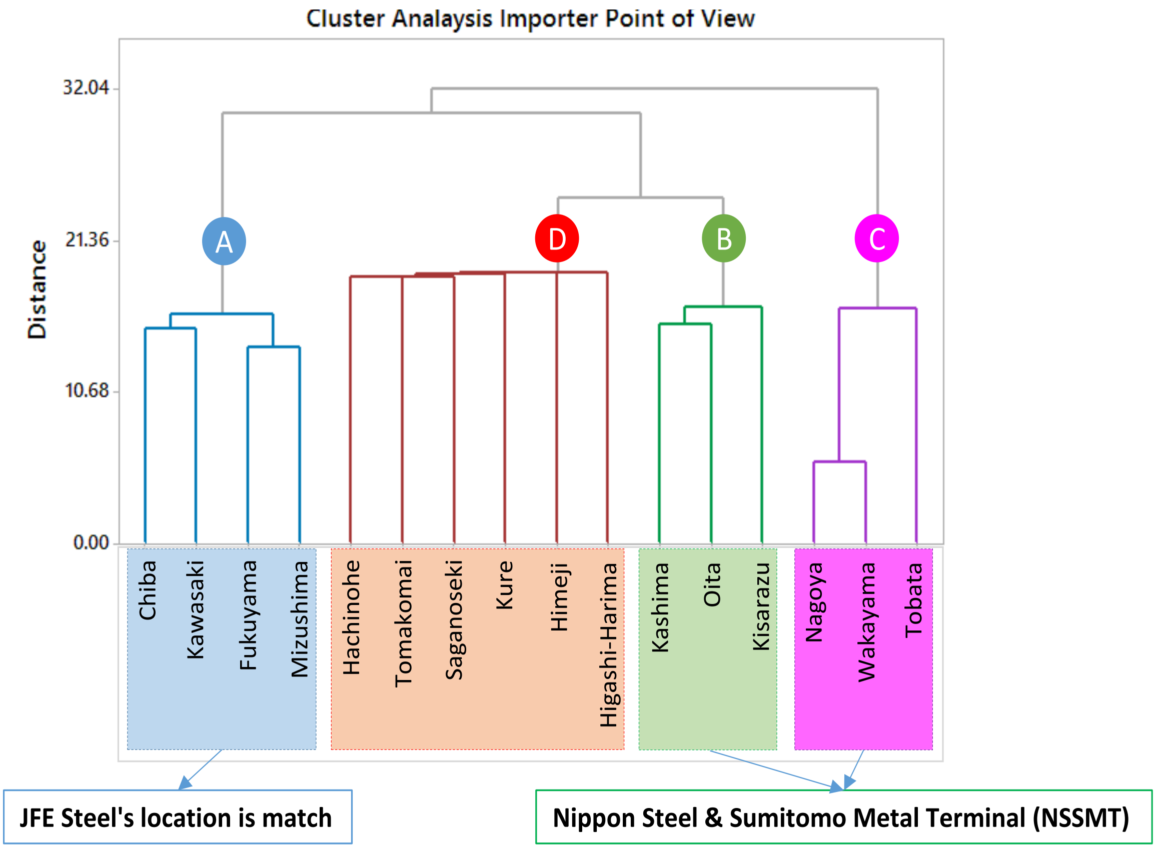


Fig. 5.11 Matching location of shipper based on Consignee point of view

In the case of the results of Shipper point of view based on Fig. 5.12, Cluster A (Port Hedland) is matched with the BHP Billington and Atlas Iron. Cluster B (Dampier, and Port Walcott) and Cluster C (Experance and Parker Point) are matched with Rio Tinto company locations. This shipper is also divided into two clusters because the port constraints in Cluster B and C are quite different. From the result can be concluded that based on the Shipper's point of view, the cluster analysis is matched with the current location of the company.

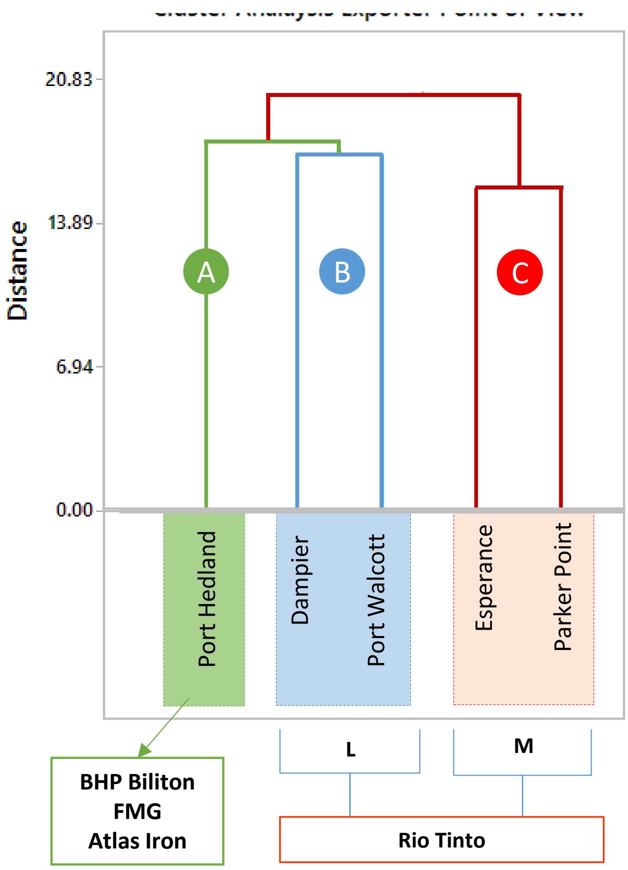


Fig. 5.12 Matching location of shipper based on Exporter point of view

5.3.2 Consistency of the cluster results

In this study, to confirm the consistency of the cluster, the cluster results from 2014 (Australia~Japan) are compared with the cluster results from 2017 (Australia~Japan). The comparison results are shown in Fig. 5.13 and Fig. 5.14 as follows:

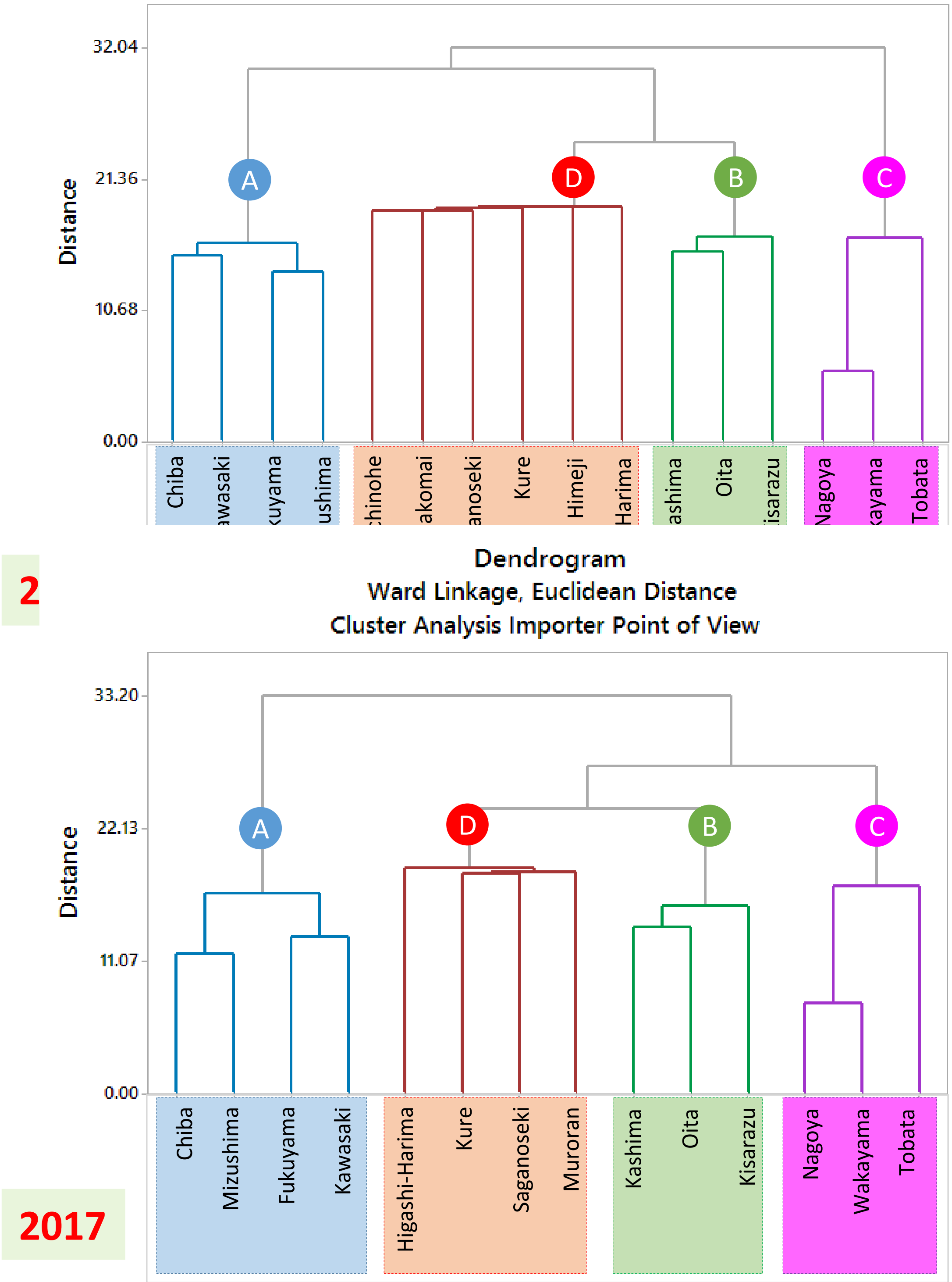


Fig. 5.13 Consistency of the cluster of 2014 and 2017 based on Consignee's point of view

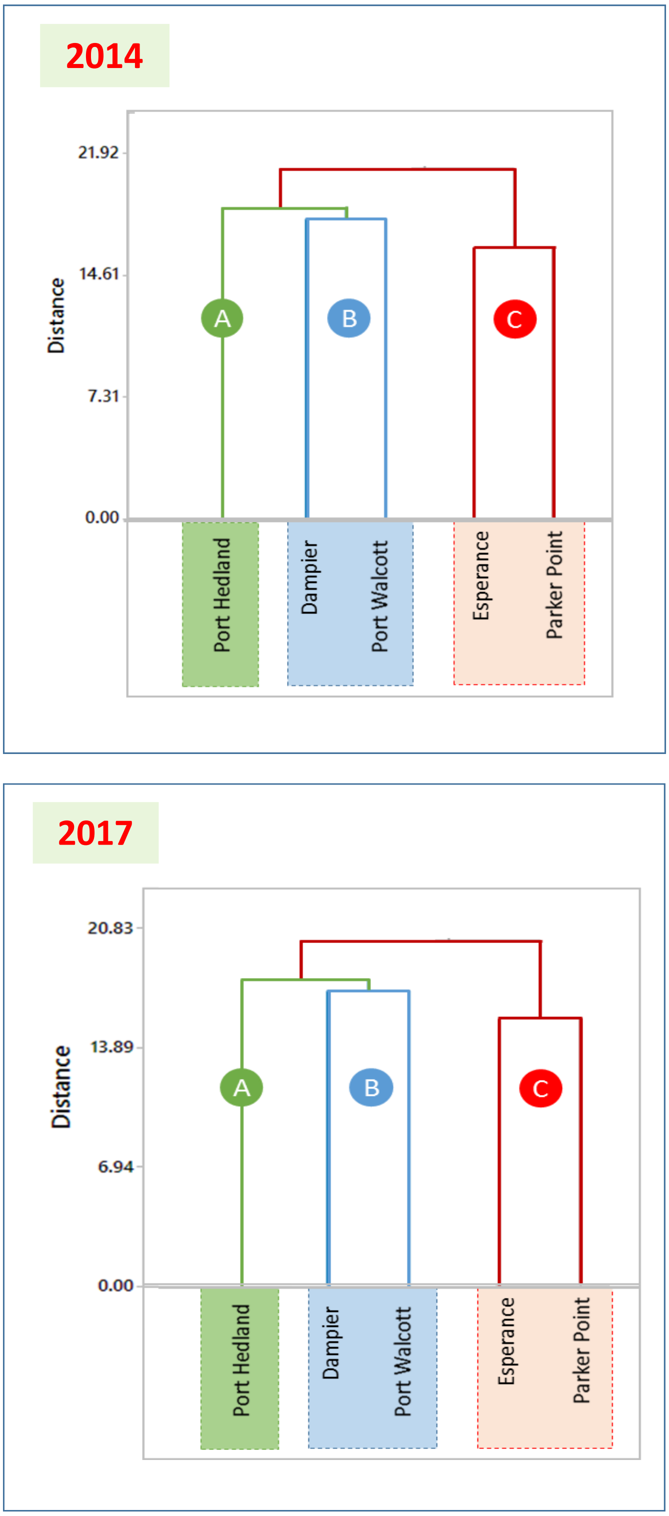


Fig. 5.14 Consistency of the cluster of 2014 and 2017 based on Consignee's point of view

Based on Fig. 5.13 and Fig.5.14, it can be concluding that the shipper from 2014 and the shipper from 2017 have remained the same. It means that the cluster results are consistent with the current situation.

5.3.3 Comparison with ship locations

Based on the result in Table 5.7, most ships operating from Australia to Japan loaded cargo from two or more ports in Australia and unloaded at two or more ports in Japan. Here, we compare the results of cluster analysis with actual operation. Some typical operations are shown in Table 6.1, where the gray represents ports in Australia, and the white represents ports in Japan.

In the case of Ship A, the cargo was loaded at Port Walcott–Port Hedland then unloaded at Mizushima, Chiba, and Fukuyama, which matches the operation of Shipper A. In contrast, in the case of Ship C, the cargo was loaded at Port Walcott, then unloaded at Kisarazu, Kashima, and Oita, matching Shipper B. As discussed in this section, cluster analysis matched actual location and ship operation conditions.

