Maritime Economics & Logistics
https://doi.org/10.1057/s41278-023-00255-z

ORIGINAL ARTICLE

Emerging practices and research issues for big data analytics in freight transportation

Michael F. Gorman[1] · **John-Paul Clarke**[2] · **René de Koster**[3] · **Michael Hewitt**[4] · **Debjit Roy**[5] · **Mei Zhang**[6]

Accepted: 20 January 2023

Abstract

Freight transportation has been experiencing a renaissance in data sources, storage, and dissemination of data to decision makers in the last decades, resulting in new approaches to business and new research streams in analytics to support them. We provide an overview of developments in both practice and research related to big data analytics (BDA) in each of the major areas of freight transportation: air, ocean, rail, and truck. In each case, we first describe new capabilities in practice, and avenues of research given these evolving capabilities. New data sources, volumes and timeliness directly affect the way the industry operates, and how future researchers in these fields will structure their work. We discuss the evolving research agenda due to BDA and formulate fundamental research questions for each mode of freight transport.

Keywords Big data analytics · Data sources · Air freight · Oceanic shipping · Rail · Trucking

✉ Debjit Roy
debjit@iima.ac.in

[1] Department of MIS, Operations and Analytics, University of Dayton School of Business, Dayton, USA

[2] The University of Texas at Austin, Austin, TX, USA

[3] Rotterdam School of Management, Erasmus University Rotterdam, Rotterdam, Netherlands

[4] Quinlan School of Business at Loyola University Chicago, Chicago, IL, USA

[5] Indian Institute of Management Ahmedabad, Ahmedabad, India

[6] Otis Elevator Company, Farmington, USA

Published online: 09 February 2023

1 Introduction

In recent years, freight transportation of all kinds (air, ocean, rail, and truck) has experienced vastly more data, available faster, often more accurate, and sometimes more unstructured. Fosso Wamba et al. (2015) describe the “Five Vs” of Big Data Analytics (BDA) as data’s growing volume, velocity, variety, veracity, and value. Kour et al. (2019) describe how BDA has become a central theme in transportation systems and developers focus on hardware architecture to facilitate the first four V’s of BDA: the collection, dissemination, integration, and storage of this wealth of data. While we specifically discuss the new data sources of freight transportation, we focus on the fifth V, value, and how this advanced data infrastructure and architecture have created needs for decision-making systems that are faster, more automated, and more frequently used, creating information for both automated and human-in-the-loop decisions.

The increase in the volume of data is driven by a variety of sources, including satellite tracking and low-cost sensors that track the location and condition (temperature, wear, vibration, etc.) of fixed assets (trucking terminals, rail yards and track, and ports) and moving assets (locomotives, rail cars, tractors, cargo planes, or containerships) all over the world. The volumes of data that are created from interconnected smart devices, often referred to as the Internet of Things (IoT), come at very low cost; before, it took human effort to collect such data, which was high-cost, time-intensive, or impossible to collect. The IoT makes data available almost instantly via wireless data transfer, eliminating the data latency resulted from lack of connectivity to data storage and dissemination technologies. Data are often more accurate and reliable from these specialized sensors, which are finely tuned and objective; the capability vastly reduces the human judgement that was often required to create and evaluate data.

Unstructured data arises from various customer and supplier websites, burgeoning video storage, and relationships between data sets that were previously uncollected or not analyzed. All of these sources require automated means for combining and cross-referencing, monitoring and making decisions on these inputs. Considerable preprocessing is required to convert the data into formats conducive to analysis. However, new data storage technologies (e.g., Hadoop) have lowered the cost of data storage, manipulation, and analysis. This data revolution has enabled the transportation industry to create value from these new and varied data sources and forms through automated descriptive, predictive, and prescriptive analytical technologies.

This is a pivotal time for big data analytics in the transportation industry. Gorman et al. (2014) note the growing levels of applied analytics in freight transportation; our` research provides insight into how big data and new methods and applications in analytics might inform future research agendas.

Dong et al. (2021) evaluate the impact of emerging technologies in freight transportation. They consider a wide range of hardware technologies including 3D printing, robotics, and autonomous vehicles among others, as opposed to our focus on data and analytics. They do identify big data (including internet of things and blockchain), cloud computing, and artificial intelligence as critical information systems and developments soon to disrupt freight transportation. They find that the volume of literature, including big data, exceeded all other emerging technologies in their study. The same authors also find AI to be among the most entrenched and well-studied technologies in freight transport, with a large and expanding future role in this sector. Our research expands notably in these two important areas.

We discuss the latest developments in the four major freight transportation modes: air, ocean, rail, and road in turn. We also review the available data and their use in practice, focusing on mobile assets (locomotives, trucks, ships, and aircraft) as well as fixed ones (terminals, hubs, airports, and other infrastructure). We subsequently discuss the evolving research agenda and formulate fundamental research questions for each mode. We conclude with observations on the commonalities and synergies between modes, and future research agendas given these new data sources and capabilities.

2 Data types, sources, and structures

The volume of data has grown significantly over the years for all modes of transport. As the number of sensors (e.g., on cargo and transport assets) and the data collection frequency have dramatically increased, the requirements for large data storage platforms have also grown. For example, aircraft data is now so voluminous that specialized platforms have emerged, required to house and manage them. As discussed by Cambier (2018), these big data platforms need to decode the raw sensor data into structured and managed data, have an architecture for fast data access and manipulation, and provide analytical and visualization capabilities. They also need to have an open design for the incorporation of new analytical and predictive models. Table 1 summarizes new data sources that offer avenues for data analytics in freight transport discussed in this section.

2.1 Asset tracking data

New devices allow tracking of transportation assets in real time. For example, global positioning system (GPS) devices that track the real-time position of the asset via satellites are widely used by all modes of transport. The railroads also leverage satellite, sensor, and video data to collect rail car location status and yard information. The American Association of Railroads asserted all trains should be tracked using satellites by 2020 (Association of American Railroads 2019). Likewise, capturing data on ocean vessels, such as the time-stamped vessel position and speed, or container-specific data such as container type, position, origin and destination at container terminals using an automatic identification system (AIS) is becoming the norm. These data are often captured automatically and continuously using sensors on containers, the vessels, or the handling equipment. Telematics are widely used in trucks to track driver behavior and prevent road accidents. Different forms of

Table 1 Data sources for freight transport operations

Data type	Applications					Data source	Data ownership
	General	Air	Ocean	Rail	Truck		
Asset location, Asset movement, and meteorological data	Asset productivity Navigation; Trip safety	Tracking handling equipment, aircraft, ULDs and equipment; Navigation and safety	Maritime clusters; Tracking vessels, containers, and equipment; Navigation and safety	Tracking locomotives, railcars, and trains; railcar utilization	Tracking truck movement, driver productivity; Navigation and safety	Transponders, GPS (all), Automatic Identification System (AIS) navigational data (Vessel), Satellite, ground stations (for weather)	Asset Owners/ 3rd Party
Asset Image and video data	Monitoring	Aircraft (and component) damage investigation; Hub monitoring	Shipping lane anomalies; Terminal monitoring; Container integrity; Accident prevention	Yard monitoring; Track condition maintenance	Freight lane monitoring; Accident prevention	Stationary cameras, Drones and Satellite	3rd Party
Component reliability data, Engine vibration data, Simulation data	Maintenance, Safety	Aircraft maintenance, Runway maintenance	Vessel maintenance, Terminal equipment maintenance	Locomotive, rolling stock, and track maintenance	Truck maintenance	IoT sensors (all), Aircraft and Operators' Business Transaction System (Air)	Asset Owners
Electronic Data Interchange (EDI), Transport Management System Data	Operations	Transaction communication between stakeholders	Transaction communication between stakeholders on stacking, stowage, delivery	Load tendering, tracking	Transaction communication between stakeholders	Transactional data	Asset Owners

Table 1 (continued)

Data type	Applications						
	General	Air	Ocean	Rail	Truck	Data source	Data ownership
Demand, capacity, and customer click data	Asset revenues	Revenue management, Freight Matching, Dynamic pricing	Contracts with freight forwarders and shippers, Freight Matching	Load acceptance; load routing	Dynamic pricing (Freight Matching, Auctions)	Web crawlers, Company specific revenue management systems	Asset Owners/ 3rd Party
Test bed data	Performance estimation	Cockpit simulations	Trip and Terminal simulations, Plan validations	Locomotive, train simulations	Truck exchange terminal simulations	Generated	Consultants
Freight and terminal statistics	Benchmarking performance	Benchmarking airports	Benchmarking ports and container terminals	Rail yard status	Benchmarking fleet owners	Actual data, Port handbooks	Consultants

in-vehicle systems, whether installed in the vehicle (Amarasinghe et al. 2015) or on smartphones (Wahlström et. al, 2018), enable greater visibility of the location of the carrier's fleet. The use of electronic logging devices (Miller et. al, 2020; Scott et. al, 2020), and the eventual integration of autonomous trucks into carrier fleets provide widespread access to location. These data sources also allow for predictive analytics, in particular estimating the time of arrival or providing a warning for potential incidents. Weather information such as the direction of wind or` temperature, captured using aircraft sensors in aircrafts or satellites for trains, vessels, and trucks guide navigation and allow one to perform predictive analytics.

2.2 Asset condition data

Sensors throughout the assets used in freight transport return real-time data on asset condition and status. Examples include aircraft engine vibration data, locomotive data, vessel data, container, and railcar data. In the case of rail, track condition data can now be monitored continuously in some cases. The airfreight industry uses high-frequency status data from sensors on aircraft, ground support equipment, containers, and packages with advanced analytics to improve efficiency, reliability, and safety. Similar devices are on trucks, oceanic vessels, and locomotives. Such frequent, accurate and timely data highlights abnormalities and can help towards condition-based maintenance.

2.3 Simulation and terminal statistics data

Simulation data help to analyze the performance of the asset movement under realistic conditions. For example, commercial modular aero-propulsion system simulation (C-MAPSS) data captured by NASA provide information on airline operations and possible events. Likewise, simulations for rail yard and movement; container terminal and navigation; and truck hub and freight movement proactively provide information for business planning. For example, which route should be chosen to reach the final customer without delays. Industry bodies publish reports on the performance of hubs with respect to the count of assets; there are reports on hub benchmarking, which are quite helpful for improving asset and hub performance.

