



## **A Review of Big Data in Road Freight Transport Modeling: Gaps and Potentials**

Downloaded from: <https://research.chalmers.se>, 2025-03-31 23:34 UTC

Citation for the original published paper (version of record):

Shoman, W., Yeh, S., Sprei, F. et al (2023). A Review of Big Data in Road Freight Transport Modeling: Gaps and Potentials. *Data Science for Transportation*, 5(2).  
<http://dx.doi.org/10.1007/s42421-023-00065-y>

N.B. When citing this work, cite the original published paper.



# A Review of Big Data in Road Freight Transport Modeling: Gaps and Potentials

Wasim Shoman<sup>1</sup> · Sonia Yeh<sup>1</sup> · Frances Sprei<sup>1</sup> · Jonathan Köhler<sup>2</sup> · Patrick Plötz<sup>2</sup> · Yancho Todorov<sup>3</sup> · Seppo Rantala<sup>3</sup> · Daniel Speth<sup>2</sup>

Received: 7 July 2022 / Revised: 28 December 2022 / Accepted: 13 February 2023  
© The Author(s) 2023

## Abstract

Road transport accounted for 20% of global total greenhouse gas emissions in 2020, of which 30% come from road freight transport (RFT). Modeling the modern challenges in RFT requires the integration of different freight modeling improvements in, e.g., traffic, demand, and energy modeling. Recent developments in 'Big Data' (i.e., vast quantities of structured and unstructured data) can provide useful information such as individual behaviors and activities in addition to aggregated patterns using conventional datasets. This paper summarizes the state of the art in analyzing Big Data sources concerning RFT by identifying key challenges and the current knowledge gaps. Various challenges, including organizational, privacy, technical expertise, and legal challenges, hinder the access and utilization of Big Data for RFT applications. We note that the environment for sharing data is still in its infancy. Improving access and use of Big Data will require political support to ensure all involved parties that their data will be safe and contribute positively toward a common goal, such as a more sustainable economy. We identify promising areas for future opportunities and research, including data collection and preparation, data analytics and utilization, and applications to support decision-making.

**Keywords** Big data · Big data analytics · Road freight transport · Transport modeling

## Introduction

The road freight transport (RFT) sector faces challenges in reducing costs and incorporating sustainability measures into operations (Kumar et al. 2019). The demand for freight has increased drastically and causes severe traffic congestion, carbon emissions, and reduced air quality (Chankaew et al. 2018). Road transport alone accounts for 20% of global total greenhouse gas (GHG) emissions (Santos 2017); 30% of this comes from RFT (Ge and Friedrich 2020). Inefficiency (e.g., between 40 and 60% empty trips), redundant trips, lack of synchronization between transport modes, and the ability to deal with disruption and risks (Kumar et al. 2019)

are challenges that the road freight sector faces. In addition, new societal challenges toward energy transition, such as electric vehicle operations and charging infrastructure allocation, require new data, modeling types, and analytics (Wang and Sarkis, 2021). The abovementioned challenges mainly concern traffic, demand, and energy modeling fields.

Emerging data sources from the drastic take-off of the internet of things (IoT) and information communication technology (ICT) have generated massive quantities of data and the opportunities to create new insights, business value, and novel revenue streams. These emerging data sources, commonly called Big Data (Tiwari et al. 2018), raise various challenges in analysis and management. Big Data provides many opportunities to gain insights into RFT operation, planning, and infrastructure development. However, despite existing and future promising applications of Big Data, its critical challenges have not been systematically reviewed. These challenges might hinder further successful applications and value generation. Reviews on Big Data in freight have been narrower and focused on specific applications (Ktrakazas et al. 2019). For example, Chung (2021) reported contributions made in the applications of Big Data

✉ Wasim Shoman  
wasim.shoman@chalmers.se

<sup>1</sup> Department of Space, Earth and Environment, Chalmers University of Technology, 41296 Göteborg, Sweden

<sup>2</sup> Fraunhofer Institute for Systems and Innovation Research ISI, Breslauer Strasse 48, 76139 Karlsruhe, Germany

<sup>3</sup> VTT Technical Research Centre of Finland, Tietotie 4C, 02100 Espoo, Finland

in improving logistics operations and transportation network efficiency. Tiwari et al. (2018) reviewed the importance and impact of Big Data and its application in supply chain management. Borgi et al. (2017) presented a survey on Big Data opportunities in improving transport and logistics operational efficiency. They present some prominent use cases with Big Data and identify future research challenges.