Table 5.8 Characteristics of actual ship operation

Ship A		Ship C	
Origin	Destination	Origin	Destination
Port Walcott	Fukuyama	Port Walcott	Kisarazu
Fukuyama	Mizushima	Kisarazu	Port Walcott
Mizushima	Port Hedland	Port Walcott	Kashima
Port Hedland	Mizushima	Kashima	Port Walcott
Mizushima	Port Hedland	Port Walcott	Oita
Port Hedland	Chiba	Oita	Port Walcott
Chiba	Mizushima	Port Walcott	Kisarazu
Mizushima	Port Hedland	Kisarazu	Port Walcott
Port Hedland	Chiba	Port Walcott	Kashima
Chiba	Fukuyama	Kashima	Port Walcott

5.4 Findings of Shipper Model

(1) Characteristics of Shipper Model Based on Owned Company Location

The characteristics of the shipper model based on the importer point of view are shown in Table 5.9, and the the characteristics of shipper model based on the exporter point of view are shown in Table 5.10.

Table 5.9 Characteristic of shipper model based on Consignee's point of view

Cluster	Port Name	DWT	L (m)	B (m)	d (m)	Cargo (MT)
JFE Steel	Chiba	220,000	300	50	18	6,7
	Fukuyama	220,000	300	50	18	8,8
	Kawasaki	260,000	340	50	18	3,8
	Mizushima	220,000	340	50	18	9,4
Nippon Steel & Sumitomo Metal Terminal (NSSMT)	Kashima	300,000	340	60	19	11,2
	Oita	400,000	450	60	25	15,8
	Kisarazu	300,000	330	60	19	11,7
Nippon Steel & Sumitomo Metal Terminal (NSSMT)	Tobata	160,000	327	43	16	0,3
	Wakayama	160,000	300	43	14	1,1
	Nagoya	110,000	300	43	16	1,4
KOBELCO Nisshin	Kure	276,000	360	45	18	3,1
	Himeji	257,000	335	47	16	0.2
	Higashi-Harima	180,000	330	47	17	6,1

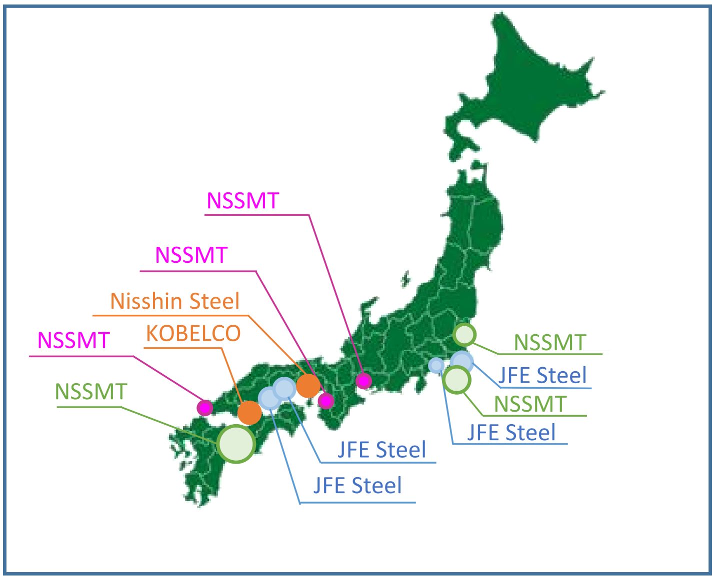


Fig. 5.15 Characteristics of shipper model (based on Consignee's point of view)

Table 5.10 Characteristic of shipper model based on Shipper's point of view

Cluster	Port Name	DWT	L (m)	B (m)	d (m)	Cargo (MT)
BHP Biliton	Port Hedland	340,000	335	60	19.5	28
RIO Tinto A	Dampier	260,000	330	55	19.2	6
	Port Walcott	250,000	340	55	19.5	38
Rio Tinto B	Esperance	220,000	300	50	18.5	1.9
	Parker Point	220,000	300	50	18	3.7

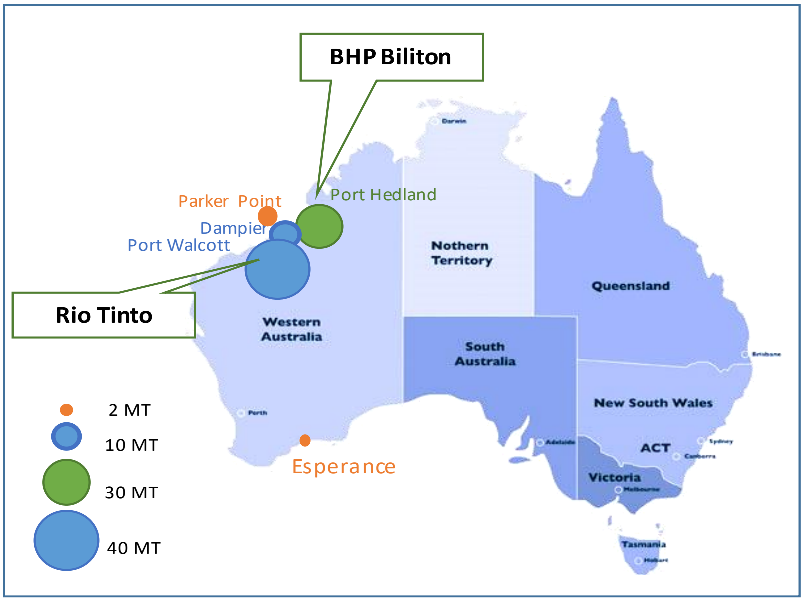


Fig. 5.16 Characteristics of shipper model based on Shipper's point of view

Based on the information above it can be concluded that based on the importer point of view the characteristics of shipper from Japan is defined as follows:

- JFE Steel used a Medium (M) size of the ship in their operation
- NSSMT is divided into two: Very Large (VL) and small ship (S)
- Kobelco Nisshin used the Large (L) ship size due to the port constraint.

However, based on the exporter point of view the characteristics of shipper from Australia is defined as follows:

- BHP Billiton used a Very Large (VL) size of the ship in their operation
- Rio Tinto A used Large (L) ship size in their operation
- Rio Tinto B used the Medium (M) ship size due to the port constraint.

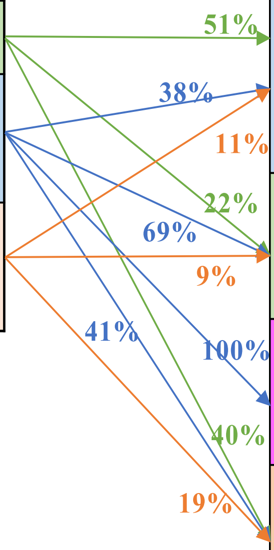
(2) Characteristics of Shipper Model Based on Port Restrictions

The characteristics of the shipper model based on the port restrictions between Australia~Japan based on Table 5.11 is shown as follows:

- Port Hedland can have covered all of the operations since the port constraints are very large.
- The constraint of cluster C on the Japan side is the most severe.
- In the case of the very large ships, the need for the cargo is covered by the port Hedland.
- Mostly, the need for the Japan cluster is covered by cluster B and cluster C from the Australian side.

Table 5.11 Characteristic of shipper model based on port restriction

Port Name	DWT	Cluster
Port Hedland	340,000	A
Dampier	260,000	B
Port Walcott	250,000	B
Esperance	220,000	C
Parker Point	220,000	C



Cluster	Port Name	DWT
A	Chiba	220,000
A	Fukuyama	220,000
A	Kawasaki	260,000
A	Mizushima	220,000
B	Kashima	300,000
B	Oita	400,000
B	Kisarazu	300,000
C	Tobata	160,000
C	Wakayama	160,000
C	Nagoya	110,000
D	Kure	276,000
D	Higashi-Harima	180,000

(3) Characteristics of Shipper Model Based on the Relationships between Shippers

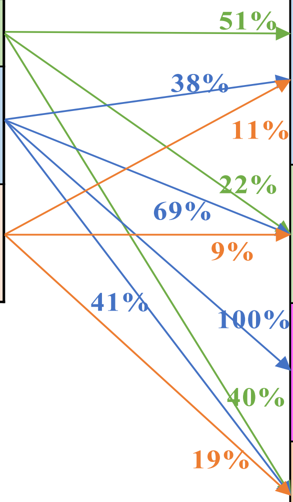
The characteristics of the shipper model based on the port restrictions between Australia~Japan based on Table 5.12 is shown as follows:

- JFE Steel has a strong relationship with BHP Billiton since the coverage of the cargo volume is the highest. Followed by the Rio Tinto in 2nd place.

- NSSMT has a very strong relationship with Rio Tinto since the coverage is reaching 78%.

Table 5.12 Characteristic of shipper model based on the relation between shippers

Port Name	DWT	Cluster
BHP Biliton	340,000	A
Rio Tinto	260,000 250,000	B
Rio Tinto	220,000 220,000	C



Cluster	Port Name	DWT
A	JFE Steel	220,000 220,000 260,000 220,000
B	Nippon Steel & Sumitomo Metal Terminal (NSSMT)	300,000 400,000 300,000
C	Nippon Steel & Sumitomo Metal Terminal (NSSMT)	160,000 160,000 110,000
D	Nisshin Steel	276,000
	KOBELCO	180,000

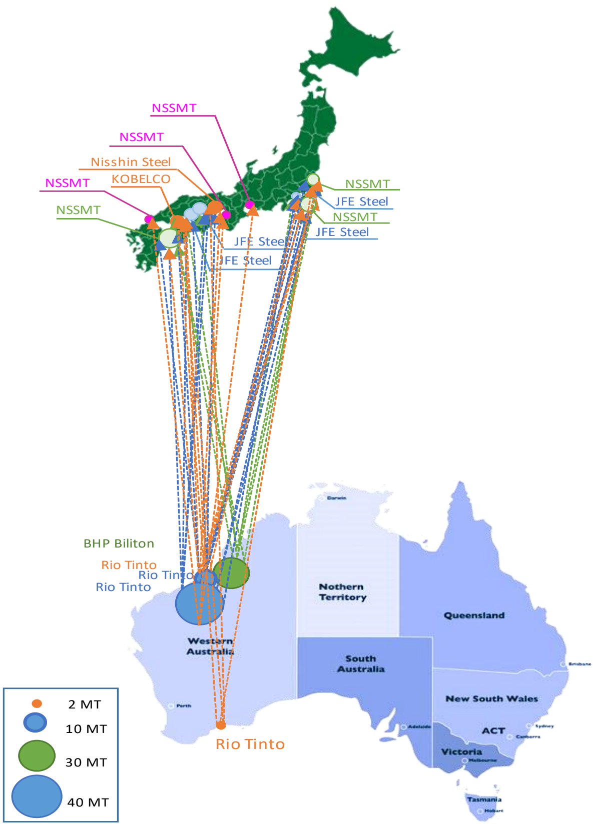


Fig. 5.17 Characteristics of shipper based on the relationship between shippers

Chapter 6

Shipowner Model

6.1 Development of Shipowner Model

The shipowner model can be used to estimate shipment days, cargo amounts, and shipment costs in response to a transportation request from an operator. To realize this model, we generated the draft rate, average speed in loading and ballast conditions, and time in port due to loading and ballast conditions using deep learning on data extracted from the MLDB.

6.1.1 Method of Shipowner Model Using Deep Learning

In this fast-growing digital world, Big Data and Deep learning are the high attention of data science. Big Data is the collection of a huge amount of digital raw data that is difficult to manage and analyze using traditional tools. As digital data is growing exponentially in different shapes, formats and sizes, therefore it is very important to manage this large volume of data according to the needs of the organization.

Deep learning methods are extensively applied to various fields of science and engineering such as speech recognition, image classifications, and learning methods in language processing. Similarly, traditional data processing techniques have several limitations of processing a large amount of data. In addition, Big Data analytics requires new and sophisticated algorithms based on machine and deep learning techniques to process data in real-time with high accuracy and efficiency.

Deep learning is an expressive machine learning technique that has recently attracted considerable attention. Machine learning is a mechanism for inputting training data into a learning machine, generating a learning model, and processing data using the learned model. The key benefit of deep learning is the analysis of massive amounts of unsupervised data, making it a valuable tool in big data analytics [146].

Deep Learning is a big data analytics tool. The advantage of deep learning is its potential to serve as the solution to data analysis and learning problems at a large rate of input data. Deep learning is capable to automatically obtain insights from complex unlabeled and unsupervised data. Deep learning algorithms are an assorted layer. Each layer comprises a nonlinear transformation of its input and produces an interpretation in its output. Deep learning aims to perform the learning of complicated and abstract insights from the input data. In a hierarchical architecture, deep learning passing the data through multiple transformation layers.

As shown in Fig 6.1, the first layer of a neural network used for deep learning is the input layer. Each node in this layer takes an input and passes its output as input to each node in the next (hidden) layer, which has no connection to the outside and is only activated by nodes in the previous layer.