2.3.1 Cloud-based transportation management systems (TMSs)

Transportation management systems (TMS; Verwijmeren 2004) are software systems that facilitate the management of shipper transportation needs for order delivery. Typically, such systems are widely adopted in road transport (trucking), but also in multimodal transport coordination at the landside of ports where trains and trucks arrive to transport the export and import containers. TMS typically connects to Enterprise Resource Planning (ERP) systems and facilitate activities; for example, building a route to deliver shipments, identifying a carrier from a select list of them engaged in that route, and paying the carrier upon completion of that route. As such, a TMS records when a shipper needed transportation services and what carrier(s) were able to provide them. The use of a TMS has enabled shippers to reduce transportation costs by approximately 5–15% (Logistics Management 2017). They are now moving away from shipper-owned software systems that often contain transactional data for only a single carrier. Many shippers now use a single cloud-based installation of a shared TMS system (Oracle 2018). While cloud-based TMS typically keeps shipper data in silos, they contain much broader visibility of both contract and spot market transactions across the entire market of transportation supply and demand than has been available to date. Software companies that make cloud-based TMS cite the breadth of the carrier network accessible via their system as a value-add to shippers. Data from such systems can be adopted to coordinate resource activities and make real-time equipment scheduling decisions to minimize asset delays at hubs.

Digital Load Boards/Freight Matching Data. Historically, shippers have engaged with carriers on the spot market either directly or through intermediaries. The internet has provided a third mechanism: electronic marketplaces (Caplice 2007; Nandiraju et al., 2008). Load boards are online marketplaces wherein shippers can post their shipments and carriers post their capacity availability. Digital freight matching tools provide additional functionality of automatically identifying potential matches for both shippers and carriers. As described in Table 1, both mechanisms provide a record of the shippers needing transportation services and of the carriers willing to provide them.

Uber Freight (Sanchez et al., 2018) is a digital load board that provides a service to shippers similar to Uber ride. Similar platforms are also emerging for ocean and air transport, to match vessel or aircraft cargo capacity with shippers' demand (e.g., TEUbooker.com, Cargo.com). Data from such platforms lead to matching algorithms that can maximize the profits for the shipper, carrier, and platform.

Data ownership and access is also an important consideration in analytics because often multiple data elements need to be integrated to analyze integrated decisions. There is much debate in, e.g., air freight amongst Original Equipment Manufacturers (OEMs), aircraft operators, and Maintenance, Repair and Overhaul (MRO) vendors about the ownership of data. The resulting compartmentalization—aircraft data is often stored in OEM proprietary formats while operators and MRO vendors typically restrict access to their operation and maintenance data—limits the use of analytics which would allow improvements in operational efficiency, reliability, and safety. Therefore, there is a need to resolve the relevant legal and competition issues.

3 Air freight transport

As shown in Table 1, the airfreight industry uses high-frequency location and status data from sensors on aircraft, ground support equipment and containers, to improve the efficiency, reliability, and safety of their operations. Aircraft and engines have long had built-in sensors to track parameters like airspeed, air pressure, fuel flow rate, fan rotation speed, and exhaust gas temperature data. Accident investigators have used the data from flight data recorders ("black boxes") to determine the causes of accidents. However, the number, sophistication, and reporting frequency of these sensors have greatly increased. Now tens of thousands of sensors can generate hundreds of gigabytes of raw sensor data per day (Bullock 2017), resulting in new uses, in virtually all aspects of air operations, hub operations, and aircraft maintenance.

3.1 Revenue management

3.1.1 Current state

Big data analytics has been at the core of innovations in air cargo revenue management. Rizzo et al. (2019) proposed a revenue management system for air-cargo that combines machine learning prediction with decision-making using mathematical optimization methods. Their solution reduces offloading costs and optimizes revenue generation by addressing the wide discrepancy between the quantity (weight or volume) a shipper will book, and the actual quantity shipped—a problem that is unique to the air-cargo business. To address this problem, American Airlines created a machine learning model that analyzes each customer's booking to predict the likelihood of a no-show shipment (Peckham 2020) using a year's worth of cargo data, i.e., half a million records of 20 variables, employing an open-source, GPU-accelerated machine learning (ML) package. Separately, IBS Software and Korean Air leveraged data analytics and ML to develop an integrated revenue management solution to better match cargo supply and demand while improving profitability (AIT News Desk 2019).

3.1.2 Research agenda

Given the increasing quantity and availability of data, there is every reason to expect that big data analytics will play an even greater role in future air cargo revenue management research. Specifically, the primary focus will be the leveraging the aforementioned and other yet to be developed ML-based approaches to improve both the accuracy and lead time of demand predictions.

3.2 Air traffic management (ATM)

Balancing demand and capacity is central to ATM; however, demand and capacity imbalances are difficult to predict due to the lack of accurate four-dimensional trajectory information (4D; three spatial dimensions plus the time dimension). Cruciol et al. (2015) used a Bayesian network to identify correlations between departing and arriving flights, to update flight plans and reduce delay costs associated with extra crew hours and fuel consumption. Pozzi et al. (2015) showed that there are significant ATM benefits to be had by utilizing burgeoning BDA. In the case of weather, Aircraft Communication, Addressing and Reporting System (ACARS) data has been used to identify events (Roy, 2020) and develop and validate weather models (Benjamin et al. 1991; Roy 2001; Schwartz and Benjamin 1995). Lee et al. (2020) developed an analytical methodology called Safety Analysis of Flight Events (SAFE) that combines correlation analysis, classification-based supervised learning, and data visualization schemas, to isolate critical parameters and eliminate tangential factors for safety events in aviation.

3.3 Aircraft maintenance

3.3.1 Current state

Big data analytics is increasingly utilized to improve aircraft maintenance operations via predictive maintenance (PM) enabled by massive aircraft sensor and operations data. PM is one of three elements in the trade-off between scheduled maintenance (SM) and reactive maintenance (RM), depicted in Fig. 1. Based on the actual condition of a component or system, PM enables determination of the optimum balance between excessive SM (which is wasteful) and insufficient SM (which results in costly RM). However, PM requires accurate and plentiful data that traditionally were difficult or impossible to obtain. Boeing's AnalytX, Airbus's Skywise, and AF/KLM's Prognos have all claimed successful applications on selected aircraft components. However, according to a survey of OEMs, airlines, and MRO vendors by Varfis (2020), it appears that most organizations are still in the process of investigating and evaluating the potential of PM.

There are two approaches to PM modeling. One is a model-based approach where modelers utilize the physical design of the system, e.g., Boeing's digital twin discussed by Mecer (2020). The second is reliability analysis of components and estimation of the Remaining Useful Life (RUL), which are traditional approaches based on failure event data and usage data. Sensor data has made Prognostics and Health Management (PHM) possible, revolutionizing component reliability analysis. For example, Sun et al. (2020) created a health index with sensor data in a degradation model for an aircraft air-conditioning system. Che et al. (2019) presented a PHM model of aircraft systems that combines multiple deep learning algorithms for condition assessment, fault classification, sensor prediction, and RUL estimation.

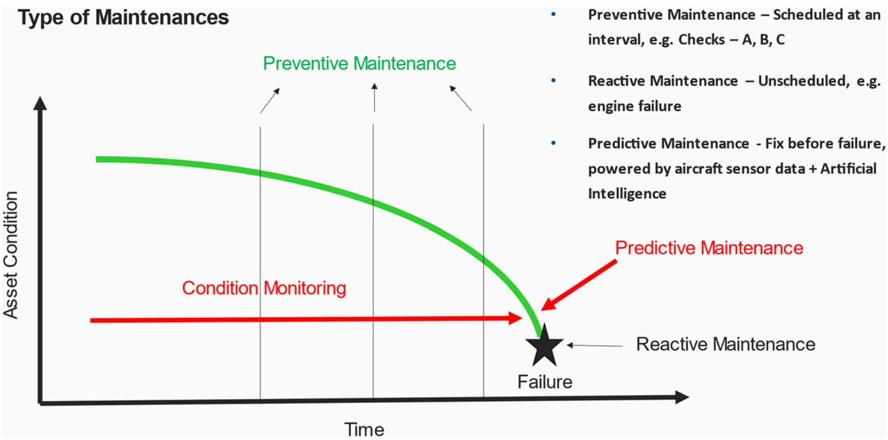

Fig. 1 The trade-offs between scheduled (preventative), reactive and predictive maintenance

Modeling aircraft engine RUL is a very active area because engines are the most expensive aircraft components and the costliest to maintain, accounting for more than 40% of overall aircraft maintenance cost (IATA 2019).

3.3.2 Research agenda

How can we predict failure events and plan for maintenance time-windows? Advancements in RUL estimation will help long-term asset planning. However, more research is needed to operationalize models that predict failure that maintenance planners can act upon. One challenge in estimating RUL is the assumption that the system failure mode is well defined something often not the case, in sophisticated systems like aircraft engines. The decision to remove an engine for maintenance is often a process conducted by a team of experts, considering all aspects of engine health, maintenance history, resource availability, and operational plan. Another challenge to operationalization is the stringent requirement of RUL accuracy, at the end of the life of a system. Nguyen et al. (2019) proposed an approach that predicts the probability of failure in time windows, instead of a precise failure time, with the objective to assist planners in ordering parts and start the preparation process.

How can we leverage engine vibration and other sensor data to detect component failure? Forest (2020) proposed a method to extract raw engine vibration data and detect anomalies. Ordóñeza et al. (2019) combined time-series modeling and ML algorithms to predict RUL for aircraft engines. Zhang (2018) discussed the engine watch list, developed at American Airlines, that predicts engine removals in next 30 days, using engine sensor data. The model is able to predict 2/3 of unscheduled removals. Meert et al. (2019) discussed probabilistic models and ensembles of learning algorithms using sensor data for troubleshooting degradation projection and fault detection. Model precision and un-explainability constitute challenges in these predictive models. Maintenance business processes also need to adjust, to adopt these technologies. Expert knowledge of the systems, operational and maintenance practices, all play important roles in understanding data, creating features that capture signals of anomalies, and developing predictions that are explainable and actionable.

How can we use event logs and text mining techniques to identify failure patterns? Unstructured text data represents a new opportunity for maintenance. Pilots and mechanics document maintenance issues, and the actions taken to fix them, with free text. There are sophisticated text mining algorithms that analyze such maintenance logs to categorize the issues and find chronic patterns. For example, Slattery (2017) reports a joint effort by IBM Watson and Korean Air which provided a Natural Language Processing solution, using historic maintenance texts, to help mechanics improve diagnostic processes.

How can we identify foreign objects using image processing techniques? Foreign objects such as debris on the runway, bird strikes, or hail can cause dents on the airframe. Visual observations are error prone and time-consuming. Bouarfa (2020) presents an image processing algorithm that analyzes pictures taken by drones, to assess dent damage. If implemented, this could dramatically reduce the preventive effort to detect and assess dent damages.

How can machine learning models identify the root causes of fleet performance deterioration? When a fleet has operational performance issues, such as higher than normal rate of delays or cancellations, organizations investigate their operations to look for potential causes. A thorough analysis of all aspects of the fleet is, however, time-consuming and cumbersome, requiring frequent coordination and collaboration from all organizations involved. ML models can easily consider all these factors together, provided data is available. Zhang (2018) discussed how ML models were used in root–cause analyses for fleet performance at American Airlines, which could be an aircraft scheduling issue that causes maintenance gaps, or problematic aircrafts that consume limited maintenance resources and leave the rest of the fleet vulnerable.