The premise of this paper is that new digital technologies in road freight offer new opportunities for analysis and management of freight transport. Digital technologies generate unprecedentedly large and highly detailed (spatially and temporally) data at individual and aggregated levels. This paper reviews the opportunities and challenges of Big Data analysis in RFT presented in the literature. To identify relevant articles, we searched the Google Scholar database with a combination of the following keywords: road freight, commercial vehicle, Big Data, and Big Data analytics within the scope of "traffic modeling," "demand modeling," and "energy modeling." The search results in around 980 manuscripts. The query was further adjusted to search the indexed articles' titles, abstracts, and keywords by devising the following inclusion and exclusion criteria. First, we considered manuscripts written in English in peer-reviewed journal articles, books, conference proceedings, and technical reports of credible institutions such as the EU commission. Then, we included only manuscripts that elaborated on Big Data's surveys, applications, technical architecture, and implementation in RFT. After further screening, 98 manuscripts were included and subsequently reviewed individually.

This review paper contributes to the literature on Big Data in RFT applications in two ways: i) it provides an overview of Big Data applications in RFT modeling, specifically concerning traffic, demand, and energy; (ii) it identifies challenges and gaps that currently exist and hinder the broader use or adoption of Big Data applications in RFT modeling. The paper is organized as follows. "Definition and types of Big Data in road freight applications" provides the definition and a review of Big Data and types of Big Data in RFT applications. "Big Data analytics and applications in RFT modeling" reviews Big Data analytics and their categories and applications in RFT modeling. We discuss the current organizational, technical, legal, and political challenges toward a broader usage of Big Data in RFT in "Challenges in using Big Data in freight modeling". We provide suggestions to address these challenges and future opportunities in "Future Opportunities". Finally, we finish with a discussion and conclusions in.

## Definition and Types of Big Data in Road Freight Applications

The study of Big Data is evolving continuously, including more than just large volumes of data. The main attributes

of Big Data include the "5 V" concept: volume, velocity, variety, verification/veracity, and value (Tiwari et al. 2018). Volume refers to the quantities of data produced. Velocity refers to the speed of data collection, transfer, and storage; and the extraction speed of discovering useful knowledge and decision-making models (Hee et al. 2018). Variety focuses on the variety of data produced by detectors, sensors, and even social media (Zhu et al. 2018). Verification/veracity refers to uncertainty regarding the quality inherent in traffic and other transportation data, such as inaccurate or incomplete data (Neilson et al. 2019). Finally, value refers to revealing underexploited revenue sources from Big Data to support decision-making (Nguyen et al. 2018).

Big data in the freight transport system comes from many sources on the road and inside the vehicle, see Table 1. Milne and Watling (2019) described the characteristics of data sources that fit the Big Data concept from a transport perspective as i) continuous monitoring with a high temporal sampling rate and without gaps, ii) data may not be owned by the data analyzer, iii) data collection may not have been designed for transport applications and iv) there may be an ability to link multiple contemporaneous data sources.

The RFT Big Data applications utilize many data sources that differ in their data collection methods, source, or entity that collects the data, and their resolution and collected information. For example, current roadside sensors, owned mainly by local transport authorities, are not new to transport professionals. These data sources already generate valuable information for traffic modeling, such as the speed of vehicles, vehicle types, and congestion (Neilson et al. 2019; Verendel and Yeh 2019).

The private sector collects most vehicle-level data from vehicle tracking devices. For example, many private freight entities have records of their businesses, e.g., global positioning systems (GPS) measurements of vehicles. However, they are usually not shared externally due to data privacy, monetary value, and commercial secrecy concerns. This data type has a high spatial and temporal resolution (often with a spatial and temporal resolution of less than one meter and one second, respectively). The data are sampled for every vehicle and contains important information for demand and energy modeling, such as visited locations, trip chains, and vehicle energy consumption.