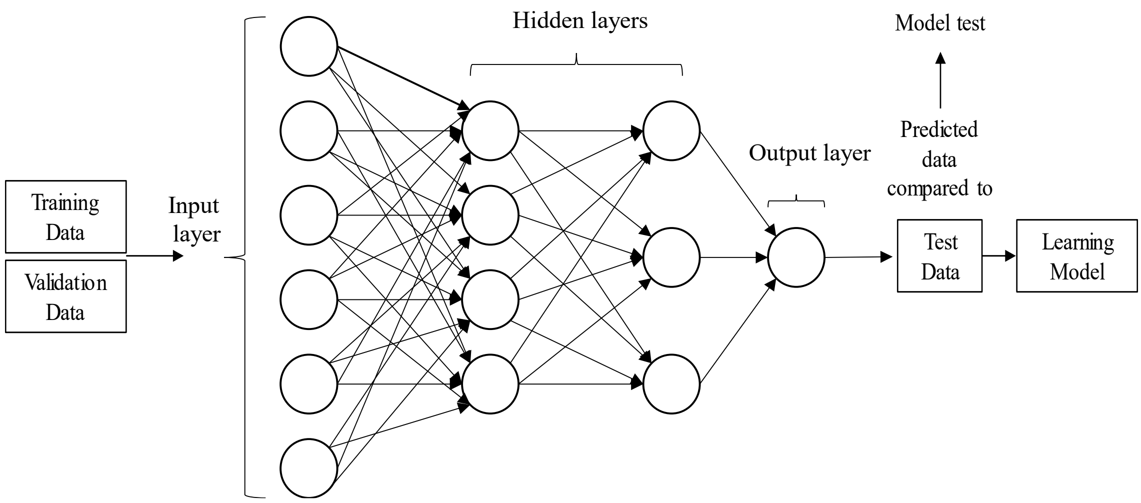


Fig. 6.1 Structure of a deep learning neural network

As shown in Fig 6.1, the deep learning architecture consists of an input layer, hidden layer, and output layer, handles the training data validation data, test data, and results from a learning model [25]. Each item will be described as follow.

a. The input layer is where the input data is received by the model. The input data (vectors) is divided into assorted nodes that are featured as input.
b. Hidden layers comprise one or more layer bridges input layer and output layer. The hidden layer aims the realization to model non-linear functions as it makes various pattern recognition possible with a layered activation function.
c. Output layer gives a prediction of real value (regression) illustrated in or probability (classification) depend on the model setup
d. The connection between layers is a feed-forward network that connects the nodes between layers. A connection is the representation of variable flow between nodes.
e. Training data is dataset comprises the example of the model parameter.
f. Validation data is a dataset used to measure the responses of learned training data. Overfitting training can be detected as the error of validation data keep increase compared to the training data error.
g. Test data is the dataset used to evaluate the final trained model. Generalization of the trained model can be assessed with comparison to test data that has never been used in the training process.

6.1.2 Deep Learning in This Study

In this study, draft rate, average voyage speed, and time in port were predicted using Deep Learning (DL) based on the following steps:

(1) Collect training data

Usually, neural networks are trained to perform single-step prediction, in which the predictor uses some available input and outputs observations to estimate a variable of interest for the timestep immediately following the latest observation. In this study, all shipping data were extracted from the MLDB. Around 75% of ship operation data from Australia to Japan in 2014 were used for training data and the remaining 25% were used for evaluation.

(2) Generate learning model

To generate a learning model, the input layer, output layer, and hyperparameters for each learning model which is consist of three distinct learning models: draught rate, average voyage speed, and port staying in time were set as follows:

a. Draught rate learning model

Ship operation data of bulk carriers operating from Australia to Japan from 2013 until 2015 is extracted from MLDB. While the only arrival in Japan ports operation that will be used, not all data can be utilized due to error or abnormal recorded draught data. The input and output layers are set as shown in Table 6.1 below.

Table 6.1 DL Input and Output Layer in Draught Rate Learning Model

<table>
<tr><th colspan="2">Deep Learning Input Layer</th></tr>
<tr><th>Data Type</th><th>Item</th></tr>
<tr><td rowspan="14">Ship</td><td>DWT</td></tr>
<tr><td>Length (L)</td></tr>
<tr><td>Breadth (B)</td></tr>
<tr><td>Depth (d)</td></tr>
<tr><td>Max Cargo (m^3)</td></tr>
<tr><td>Ship Design Speed (kt)</td></tr>
<tr><td>M/e Power (kW)</td></tr>
<tr><td>Built Year</td></tr>
<tr><td>IMO</td></tr>
<tr><td>Owner</td></tr>
<tr><td>Operator</td></tr>
<tr><td>Flag</td></tr>
<tr><td>Classification</td></tr>
<tr><td>Built Country</td></tr>
<tr><td rowspan="5">Port</td><td>Arr. Port Max. DWT</td></tr>
<tr><td>Dep. Port Max DWT</td></tr>
<tr><td>Arr. Port Max L, B, d (m)</td></tr>
<tr><td>Dep. Port Max L, B, d (m)</td></tr>
<tr><td>Dep. Port -Arr. Port Name</td></tr>
<tr><td>Route</td><td>Dep. Port – Arr. Port Distances</td></tr>
<tr><td>Train Target</td><td>Ship Draught Rate (%)</td></tr>
</table>

Deep Learning Output Layer	
Data Type	**Item**
Ship	Draught Rate (%)

b. Average voyage speed learning model

In the following step from the draught rate learning model, the data used is evaluated again focusing on average voyage speed and resulted from sailing days. The sailing days needed for the ship operating from Australia to Japan can be understood by the recorded departure date from Australia and recorded arrival date in Japan. The actual sailing days of the ship will be compared with the estimated sailing days with the ship design speed which is expected to be the speed at which the ship operates. Based on the selected data, the arrangement of items for the input layer and output layer is shown below in Table 6.2.

Table 6.2 DL Input and Output Layer in Average Voyage Speed Learning Model

Deep Learning Input Layer	
Data Type	**Item**
Ship	DWT
	Length (L)
	Breadth (B)
	Depth (d)
	Max Cargo (m^3)
	Ship Design Speed (kt)
	M/e Power (kW)
	Built Year
	IMO
	Owner
	Operator
	Flag
	Classification
	Built Country
Port	Arr. Port Max. DWT
	Dep. Port Max DWT
	Arr. Port Max L, B, d (m)
	Dep. Port Max L, B, d (m)
	Dep. Port -Arr. Port Name
Route	Dep. Port – Arr. Port Distances
Predicted	Ship Draught Rate (%)
Train Target	Average Voyage Speed (kt)

Deep Learning Output Layer	
Data Type	**Item**
Ship	Average Voyage Speed (kt)

c. Port staying time learning model

The later stage from the average voyage speed learning model is to evaluate port staying time spent by ship in Japan port. The port staying time needed for the ship operating from Australia to Japan can be drawn by the recorded arrival date in Japan and recorded departure date from Japan. The arrangement of items for the input layer and output layer is shown below in Table 6.3.

Table 6.3 DL Input and Output Layer in Port Staying Time Learning Model

Deep Learning Input Layer	
Data Type	**Item**
Ship	DWT
	Length (L)
	Breadth (B)
	Depth (d)
	Max Cargo (m^3)
	Ship Design Speed (kt)
	M/e Power (kW)
	Built Year
	IMO
	Owner
	Operator
	Flag
	Classification
	Built Country
Port	Arr. Port Max. DWT
	Dep. Port Max DWT
	Arr. Port Max L, B, d (m)
	Dep. Port Max L, B, d (m)
	Dep. Port -Arr. Port Name
Route	Dep. Port – Arr. Port Distances
Predicted	Ship Draught Rate (%)
Predicted	Average Voyage Speed (kt)
Train Target	Port Staying Time (h)

Deep Learning Output Layer	
Data Type	**Item**
Ship	Port Staying Time (h)

For all of the steps, the hyperparameters must be set as priors to optimize the model by minimizing the cost function of learning from the dataset. The hyperparameters used to generate the three distinct deep learning models are shown in Table 6.4

Table 6.4 Deep learning hyperparameters

Nodes in Hidden Layer	20
Hidden Layers	40
Activating Function	Max Out Function
Drop Out Rate	0.01
L_1 Regularization	0.001
L_2 Regularization	0.001

6.2 Result of Shipowner Model

In the shipowner model, we estimated the draft rate, average service speed, and time in port using deep learning analysis. As an example, the results of the three distinct deep learning models are shown in below Table 6.5. Hence, the results of the deep learning errors are illustrated in Table. 6.6.

Table 6.5 Example of prediction result by deep learning

Freight Shipping Route	Ship	Draft rate	Average Voyage speed (loaded)	Average Voyage speed (unloaded)	Staying Time (Loaded)	Staying Time (Unload)
Port Hedland -Chiba —One round Trip	A	88%	11.6Knot	11.7Knot	3.1 days	1.9 days
	B	90%	11.3Knot	12.4Knot	4.1 days	1.9 days
	C	85%	11.0Knot	11.6Knot	3.4 days	1.6 days

As shown in Table 6.6, the average draft rate error using the deep learning method is 3.4% for draught rate, 0.2% knots for average service speed, and 0.9 days for port staying time.

Table 6.6 Result of the shipper model

Method	Draught Rate	Average Service Speed	Port Staying Time
Deep learning	3.4%	0.2 knots	0.9 days

We next used deep learning estimation results to calculate the number of shipment days, amount of cargo, and shipment cost. Shipment days were calculated by considering the route distance, navigation speed, and time in port. The cargo transport volume was calculated based on the method developed by Kigure et al. [147], and the shipment cost was calculated using the method from Aoyama et al. [148] using the generated data.

Table 6.7 Examples of estimation results by each owner using prediction results

Cargo shipping request	Ship	Availability	Annual Cost	Annual Transport Volume
2014 Port Hedland -Chiba Iron ore	A	No	-	-
	B	Yes	\$2,115 x10^4	147 x10^4 ton
	C	Yes	\$2,570 x10^4	130 x10^4 ton

6.3 Evaluation of Shipowner Model

To confirm the effectiveness of the shipowner model, we compared the standard deviation of the estimation result of deep learning analysis with that of the response surface method.

6.3.1 Comparison using response surface method

To confirm the draft rate, we used the response surface method; like a deep learning analysis method, the following input and output were set:

- Input: ship DWT, length, breadth, depth, draft, service speed, horsepower, year built, the distance between routes, operator, shipyard, maximum draft, arrival limit, departure limit, new construction shipbuilding price index, and constraints of loading and unloading ports e.g. Max DWT, max LOA (m), B (m), and D (m)
- Output: draught rate during navigation (loading and unloading), etc.

As shown in Table 6.8, the average draft rate error using the response surface method is 5.9%, higher than the result using deep learning analysis.

Table 6.8 Comparison of deep learning vs response surface

Method	Draught Rate	Average Service Speed	Port Staying Time
Deep learning	3.4%	0.2 knots	0.9 days
Response Surface	5.9%	-	-

6.3.2 Threshold of the estimation

The draft rate, average service speed, and time in port are different even when the same ship operates on the same route. In this paper, the standard deviation of such a case is set as the threshold of the estimation. The threshold is also known in Table 6.9. These are calculated by using the actual data of bulk carriers that operate between Australia and Japan from 2013 to 2015. As shown in Table 6.9, the estimation result using deep learning is better than the threshold although that of the response surface method is worse.

Table 6.9 Comparison of deep learning, response surface, and threshold

Method	Draft Rate	Average Service Speed	Port Staying Time
Deep learning	3.4%	0.2 knots	0.9 days
Response surface	5.9%	-	-
Threshold	3.5%	0.9 knots	1.2 days

Chapter 7

Operator Model

7.1 Development of Operator Model

The operator model collects estimation results from the shipowner model. The procedure to determine ship allocation is as follows:

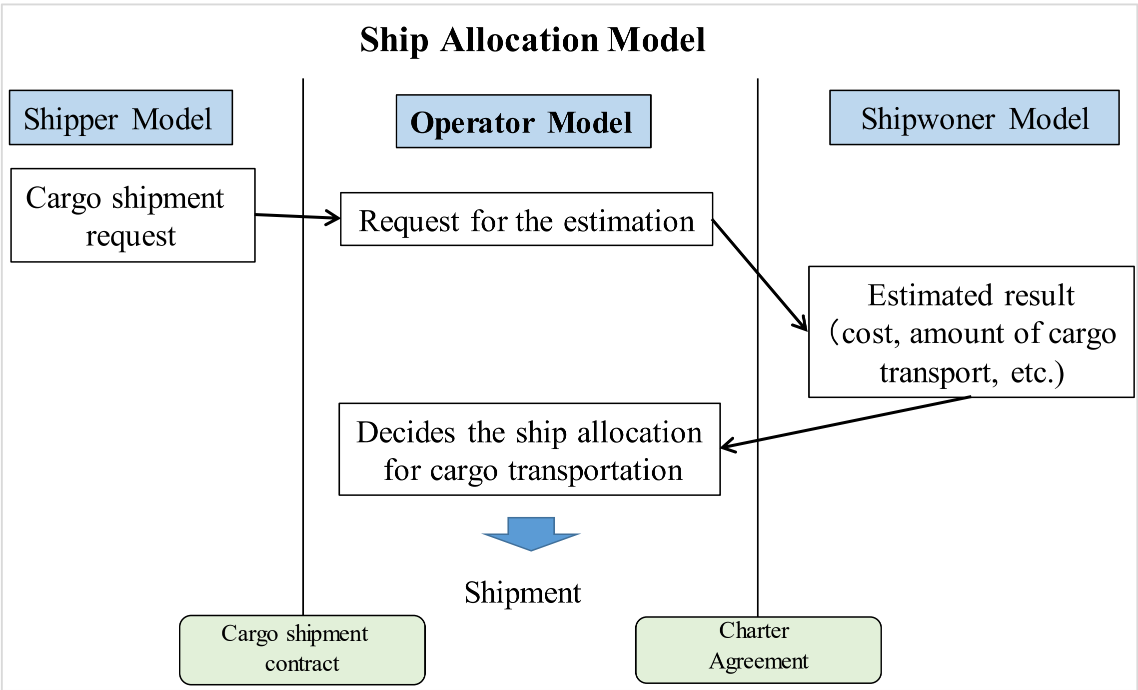


Fig. 7.1 Illustration of Operator Model

(1) Calculate the total cost and cargo volume

As shown in Table 7.1 (1), shipowners bid for all shipment requests (Ships A–D) based on each selected route (Routes A–B) from Shippers A and B. The cost per unit transport volume was calculated by considering the total operation cost and total transportation volume *t*. Operation cost-benefit is calculated using the following formula described in Equation (5) until Equation (10).

$$FC_i(ton) = Eng.Power(kW) \cdot SFOC\left({}^{g}/_{kWh}\right) \cdot Sailing.Days(day) \cdot \frac{24({}^{h}/_{day})}{1000{,}000({}^{ton}/_{g})} \qquad (5)$$

Where *Eng Power (kW)* is the ship maximum engine power, $SFOC_i$ $({}^{g}/_{kWh})$ is assumed ship fuel oil consumption, *Sailing Days*$_i$ *(day)* is the ship sailing days, and FC_i *(ton)* is the total fuel consumption of each ship.

$$Fuel.Cost.at.Sea_i(US\$) = FC_i(ton) \cdot Fuel\ \ Price\left({}^{US\$}/_{ton}\right) \qquad (6)$$