4 Ocean freight transport

Ocean freight is the largest contributor to all freight transported, measured in tonne-kilometres. Ships transport over 80% of the volume of international trade in goods (Unctad, 2021). The availability of correct and timely information is crucial to improve responsiveness, to increase a vessel's potential capacity, and to reduce costs. Traditionally, data sources have often been restricted to historical ship and port operations. However, recently, real-time data have given rise to several analytics opportunities in oceanic freight. We discuss how data on routes, demand forecasts, and port development choices can support decisions like vessel selection, or how data from sensors can improve dynamic decision-making on vessels, like path planning as well as cargo placement at terminals.

4.1 Sea operations

Optimizing running costs through a better service network design and voyage analytics is crucial to the success of liner shipping and sea trade. The role of real-time data availability is having a significant impact on maritime operations, for example improving vessel movements, optimizing routes, and reducing fuel consumption.

4.1.1 Maritime network design, surveillance, and voyage analytics

4.1.1.1 Current state At a strategic level, shipping lines must design their shipping schedules which are published in a timetable. The latter includes the ports of call, the sequence of visits, the frequency of the service (number of visits per port per month) and the type and number of vessels on the loop. Some ports within the network may serve a hub role, with freight supplied from and to other ports in the vicinity by smaller vessels or barges in a feeder subnetwork. The number of options and combinations is vast; thus, the problem is typically split into several subproblems, which are solved sequentially; following this, the solutions are aggregated. Multiple actors, including the World Trade Organization, port authorities in different countries, and the shipping lines and the alliances in which they operate, capture the data underlying these problems. Forecasts predict the flows per transport leg. Ocean alliances help reduce voyage and running costs by sharing resources, in particular the vessels, networks, and the terminals along the routes. Optimization frameworks such as integer and dynamic programming evaluate network design (with and without accounting for disruption), port selection and timetable problems. For examples, see Brouer et al. (2014), Mulder and Dekker (2014), Wang and Meng (2012a, 2012b). Given a set of demands (defined by origin, destination, time limit) and a set of vessels with variable capacity and cargo carrying capabilities, the task is to design a set of weekly services using vessel assignment rules such that the cargo arrives within the stated time constraints. Due to the strategic nature of the decision problem such as setting up hubs and determining the trade routes, the usage of real-time, highly granular data in this area has been limited.

Usage of dynamic data is now commonly observed in maritime surveillance and speed optimization. In 2002, Regulation 19 of the International Convention for the Safety of Life at Sea (SOLAS) was enacted, obligating large cargo vessels to use automatic identification systems (AIS) for reporting real-time ship data like a ship's identification, draught, location, course and speed (Bhattacharjee 2021); data provided via satellite during specified time intervals. Route monitoring is required for regulatory compliance and it is critical to the safety of the vessel.

Research on maritime traffic using rich AIS data such as on major trade lanes and associated ship movement corridors has picked up recently. Way point clustering (WPC) and trajectory segment clustering (TSC) are the two methods for identifying transport lanes. WPC clusters distinct points and the generated nodes are then connected with straight paths. Using a threshold value of distance among routes, TSC groups segments with similar position and heading, to generate shipping corridors along the lanes (Gonzalez et al. 2014).

Deviations from the predicted shipping paths can help to detect real time anomalies and to improve compliance. Kontopoulos et al. (2020a) utilized sparse historic AIS data and a Lagrange polynomial interpolation technique to extract shipping lanes. To obtain coherent trajectory clusters, they developed a density-based clustering algorithm. Implementation of their approach shows that their method is accurate; more than 90% of any future vessel sailing path lies in the compact convex hull formed using the extracted shipping lanes. Kontopoulos et al. (2020b) proposed an alternative to the density-based clustering algorithm. They use a combination of three variables: speed, course, and position to estimate the distance between two consecutive vessel positions. Using this method, they were better able to estimate the spatial distance, and to obtain dominant travel path clusters with interesting properties. For example, using vessel speed, heading distance, and location, all vessels can be clustered based on their trajectories. If some vessels do not belong to any cluster due to deviations from the cluster threshold values, this can indicate anomaly events such as AIS spoofing and illegal activities. The also suggest that the enriched network model can be processed and further examined with data mining techniques in an unsupervised manner to identify anomalies in vessel trajectories. In partnership with IBM, Yeo et al. (2019) use ML based models in the SAFER system that identifies vessel entities of interest, forecasts vessel arrival times, and potential traffic hot spots within port waters. Using an illegal bunkering detection module, SAFER can also detect abnormal vessel behavior, such as transfer of marine fuel to ships using mass flow meter data.

Huang et al. (2020) developed a real-time emissions monitoring application using real-time ship AIS data to estimate ship fuel consumption and emissions. AIS messages are divided into continuous data blocks and go through a series of preprocessing steps, including vessel path trajectory extraction, linking path trajectories (association), and interpolation. Regression analysis is effectively used on ship attributes to estimate exhaust emissions in a case study monitoring exhaust emissions in the port of Shenzhen.

AIS data can also be used in combination with weather data to optimize vessel sailing speed and fuel costs dynamically, based on ship location. Mao et al. (2016) developed a speed prediction program based on statistical models, which can be used to plan the expected time of arrival (ETA) and to determine the optimal route plan. Coraddu et al. (2017) used data, measured from the onboard automation systems, to predict fuel consumption and optimize vessel trim. Mulder et al. (2019) used these data to dynamically optimize vessel speed to meet a given timetable. In addition, the authors used this method to allocate buffer times to timetables.

4.1.1.2 Research agenda Port terminals currently use information technology systems to manage data and movements of containers. Electronic Data Interchange for Administration Commerce and Transport is the most common standard to exchange information among participants. With sensors, ports exchange data with ships to create a bay plan, optimize yard position, define the load/discharge sequences (stowage plan), and track containers in real-time. Conca et al. (2018) found that real-time data sharing greatly improves operational planning such as stowage planning. Real-time data sharing lays the foundation for future autonomous shipping (see Munim and Haralambides 2022).

How can a charterer/shipper use data analytics choose a vessel? Real-time data sharing can help determine which ship (and price) is right for the specific cargo. Vessel quality parameters including safety management and navigation, maintenance data, and other qualitative feedback from various actors, like inspectors, terminals, and port authorities, are useful to assess the fit of the vessel to cargo. Data analytics can help both charterers and vessel operators analyze the different sources of information and select the right vessel with the lowest level of risk. However, building trust among the stakeholders and sharing data is a challenge. Smart ports and terminals are the future. Digitization platforms such as blockchain (like Tradelens, an IBM-Maersk blockchain initiative; cf. Lal and Johnson 2018) and IoT applications (e.g., as implemented by the Port of Rotterdam in collaboration with partners like IBM, Cisco, Esri and Axians) are helping to eliminate inefficiencies in shipping operations (Port of Rotterdam, 2019). The overall goal is to make the maritime platform adhere to the highest safety, reliability, efficiency, and standardization standards.

In the container logistics industry, multi-sided digital platforms have emerged that offer new sources of matching spot demand with supply of container transport. TEUbooker.com is an example. Digital platforms (e.g., Matchbacksystems.com and Boxreload.com) offer real-time matching and container tracking services. Another example is Portbase.com, an open platform that offers a common database and a variety of services to safely and efficiently exchange information between all stakeholders (including government and customs) related to the movements of containers, dry and liquid bulk, and general cargo in and between all Dutch ports.

4.1.2 Maritime safety and maintenance

4.1.2.1 Current state Studies like the one by Zheng et al. (2016) use vessel accident data to identify the determinants of crew injuries in container vessel accidents. Data from the US Coast Guard container vessel accident database are used to model and identify abnormal vessel movements. Venskus et al. (2019) analyzed multivariate, heterogeneous sensor data using neural networks to make proper and timely decisions on vessel movements. Wu et al. (2017) used statistical models that are also adopted to identify and predict cargo loss at sea.

Hundreds of vessel components can be monitored using AIS data, generating real time alarms and status information updates. The Naval Surface Warfare Center Philadelphia Division conducts PM by analyzing data collected from ship components to determine servicing frequency, oil and filter changes, and other maintenance tasks. Jimenez et al. (2020) developed a predictive maintenance tool for a Norwegian shipowning company, based on operational data such as vibration data obtained from the engine and rotating components of the vessel. The authors showed that 89% of all failures were not age-related and a traditional preventative maintenance approach was inadequate to avoid failure. Jimenez et al. (2020) also report a US Navy study, which attributes age-related failures to continuous degradation by corrosion when the vessel operates in a saline environment. Because most of the other failures were random, a data-driven PM focus was required.

4.1.2.2 Research agenda *Can prescriptive models be developed to optimize the schedule of maintenance activities and minimize costs?* While real time data is used for PM such as component inspection, cleaning, and reassembly, it is not clear how to develop and evaluate models for prescriptive maintenance. Such models should include the possibility to overhaul or replace multiple components at the same time so that the vessel would not need to again undergo maintenance within a short span.

4.2 Port operations

While data availability improves maritime operations at sea, the availability of such data can immensely improve and optimize the planning operations in port. In particular, we discuss how data availability can reduce vessel sojourn times in port and minimize congestion at the landside.

4.2.1 Vessel waiting and container dwell times

4.2.1.1 Current state BDA plays a significant role in reducing both vessel waiting time to access a berth and vessel berth time to load and unload containers; something of great importance to carriers, terminal operators and port authorities. Short waiting and berth times mean more turnover of the vessel; long waiting and berth times can lead to penalties. Vessel berth times are affected by several terminal design parameters such as the type of available equipment and the number of resources used (e.g., quay cranes, and stack cranes at a container terminal, or the number of loading arms at a liquid bulk terminal). The technology choice of the equipment and layout of the terminal also affect ship berth times (Roy et al. 2020). Bierwirth and Meisel (2010, 2015) show that optimal assignment of berth and handling equipment to vessels and storage locations in the yard can minimize vessel delays. Recently, data-driven optimization techniques are developed for berth scheduling and quay crane assignments. For example, Koley et al. (2022) proposed four different Machine Learning techniques, comprising linear regression and artificial neural network, to predict vessel arrival times. Using AIS-based forecasts, they estimate dynamic time buffers (DTBs) and improve the robustness of berth schedules.

To load a container vessel requires a stowage plan. Typically, the problem is approached in a static and deterministic fashion, where a 3D bin packing problem must be solved meeting certain constraints. Zhang et al. (2008) and Gharehgozli et al. (2016) give an overview. However, many stakeholders need to provide information for efficient stowage and changes in cargo pick-ups and drops can occur during a voyage. Conca et al. (2018) claim benefits can be enjoyed by using actual dynamic data, shared between the stakeholders (carriers and terminals), improving the quality of stowage. A ship's stowage plan changes from one port call to another, due to discharging and loading containers. Thus, the timely exchange of the latest status updates with the next terminal allows timely creating a new stowage plan which speeds up the terminal processes.