Other data types are not explicitly collected for transport applications nor by transport entities. However, they could be converted to relevant transport information. An example is data from aerial sensors that initially come from imprints of some physical phenomenon, such as sound or light, acquired by a specific sensor and converted into a digital signal. Sensing devices may include microphones, cameras, thermal sensors, etc. These can be video streams from in-vehicle, surveillance, or roadside cameras (e.g., CCTV cameras for road traffic monitoring), motion (accelerometer

**Table 1** Sources, examples, and characteristics of Big Data types in road freight transport modeling

Data collection method or devices	Examples of collection devices	Spatial and temporal resolution and sampling rate	Examples of collected data
Road mounted sensors	Roadside sensors (e.g., Bluetooth-enabled devices and traffic counting devices), and tolls	Spatial resolution depends on the mounted sensors (e.g., the number of sensors mounted on roads). High temporal resolution and sampling rate	Active vehicle numbers on the road and specifications and traffic information (e.g., average speed, traffic count)
Vehicle tracking devices	GPS measurements, and on-board diagnostics (OBD) from the automotive electronic system	High temporal and spatial resolution and sampling rate	Individual vehicle movements, stop locations, and energy consumption profiles
Aerial sensors	CCTV cameras, ambient gauges and aerial/satellite cameras and sensors	Spatial and temporal resolution and sampling rates depend on the data provider and collection method	Multimedia data (e.g., images and videos), land use thematic maps, remotely sensed data, e.g., imageries, and temperature
Shipment identifiers "reported multidimensional data"	Radio frequency identification and barcode scanners	The sampling rate is vehicle or cargo based. Spatial and temporal resolutions are not relevant	Supplementary data about cargos and shipment status from warehouses
Communication media	Phones, smartphones, computers, and tablets	Spatial and temporal resolutions and sampling rates depend on the data provider and collection method	Text such as SMS and emails, human language, RSS feeds, and social media

data), or electromagnetic reflectance of objects (Borgi et al. 2017).

Many untapped Big Data types are collected from communication media devices, such as text data from call center applications, semi-structured data from various business-to-business processes, and geospatial data in logistics (Misra and Bera 2018). The current Big Data analytics techniques open doors to utilize unstructured data, e.g., text and human language, and semi-structured data, e.g., XML, RSS feeds, and social media. The unstructured data can be joined with structured data, which previously held unchallenged hegemony in analytics, to gain additional insights. For instance, social media data from Twitter and Facebook posts are exploited for new insights derived from freight transport applications (Singh et al. 2018; Wang and Sarkis, 2021).

Both data from the analog and communication devices have a range of spatial and temporal resolutions and sampling rates. These resolutions depend on the ability of the data provider to deliver details without revealing the business's or user's personal information. It also depends on the data collection method, such as cell towers, phone sightings, or phone apps, when tracking drivers' movements.

Another data type is the shipment data reported from warehouses, workers, or drivers by scanning barcodes or via radio frequency identification. This data can add historical context to other collected data or supplementary information (Jensen et al. 2004). The sampling rate of this data depends on the delivery to depots or change to transport mode.

### Big Data Analytics and Applications in RFT Modeling

Big Data analytics applies advanced techniques, including artificial intelligence (AI), agent-based modeling, and optimization of big datasets to convert them into new knowledge (Tiwari et al. 2018; Zhu et al. 2018). Its principal function is to examine and analyze large amounts of structured and unstructured data or to fuse different data types to uncover hidden patterns such as correlations, trends, and other valuable information and create new knowledge.

Based on the application's aim, Big Data analytics techniques in freight transport can be categorized into descriptive, predictive, and prescriptive analytics, see Table 2. Descriptive analytics answer what happened in the past or happens in the present. For instance, descriptive real-time information about the location and quantities of goods provides managers with information to adjust delivery schedules, place replenishment orders and emergency orders, change transportation modes, etc. Predictive analytics make forecasts based on past trends and answer the question of what will happen given certain assumptions (Blackburn et al. 2015; Mesgarpour and Dickinson 2014). Prescriptive analytics derive decision recommendations based on descriptive, predictive analytics, and mathematical models.