Where FC_i *(ton)* is the total fuel consumption of each ship, *Fuel Price* $({}^{US\$}/_{ton})$ is the average fuel price per ton of selected year, and *Fuel Cost at Sea*$_i$ *(US$)* is the total fuel cost at sea of each ship.

$$Fuel.Cost.at.Port_i(US\$) = \frac{Fuel.Cost.at.Sea_i(ton)}{Sailing.Days(day)} \cdot 0.1 \cdot \frac{Port\ \ Staying\ \ Time_i(h)}{24({}^{h}/_{day})} \qquad (7)$$

Where *Fuel Cost at Sea*$_i$ *(US$)* is total fuel cost at sea of each ship, *Sailing Days*$_i$ *(day)* is the ship sailing days, *Port Staying Time*$_i$ *(h)* is the ship port staying time, and *Fuel Cost at Port*$_i$ *(US$)* is total fuel cost at the port of each ship. Within this formula, assumed 10% of port staying time, such as waiting to enter the port area or exit the port area, a ship used its fuel before or after loading/unloading activity.

$$Total.Fuel.Cost_i(US\$) = Fuel.Cost.at.Port_i(US\$) + Fuel.Cost.at.Sea_i(US\$) \qquad (8)$$

Where *Fuel Cost at Port*$_i$ *(ton)* is total fuel cost at the port, *Fuel Cost at Sea*$_i$ *(ton)* is total fuel cost at sea, and *Total Fuel Cost*$_i$ *(US$)* is total fuel cost in the single laden voyage.

$$Cargo\ \ Income_i(US\$) = Transportable_i(times) \cdot V_i(ton) \cdot Fair_i\left({}^{US\$}/_{ton}\right) \qquad (9)$$

Where *Transportable*$_i$ *(times)* is transportable times of a ship to a certain port, V_i *(ton)* is the cargo volume estimation from Equation 2, *Fair*$_i$ $\left({}^{US\$}/_{ton}\right)$ is freight rates per ton LNG cargo, and *Cargo Income*$_i$ *(US$)* is the estimated total income earned.

$$Cost\text{-}benefit_i(US\$) = Cargo\ \ Income_i(US\$) - \ \ Total.Fuel.Cost_i(US\$) \qquad (10)$$

Where *Cargo Income*$_i$ *(US$)* is total profit earned by a ship, *Total Fuel Cost*$_i$ *(US$)* is total fuel cost in the single laden voyage, and *Cost-benefit*$_i$ *(US$)* is estimated profit earned by a ship.

(2) Calculate the standard deviation

Based on the cost per unit transport volume from the previous step, the standard deviations of some ships were calculated for each route.

$$Deviation\ \ Value_i = \frac{Average\ \ of\ \ Cost\text{-}benefit_i(US\$)}{Standard\ \ Deviation\ \ of\ \ Cost\text{-}benefit_i(US\$)} \qquad (11)$$

Where *Average of Cost-Benefit*$_i$ *(US$)* is the average cost-benefit of a certain route, *Standard Deviation of Cost-Benefit*$_i$ *(US$)* is the standard deviation of a certain route, and *Deviation Value*$_i$ is the deviation value of the selected route.

The average and standard deviation of cost-benefit are including the operation which has 0 as cost-benefit (transportable times is 0). This is based on the idea that the value of routes which possible to be served by fewer ships is higher rather than the route which possible to be served by more ships. Therefore, the certain route which possible to be served by 1 ship is more valuable rather than another route which possible to be served by 10 ships. The deviation value is an index for judging which ship is good for transporting a given cargo type on a certain route. Table 7.1 (2) shows a sample calculation of deviation values.

(3) Ship assignment

Ship assignment decides which kind of ship to charter regularly by considering standard deviation values. Ships with the highest standard deviation values are assigned to a shipment on the selected route. For example, as shown in Table 7.1 (2), Ship B is assigned to Route A2.

Table 7.1 Ship allocation process

(1) Calculate the total cost and cargo volume						
Shipper	Route	Cargo Volume (t)	Ship A ($/t)	Ship B ($/t)	Ship C ($/t)	Ship D ($/t)
A	A1	3.5×10^6	14.8	14.1	16.9	19.9
	A2	2.0×10^6	14.7	13.9	16.4	19.4
B	B1	4.7×10^6	13.6	13	15.1	18.3
	B2	6.0×10^6	13.1	12.6	14.5	18.2
(2) Calculate the standard deviation and ship assignment						
Shipper	Route	Cargo Volume (t)	Deviation Value			
			Ship A	Ship B	Ship C	Ship D
A	A1	3.5×10^6	57.2	60.31	47.9	34.59
	A2	**2.0×10^6**	56.64	**60.43**	45.58	34.35
B	B1	4.7×10^6	56.81	59.74	49.51	33.92
	B2	6.0×10^6	56.84	59.12	50.45	33.58
(3) Recalculate the amount of cargo shipment request						
Shipper	Route	Cargo Volume (t)	Ship A ($/t)	Ship B ($/t)	Ship C ($/t)	Ship D ($/t)
A	A1	3.5×10^6	14.8	—	16.9	19.9
	A2	**0.6×10^6**	14.7	—	16.4	19.4
B	B1	4.7×10^6	13.6	—	15.1	18.3
	B2	6.0×10^6	13.1	—	14.5	18.2

(4) Recalculate the amount of cargo shipment requests

When a shipment is assigned to a selected route, as shown in step 3, the remaining cargo shipment is calculated by subtracting the amount of cargo shipment requested by the shipper. Therefore, after assignment, the amount of cargo to ship is updated and Steps (1–3) are repeated until all cargo is successfully transported.

7.2 Evaluation of Ship Allocation Model

The confirmation of the ship allocation model proposed in this study is evaluated by checking the reproducibility of the proposed models i.e. shipper, shipowner, and operator. The evaluation of the ship allocation model is described in detail as follows:

7.2.1 Problem definition

To evaluate the reproducibility of the proposed model, we simulated ship allocation. Trade condition (i.e. trade volume, trade routes, and fuel price), allowable ship specification, number of ships, and port constraints were set as inputs. Then, the result of the allocation model was compared with actual ship allocation. Moreover, all information for simulating ship allocation was extracted from the MLDB. Operation from Australia to Japan in 2014 was taken as an example.

7.2.2 Simulation results

Fig.7.2 showed the ship allocation for each shipper using the proposed model. As explained in the previous section, there are four Shippers (A–D). The vertical axis shows the number of operations (shipments). The horizontal axis shows ship size (in 10^3 DWT). Using cluster analysis, the ships are grouped into six clusters: 100, 170, 210, 230, 250, and 300 (10^3 DWT). The actual and simulation results are shown together to validate the proposed model.

7.2.3 Discussions

As shown in Fig. 7.2, the simulation results generally agree with actual conditions. In this section, the allocation process was evaluated:

(1) Ports Constraints

In the MLDB, port constraints were generated in two steps:

- Step 1: Extract the constraints from port information. First, port constraints were obtained from port information. However, some constraints were unavailable or did not match actual conditions.
- Step 2: Modification using operating data. Port constraints in Step 1 were compared with actual operations. When the two did not match or when some

constraints were not available, we modified the port constraints using operating ship specifications.

- Some port constraints are shown in Table 7.2, in which white represents data from Step 1 and gray represents data modified based on the actual operation (Step 2).

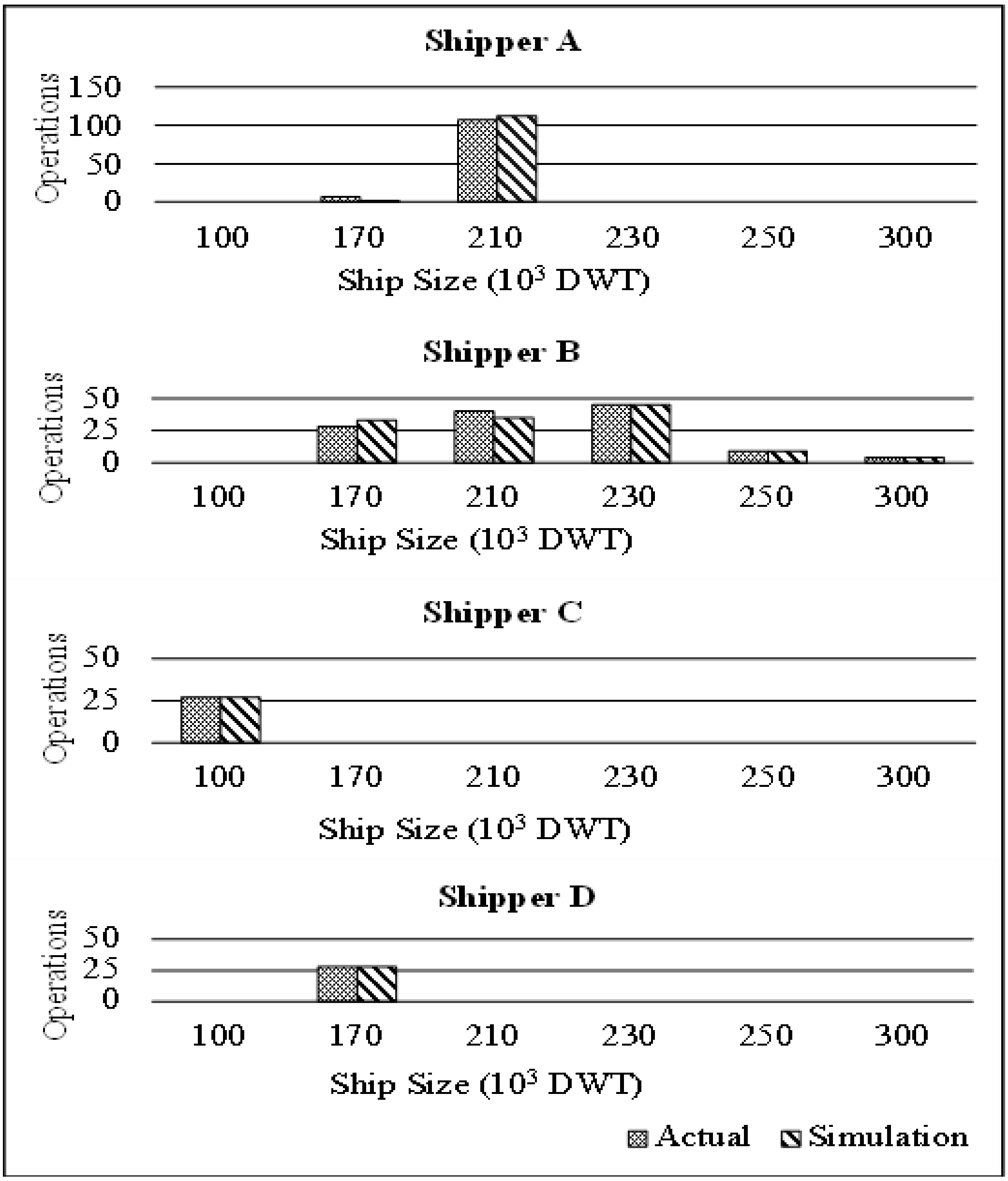


Fig. 7.2 Comparison of actual and simulation results

(2) Ship Specifications

By examining the actual operation extracted from the MLDB, we identified typical ship specifications for each ship size, which are shown in Table 7.3.

Table 7.2 Port constraints

Shipper	Port Name	DWT	L (m)	B (m)	d (m)
A	Fukuyama	220,000	300	50	18
	Chiba	220,000	300	50	18
	Mizushima	260,000	340	50	18
	Kawasaki	220,000	340	50	18
B	Kashima	300,000	340	60	19
	Kisarazu	300,000	330	60	19
	Oita	400,000	450	60	25
C	Nagoya	110,000	300	43	16
	Tobata	160,000	327	43	16
	Wakayama	160,000	300	43	14
D	Higashi-Harima	180,000	330	47	17
	Kure	276,000	360	45	18
	Himeji	257,000	335	47	16

Table 7.3 Typical ship specifications

DWT	L (m)	B (m)	D (m)	d(m)	HP
106,507	255	43	19	13	16,680
177,855	292	45	25	18	22,920
210,036	300	50	25	18	21,808
229,013	320	54	24	18	30,499
250,813	330	57	25	18	29,789
297,736	325	55	29	21	30,808

(3) Allocation process

By Ship, the allocation was started from Shipper C, because its port constraints were most severe. Hence, $B \leq 43$ m and $d \leq 14$ m became active constraints, and ships with 100,000 DWT ($B \leq 43$ m) were selected. Next, ships for Shipper D were allocated where B (m) should be less than 45 m. After that, ships for Shipper A were allocated because its port constraints were more severe than those of Shipper B. Finally, the remaining ships were allocated to Shipper B.

Based on Fig. 7.2, we see that the simulation results for all shippers generally agreed with the actual results. Moreover, as shown in Fig. 7.2, Shipper A mostly used 210,000 DWT ships for their operation. Shipper B used various kinds of ships (170,000–300,000 DWT). Since the simulation of ship allocation matched actual ship allocation, and the ship specifications agreed with port constraints, we confirmed the effectiveness and reproducibility of the proposed models.