Low average quay crane (QC) processing rates are associated with long vessel sojourn times at the terminal and can lead to port congestion. Linn et al. (2007) developed an artificial neural network model to predict the QC rates for the next planning period. Maldonado (2019) developed a two-stage analytics framework to minimize the number of rehandles (i.e., restacking) at a terminal by finding appropriate stacking locations. They predict the dwell time of each container arriving at the port using multiple linear regression, decision trees, and random forests, employing a large sample of import container data. Container features such as size and height, customs clearance requirements, empty/full status, weight, consignee name, and month are used in the prediction models. The authors' results show that models using dwell time prediction can substantially reduce the number of container rehandles by appropriate stacking. Recently, Verma et al. (2019) used reinforcement learning to determine the optimal sequence of container movements so that the rearrangement of containers in the yard is minimized.

4.2.1.2 Research agenda *How can we use real-time vessel arrival data along with container transshipment information to stack the containers appropriately in the* *yard?* Optimal yard assignment will reduce the time to remarshal the container for its onward travel and therefore lead to lower costs and speedier operations.

4.2.2 Landside operations

4.2.2.1 Current state Trucks, trains or barges drop off export containers and pick up import containers in landside operations. Truck terminal appointment systems are commonly used to schedule truck arrivals and reduce congestion. While optimization methods are commonly used to decide the number of optimal time-slots and assign trucks to a particular service slot, data mining techniques can be used to pair import containers with export containers such that empty truck trips are reduced and truck service levels are improved. Caballini et al. (2020) used a hierarchical clustering technique to combine import and export container loads based on multiple features such as container size, weight, type (IMO, Reefer, Normal), vessel departure time, and then assign the trucks to a preferred time slot to minimize their turnaround time. Li et al. (2022) use a bi-objective mixed integer programming model to ensure a smooth turnaround of trucks performing dual transactions at a landside by optimizing the allocation of appointment quotas along with the deployment of yard handling equipment. Using real case studies from Mexican and Italian container terminals, they show that their two-phase approach could reduce empty-truck trips by up to 34%.

4.2.2.2 Research agenda *How can new data sources that provide real-time vessel location information be leveraged to improve port operations planning and scheduling activities, and reduce port congestion?* Non-profit organizations have developed standard data exchange formats that would permit real-time data sharing and collaboration among stakeholders. For example, the Digital Container Shipping Association (DCSA, 2020) published the just-in-time arrival guide standards which would enable containerships to optimize their sailing speed based on the current congestion levels at the ports, thereby lowering fuel consumption and reducing CO2 emissions (IPCDMC.org). For example, a simulation exercise carried out by Wärtsilä on the Port of Singapore's container operations revealed that optimized arrival times can reduce CO2 emissions by 1,6 M tonnes annually (https://www.wartsila.com/insights/article/benefits-of-just-in-time-sailing-how-to-take-port-operations-to-new-heights). Further benefits include increase in navigational safety and better fleet planning and scheduling. Likewise, The Dutch TNO (tno.nl) and the Port of Rotterdam Authority analyzed all container vessels berthing at the port in 2017 and estimated that by supplying accurate information to ships, about 4 percent – or 134,000 tonnes – of CO2 emissions can be saved every year (https://www.portofrotterdam.com/en/news-and-press-releases/just-time-sailing-saves-hundreds-thousands-tonnes-co2). The vessels would need to adjust their sailing speed by an average of 5 percent, and still arrive at the planned arrival time. Additional savings could be attainable if ships were better informed twelve hours before arrival at the port.

The Swedish Maritime Administration's Sea Traffic Management (seatraffic-management.info; STM) project creates a new paradigm for maritime information sharing in real time. STM creates standards that allow interoperability among stakeholders and allow information holders to retain their valuable data and choose how they share it. STM has validated the infrastructure, software, and services on 300 ships in nine ports and in five shore centers. The results show improvements in safety, efficiency, and ecology.

Real-time data (both actual and estimated voyage and port-centric time-stamps) allow researchers to show the value of real-time schedule optimization. The value of BDA lies in integrating data and coordinating stakeholders from different modes. For example, the GPS data from trucks can be adopted to initiate container handling operations before the truck arrives. Further, *how can we appropriately pair (match) an import with an export truck, such that the empty leg of the trucks can be reduced, saving both costs and reducing emissions?* Real-time integration of data across different modes of transport and the port status would also enable synchromodality. Acero et al. (2022) define synchromodality as "a multimodal transportation planning system, wherein the different agents involved in the supply chain work in an integrated and flexible way that enables them to dynamically adapt the transport mode they use based on real-time information from stakeholders, customers, and the logistic network". With real-time data integration (demand across modes, state of the modal assets, and port resource status), we can predict delays in transport related to the due times and facilitate adaptive modal choice. Such integrations may prevent transport delays during disruptions. At a strategic level, data from the terminals can help predict shipper on-time performance, booking and cancellation behavior, and capacity utilization. Shipper trend analysis and forecasting can help carriers to maximize vessel utilization and improve business profitability.

5 Rail Freight Transport

This is a pivotal time for establishing new lines of freight rail research based on new data sources (Table 1), hardware, and computer system architecture developments. Some authors talk about the structural changes required to create, disseminate and store these new data sources. Land et al. (2019) describe some of the general infrastructural advancements and developments in freight rail data systems and their effect on decision-making and operations. Xin and Xiaoning (2020) discuss the architecture requirements for data collection and dissemination in such systems, and McMahon et al. (2020) focus on railway asset data collection and data management requirements in a big data environment. Thaduri et al. (2015) discuss the insights that can be gleaned from big data for all phases of rail asset management, including maintenance, safety and operations in passenger rail, though many of the concepts apply to freight rail as well.

These new data sources have led to analytical research stemming from the new capabilities. In a survey of big data and rail analytics, Ghofrani et al. (2018) note that two-thirds of the rail research articles in that survey were categorized as "big data" articles, and nearly half of the papers appeared in the last 3 years (2015–2017) of a fifteen-year span (2003–2017). Ghofrani further notes that nearly half of the BDA applications are applied to maintenance, just under one third in operations, and just over one fifth in safety topics. The new data sources described above have led to entirely new and somewhat radical opportunities to use this high-frequency, high-volume data to create advanced analytical techniques, to improve operations and locomotive, railcar and truck management and maintenance. Below we discuss each area in turn.

5.1 Rail operations

5.1.1 Current state

Historically, when a train passes a fixed signal, a fixed block of track is allocated to only that train for safety reasons. The location of a train is only known to be within some fixed area of the rail tracks, called a track "block", which can be from 10 to over 150 km long.

In 2008, the US government mandated that all major freight railroads in the USA implement satellite-based tracking systems for their trains over the 76,000 miles of mainline track (Rail Safety Improvement Act of 2008). The satellite tracking of trains gives pin-point accuracy of train locations, overlaying or replacing the fixed track block and signaling system. Newly implemented Positive Train Control (PTC; AAR, 2020) systems can leverage improved location information to reduce human error. These systems are overlay systems that work alongside existing track-side monitoring systems and movement policies based on fixed block train control policies. Despite traditional regulations and norms on dispatching, PTC allows for improved approaches, augmented by new analytical methods.

Augmentations to PTC, known as Communications-Based Train Control (CBTC), frequently used in passenger rail, allow for more aggressive use of track, allowing the leading and trailing distance between trains to shrink, using "dynamic blocks" or "moving blocks" based on the train stopping distance and its precise location not on fixed, static track blocks, and perhaps eventually allow for autonomous trains. These developments lead the way for entirely new dispatching algorithms that leverage the enormous amount of real time location information and drastically expanded track capacity from dynamic blocks. Similar advances have been made in China, resulting in cloud-based dispatching and freight management tools handled in distributed systems using Tashi and Hadoop and Map-Reduce search technology (Zhang, et al. 2009).

5.1.1.1 Research agenda *How will dispatching algorithms work under a moving block rail dispatching regime?* The new PTC/CBTC capabilities require a reconstructed view on rail dispatching research to accommodate a new set of capabilities and a relaxed set of assumptions. Stephens (2021) reports that BNSF railways received a patent for a "moving block" system that would allow dispatching based on GPS location rather strictly on signaling. No current dispatching algorithms are based on fixed block. Dingler et al. (2010) discuss increases in track capacity due to moving block capabilities and suggest moving blocks increase a track's capacity by as much as 25% but require new dispatching methods and algorithms. Such capabilities require significant analysis for improved train dispatching practices in a number of ways. Zhao and Ioannou (2015) discuss improved dynamic headway dispatching policies and how these relate to the new source of data. Finally, Diaz de Rivera, et al. (2020) discuss changing train fleeting policies given dynamic blocking.

5.1.2 Locomotive maintenance

5.1.2.1 Current state With all of its assets, railroads strive to replace RM (after a part failure) with SM (before part failure). A major challenge in failure prediction came from the infrequency and inaccuracy of data. New sensors on tracks, cars, and locomotives provide near real-time information on the condition of these assets, changing the need for scheduled inspections, which are costly and often remove the asset from productive service. Ghofrani (2017) found that in the maintenance area, BDA applications focused on PM and condition-based maintenance. Railroads' maintenance planning processes are moving away from periodic inspections to continuous status monitoring, which changes the required prediction tools. McMahon et al. (2020) provide a review of the requirements and challenges in big data analytics applications to asset management. Locomotives are highly complex pieces of equipment, now equipped with computer and communications technology for recording operating statistics (such as fuel burn and various torque measures), condition reporting (temperatures, liquid pressures, vibration levels), cab video, and location information.

5.1.2.2 Research agenda *What new analytical methods can be developed to take advantage of the near real-time flow of highly accurate and voluminous data?* There are many locomotive components, each with its own pattern of wear and criticality. For example, sensors on locomotives monitor operating practices of engineers and can be used to optimize speed/fuel trade-offs remotely. Others measure vibration, temperature, emissions and more. Each sensor may require different approaches and requirements for speedy identification and resolution. For example, Lei et al. (2016) describe an unsupervised ML application for evaluating on a continuous basis the condition of motor bearings.

5.1.3 Railcar maintenance and management

5.1.3.1 Current state Railcar status has long been monitored on a frequent basis from way side detectors, but these report risk conditions and failures which require the train to stop and remove the car in disrepair from the train at which time corrective maintenance can be conducted. Train inspections occur on a scheduled basis (e.g., inspection points every 500 miles) during a trip. Often, the costly and time-consuming inspection is premature and reveals no defects; other times, the inspection does not identify the car at-risk before a break down occurs after the train departs.

A plethora of new sensors now help to identify the location, mechanical status, and current availability of railcar equipment. Coupled with IoT, data is available for decisions immediately. The sensors are not only valuable for railcar preventive maintenance. Rail and container yards contain thousands of cars and containers at any one time. Equipment status and availability can be evaluated by sensors which discern whether the equipment is loaded or empty and therefore available for use. Precise information on location helps reduce lost search time and allows for more efficient car and container movement plans.