**Table 2** Categories of Big Data analytics applications in road freight transport modeling

Analytical type	Scope of transport modeling	Applications	Utilized methods
Descriptive (What has happened)	Traffic	Shared real-time traffic information (e.g., travel time, route choice, traffic count, congestion, and safety analysis) (Tiwari et al. 2018) and object detection (Kaffash et al. 2021; Yuan et al. 2017)	Online analytical processing and visualization tools supported by real-time information and reporting technology (Thakur et al. 2015; Gingerich et al. 2016; Tiwari et al. 2018)
	Demand	Driving behavior (e.g., start-stop locations) (Forrest et al. 2020; Gingerich et al. 2016), Origin–Destination (OD) matrix extraction (Ma et al. 2011; Zanjani et al. 2015) and trip chain generation (Alho et al. 2018; Joubert and Axhausen 2013; Khan and Machemehl 2017; Laranjeiro et al. 2019; Romano Alho et al. 2019; Thakur et al. 2015; Venkatarahan et al. 2020; Yang et al. 2014)	
	Energy	Vehicle power or fuel consumption (Cho and Choi 2017; Wei et al. 2022)	
Predictive (What will happen)	Traffic	Traffic control, management and optimization (Costa et al. 2017; Kaffash et al. 2021; Kim et al. 2017; Lv et al. 2015; Mehmood et al. 2017; Mou 2020), route optimization, and travel time (Chen, 2020; Hess et al. 2015)	Statistical analysis (Costa et al. 2017; Mehmood et al. 2017; Yves R. et al., 2018), simulation (Naumov et al. 2022; Perrotta et al. 2019) and optimization (Basso et al. 2021; Chen, 2020; Mou 2020)
	Demand	Demand forecast (Lee 2017; Yves R. et al., 2018) and supply chain (Chen, 2020; Giusti et al. 2019)	
	Energy	Vehicle energy or fuel demand projections (Basso et al. 2021; Cho and Choi 2017; Dimokas et al. 2020; Du et al. 2022; Gao et al. 2017; Perrotta et al. 2019)	
Prescriptive (What should be done)	Traffic	Identifying traffic safety measures on roads (Shi and Abdel-Aty 2015)	Statistical analysis, mathematical optimization (Shi and Abdel-Aty 2015) (Figueiras et al. 2019), simulation, and multi-criteria decision-making (Shi and Abdel-Aty 2015)
	Demand	Connected vehicles/autonomous vehicles routing (Dimokas et al. 2021; Parmar et al. 2019) and dynamic toll charging system (Figueiras et al. 2019)	
	Energy	Electric charging infrastructure and optimization for supply chains to reduce emissions & energy consumption (Amol A Phadke et al. 2021; Plötze and Speth 2021; Speth et al. 2022)	

It answers the question of what should be done to achieve specific goals (Souza 2014; Wang et al. 2016). Predictive and prescriptive analytics are vital in helping freight-related entities make effective decisions on the organization's strategic direction (Munizaga 2019). They can be applied to address problems related to the changes in organizational culture, sourcing decisions, and the development of products or services.

Here, we further elaborate on these analytical categories, showing their importance to RFT with examples of relevant applications. We also discuss potential issues for some of these applications.

## Descriptive Applications

Descriptive analytics are helpful in illustrating the current status quo, such as total stock in inventory, average money spent per customer, year-to-year changes in sales (Tiwari et al. 2018), and total energy consumption. In RFT, two major descriptive applications are origin–destination (OD) matrix extraction and trip chaining in modeling freight demand. Trip chains provide new perspectives for analyzing freight systems (Duan et al. 2020). A trip chain for a vehicle denotes a set of trips between 'significant' locations (e.g., depots, shops, etc.) before ending its journey. A complete trip chain determines the travel patterns of a higher spatial and temporal resolution table of the OD matrix. It captures the behavior of freight vehicles, including the locations of vehicles' activities, the duration of activities, the frequency of visiting these locations, and the sequence in which they are visited (Joubert and Meintjes 2015). For example, trade-offs between customer demand and logistics costs and long operational hours in goods shipments often demand multiple tours by a single commercial vehicle in its daily operation (Ruan et al. 2012).