Chapter 8

Simulations

8.1 Case Studies

To develop the new ship allocation and examine which specification of the ship is effective, by considering the present condition based on the data extracted from MLDB and the future scenario such as fuel price, fuel oil consumption, etc as shown in Fig. 8.1, we execute the following simulations:

- Examination of supply-demand balance
- Examination of supply-demand balance with effective ship size
- Examination of supply-demand balance with the influence of fuel efficiency on demand

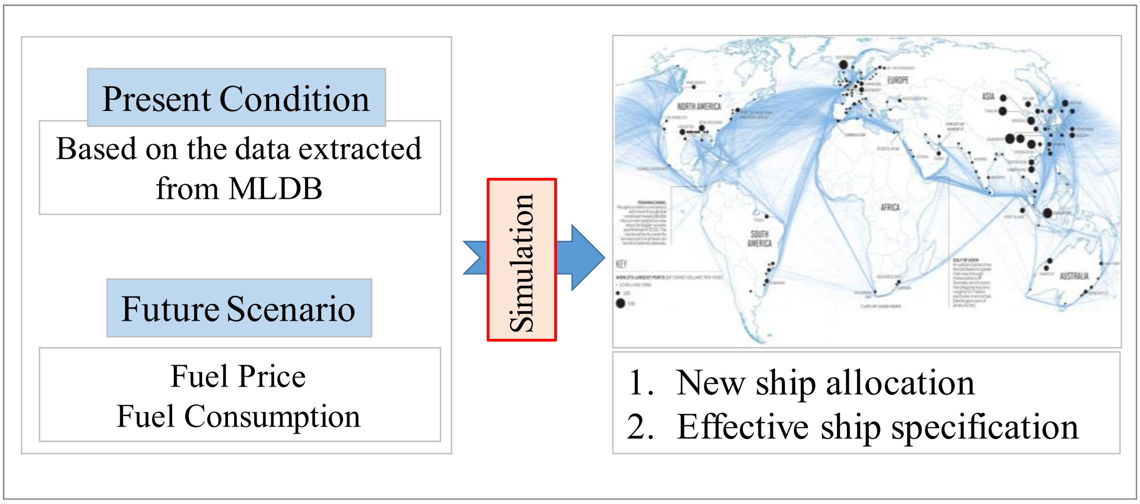


Fig. 8.1 Illustration of the simulation

8.2 Examination of supply-demand balance

8.2.1 Problem definition

In this study, the supply-demand balance of bulk carriers that operated between Australia and Japan in 2014 carrying iron ore was examined. In contrast with the simulation conducted in Section 7.2, we carried out ship simulation without constraints, meaning that there was no limit to the number of ships per year of operation. The operator shipped cargo shipments with freely selectable ships. In this restricted example (using the actual number and ship types used in 2014 between Australia and Japan), the simulation result is defined as supply. In the unrestricted case, the simulation result is defined as demand and illustrated as shown in Fig.8.2.

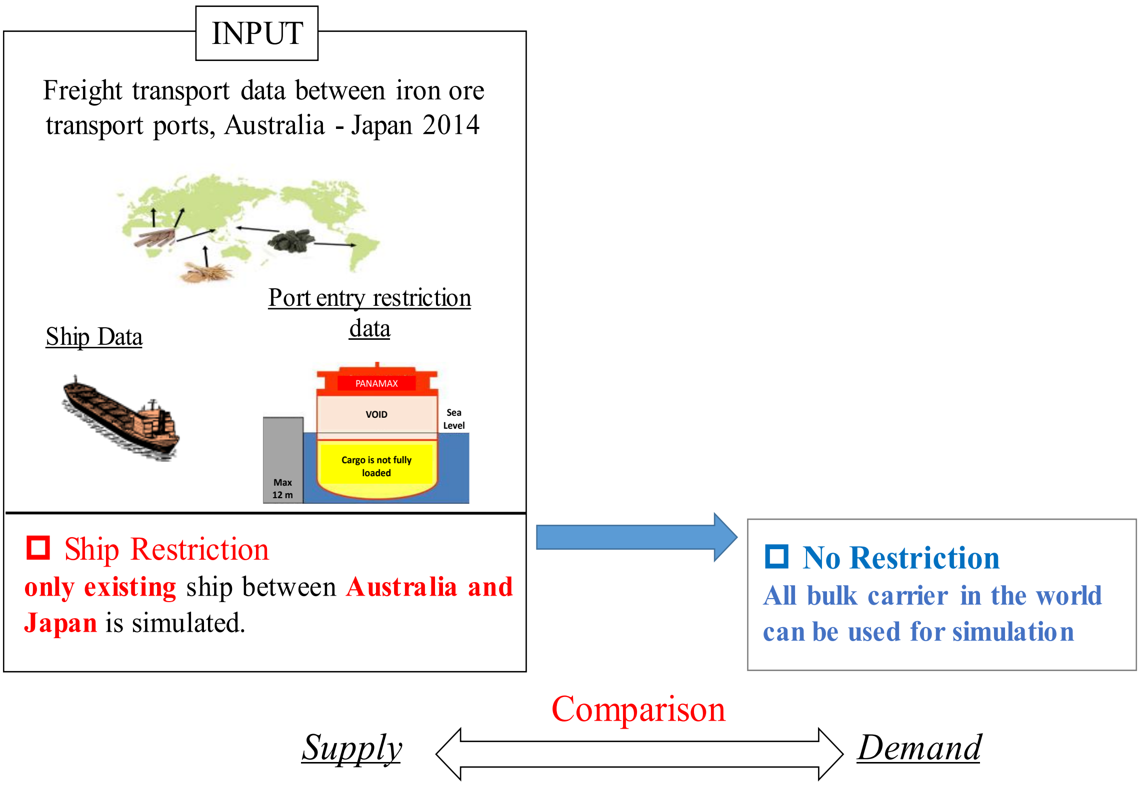


Fig. 8.2 Illustration of the supply-demand balance simulation

8.2.2 Simulation results

Fig. 8.3 shows the difference in ship allocation results using constraints (supply) and without constraints (demand). These results were compared to evaluate ship supply-demand balance and determine the kind of ship likely to be in demand in the future. The vertical axis is the number of operations (shipments). The horizontal axis is the size of the ships (in 10^3 DWT).

8.2.3 Discussions

As shown in Fig. 8.3, without constraints, the allocation of 210,000, 250,000, and 300,000 DWT ships increased. However, the allocation of 170,000 and 230,000 DWT ships decreased. Therefore, 170,000 and 230,000 DWT ships were not very competitive for shipments between Australia and Japan. Meanwhile, there were an insufficient supply of 210,000, 250,000, and 300,000 DWT ships. Hence, these ships are competitive for shipments from Australia to Japan and are expected to be in demand in the future. This result can be understood from the port constraints shown in Table 7.2.

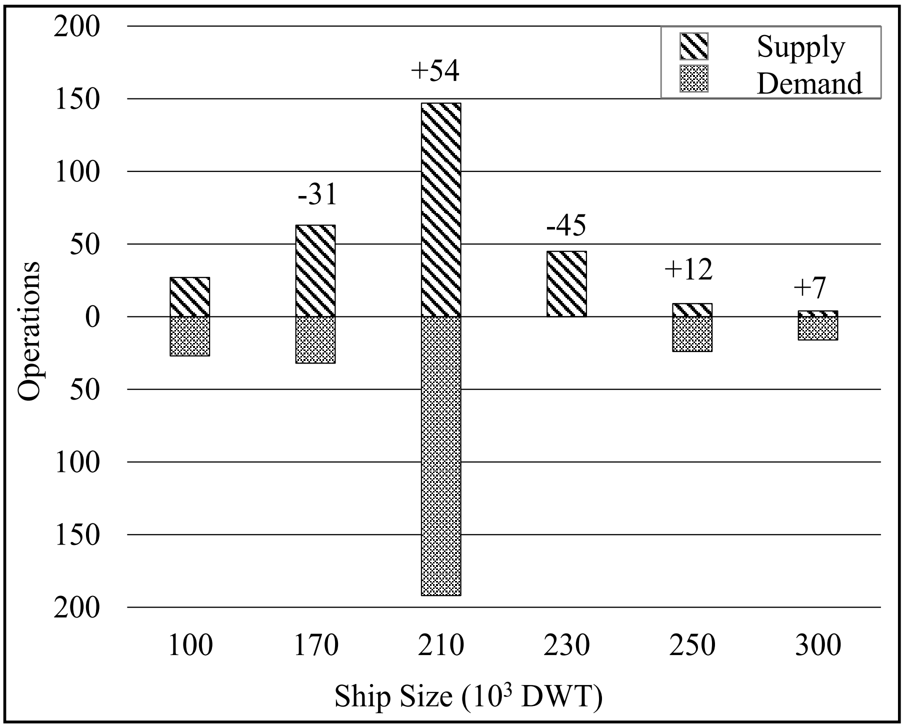


Fig. 8.3 Comparison of actual and simulation results

8.3 Examination supply-demand balance with new ship specification

8.3.1 Problem definition

Based on the discussion in the previous section, 210,000, 250,000, and 300,000 DWT ships are expected to be in demand. Thus, it is necessary to examine the influence of ship size on allocation and examine the distribution of ships (210,000, 250,000, and 300,000 DWT) for which increased demand is expected on the selected route (Australia to Japan). In this simulation, we accounted for the depreciation value of a new ship and ignored the depreciation value of existing ships. The useful life of a ship was set to 15 years [149], and depreciation value was calculated based on reference [150].

8.3.2 Simulation results

The principal particulars of 210,000, 250,000, and 300,000 DWT ships are shown in Table 8.1. Using the proposed method, we examined the principal particulars of the ships. Moreover, by conducting this simulation we could identify the number of routes and ships that could be allocated for the selected route.

Table 8.1 Ship specifications

DWT	L (m)	B (m)	D (m)	d(m)	HP
106,507	255	43	19	13	16,680
177,855	292	45	25	18	22,920
210,036	300	50	25	18	21,808
229,013	320	54	24	18	30,499
250,813	330	57	25	18	29,789
297,736	325	55	29	21	30,808

As shown in Fig. 8.4, 300,000 DWT ships were in demand on a single route. In contrast, 250,000 DWT and 210,000 DWT ships can be expected to be in demand on multiple routes.

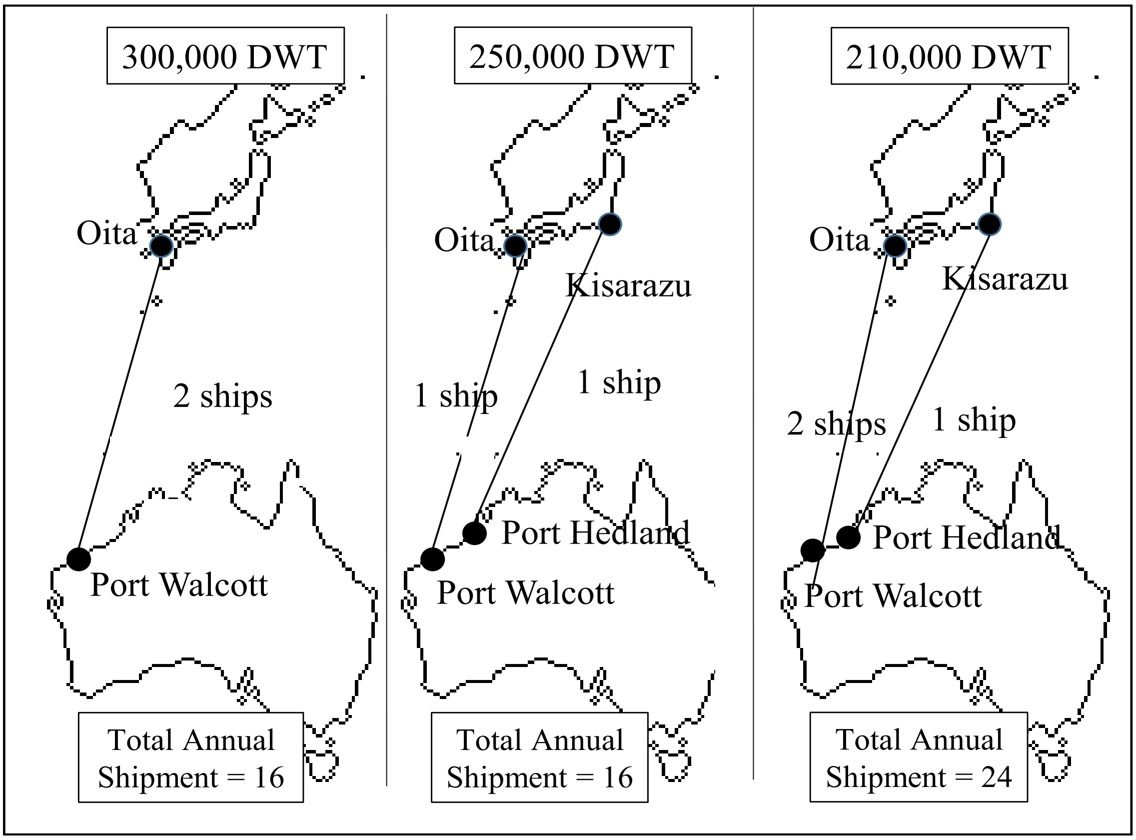


Fig. 8.4 Ship distribution by size

8.3.3 Additional simulations

From the simulation result shown in Fig. 8.4, 250,000 DWT and 210,000 DWT ships are in demand. To clarify which is preferred, we executed an additional simulation in which the fuel efficiency of 250,000 DWT and 210,000 DWT ships increased by 10%. The results are shown in Fig. 8.5. In Fig. 8.5, the number of allocated ships and routes for 210,000 DWT ships increased rapidly (five additional ships and three additional routes) as shown in Table 8.2. However, only two additional ships and one additional route were called for in the 250,000 DWT ship simulation. Therefore, 210,000 DWT ships have the highest potential in iron ore transportation between Australia and Japan.

This result is affected by the port constraints shown in Tables 7.2 and 8.3. Ship of 210,000 DWT can enter all main ports in Australia and Japan and 250,000 DWT ships cannot enter some key ports.