5.1.4 Research agenda

What preventative maintenance algorithms are needed to make use of new locomotive sensor data? The sensors provide more accurate and timely data from which to make maintenance decisions. Whereas manual locomotive and car inspections take place on a time interval and mileage basis, sensors provide instead a continuous flow of information. This helps avoid long intervals of underperformance or equipment deterioration while operating defectively, and coupled with new analytical approaches, can enable better maintenance scheduling and planning. For example, Tarawneh et al. (2018) discuss the use of sensors which continuously evaluate and disseminate information, such as temperature, vibration, load and impact stress on axles and wheels. In another example, Albakay et al. (2019) discuss PM-based BDA for improved performance in train safety, availability, and reliability.

5.1.5 Track maintenance

5.1.5.1 Current practice Generally, there are two types of track inspections: geometry and structural. Geometry inspections focus on the gauge, profile and alignment of the rails. Structural inspections focus on the condition of the rail, ties, and ballast. Historically, teams of engineers would ride in telemetry cars, slowly traversing the track, taking measurements. The need for analysis was relatively low because the slow and costly manual data collection of track inspection was infrequent. Of course, this method meant that track segments were measured quite infrequently, and the need for analytical methods was low.

The increase of low-cost data collection tools such as track monitors on trains (Saki et al. 2019) and drones with video capture reduce the time and expertise required to inspect the tracks and have increased the inflow rate of track data dramatically. With an increased flow of track data, faster analysis of track geometry and conditions is required. Salierno et al. (2020) focus on the data collection and preparation process for defect detection in a big data environment. Gerum et al. (2019) discuss the importance and improvements in data accuracy and precision of track failure based on track conditions. Once track conditions are known, track maintenance crews must be scheduled. Consilvio et al. (2019) discuss the entire process from data collection and manipulation to optimization of scheduling maintenance crews. New track predictive maintenance algorithms will be needed due to high-volume data now regularly available.

5.1.6 Research agenda

5.1.6.1 How can we leverage video and sensor information to allow for better predictive maintenance of tracks? Artificial intelligence methods can process video of the track, increasing the speed and frequency of analysis to match the increased flow of data collection (Pall et al. 2014). For example, Martey et al. (2017) propose a method to predict geometry defects in a big data environment.

5.1.7 Safety

5.1.7.1 Current state A number of new data sources help with safety. Of course, well-maintained locomotives, railcars, and tracks are paramount for safety, but there are others; for example, satellite tracking and control of locomotives as a second layer of protection from collisions, and videos of engineers measuring fatigue. These data cannot be economically or effectively monitored by human eye, and new analytical methods are needed.

5.1.7.2 Research agenda *How can we leverage video data analytics in locomotive cabs to monitor train operators to improve safety of operations?* Fixed cameras on locomotives provide continuous, real-time data. Coupled with AI-based automated surveillance technology, real-time preventative alerts will identify fatigue as it happens.

6 Trucking-based freight transport

For many countries, trucking accounts for the vast majority of freight market share. In the USA, it has been estimated that 72.5% of all freight is moved by truck (American Trucking Association 2020). From an economic perspective, the size of the US trucking industry in 2017 was estimated at more than $700 billion (Business Insider, 2019); globally, this is estimated to be more than $4 trillion (FreightWaves, 2020).

In this section, we propose a research agenda for trucking-based freight transportation that is enabled by the dramatic increase in available data regarding operations in the broader transport sector. We outline two different streams of research within that agenda. For each stream, we describe the current operational issues motivating the research and the data sources that we believe enable such research to be impactful.

6.1 Market-oriented research agenda

6.1.1 Current state

The truckload (TL) carrier market in the USA is large and notoriously fragmented. The number of for-hire carriers in the country is often reported to consist of over 900,000 companies, with over 90% of them having fewer than six vehicles (American Trucking Association 2020). In Europe instead, over 90% of trucking companies have less than 10 trucks (trans.info, 2020). Even the largest TL carrier in the USA has less than 5% market share (Logistics Management, 2020). In addition, truckload transportation capacity is often established in response to shipper requests rather than with advanced scheduling. As a result, the supply of truck-based freight transportation is typically dynamic, volatile, and only partially observed. Knowing what trucking capacity a carrier can provide where it is needed is a constant problem for shippers (McCrea 2020).

The demand for trucking capacity has similar characteristics. Shippers procure transportation capacity from carriers via two mechanisms (Jothi Basu et al., 2020; Lafkihi et al. 2019). Formal, long-term, contracts with one or more carriers that typically specify the rate at which that carrier will provide capacity between an origin and destination. These contracts are often established via auctions (Lim et al. 2008) and account for about 80% of the transportation procured in the USA (Howe, 2020). Second, spot markets are used for the remaining 20%. However, relying on the spot market is risky, as spot prices can exhibit tremendous volatility (Budak et al. 2017). Sometimes, that volatility forces shippers to turn to the spot market; it has been reported (Robinson 2020) that 20% of the time carriers reject requests from contracted shippers to achieve high spot market rates when available market capacity is low. Like supply, the demand for trucking capacity is volatile; the average tender lead time (the time between when a shipper announces a need for transportation and when that transportation is to occur) is often less than three days (FreightWaves, 2019).

While transportation procurement has long been an area of opportunity for analytics (Lafkihi et al. 2019), researchers have noted (Caplice et al., 2003; Lafkihi et al. 2019) that most of the contract-focused procurement activities have relied less on analytical models and more on negotiation. More specifically, much of the procurement-related literature (e.g., Lafkihi et al. 2019) has focused either on mechanisms for establishing transportation contracts (Zhang et al. 2015) or acquiring capacity via the spot market (Lim et al. 2008; Lindsey et al., 2017). Relatedly, shippers often resort to the spot market because they cannot promise sufficiently high, repeated, day-of-week specific volume to a carrier. Shippers are thus subject to volatile and uncertain spot market pricing that results from inventory and production planning decisions that are made without visibility of their impact on transportation options and costs.

Different carriers will follow different strategies with respect to accepting transportation requests from shippers. Carriers who dedicate a portion of their fleet to a single shipper, or seek to provide a high level of service to an established customer base, will likely accept every or nearly every request made. Other, for-hire, carriers can accept or reject shipper requests for transportation. The price a shipper is willing to pay is one of many factors a carrier will consider when accepting a request. Others include whether the shipper has goods ready for pickup when promised, the ease with which goods can be loaded at their origin and/or unloaded at their destination, and shipper promptness with respect to payment. More specifically, whether accepting a request is profitable for a carrier also depends on whether doing so enables the carrier to service other requests. For example, if a driver must wait many hours before they can complete delivery at the shipper destination, driver hours of service regulations may prohibit that driver from serving more requests that day. Technology platforms such as Uber Freight that provide carrier reviews of experiences with shippers provide greater visibility into how easy a shipper is to do business with.

To summarize, trucking-based freight transportation is a market that consists of many players on both the shipper and carrier sides. There is also significant volatility on both sides of the market that complicates planning and operations, largely because carrier and shipper visibility in this market is limited. Carriers have only knowledge of the shippers for whom they have transported loads and shippers have only knowledge of the small number of carriers that have transported their loads. However, the data sources discussed previously will provide both shippers and carriers greater visibility of this market. Thus, we next propose a research agenda to realize market efficiencies from that visibility.

6.2 Research agenda

We next outline research questions regarding matching the demand for freight transportation with its supply. We begin with a question relevant to carriers. We then turn to questions relevant to shippers.

How can carriers use competition and spot market price data to decide which transportation service request to accept? ADP-based methods (Godfrey et al., 2002), developed for route optimization-oriented fleet management problems, typically assume carriers accept all transportation service requests. As noted above, this assumption may be reasonable for some (e.g., dedicated) carriers. However, even dedicated carriers may want to opportunistically accept requests from outside their customer base. Greater demand visibility enables both dedicated and non-dedicated carriers to intelligently pick and choose transportation requests to service. Adapting ADP-based methods to capture selective service opportunities presents several new research challenges. One is to recognize the presence of competition from other carriers, a second is to recognize the volatility in spot market prices, and a third is to inform Approximate Dynamic Programming (ADP)-based methods with data derived from free text user reviews regarding shipper readiness and reliability.

How can shippers determine the optimal mix of contracted and spot market shipments using real-time market datasets? A shipper may establish contracts for frequent and high-volume transportation needs and rely on the spot market for moves that are low volume. Greater visibility of contract and spot markets creates more comprehensive data sets and will improve the calibration of analytical procurement models. Examples include models that can help a shipper value a transportation contract (Tsai et al. 2011), determine whether to sign a contract or resort to the spot market on a single lane (Boada-Collado et al. 2020), and determine which lanes should be collectively outsourced to a third-party transportation provider (Rajapakshe et al. 2014). Boada-Collado et al. (2020) considered this question, but did so for a setting consisting of a single lane and constant, known, spot market prices; new spot market data allow researchers to relax this assumption.

How can shippers inform production planning models with explicit recognition of the impact of well-timed production and inventory decisions on transportation options along with representation of uncertainty and volatility in spot market rates? The production routing problem (Adulyasak et al. 2015) jointly determines production, inventory, and transportation decisions. While the model assumes transportation costs and capacities are known and constant, there is often volatility in both. Research into this problem has not recognized that the timing of production and inventory decisions can impact the transportation options available to the shipper. Nor has research recognized that there may be uncertainty regarding those options. Generally, *full truckload transportation* is cheaper than *less-than-truckload transportation* on a per-unit-of-weight measure. Well-timed sourcing decisions may enable a shipper to use full truckload transportation for their inbound transportation of materials. Well-timed production decisions can do the same for outbound transportation of finished goods. Similarly, a carrier is often willing to charge a lower per-unit-of-distance rate on two complementary moves than it would on each individual move. Well-timed and coordinated sourcing and production decisions can enable a shipper to achieve savings by offering both inbound and outbound transportation moves to the same carrier. When there is volatility in the transportation market, achieving savings via the mechanisms described above requires forecasting models that accurately predict transportation options and prices. Insights into the contract and spot markets afforded by new data sources will enable development and calibration of such forecasting models.