Both OD matrix and trip chain applications require large amounts of data and both can be combined with auxiliary sources. Ma et al. (2011) developed a performance measures program to monitor the travel generated by 2500 trucks in Washington, USA, for more than a year. The trucks responded to shippers' business needs by picking up goods and dropping them off at their destinations. The program extracts OD pairs from the GPS readings of the individual truck to quantify truck travel characteristics and performance. However, they also raise important data and analytics issues, including the lack of representativeness for the entire fleet, data being expensive to acquire, potential privacy issues, and the need for anonymization. Moreover, truck GPS data from different vendors use different formats and communication methods. One of the largest uncertainties comes from the arbitrary thresholds (e.g., durations to identify stops) used to determine the depots for the OD identification due to the lack of other data sources, such as travel

diaries that could identify or verify stop locations (Gingerich et al. 2016).

Romano Alho et al. (2019) demonstrated the importance of carefully selecting and disclosing the methods for data processing by illustrating that the predictions of tours, tour types, and tour chains are highly dependent on assumptions. They implement high-resolution tour-chain algorithms for a large-scale GPS-based survey and use stop-level data from a driver survey conducted from 2017 to 2019 in Singapore. The GPS data on its own is not enough to identify the fleet's stops and activity types. The integration of stop-level and point of interest (POI) data offers a different way to determine the base locations. Like other studies, they use arbitrary thresholds, e.g., durations, to identify depots or stops, activities, and chaining of trips, i.e., tours.

Duan et al. (2020) proposed a multiscale depot-identified method based on the density-based spatial clustering of applications with noise algorithms using large-scale vehicle trajectory datasets from GPS equipment, POI, and other spatial data. They extract the travel patterns of 1852 freight vehicles over one month collected from multiple provinces in China. The study implements data mining algorithms to identify the base depots and trip ends for multiday freight trip chains. The multi-features of the trip chain quantitatively describe each vehicle, including the number of average trip chains, the number of average stops per trip chain, the average dwell time per stop, and the average travel time per trip. The results show that some travel patterns are limited to specific vehicle types. The study reports that limitations include temporal resolution, small quantity, and noncomplete trip chains in some cases. They also note that traditional travel surveys and diaries can be combined with ground-truth data and land use structures to validate the accuracy of the results and reduce errors.

## Predictive Applications

Many freight professionals are keen to improve demand forecasting and production planning with Big Data (Chase 2013). Predictive analytics make predictions for decision support by exploring data patterns, such as signaling demand changes, determining optimal prices, and tracing consumer loyalty data. These improvements help detect new market trends and determine the root causes of failures, issues, and defects (Tiwari et al. 2018). Forecasting involves identifying relevant information and sources and selecting appropriate methods for computing the projected demand. Given the volatile, uncertain, complex, and ambiguous business environment, greater research attention is needed in demand forecasting.

Predictive analytics explore data patterns to make decision-support predictions (Tiwari et al. 2018). Regression, time-series analysis, and machine-learning techniques are

commonly used in forecasting demand and energy (Borgi et al. 2017; Seyedan and Mafakheri 2020). Forecasting demand includes studying customer behavior, purchasing patterns, and identifying trends in sales activities that are associated with the demand. Transport demand forecasting is critical to managing supply chains, as tactical and operational decisions on production planning, inventory levels, logistics, and scheduling are based on anticipated demand development. Blackburn et al. (2015) presented a time series model that considers historical and business-related Big Data, such as economic or industry-related information. The study utilizes different data sources, e.g., national statistics institutions and national input–output tables, to describe the links between supply and customer industry and expected trends. The study shows that integrating different information sources improves predictions and substantially decreases uncertainty. The study also notes important issues that impact the predictive analytics application: the availability of data, adequate analytical methods, expertise from various functions, and top management support. Furthermore, the accuracy is sensitive to the prediction model.