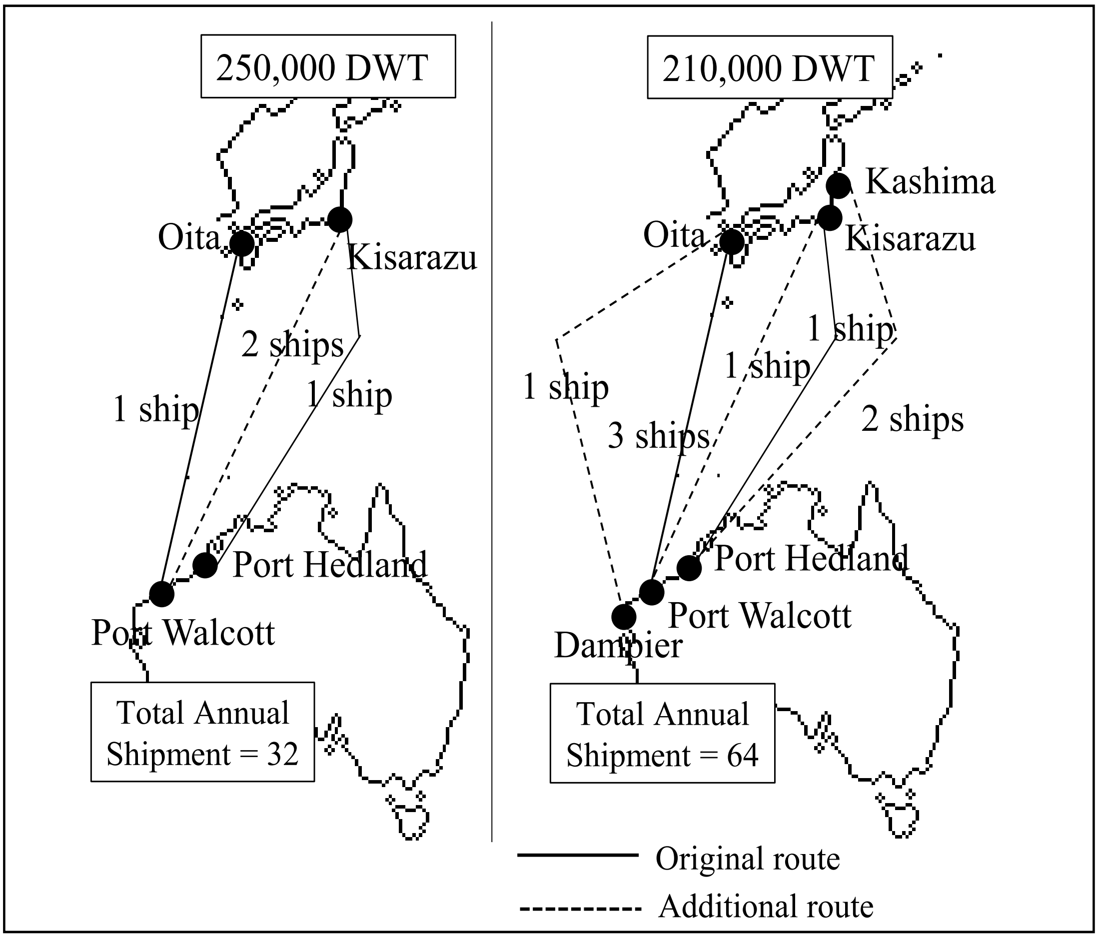


Fig. 8.5 Ship distribution when fuel efficiency increases by 10%

Table 8.2 Results of ship distribution by efficiency increases 10%

Ship size (DWT ton)	Route		Ship	
	0%	10%	0%	10%
250000	2	**3**	2	**4**
210000	2	**5**	3	**8**

Table 8.3 Port constraints in Australia

Port Name	DWT	L (m)	B (m)	d (m)
Port Hedland	260,000	330	55	19.2
Dampier	250,000	340	55	19.5
Port Walcott	340,000	335	60	19.5
Parker Point	220,000	300	50	18.5
Esperance	220,000	300	50	18

8.4 Examination of Supply-Demand Balance with the influence of Fuel Efficiency

8.4.1 Problem definition

To examine the influence of fuel efficiency on ship demand and to draw future development targets, we simulated increasing fuel efficiency by 5%, 10%, and 15%. A ship with no fuel efficiency change is defined as S_0. Ships with fuel efficiency increases of 5%, 10%, and 15% are denoted by S_1, S_2, and S_3, respectively. The 210,000 DWT ships were simulated as they were the most competitive. As in the simulation in Section 8.3, we considered depreciation values. To evaluate ship effectiveness, we compared the simulation result (ship replacement) with actual ship allocation.

8.4.2 Simulation results

Table 8.4 shows the simulation result of ship allocation on the intended route (Australia to Japan) after modifying the fuel efficiency of S_0, S_1, S_2, and S_3. The table shows the number of allocated routes, operations, and ships.

Table 8.4 Replacement 210,000 DWT ships

	Allocated Routes	Operations	Ships
S_0	3	40	5
S_1	4	47	6
S_2	5	64	8
S_3	10	125	16

As shown in Table 8.4, the number of operations and ships for S_2 and S_3 increased greatly. However, only a small increase occurred for S_0 and S_1. Therefore, we focused on ships S_2 and S_3. The simulation result of increasing fuel efficiency by 10% and 15% is shown in Fig. 8.6.

The vertical axis shows an operation number. The horizontal axis shows the ship size. In Fig.8.6, the simulation results of ships S_2 and S_3 are compared with actual ship allocation.

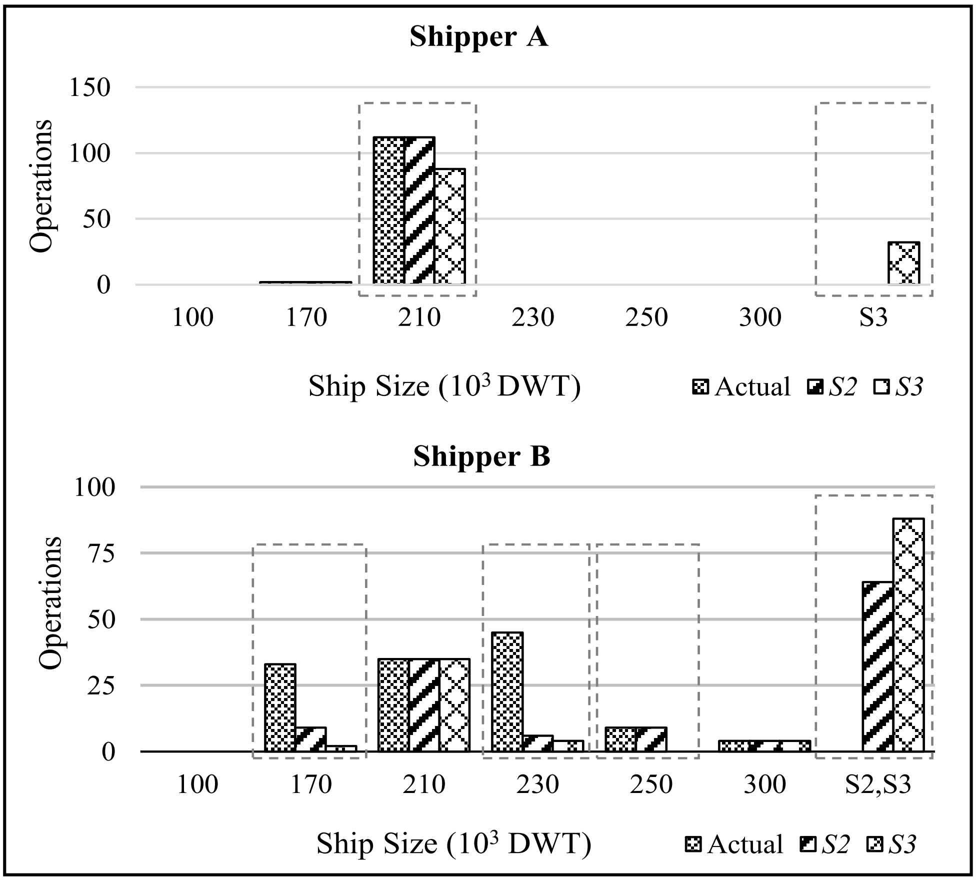


Fig. 8.6 Comparison of simulation result by increasing fuel efficiency 10% (S_2) and 15% (S_3) with actual ship allocation

7.4.3 Discussions

As shown in Fig. 8.6, when fuel efficiency increases by 10% (S_2), the replacement by 210,000 DWT ships of 170,000, 210,000, 230,000, and 250,000 DWT ships has not occurred for Shipper A, since the simulation result shows the same number of operations compared with actual ship allocation. Hence, Shipper A will not buy a new ship. However, in contrast, Shipper B will buy a new ship because the operation number increased by 64.

By improving fuel efficiency by 15%, Shippers A and B might each buy a new ship, as significant replacement occurred for both shippers. For Shipper A, the number of the operation increased by 35, and for Shipper B, the number of the operation increased by 90. Moreover, as shown in Table 8.4, the total operation number is increased from 64 to 125, and the number of ships from 8 to 16.

In summary, using the proposed model, we simulated ship supply and demand. Moreover, the principal particulars of the ships expected to be in demand were identified. In addition, we obtained the impact of fuel efficiency on ship demand.

Chapter 9

Conclusions and Future Tasks

9.1 Conclusions

This study presents a ship basic planning support system using maritime logistics big data. Recently, the era of shipping records big data has been started along with mandatory digitalized ship movement with AIS (Automatic Identification System). Substantial data amount is generated by AIS, such as the ship's unique identification of international maritime organization number (IMO number), position, course, speed, and destination. Moreover, maritime logistics big data, such as ship and port specification data, route data, international trade data, and data provided by AIS, are currently available and can be used. Based on that, some studies have been conducted on big data utilization to improve ship construction, ship operations, and ship maintenance. However, studies regarding the utilization of such maritime logistics big data for ship basic planning or ship design are limited. Therefore, in this study, we developed marine logistics database to realize the ship allocation by conducted a ship allocation model. The MLDB consists of the latest marine logistics data, i.e., operation information from AIS, ship, port, route, and international trading information. The data are managed, integrated, and structured to derive valuable insights from information buried in marine logistics data. Based on the MLDB, the extracted data are used to construct a ship allocation model. The ship allocation model constructed aims to replicate the actual ship allocation.

The actual shipping market has the essential affection in the actual ship allocated. The ship allocation model consists of three specified models, shipper model, shipowner model, and operator model. The shipper model creates the clusters of exporter and importer ports, and its cargo demand is to be transported in the selected year. Later, the demand estimated will be used in the operator model as a freight transport request to the shipowner model. Following the request, the shipowner bid the cost and transportable volume on the route where the shipowner ship operates. To finalize the charter contract between operator and shipowner, the operator will choose the served bid in economical consideration and create ship allocation. The next important stage of this study has conducted the simulation. The simulation was examined supply-demand balance, examined effective ship size, and the influence of fuel efficiency on demand.

The conclusions of this dissertation are summarized as follows:

1. Marine Logistic Database (MLDB) manages the available data from maritime industry big data.
2. The error cleaning is performed to maintain the quality of data to developed MLDB and validation is applied to all the data.
3. The ship allocation model composed of distinct shipper, shipowner, and operator models was developed. This model was effective in estimating ship supply and demand, the influence of ship size and fuel efficiency on ship allocation, and the principal particulars of ships for which demand is expected to increase.
4. The results of the shipper model show that the cluster results compared between 2014 and 2017 remained the same. Moreover, based on the exporter's point of view and importer's point of view, there are no shipper cluster has changed. The only change is the amount of cargo transport volume.
5. By using the proposed model, we confirmed the reproducibility of the ship allocation model. The supply-demand balance, effective ship specifications, and influence of

ship efficiency on demand could be realized using the ship allocation model proposed in this study.

6. The ship with the most competitive demand on the selected route (Australia to Japan) for iron ore was the 210,000 DWT ship. In the future, we plan to automate the ship allocation model to simulate worldwide ship allocation for various cargos, ship sizes, and ship types.

9.2 Future Works

In this dissertation, the constructed MLDB, ship allocation model, and simulations deliver acceptable results, several improvements are possible to be acquired in the future. Followings are the prospective future works that can be adopted to complement this study:

1. The target ship of this study is the bulk carriers which are limited for iron ore from Australia to Japan, and we confirmed the reproducibility of the ship allocation of it. It is a big challenge to adopt this method for all the bulk carriers that operated from all over the world with the other cargo type to evaluate the characteristics of the ship operation, the real ship allocation, and predict the future ship specification in demand to be expected.
2. To be more effective, the automation system for the support system of ship basic planning support using MLDB should be executed.
3. For the allocation process accuracy, the more detailed port data will give more advantages. Because in this study the port data information is limited in the port-level specification. With detailed port data information, it will strongly increase the port limitation accuracy of the ship allocation process.
4. The future scenario inputted to the developed ship allocation model has no specific period. Therefore, it will be advantageous for the ship allocation to be dynamically applicable for the specifically targeted variable in the long-span period, for example in the next 10 years, 20 years, or further. For that consideration, the mathematical correlation of related variables and suitable approach is necessary to be confirmed.

Bibliography

[1] Jović, M. et al. (2019) 'The Transition of Croatian Seaports into Smart Ports', in *42nd international convention on information and communication technology, electronics and microelectronics*. Opatija, Croatia, pp. 1618–1622.

[2] Marshall University (2019) 'Big Data, Big Data Analytics'. Available at: https://www.marshall.edu/bigdata/ (Accessed: 1 June 2021).

[3] Zaman, I. et al. (2017) 'Challenges and opportunities of big data analytics for upcoming regulations and future transformation of the shipping industry', *Procedia Engineering. Elsevier B.V.*, 194, pp. 537–544. DOI: 10.1016/j.proeng.2017.08.182.

[4] Brouer, B. D., Karsten, C. V. and Pisinger, D. (2016) 'Big Data Optimization: Recent Developments and Challenges (Big Data Optimization in Maritime Logistics)', in. *Springer International Publishing*.

[5] Big Data in the Shipping Industry (2018). https://www.morethanshipping.com/big-data-in-the-shipping-industry/ (Accessed: 1 June 2021).

[6] Acharjya, D. P. and Ahmed, K. (2016) 'A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools', *International Journal of Advanced Computer Science and Applications*, 7(2). DOI: 10.14569/ijacsa.2016.070267.

[7] Voss, S., Sebastian, H.-J. and Pahl, J. (2017) 'Introduction to Intelligent Decision Support and Big Data for Logistics and Supply Chain Management Minitrack', in *Proceedings of the 50th Hawaii International Conference on System Sciences*, pp. 1516–1517. DOI: 10.24251/hicss.2017.159.

[8] Heilig, L., Schwarze, S. and Voss, S. (2017) 'An Analysis of Digital Transformation in the History and Future of Modern Ports', in *50th Hawaii International Conference on System Sciences*. Hawaii, pp. 1341–1350. DOI: 10.24251/HICSS.2017.160.

[9] What is big data analytics? Everything you need to know | InfoWorld (2018). Available at: https://www.infoworld.com/article/3220044/what-is-big-data-analytics-everything-you-need-to-know.html (Accessed: 1 June 2021).

[10] Nita, S. L., and Mihailescu, M. I. (2017) 'Importance of Big Data in Maritime Transport', *Scientific Bulletin of Naval Academy*, XX (1), pp. 2–6. DOI: 10.21279/1454-864x-17-i1-079.