6.3 Carrier operations-oriented research agenda

6.3.1 Current state

For carriers, fleet management has long been an opportunity area for analytical models (Gorman et al. 2014). A critical issue in fleet management is determining which vehicle to assign to serve an accepted transportation request. In the absence of autonomous vehicles, a related issue is determining the driver to drive that vehicle. These allocation problems are challenging because requests occur over time and, in many countries, drivers can only work a limited number of hours a day. Poor choices can lead to excessive empty miles, higher costs, lower margins, and lower equipment utilization. For example, it has been reported that 15–30% of freight truck trips in Europe concern empty miles (Eurostat 2017). Often, these empty miles are a result of a vehicle and driver completing a transportation request and there not being a next request nearby for them to serve. Thus, effectively allocating vehicles and drivers to transportation requests requires anticipating when and where future requests will occur and how much competition there may be for those requests from other carriers. Regarding transportation requests, researchers have tried to utilize GPS data to forecast commodity movements (Akter et al., 2019), providing more insight into freight movement. Greater visibility of historical transportation demands enables better forecasting of future freight transportation demand (Garrido et al., 2000; Xiao et al. 2020). Regarding carrier competition, researchers have collected GPS truck data from multiple carriers to generate forecasts of freight truck movements across carriers (Bassok et al. 2011; Ben-Akiva et al. 2016; Flaskou et al. 2015; Ma et al. 2016). An improved ability to forecast both supply and demand in the freight transportation market can lead to better allocation decisions and fewer empty miles (Ichoua et al. 2006; Miller et al. 2020a, b).

Another critical issue for carriers in the supply of transportation capacity is the chronic shortage of drivers; in 2018, there was an estimated shortage of over 60,000 drivers in the USA (TB&P 2020), and the shortage is expected to grow in magnitude because the average age of a truck driver in the USA is over 50 years. However, some expect autonomous trucks to be a mitigating factor. In any case, effective driver utilization is critical. Given the rules and regulations regarding how long drivers can be on the road and away from their domicile terminal, the allocation of drivers to domicile terminals has a critical impact on their potential to meet shipper demands for transportation and the effective utilization of drivers.

Detention, or a truck waiting more than two hours while loading or unloading at a dock, may cause costly congestion. Reportedly, 63% of drivers are detained for more than three hours per stop *(*DAT Freight and Analytics 2016). Such time reduces the amount of capacity a driver (and truck) can provide. While carriers frequently cite detention as a major problem, shippers do so much less often (DAT Freight and Analytics 2016). Telematics can provide carriers with data to justify the need of shippers to change *those* practices which lead to excessive detention times. One source of detention is shipments that are not ready for loading when the vehicle arrives. Encouraging a shipper to communicate more frequently with the driver, to get an up-to-date estimated time of arrival, can help ensure that goods are ready for loading at the right time. Another source is a vehicle arriving but there not being an available parking spot at the dock for loading. Encouraging a shipper to adopt technology solutions like a yard management system can help ensure that when a vehicle arrives there is a place for it to park and receive its load. Anticipating detentions at shipper destinations can enable a carrier to better allocate drivers to such requests.

6.4 Research agenda

How can carriers leverage telematics data to decide which vehicle and driver should be assigned to which transportation request? Determining which vehicle and/or driver to assign to serve a transportation request often falls under the broad umbrella of route optimization. Analytical techniques for route optimization often involve some form of Approximate Dynamic Programming (ADP; Powell 2007; Ulmer et al. 2019), where near-term decisions are made based upon statistical estimates of key performance indicators which are, however, realized after decisions are implemented. The quality of transportation request forecasts drives the quality of these estimates; new data sources can improve their performance and open up new research directions. Applications of ADP methods to truckload fleet management require real-time knowledge of vehicle locations. Documented applications of these methods in the past (Simão et al. 2009, 2010) have only existed for large carriers that have the resources to invest in new data sources. The vehicle telematics data sources outlined above make such knowledge also available to smaller carriers. Similarly, implementing and executing these methods has required computing resources well beyond the financial reach of many carriers. However, the advent of on-demand cloud data storage and packaged computing services reduces those requirements, expanding their use in practice.

How should statistical distributions be embedded in driver domiciling and fleet sizing models to reduce empty miles and increase asset utilization? Determining the domicile terminal for a fleet of drivers has received some attention (Erera et al. 2009). Fleet sizing in a truckload context (Baykasoğlu et al. 2019), a related problem, has as well. However, none of the existing approaches have recognized the potential to move loads acquired via a spot market, whereby other carriers may compete too. Richer data sets regarding shipper needs for spot market transportation will enable the development of statistical distributions of such demands.

7 Conclusions

We have shown that freight transportation via air, water, rail, and truck has been radically changed by big data analytics, which has created a new world in freight transportation. New data sources, volumes and timeliness directly affect the way the industry operates, and how future researchers in these fields structure their work. From the condition data sources, we see a dramatic improvement in asset maintenance and management across transport modes. From improved data accuracy and timeliness, we see improved network-wide data and forecasts that allow for better planning, scheduling, and supply chain coordination between market participants. Availability of market-wide transactions affects how some transportation services are bought and sold.

As the freight transportation industry rolls out its approaches to maintenance, operations, supply chain and market transactions, so must research do to support them. We have described the new and evolving methods in use, and in need, across the freight transportation landscape. Future research agendas should pay heed to the rapidly changing conditions in the freight transportation industry. For rail, we discuss the developments mostly with respect to the USA. Extending the study to include other countries would be a potential opportunity. We have formulated a number of important research questions for each transport mode related to the new data sources and possibilities. These include:

(1) Maintenance: How can we use component reliability data obtained from IoT sensors for better preventive and condition-based maintenance strategies? *(2) Planning and scheduling*: How can we use the improved availability of asset tracking data across modes for better resource planning and scheduling decisions, and improving transportation speed? *(3) Matching demand and supply*: How can we use the multi-sided platforms (especially in road and ocean freight) for better matching of supply and demand, with better capacity utilization, and pricing?

Decision models in all three areas are static in nature due to lack of real-time data availability (planning decisions are taken over a longer horizon and often heuristics are adopted to obtain a solution). New real-time and dynamic information allow the decisions to be made online, and the models can be run on a real-time basis. This provides new data-driven opportunities to both research and practice in freight transportation.

Acknowledgements We are grateful to the journal Editor-in-Chief, handling editor, and reviewers for their comments and suggestions that greatly improved the quality of this paper.

References

Acero, B., M.J. Saenz, and D. Luzzini. 2022. Introducing synchromodality: One missing link between transportation and supply chain management. *Journal of Supply Chain Management* 58: 51–64.

Adulyasak, Y., J.F. Cordeau, and R. Jans. 2015. The Production Routing Problem: A Review of Formulations and Solution Algorithms. *Computers & Operations Research* 55: 141–152.

AIT News Desk. 2019. IBS Software Has Launched an Integrated Revenue Management System at Korean Air to Boost Cargo Profitability. *AITHORITY*. https://aithority.com/machine-learning/ibs-software-has-launched-an-integrated-revenue-management-system-at-korean-air-to-boost-cargo-profitability/

Akter, T., S. Hernandez. 2019. Truck Activity Pattern Classification Using Anonymous Mobile Sensor Data. *Innovations in Freight Data Workshop*.

Albakay, N., M. Hempel, H. Sharif. 2019. Big Data Analytics for Proactively Optimizing Rolling Stock Maintenance:2019 Joint Rail Conference. *American Society of Mechanical Engineers Digital Collection*.

Amarasinghe, M., A.S. Kottegoda, S. Arachchi, H. Muramudalige, D. Bandara, and A. Azeez. 2015. Cloud-based driver monitoring and vehicle diagnostic with OBD2 telematics. *Fifteenth International Conference on Advances in ICT for Emerging Regions (ICTer)* 2015: 243–249.

American Trucking Association. 2020. Economics and Industry Data. https://www.trucking.org/economics-and-industry-data

Association of American Railroads. 2019. Positive Train Control (PTC) Backgrounder. https://www.aar.org/wp-content/uploads/2018/04/AAR-Positive-Train-Control.pdf.

Bassok, A., E. D. McCormack, M. L. Outwater, C. Ta. 2011. Use of truck gps data for freight forecasting. *Transportation Research Board*, No. 11–3033.

Baykasoğlu, A., K. Subulan, A. Serdar Taşan, and N. Dudaklı. 2019. A review of fleet planning problems in single and multimodal transportation systems. *Transportmetrica a: Transport Science* 15 (2): 631–697.

Ben-Akiva, M., T. Toldeo, J. Santos, N. Cox, F. Zhao, Y. Lee, and V. Marzano. 2016. Freight data collection using GPS and web-based surveys: Insights from US truck drivers' survey and perspectives for urban freight. *Case Studies on Transport Policy* 4 (1): 38–44.

Benjamin, S.G., K.A. Brewster, R. Brümmer, B.F. Jewett, T.W. Schlatter, T.L. Smith, and P.A. Stamus. 1991. An Isentropic Three-Hourly Data Assimilation System Using ACARS Aircraft Observations. *Monthly Weather Review* 119: 4.

Bhattacharjee, S. 2021. Automatic Identification System (AIS): Integrating and Identifying Marine Communication Channels. Marine Navigation, URL: https://www.marineinsight.com/marine-navigation/automatic-identification-system-ais-integrating-and-identifying-marine-communication-channels/

Bierwirth, C., and F. Meisel. 2010. A survey of berth allocation and quay crane scheduling problems in container terminals. *European Journal of Operational Research* 202 (3): 615–627.

Bierwirth, C., and F. Meisel. 2015. A follow-up survey of berth allocation and quay crane scheduling problems in container terminals. *European Journal of Operational Research* 244 (3): 675–689.

Boada-Collado, P., S. Chopra, and K. Smilowitz. 2020. Partial Demand Information and Commitment in Dynamic Transportation Procurement. *Transportation Science* 54 (3): 588–605.

Bouarfa, S. 2020. Automated Aircraft Visual Inspection. *Proceedings of Agifors Aircraft Maintenance Operations Special Session 2020*.

Brouer, B., J.C. Alvarez, C. Plum, D. Pisinger, and M. Sigurd. 2014. A base integer programming model and benchmark suite for liner shipping network design. *Transportation Sci.* 48 (2): 281–312.

Budak, A., A. Ustundag, and B. Gulogu. 2017. A forecasting approach for truckload spot market pricing. *Transportation Research Part a: Policy and Practice* 97: 55–68.

Bullock, C. 2017. Data analysis is keeping planes flying. https://www.raconteur.net/technology/data-analytics/data-analysis-is-keeping-planes-flying/

Caballini, C., M.D. Gracia, J. Mar-Ortiz, and S. Sacone. 2020. A combined data mining – optimization approach to manage trucks operations in container terminals with the use of a TAS: Application to an Italian and a Mexican port. *Transportation Research Part e: Logistics and Transportation Review* 142: 102054.