[11] Mauro, A. De et al. (2014) 'What is Big Data? A Consensual Definition and a Review of Key Research Topics', *in AIP conference proceedings*, pp. 97–104. DOI: 10.13140/2.1.2341.5048.

[12] Hammad, K. A. I. et al. (2015) 'Big Data Analysis and Storage', in *International Conference on Operations Excellence and Service Engineering*. Orlando, Florida, USA, pp. 648–659.

[13] Taylor-Sakyi, K. (2016) Big Data: Understanding Big Data. Available at: http://arxiv.org/abs/1601.04602 (Accessed: 6 June 2021).

[14] Charles W. Chase, Jr., 2013." Using Big Data to Enhance Demand-Driven Forecasting and Planning," *Journal of Business Forecasting*, Volume 32, Issue 2, 2013.

[15] Hideyuki Ando, MTI (NYK Group), 2016. “The Application of Big Data for Ship Operational Efficiency,” *Japan –Norway Workshop Future Technology and Finance in the Maritime Sector*, 12th February 2016

[16] Martin Kits van Heyningen., 2015. “Ship Insight-The Buzz of Big Data,” The Quarterly Journal |Summer.

[17] Hossein H. and Emmanuel S.S., 2015. “Forecasting with Big Data: A Review,” Ann. Data. Sci. (2015) 2(1):5–19 DOI 10.1007/s40745-015-0029-9

[18] Hossein H. and Emmanuel S. S., Forecasting with Big Data; A Review, Springer, pp. 5–19, 2015.

[19] IHS Maritime & Trade, Sea-web Ports. http://maritime.ihs.com/

[20] IHS Maritime & Trade, Sea-web Ships. http://maritime.ihs.com/

[21] IHS Fairplay Ports & Terminals Guide, CD-ROM, 2013

[22] United Nations Commodity Trade Statistics Database. https://comtrade.un.org/data/

[23] Market Intelligence Network, MINT. https://www.marketintelligencenetwork.com/

[24] Koga, S. (2015) Major challenges and solutions for utilizing big data in the maritime industry. Available at: http://commons.wmu.se/all_dissertationshttp://commons.wmu.se/all_dissertations/490 (Accessed: 3 June 2021).

[25] Mirović, M., Miličević, M. and Obradović, I. (2018) ‘Big data in the maritime industry’, Nase More, 65(1), pp. 56–62. DOI: 10.17818/NM/2018/1.8.

[26] SOLAS Guidance on Chapter V – Safety of Navigation - Regulation 20 (2002). Available at: http://solasv.mcga.gov.uk/regulations/regulation20.htm (Accessed: 23 May 2021).

[27] Available at: http://www.imo.org/en/KnowledgeCentre/IndexofIMOResolutions/Documents/MSC - Maritime Safety/333(90).pdf (Accessed: 25 May 2021).

[28] AISHub (2019) AIS data sharing and vessel tracking by AISHub. Available at: https://www.aishub.net/ (Accessed: 25 May 2021).

[29] Lensu, M. and Goerlandt, F. (2019) 'Big maritime data for the Baltic Sea with a focus on the winter navigation system', *Marine Policy. Elsevier Ltd*, 104(January), pp. 53–65. DOI: 10.1016/j.marpol.2019.02.038.

[30] SOLAS Guidance on Chapter V – Safety of Navigation - Annex 17 (2002). Available at: http://solasv.mcga.gov.uk/Annexes/Annex17.htm (Accessed: 6 August 2019).

[31] Watson, R., Lind, M. and Haraldson, S. (2017) 'Physical and Digital Innovation in Shipping: Seeding, Standardizing, and Sequencing', in *Proceedings of the 50th Hawaii International Conference on System Sciences* (2017). DOI: 10.24251/hicss.2017.579

[32] Bošnjak, R., Šimunović, L. and Kavran, Z. (2012) Automatic Identification System in Maritime Traffic and Error Analysis, Transactions on Maritime Science. Available at: https://hrcak.srce.hr/file/132144 (Accessed: 24 May 2021).

[33] Lázaro, F. et al. (2019) 'VHF Data Exchange System (VDES): an enabling technology for maritime communications', *CEAS Space Journal*, 11(1), pp. 55–63. Available at: https://link.springer.com/article/10.1007/s12567-018-0214-8.

[34] IALA (2019) VDES FAQ. Available at: https://www.iala-aism.org/technical/connectivity/vdesvhf-data-exchange-system/ (Accessed: 24 May 2021).

[35] Valčić, S., Žuškin, S. and Brčić, D. (2018) 'VHF Maritime Mobile Band - A New System to Declutter AIS Channels', Sea Technology, 59(7), pp. 24–27.

[36] Žuškin, S., Grubišić, N. and Sumner, M. (2015) 'Shipowner management in accordance with mutual agreement', Pomorstvo 29(1):69-74

[37] Trelleborg Marine Systems (2018) Use of big data in the maritime industry. Available at: https://www.patersonsimons.com/wpcontent/uploads/2018/06/TMS_SmartPort_InsightBee_Report-toGUIDE_01.02.18.pdf (Accessed: 24 May 2021).

[38] Ship Tracking (2019). Available at: http://www.freightmetrics.com.au/Marine/tabid/402/Default.aspx (Accessed: 6 August 2019).

[39] Wada Y, Hamada K, Hirata N, Yui, I, Arifin, M.D, Shohei, M, Kazutaka S: A study on system dynamics model for the demand forecasting of shipbuilding. *12th IMDC* (2), 167-182 (2015).

[40] Press, G. (2014). Big Data Definitions: What's Yours? Forbes. Retrieved June 27, 2015, from http://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-your s/

[41] Friedman, U. (2012, October 8). Big Data: A Short History. Foreign Policy. Retrieved July 8, 2015, from http://foreignpolicy.com/2012/10/08/big-data-a-short-history/

[42] IBM. (n.d.). What is Big Data (Unpublished master's thesis). Retrieved June 28, 2015, from http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

[43] H. Segercrantz, “BIG DATA & Big Savings for Maritime Ops,” Available at <http://www.marinelink.com/news/maritime-savings-data405723.aspx>, [Accessed June 8th, 2016]

[44] S. Shukla, V. Kukade and A. Mujawar, “Big Data Concept, Handling, and Challenges: An Overview,” *International Journal of Computer Applications* (0975 –8887) Volume 114 –No. 11, March 2015

[45] Press, G. (2014). 12 Big Data Definitions: What's Yours? Forbes. Retrieved June 27, 2015, from http://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-your s/

[46] Friedman, U. (2012, October 8). Big Data: A Short History. Foreign Policy. Retrieved July 8, 2015, from http://foreignpolicy.com/2012/10/08/big-data-a-short-history/

[47] IBM. (n.d.). The Four V's of Big Data. Retrieved June 29, 2015, from http://www.ibmbigdatahub.com/infographic/four-vs-big-data

[48] Ben Ayed, A., Ben Halima, M., Alimi, A. M., 2015, “Big Data Analytics for Logistics and Transportation”, *4th IEEE International Conference on Advanced Logistics and Transport*, pp. 311-316. https://doi.org/10.1109/ICAdLT.2015.7136630

[49] Nemschoff, M., 2014, “Why the Transportation Industry is Getting on Board with Big Data & Hadoop”. Available: https://mapr.com/blog/whytransportation-industry-getting-board-big-data-hadoop/(12.5.2017.)

[50] Vlahogianni, E. I., 2015, “Computational Intelligence and Optimization for Transportation Big Data: Challenges and Opportunities”, Computational Methods in Applied Sciences, 38, pp. 107-128.https://doi.org/10.1007/978-3- 319-18320-6_7

[51] Jain, V. K., 2017, “Big Data and Hadoop”, Khanna Publishing, New Delhi

[52] Vlahogianni, E. I., 2015, “Computational Intelligence and Optimization for Transportation Big Data: Challenges and Opportunities”, Computational Methods in Applied Sciences, 38, pp. 107-128.https://doi.org/10.1007/978-3- 319-18320-6_7

[53] Perera, L. P., Mo, B., Kristjánsson, L. A., 2015, “Identification of Optimal Trim Configurations to improve Energy Efficiency in Ships”, *9th IFAC Conference on Manoeuvring and Control of Marine Craft* (MCMC 2015), pp. 267-272.

[54] Van Rijmenam, M., 2014, “Think Bigger: Developing a Successful Big Data Strategy for Your Business”, Amacom New York, NY.

[55] Sivan, A. P., Johns, J., Venugopal, J., 2014, “Big Data Intelligence in Logistics Based On Hadoop and Map Reduce”, *IJIRSET*, 3, pp. 2634-2640.

[56] H. Segercrantz, “BIG DATA & Big Savings for Maritime Ops,” Available at, [Accessed June 8th, 2016]

[57] S. Shukla, V. Kukade and A. Mujawar, “Big Data Concept, Handling, and Challenges: An Overview,” *International Journal of Computer Applications* (0975 –8887) Volume 114 –No. 11, March 2015

[58] MARINET (2019) e-Navigation testbed and development of equipment for e-Navigation. Available at: https://marinet.org/e-navigation-testbed-and-development-of-equipment-for-e-navigation/ (Accessed: 24 May 2021).

[59] International Maritime Organisation (IMO) (2014) NCSR 1/28: E-Navigation Strategy Implementation Plan. London: IMO. Available at: http://www.imo.org/en/OurWork/Safety/Navigation/Documents/enavigation/SIP.pdf (Accessed: 24 May 2021).

[60] HARBOURS REVIEW (2019) Launch of the Maritime Connectivity Platform Consortium (MCC). Available at: http://harboursreview.com/launch-of-the-maritime-connectivity-platform-consortium-(mcc).html (Accessed: 24 May 2021

[61] Maritime Connectivity Platform (MCP) (2019). Available at: https://maritimeconnectivity.net/ (Accessed: 24 May 2021).

[62] Lee, H. et al. (2018) 'A decision support system for vessel speed decision in maritime logistics using weather archive big data', *Computers and Operations Research. Elsevier Ltd*, 98, pp. 330–342. DOI: 10.1016/j.cor.2017.06.005.

[63] Perera, L. P., and Mo, B. (2016) 'Machine intelligence for energy-efficient ships: A big data solution', in *Proceedings of the 3rd International Conference on Maritime Technology and Engineering* (MARTECH 2016). Lisbon, Portugal, pp. 143–150. DOI: 10.1201/b21890-21.

[64] FastSeas (2019) What is FastSeas? Available at: https://fastseas.com/#what (Accessed: 24 May 2021).

[65] Lambrou, M. (2016) 'Innovation Capability, Knowledge Management, and Big Data Technology: A Maritime Business Case', *International Journal of Advanced Corporate Learning* (iJAC), 9(2), p. 40. DOI: 10.3991/ijac.v9i2.6010.

[66] Schmidt, D. (2019) Weather Routing Software for Sailors. Available at: https://www.cruisingworld.com/weather-routing-software-for-sailors/ (Accessed: 24 May 2021).

[67] Network Mi News (2019) Wärtsilä Navi-Planner Lifts Voyage Planning and Optimisation To Unprecedented Levels. Available at: https://www.marineinsight.com/shipping-news/wartsila-navi-planner-lifts-voyage-planning-and-optimisation-to-unprecedented-levels/ (Accessed: 28 May 2021).

[68] Marine Advisory System - OCTOPUS (2019). Available at: https://new.abb.com/marine/systemsand-solutions/digital/advisory (Accessed: 28 May 2021).

[69] Bon Voyage System (2019). Available at: http://www.ships-for-sale.com/bon_voyage_weather_forcasting.htm (Accessed: 28 May 2021).

[70] Vessel Traffic Service (VTS) (2019). Available at: https://www.sheltermar.com.br/en/vts/.

[71] Society for Gas as a Marine Fuel (SGMF) (2017) Gas as a marine fuel - Safty Guidelines Bunkering FP07-01. Version 2.0. Available at: https://www.sgmf.info/shop/preview/9 (Accessed: 24 May 2021).

[72] ORBCOMM (2019) Networks: Satellite AIS. Available at: https://www.orbcomm.com/en/networks/satellite-ais (Accessed: 24 May 2021).

[73] Šakan, D. et al. (2018) 'Near real-time S-AIS: Recent developments and implementation possibilities for global maritime stakeholders', Pomorstvo, 32(2), pp. 211–218. DOI: 10.31217/p.32.2.6.

[74] SVB (2019) AIS Transceivers. Available at: https://www.svb24.com/en/category/electronics-gps- a-i-s-devices-ais-transceivers.html (Accessed: 28 May 2021).

[75] IALA (2013) IALA Guideline No. 1095 On Harmonised implementation of Application-Specific Messages (ASMs) - Edition 1. Available at: https://www.e-navigation.nl/sites/default/files/1095 Implementation of Application-Specific Messages.pdf (Accessed: 28 May 2021).

[76] Bonham, C. et al. (2018) Analysing port and shipping operations using big data. Available at: https://datasciencecampus.ons.gov.uk/wp-content/uploads/sites/10/2018/06/here.pdf (Accessed: 22 May 2021).

[77] Mahapatra, N. (2019) The Use of Big Data in Ports and Terminals. Available at: https://www.academia.edu/27232641/The_Use_of_Big_Data_in_Ports_and_Terminals (Accessed: 24 May 2021).

[78] safety4sea (2016) Bahri relying on Big Data to enhance productivity. Available at: https://safety4sea.com/bahri-relying-on-big-data-to-enhance-productivity/ (Accessed: 24 May 2021).

[79] MI News Network (2019) The Long-Range Tracking and Identification (LRIT) System: Tracking and Monitoring Ships. Available at: https://www.marineinsight.com/maritime-law/the-long-range-tracking-and-identification-lrit-system-tracking-and-monitoring-ships/ (Accessed: 24 May 2021).

[80] EMSA (2019) VESSEL TRACKING GLOBALLY - Understanding LRIT. Available at: http://www.emsa.europa.eu/lrit-home/download/452/256/23.html (Accessed: 24 May 2021).

[81] Obradović, I., Miličević, M., Žubrinić, K., 2014, "Machine Learning Approaches to Maritime Anomaly Detection", International Journal of Maritime Science & Technology "Our Sea", 61 (5-6), pp. 96-101.

[82] Brandsæter, A., Manno, G., Vanem, E., Glad, I. K., 2016," An application of sensor-based anomaly detection in the maritime industry", *IEEE International Conference on Prognostics and Health Management* (ICPHM), pp. 1-8. https://doi.org/10.1109/ICPHM.2016.7811910, https://doi.org/10.1109/ICPHM.2016.7754713

[83] Alevizos, E., Artikis, A., Patroumpas, K., Vodas, M., Theodoridis, Y., Pelekis, N., 2015, "How not to drown in a sea of information: An event recognition approach", 2015 *IEEE International Conference on Big Data* (Big Data), pp. 984-990. https://doi.org/10.1109/BigData.2015.7363849

[84] Fernández, P. et al., 2017, “Web-Based GIS Through a Big Data Open Source Computer Architecture for Real-Time Monitoring Sensors of a Seaport”, The Rise of Big Spatial Data, pp. 41-53. https://doi.org/10.1007/978-3-319-45123- 7_4

[85] Kurashiki, T., 2016, “Maritime Big Data Solutions by ClassNK”. Available: http://www.mpa.gov.sg/web/wcm/connect/www/c3f96729-2bac4c77-9cad-61954a83e2d/Presentation++Toshio+Kurashiki.pdf?MOD=AJPERES(14.5.201.)

[86] Chen, J., Lu, F., Li, M., Huang, P., Liu, X., Mei, Q., 2016, “Optimization on Arrangement of Precaution Areas Serving for Ships’ Routeing in the Taiwan Strait Based on Massive AIS Data”, DMBD 2016: Data Mining and Big Data, pp. 123-133. https://doi.org/10.1007/978-3-319-40973-3_12

[87] Alessandrini, A. et al., 2016, “Mining Vessel Tracking Data for Maritime Domain Applications”, 2016 *IEEE 16th International Conference on Data Mining Workshops* (ICDMW), pp. 361-367. https://doi.org/10.1109/ICDMW.2016.0058

[88] Perera, L. P., Mo, B., 2016, “Machine Intelligence for Energy Efficient Ships: A Big Data Solution”, *3rd International Conference on Maritime Technology and Engineering* (MARTECH 2016), pp. 143-150. https://doi.org/10.1201/b21890-21

[89] SOLAS, Guidance on Chapter V – Safety of Navigation (Revision), 2007, Annex 17 - Automatic Identification Systems (AIS). http://solasv.mcga.gov.uk/annexes/ Annex17.htm, last accessed 2019/07/20.

[90] Yoshida T: Utilizing big data and the internet of things in shipping. Sea Japan 2016 MTI, 1-32 (2016).

[91] Aoyama K, Liu, J, Hiro, Y: Monitoring system for advanced shipbuilding construction management: Extracting and utilizing monitoring data by considering the reliability of monitored data. ICCAS, 143-154 (2015).

[92] Perera L.P: Statistical filter-based sensor and DAQ fault detection for onboard ship performance and navigation monitoring systems. IFAC, 323-328 (2016).

[93] UNCTAD: Review of maritime transport. United Nations Conference on Trade and Development, Geneva 1-116 (2018).

[94] Rahm, Erhard and Hong Hai Do. (2000) “Data Cleaning: Problems and Current Approaches.” *IEEE Bulletin of the Technical Committee on Data Engineering* (23): 3-13.

[95] Li, Lin. (2012) “Data Quality and Data Cleaning in Database Applications.” [doctoral dissertation], School of Computing, Edinburgh Napier University.

[96] Someswararao, Chinta, J. Rajanikanth, V. Chandra Sekhar, and Bhadri Raju M. S. V. S. (2012) “Data Cleaning: A Framework for Robust Data Quality In Enterprise Data Warehouse.” *International Journal of Computer Science and Technology* 3 (3): 36-41.

[97] Saha, Barna, and Divesh Srivastava. (2014) “Data Quality: The other face of Big Data.” in 2014 IEEE 30th *International Conference on Data Engineering*. pp. 1294-1297.

[98] Shneiderman, Ben, and Catherine Plaisant. (2015) “Sharpening Analytic Focus to Cope with Big Data Volume and Variety.” *IEEE Computer Graphics and Applications* 35 (3): 10-14.

[99] Fellegi, I.P., & Sunter, A.B. (1969). A Theory for Record Linkage. Journal of the American Statistical Association,64.

[100] Wang, R.Y., Storey, V.C., & Firth, C.P. (1995). A framework for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering*, 7(4), 623-640.

[101] Orr, K. (1998). Data Quality and Systems Theory. Communications of the ACM, 41, 66-71.

[102] Hipp, J., Güntzer, U., & Grimmer, U. (2001). Data Quality Mining – Making a Virtue of Necessity. *Proceedings of the 6th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, Santa Barbara, California, 52-57

[103] Wang, R.Y., & Strong, D.M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 1(12), 5-34.

[104] Loshin, D. (2009). Monitoring Data Quality Performance Using Data Quality Metrics. Retrieved March 1, 2012, from http://www.it.ojp.gov/documents/Informatica_Whitepaper_Monitoring_DQ_Using_Metrics.pdf

[105] Müller, Heiko, and Johann-Christoph Freytag. (2003) Problems, Methods, and Challenges in Comprehensive Data Cleansing, Humboldt University Berlin.

[106] Gu, Randy Siran. (2010) "Data Cleaning Framework: An Extensible Approach to Data Cleaning." [master's thesis], University of Illinois, Urbana, Illinois.

[107] Khayyat, Zuhair, Ihab F. Ilyas, Alekh Jindal, Samuel Madden, Mourad Ouzzani, and Paolo Papotti. (2015) "BigDansing: A System for Big Data Cleansing", in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, Melbourne, Victoria, Australia

[108] Hernandez, M, & Stolfo, S. (1998). Real-world Data is Dirty: Data Cleansing and The merge/purge problem. Data Mining and Knowledge Discovery, 2, 9-37.

[109] Elmagarmid, A.K., Ipeirotis, P.G., & Verykios, V.S. (2007). Duplicate record detection: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 19, 1-16.

[110] Lee, M. L., Ling, T.W., & Low, W.L. (2000). IntelliClean: a knowledge-based intelligent data cleaner. Sixth ACM SIGKDD *International Conference on Knowledge Discovery and Data Mining*, Boston, MA, USA 290-294, 20-23.

[111] Cadot, M., & Martino, J. (2003). A data cleaning solution by Perl scripts for the KDD Cup 2003 task 2. ACM SIGKDD Explorations Newsletter, 5, 158-159.

[112] Sung, S. Y., Li, Z., & Peng, S. (2002). A fast filtering scheme for large database cleansing. *Proceedings of Eleventh ACM International Conference on Information and Knowledge Management*, 76-83.

[113] Feekin, A., & Chen, Z. (2000). Duplicate detection using the k-way sorting method. *Proceedings of ACM Symposium on Applied Computing*, Como, Italy, 323-327.

[114] Galhardas, H. (2001). Data Cleaning: Model, Language, and Algorithms. Ph.D, University of Versailles, Saint-Quentin-En-Yvelines.

[115] Zhao, L., Yuan, S.S., Peng, S., & Wang, L.T. (2002). A new efficient data cleansing method. *Proceedings of 13th International Conference on Database and Expert Systems Applications*, 484-493.

[116] Simoudis, E., Livezey, B., & Kerber, R. (1995). Using Recon for Data Cleaning. Advances in Knowledge Discovery and Data Mining.

[117] Redman, T. (1998). The Impact of Poor Data Quality on the Typical Enterprise. Communications of the ACM, 41, 79-82.

[117] Maletic, J., & Marcus, A. (2000). Data Cleansing: Beyond Integrity Analysis. *Proceedings of The Conference on Information Quality* (IQ2000), Massachusetts Institute of Technology, 200-209.

[118] Maletic, J., Marcus, A., & Lin, K.L. (2001). Ordinal Association Rules for Error Identification in Data Sets. *Proceedings of Tenth International Conference on Information and Knowledge Management* (CIKM 2001).

[119] Fox, C., Levitin, A., & Redman, T. (1994). The Notion of Data and Its Quality Dimensions. *Information Processing and Management*, 30, 9-19.

[120] Levitin, A., & Redman, T. (1995). A Model of the Data (Life) Cycles with Application to Quality," *Information and Software Technology*, 35, 217-223.

[121] Strong, D.M., Wang, R.Y., & Lee, Y.W. (1997). Data Quality in Context Communications of the ACM,40(5), 103-110.

[122] Svanks, M.I. (1988). Integrity Analysis: Methods for Automating Data Quality Assurance. *Information and Software Technology*, 30(10).

[123] Müller, H., & Freytag, J.C. (2003). Problems, Methods, and Challenges in Comprehensive Data Cleansing. Tech. Rep, HUB-1B-164.

[124] Rahm, E., & Do, H. (2000). Data Cleaning: Problems and Current Approaches. *IEEE Bulletin of the Technical Committee on Data Engineering*, 23(4), 3-13.

[125] Heather M. Perez, Roger Chang, and Richard Billings: Automatic Identification Systems (AIS) Data Use in Marine Vessel Emission Estimation, *Eastern Research Group, Inc.*

[126] Svanberg, M, Santen, V, Horteborn, A, Holm, H, Finnsgard, C: AIS in maritime research. Marine Policy 106 (2019).

[127] Mirović M, Miličević M, Obradović I: Big data in the maritime industry. Naše more 65(1), 56-62 (2018).

[128] Mahamed Omran, Andries Engelbrecht, Ayed A. Salman: An overview of clustering methods. *Intelligent Data Analysis* 11(6):583-605 (2017), DOI: 10.3233/IDA-2007-11602

[129] A. Jain, M. Murty, and P. Flynn. Data Clustering: A Review. *ACM Computing* Surveys, vol. 31, no. 3, pp. 264-323,1999.

[130] A. Jain, R. Duin and J. Mao. Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no.1, pp. 4-37, 2000.

[131] G. Hamerly and C. Elkan. Learning the K in K-means. In *The Seventh Annual Conference on Neural Information Processing Systems*, 2003.

[132] G. Hamerly. Learning Structure and Concepts in Data using Data Clustering, Ph.D. Thesis. University of California, San Diego, 2003.

[133] G. Coleman and H. Andrews. Image Segmentation by Clustering. In *Proceedings of IEEE*, vol. 67, pp. 773-785, 1979.

[134] A. Jain and R. Dubes. Algorithms for Clustering Data. Prentice-Hall, New Jersey, USA, 1988.

[135] R.H. Turi. Clustering-Based Colour Image Segmentation, Ph.D. Thesis. Monash University, Australia, 2001.

[136] T. Kaukoranta, P. Fränti and O. Nevalainen. A New Iterative Algorithm for VQ Codebook Generation. *International Conference on Image Processing*, pp. 589-593, 1998.

[137] S. Baek, B. Jeon, D. Lee and K. Sung. Fast Clustering Algorithm for Vector Quantization. Electronics Letters, vol. 34, no. 2, pp. 151-152, 1998.

[138] Z. Xiang. Color Image Quantization by Minimizing the Maximum Intercluster Distance. *ACM Transactions on Graphics*, vol. 16, no. 3, pp. 260-276, 1997.

[139] D. Judd, P. Mckinley, and A. Jain. Large-scale Parallel Data Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 871-876, 1998.

[140] H. Abbas and M. Fahmy. Neural Networks for Maximum Likelihood Clustering. Signal Processing, vol. 36, no.1, pp. 111-126, 1994.

[141] C. Carpineto and G. Romano. A Lattice Conceptual Clustering System and Its Application to Browsing Retrieval. Machine Learning, vol. 24, no. 2, pp. 95-122, 1996.

[142] C. Lee and E. Antonsson. Dynamic Partitional Clustering Using Evolution Strategies. In *The Third Asia-Pacific Conference on Simulated Evolution and Learning*, 2000.

[143] A. Jain, R. Duin and J. Mao. Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no.1, pp. 4-37, 2000.

[145] Yim O., Ramdeen K. T., Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data, The Quantitative Methods for Psychology, 11(1), pp. 8–21, 2015.

[146] Najafabadi M. M., et al., Deep Learning Application, and Challenges in Big Data Analytics, *Journal of Big Data*, pp. 1–21, 2015

[147] Kigure, Research on Basic Planning Support of Ships Using Maritime Logistics Big Data, Proceeding of JASNAOE, CD-ROM, 2016 (in Japanese).

[148] Oizumi U., and Aoyama K., Assessment Method of CO_2 Reduction Effect of New Technology and International Regulation on Maritime Logistics, *6th EcoDesign*, CD-ROOM, 2009.

[149] Ministerial Ordinance on Ministry of Finance Ordinance no. 27, Depreciation Assets Life Period, etc. Final Revision March 31, 2016.

[150] National Tax Homepage https://www.nta.go.jp/index.htm mutual agreement', Pomorstvo: *scientific journal of maritime research*, 29(1), pp. 69–74

Acknowledgment

I felt I have learned a lot from writing this dissertation searching for the truth of science and life. This is the great treasure I will cherish not only in my future academic career but in my whole life.

I would like to take the opportunity to express my immense gratitude to all those persons who have given their invaluable support and assistance.

First, I would like to express my utmost gratitude to my research supervisor Prof. Kunihiro Hamada, for his sincere and continuous support, prompt and useful advice, and patience during my doctoral study. His immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life.

Second, I would like to express my sincere gratitude and my thank fullness to my sub-adviser, Assistant Professor Noritaka Hirata, and Prof. Mitsuru Kitamura.

My stay and studies in Japan have been supported by the scholarship of the Japanese Ministry of Education, Science, Sport and Culture MEXT-ASCOJA, for which I am very much grateful.

I very much acknowledge the help of the staff of the offices of the Graduate School of Engineering. As well, I am thankful to all my fellow students in the Laboratory of Transportation System Innovation.

Finally, I would like to express my gratitude to my parents, my big family for the support and continuing prays so far. I am very much thankful to my wife Yorida Agustin Kurniayu and my son Keindra Arsakha Shakeel Althafunizam for their love. Without their tremendous understanding and encouragement in the past few years, it would be impossible for me to complete my study. Love you all.