Cambier, Y. 2018. ICF Big Data: Racing to platform maturity. https://www.aircraftit.com/articles/big-data-racing-to-platform-maturity/
Caplice, C. 2007. Electronic Markets for Truckload Transportation. *Production and Operations Management* 16: 423–436.
Caplice, C., and Y. Sheffi. 2003. Optimization-based procurement for transportation services. *Journal of Business Logistics* 16 (4): 423–436.
Che, C., H. Wang, Q. Fu, and X. Ni. 2019. Combining multiple deep learning algorithms for prognostic and health management of aircraft. *Aerospace Science and Technology* 94: 105423.
Conca, A., A. Di Febbraro, D. Giglio, and F. Rebora. 2018. Automation in freight port call process: Real time data sharing to improve the stowage planning. *ScienceDirect, Transportation Research Procedia* 30: 70–79.
Consilvio, A., P. Sanetti, D. Anguìta, C. Crovetto, C. Dambra, L. Oneto, N. Sacco. 2019. Prescriptive maintenance of railway infrastructure: From data analytics to decision support. *IEEE 2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*,1–10.
Cruciol, L. L. B. V., L. Weigang, J.-P. Clarke, L. Li. 2105. Air Traffic Flow Management Data Mining and Analysis for In-flight Cost Optimization. *Engineering and Applied Sciences Optimization, Computational Methods in Applied Sciences* 38, ed. N.D. Lagaros and M. Papadrakakis, Springer International Publishing, Switzerland.
Coraddu, A., L. Oneto, F. Baldi, D. Anguita. 2017. Vessels fuel consumption forecast and trim optimization: A data analytics perspective. *Ocean Engineering,* 130, 351–370. DCSA (2020), https://dcsa.org/
DAT Freight and Analytics. 2016. 63 Percent of Drivers are Detained for More Than 3 Hours Per Stop. https://www.dat.com/company/news-events/news-releases/63-percent-of-drivers-are-detained-for-more-than-3-hours-per-stop-dat-solutions
Diaz de Rivera, A., C.T. Dick, and L.E. Evans. 2020. Improving railway operational efficiency with moving blocks, train fleeting, and alternative single-track configurations. *Transportation Research Record* 2674 (2): 146–157.
Dingler, M.H., Y.C. Lai, and C.P.L. Barkan. 2010. Effects of Communications-Based Train Control and Electronically Controlled Pneumatic Brakes on Railroad Capacity. *Transportation Research Record: Journal of the Transportation Research Board* 2159: 77–84.
Dong, Chuanwen, et al. "The impact of emerging and disruptive technologies on freight transportation in the digital era: current state and future trends." *The International Journal of Logistics Management* (2021).
Erera, A., M. Hewitt, B. Kuracik, and M. Saveslbergh. 2009. Locating drivers in a trucking terminal network. *Transportation Research Part e: Logistics and Transportation Review* 45 (5): 988–1005.
Eurostat. 2017. Road freight transport by journey characteristics. *Eurostat*, https://ec.europa.eu/eurostat/statisticsexplained/index.php/Road_freight_transport_by_journey_characteristics#Road_transport_by_type_of_operation.
Flaskou, M., M.A. Dulebenets, M.M. Golias, S. Mishra, and R.M. Rock. 2015. Analysis of Freight Corridors Using GPS Data on Trucks. *Transportation Research Record* 2478 (1): 113–122.
Forest, F., Q. Cochard, C. Noyer, A. Cabut, M. Joncour, J. Lacaille, M. Lebbah, H. Azzag. 2020. Large-scale Vibration Monitoring of Aircraft Engines from Operational Data using Self-organized Models. *Proceedings of AGIFORS Annual Symposium 2020.*
Fosso Wamba, S., S. Akter, A. Edwards, G. Chopin, and D. Gnanzou. 2015. How "big data" can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics* 165: 234–246.
Garrido, R.A., and H.S. Mahmassani. 2000. Forecasting freight transportation demand with the space–time multinomial probit model. *Transportation Research Part b: Methodological* 34 (5): 403–418.
Gerum, P.C.L., A. Altay, and M. Baykal-Gürsoy. 2019. Data-driven predictive maintenance scheduling policies for railways. *Transportation Research Part c: Emerging Technologies* 107: 137–154.
Gharehgozli, A., D. Roy, and R. de Koster. 2016. Sea container terminals: New technologies and OR models. *Maritime Economics and Logistics* 18: 103–140. https://doi.org/10.1057/mel.2015.3.
Ghofrani, F., et al. 2018. Recent applications of big data analytics in railway transportation systems: A survey. *Transportation Research Part c: Emerging Technologies* 90: 226–246.
Godfrey, G.A., and W.B. Powell. 2002. An adaptive dynamic programming algorithm for dynamic fleet management, I: Single period travel times. *Transportation Science* 36 (1): 21–39.

Gonzalez, J., G. Battistello, P. Schmiegelt, J. Biermann. 2014. Semi-Automatic extraction of ship lanes and movement corridors from AIS data. *KFIE Fraunhofer Institute*.

Gorman, M.F., J.-P. Clarke, A. Hossein Gharehgozli, M. Hewitt, R. de Koster, and D. Roy. 2014. State of the Practice: A Review of the Application or OR/MS in Freight Transportation. *Interfaces* 44 (6): 535–554.

Huang, L., Y. Wen, Y. Zhang, C. Zhou, F. Zhang, T. Yang. 2020. Dynamic calculation of ship exhaust emissions based on real-time AIS data. *Transportation Research Part D: Transport and Environment*, 80.

IATA. 2019. Airline Maintenance Cost Executive Commentary.
https://www.iata.org/contentassets/bf8ca67c8bcd4358b3d004b0d6d0916f/mctg-fy2018-report-public.pdf

Ichoua, S., M. Gendreau, and J.Y. Potvin. 2006. Exploiting knowledge about future demands for real-time vehicle dispatching. *Transportation Science* 40 (2): 211–225.

11 Incredible facts about the $700 billion US trucking industry. *Business Insider*, https://markets.businessinsider.com/news/stocks/trucking-industry-facts-us-truckers-2019-5-1028248577#in-2017-the-american-trucking-industry-posted-revenues-higher-than-the-gdp-of-more-than-150-nations-1

Jimenez, V.J., Bouhmala, N. and Gausdal, A.H., 2020. Developing a predictive maintenance model for vessel machinery. *Journal of Ocean Engineering and Science*, 5(4), pp.358–386. John, S. 2019.

Jothi Basu, R., N. Subramanian, and N. Cheikhrouhou. 2015. Review of Full Truckload Transportation Service Procurement. *Transport Reviews* 35 (5): 599–621.

Kolley, L., N. Rückert, M. Kastner, C. Jahn, and K. Kathrin Fischer. 2022. Robust berth scheduling using machine learning for vessel arrival time prediction. *Flexible Services and Manufacturing Journal*. https://doi.org/10.1007/s10696-022-09462-x.

Kontopoulos, I., I. Varlamis, K. Tserpes. 2020a. A distributed framework for extracting maritime traffic patterns. *International Journal of Geographical Information Science*, 1–26.

Kontopoulos, I., I. Varlamis, K. Tserpes. 2020b. Uncovering Hidden Concepts from AIS Data: A Network Abstraction of Maritime Traffic for Anomaly Detection. In: Tserpes K., Renso C., Matwin S. (eds) *Multiple-Aspect Analysis of Semantic Trajectories*. MASTER 2019. Lecture Notes in Computer Science, 11889. Springer, Cham.

Kour, R., A. Thaduri, S. Singh, A. Martinetti. 2019. Big data analytics for maintaining transportation systems. *Transportation systems* (pp. 73–91). Springer, Singapore.

Lafkihi, M., P. Shenle, and E. Ballot. 2019. Freight transportation service procurement: A literature review and future research opportunities in omnichannel E-commerce. *Transportation Research Part e: Logistics and Transportation Review* 125: 348–365.

Lal, R., S. Johnson. 2018. Maersk betting on blockchain, *HBS* 9–518–089.

Land, A., A. Buus, and A. Platt. 2019. Data Analytics in Rail Transportation: Applications and Effects for Sustainability. *IEEE Engineering Management Review* 48 (1): 85–91.

Lee, H.K., S. Madar, S. Sairam, T.G. Puranik, A.P. Payan, M. Kirby, O.J. Pinon, and D.N. Mavris. 2020. Critical Parameter Identification for Safety Events in Commercial Aviation Using Machine Learning. *Aerospace* 7 (6): 73.

Lei, Y., F. Jia, J. Lin, S. Xing, and S.X. Ding. 2016. An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Transactions on Industrial Electronics* 63 (5): 3137–3147.

Lim, A., B. Rodrigues, Z. Xu, and Z. 2008. Transportation Procurement with Seasonally Varying Shipper Demand and Volume Guarantees. *Operations Research* 56 (3): 751–771.

Li, N., Haralambides, H., Sheng, H., Jin, Z. 2022. A new vocation queuing model to optimize truck appointments and yard handling-equipment use in dual transactions systems of container terminals, Computers & Industrial Engineering, Volume 169, 2022, https://doi.org/10.1016/j.cie.2022.108216.

Lindsey, C., and H. Mahmassani. 2017. Sourcing Truckload Capacity in the Transportation Spot Market: A Framework for Third Party Providers. *Transportation Research Part a: Policy and Practice* 102: 261–273.

Linn, R., J. Liu, Y.-W. Wan, C. Zhang. 2007. Predicting the Performance of Container Terminal Operations using Artificial Neural Networks. *Risk Management in Port Operations, Logistics and Supply Chain Security*, 1st Edition, 2007, Informa Law from Routledge.

Logistics Management. 2017. Making the Case for Transportation Management Systems,. *Logistics Management*, www.logisticsmgmt.com/article/making_the_case_for_transportation_management_systems.

Ma, X., Y. Wang, E. McCormack, and Y. Wang. 2016. Understanding Freight Trip-Chaining Behavior Using a Spatial Data-Mining Approach with GPS Data. *Transportation Research Record* 2596 (1): 44–54.

Maldonado, S., R.G. González-Ramírez, F. Quijada, and A. Ramírez-Nafarrate. 2019. Analytics meets port logistics: A decision support system for container stacking operations. *Decision Support Systems* 121: 84–93.

Mao, W., I. Rychlik, J. Wallin, and G. Storhaug. 2016. Statistical models for the speed prediction of a container shop. *Ocean Engineering* 126: 152–162.

Martey, E. N., L. Ahmed, N. Attoh-Okine. 2017. Track geometry big data analysis: A machine learning approach. *IEEE International Conference on Big Data (Big Data)*, 3800–3809

McCrea, B. 2020. The Future of Motor Freight. *Logistics Management*, https://www.logisticsmgmt.com/article/the_future_of_motor_freight

McMahon, P., T. Zhang, and R. Dwight. 2020. Requirements for Big Data Adoption for Railway Asset Management. *IEEE Access* 8: 15543–15564.

Mecer, D. 2020. Maintaining a Digital Twin for Aircraft during Operation. *Proceedings of Agifors Aircraft Maintenance Operations Special Session 2020*.

Meert, W. 2019. Machine Learning for Airplane Maintenance. *Proceedings of Agifors Aircraft Maintenance Operations Special Session 2019*.

Miller, J.W., Y. Bolumole, and M.A. Schwieterman. 2020a. Electronic Logging Device Compliance of Small and Medium Size Motor Carriers Prior to the December 18, 2017. *Journal of Business Logistics* 41: 67–85.

Miller, J., Y. Nie, and X. Lui. 2020b. Hyperpath Truck Routing in an Online Freight Exchange Platform. *Transportation Science* 54 (6): 1676–1696.

Mulder, J., and R. Dekker. 2014. Methods for strategic liner shipping network design. *European Journal of Operational Research* 235 (2): 367–377.

Mulder, J., W. van Jaarsveld, and R. Dekker. 2019. Simultaneous Optimization of Speed and Buffer Times with an Application to Liner Shipping. *Transportation Science* 53 (2): 365–382.

Munim, Z.H., and H. Haralambides. 2022. Advances in maritime autonomous surface ships (MASS) in merchant shipping. *Maritime Economics and Logistics* 24: 181–188.

Nandiraju, S., A. Regan. 2008. Freight transportation electronic marketplaces: a survey of the industry and exploration of important research issues. *UC Berkeley: University of California Transportation Center*, https://escholarship.org/uc/item/9fj2c4jw.

Nguyen, K.T.P., and K. Medjaher. 2019. A new dynamic predictive maintenance framework using deep learning for failure prognostics. *Reliability Engineering and System Safety* 188: 251–262.

Oracle (2018). Oracle Transportation Management Cloud. *Oracle*, http://www.oracle.com/us/products/applications/046953.pdf

Ordóñeza, C., F.S. Lasherasb, J. Roca-Pardiñasc, and F.J. de CosJueza. 2019. A hybrid ARIMA–SVM model for the study of the remaining useful life of aircraft engines. *Journal of Computational and Applied Mathematics* 346 (15): 184–191.

Pall, E., K. Mathe, L. Tamas, L. Busoniu. 2014. Railway track following with the AR. Drone using vanishing point detection. *Proceedings of 2014 IEEE International Conference on Automation, Quality and Testing, Robotics,* AQTR 2014.

Peckham, O. 2020. For American Airlines, Machine Learning Solves an Air Cargo Conundrum. *Datanami*. https://www.datanami.com/2020/05/14/for-american-airlines-machine-learning-solves-an-air-cargo-conundrum/

Port of Rotterdam puts Internet of Things platform into operation. 2019. https://www.portofrotterdam.com/en/news-and-press-releases/port-of-rotterdam-puts-internet-of-things-platform-into-operation.

Powell, W. B. 2007. *Approximate Dynamic Programming: Solving the curses of dimensionality* (Vol. 703). John Wiley & Sons.

Pozzi, S., C. Valbonesi, V. Beato, R. Volpini, F. M. Giustizieri, F. Lieutaud, A. Licu. 2015. Safety Monitoring in the Age of Big Data. *Ninth USA/Europe Air Traffic Management Research and Development Seminar.*

Rail Safety Improvement Act of 2008, H. R. 2095, 110th Congress, 2nd session, 2008.

Rajapakshe, T., M. Dawande, S. Gavirneni, C. Sriskandarajah, and P. Rao. 2014. Dedicated Transportation Subnetworks: Design, analysis, and Insights. *Production and Operations Management* 23: 138–159.

Robinson, A. 2020. Sonar Indexes & Insights: What is the outbound tender reject index (OTRI) & What is it saying about freight markets? *FreightWaves*, https://sonar.freightwaves.com/freight-market-blog/outbound-tender-reject-index.

Rizzo, S. G., J. Lucas, Z. Kaoudi, J.-A. Quiane-Ruiz, S. Chawla. 2019. AI-CARGO: A Data-Driven Air-Cargo Revenue Management System. arXiv:1905.09130 [cs.AI]

Roy, A. (2001). Secure aircraft communications addressing and reporting system (ACARS). 20th DASC. 20th Digital Avionics Systems Conference (Cat. No.01CH37219), Daytona Beach, FL, USA, 2001, pp. 7A2/1–7A2/11 vol.2.

Roy, D., R. De Koster, and R. Bekker. 2020. Modeling and Design of Container Terminal Operations. *Operations Research* 68 (3): 686–715.

Saki, M., M. Abolhasan, and J. Lipman. 2019. A Novel Approach for Big Data Classification and Transportation in Rail Networks. *IEEE Transactions on Intelligent Transportation Systems* 21 (3): 1239–1249.

Salierno, G., S. Morvillo, L. Leonardi, G. Cabri. 2020. An Architecture for Predictive Maintenance of Railway Points Based on Big Data Analytics. *International Conference on Advanced Information Systems Engineering*, 29–40. Springer, Cham.

Sanchez, I. H., L. P. H. Mukti. 2018. Uberization Effects on Freight Procurement. Master Thesis, Massachusetts Institute of Technology, http://hdl.handle.net/1721.1/118121

Schwartz, B., and S.G. Benjamin. 1995. A Comparison of Temperature and Wind Measurements from ACARS-Equipped Aircraft and Rawinsondes. *Weather and Forecasting* 10 (3): 528–544.

Scott, A., A. Balthrop, J. Miller. 2020. Unintended responses to IT-enabled monitoring: The case of the electronic logging device mandate. *Journal of Operations Management*, 1– 30.

Simão, H.P., J. Day, A.P. George, T. Gifford, J. Nienow, and W.B. Powell. 2009. An approximate dynamic programming algorithm for large-scale fleet management. *Transportation Science* 43 (2): 178–197.

Simão, H.P., A. George, W.B. Powell, T. Gifford, J. Nienow, and J. Day. 2010. Approximate dynamic programming captures fleet operations for Schneider National. *Interfaces* 40 (5): 342–352.

Slattery, A. 2017. AI takes to the sky with IBM Watson. *Computer Business Review, Oct. 2017*.

Stephens, B. 2021. BNSF receives patent for moving block system. *Trains.* February 9, 2021

STM (2020), https://www.seatrafficmanagement.info/.

Sun, J., F. Wang, and S. Ning. 2020. Aircraft air conditioning system health state estimation and prediction for predictive maintenance. *Chinese Journal of Aeronautics* 33 (3): 947–955.

Tarawneh, C., J. Ley, D. Blackwell, S. Crown, B. Wilson. 2018. Onboard Load Sensor for Use in Freight Railcar Applications. *International Journal of Railway Technology*

Thaduri, A., D. Galar, and U. Kumar. 2015. Railway assets: A potential domain for bigdata analytics. *Procedia Comput. Sci.* 53: 457–467.

TB&P. 2020. Driver shortage persists amid COVID-19 pandemic, mixed freight demand. https://talkbusiness.net/2020/12/driver-shortage-persists-amid-covid-19-pandemic-mixed-freight-demand/

Tsai, M.-T., J.D. Saphores, and A. Regan. 2011. Valuation of Freight Transportation Contracts Under Uncertainty. *Transportation Research Part e: Logistics and Transportation Review* 47: 920–932.

Unctad 2021. Review of Maritime Transport 2021. https://unctad.org/webflyer/review-maritime-transport-2021

Ulmer, M.W., J.C. Goodson, D.C. Mattfeld, and M. Henning. 2019. Offline-Online Approximate Dynamic Programming for Dynamic Vehicle Routing with Stochastic Requests. *Transportation Science* 53 (1): 185–202.

Varfis, G. 2020. Predictive Maintenance (PDM) is not an emerging technology- APSYS. *Proceedings of Agifors Aircraft Maintenance Operations Special Session 2020.*

Venskus J., P. Treigys, J. Bernataviciene, G. Tamulevicius, and V. Medvedev. 2019. Real-Time Maritime Traffic Anomaly Detection Based on Sensors and History Data Embedding. *Sensors.*

Verma, R., S. Saikia, H. Khadilkar, P. Agarwal, G. Shroff, and A. Srinivasan. 2019. A Reinforcement Learning Framework for Container Selection and Ship Load Sequencing in Ports. In Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '19). International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2250–2252.

Verwijmeren, M. 2004. Software component architecture in supply chain management. *Computers in Industry* 53 (2): 165–178.

Wahlström, J., I. Skog, and P. Händel. 2017. Smartphone-based vehicle telematics: a ten-year anniversary. *IEEE Transactions on Intelligent Transportation Systems* 18 (10): 2802–2825.

Wang, S., and Q. Meng. 2012a. Liner ship route schedule design with sea contingency time and port time uncertainty. *Transportation Research Part b: Methodology* 46 (5): 615–633.
Wang, S. and Q Meng. 2012b. Robust schedule design for liner shipping services. *Transportation Res. Part E Logist. Transportation Rev,* 48(6), 1093–1106.
Wang, F., J. Sun, and X. Liu. 2019. Aircraft auxiliary power unit performance assessment and remaining useful life evaluation for predictive maintenance. *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*.
Wang, Y., and Q. Meng. 2019. Integrated methods for forecasting container slot booking in intercontinental liner shipping service. *Flexible Services and Manufacturing Journal* 31: 653–674.
Wu, P.-J., M.-C. Chen, and C.-K. Tsau. 2017. The data-driven analytics for investigating cargo loss in logistics systems. *International Journal of Physical Distribution & Logistics Management* 47 (1): 68–83.
Xiao, W., C. Xu, H. Liu, H. Yang, and X. Lui. 2020. Short-term truckload spot rates' prediction in consideration of temporal and between-route correlations. *IEEE Access* 8: 81173–81189.
Xin, L., S. Tianyun, and M. Xiaoning. 2020. Research on the Big Data Platform and Its Key Technologies for the Railway Locomotive System. *Proceedings of the 2020 5th International Conference on Big Data and Computing,* pp. 6–12.
Yeo, G., S.H. Lim, L. Wynter, and H. Hassan. 2019. MPA-IBM Project SAFER: sense-making analytics for maritime event recognition. *INFORMS Journal on Applied Analytics* 49 (4): 269–280.
Zhang, B., T. Yao, T.L. Friesz, and Y. Sun. 2015. A tractable two-stage robust winner determination model for truckload service procurement via combinatorial auctions. *Transportation Research Part b* 78: 16–31.
Zhang, L., Q. Meng, and F.F. Tien. 2019. Big AIS data based spatial-temporal analyses of ship traffic in Singapore port waters. *Transportation Research Part E* 129: 287–304.
Zhang, M. 2018. Big Data / Machine Learning at American Airlines Tech. Ops. *Proceedings of Agifors Annual Symposium 2018*.
Zhang, W.Y., Y. Lin, Z.S. Ji, and G.F. Zhang. 2008. Review of containership stowage plans for full routes. *Journal of Marine Science and Applications* 7: 278–285.
Zhao, Y., and P. Ioannou. 2015. Positive train control with dynamic headway based on an active communication system. *IEEE Transactions on Intelligent Transportation Systems* 16 (6): 3095–3103.
Zheng, Y., W.K. Talley, D. Jin, and M. Ng. 2016. Crew injuries in container vessel accidents. *Maritime Policy & Management* 43 (5): 541–551.
Zhang, B., Zhang, N., Li, H., Liu, F., & Miao, K. 2009. An efficient cloud computing-based architecture for freight system application in China railway. In *IEEE International Conference on Cloud Computing* (pp. 359–368). Springer, Berlin, Heidelberg.