Electric vehicles (EVs) in freight can reduce fossil fuel consumption and lower GHG emissions. One significant challenge in their development is measuring the vehicles' energy demand for electricity. Most existing energy consumption data use laboratory-based engine dynamometer tests based on steady-state modal tests (Christopher Frey and Kim 2006; Rosero et al. 2021). Such tests may not represent real-world duty cycles and conditions, especially when driving in urban areas (Lárusdóttir and Ulfarsson 2015). Some studies implement a vehicle energy simulator to estimate fuel consumption (Wang and Rakha 2018). However, it is time-consuming to construct numerous input profiles through a graphical user interface. Overall, operating conditions can be categorized based on (i) the vehicle design; (ii) driver characteristics; (iii) travel conditions; (iv) traffic flow conditions (e.g., congestion); (v) road conditions; (vi) ambient conditions and (viii) regenerative braking (Bucher and Bradley 2018; Dias et al. 2019; Nicolaides et al. 2019). For instance, a number of studies (Cicconi et al. 2016; Kretzschmar et al. 2016; Rosero et al. 2021; Xu et al. 2015) use GPS measurements to model energy consumption for each vehicle as a function of operating conditions (e.g., travel conditions and driver characteristics) (Kretzschmar et al. 2016; Xu et al. 2015). Several operational factors or conditions can influence energy consumption prediction in actual traffic conditions. Energy consumption could be estimated more accurately from GPS measurements, portable emissions, and vehicle energy measurements.

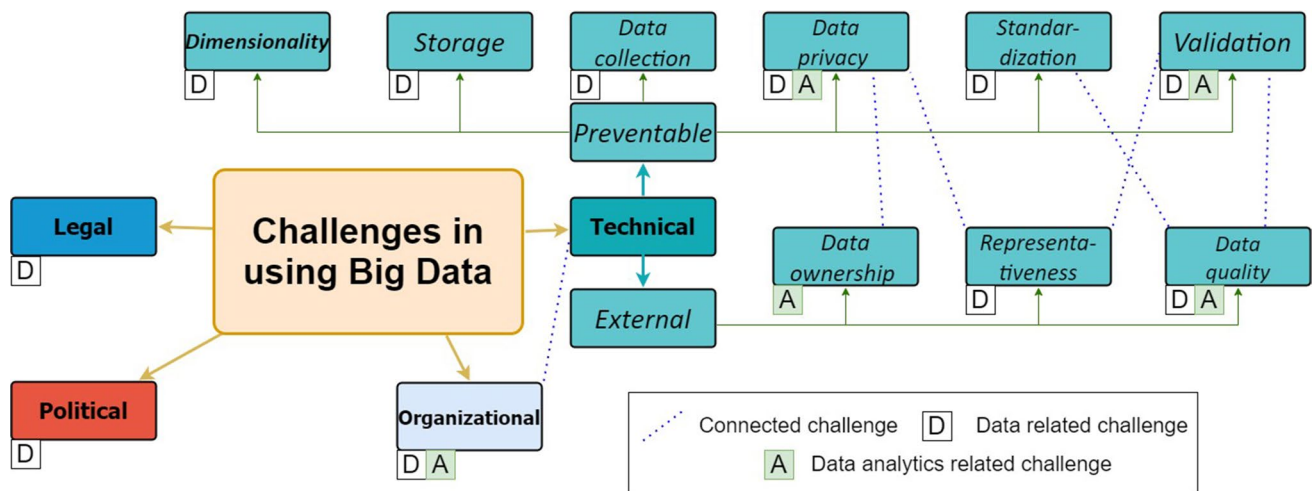
## Prescriptive Applications

Prescriptive analytics define goals such as strategic business or environmental targets within a specified time frame (Franzen et al. 2019). Prescriptive analysis drives decisions based on descriptive and predictive analytics using mathematical optimization, simulation, or multi-criteria decision-making techniques (Tiwari et al. 2018).

In energy applications, a growing prescriptive analysis application is the planning of charging infrastructure, i.e., the deployment of charging facilities (e.g., stationary chargers) according to current and predicted demand (Kavianipour et al. 2021). Arias and Bae (2016) provided a simulation model to estimate electric vehicle charging demand using real-world traffic distribution data and weather conditions in South Korea. The model includes classifying traffic patterns, pointing out influential factors such as charging starting time, charger power type, and the initial state of charge, and establishing classification criteria. Recent prominent examples are electric truck charging location placement studies in the EU by (ACEA and T&E, 2021; Plötz et al. 2021). Both decide the locations of future charging infrastructure by matching the existing truck stop's locations extracted from the on-board GPS measurements. Individual locations of each truck are clustered over many trucks and certain observation periods, i.e., 1 year. The most populated truck stops, i.e., those where most trucks stop or where many trucks stop frequently, could serve as ideal locations for constructing truck charging infrastructure.

The new connected vehicle technology is also an excellent example of new perspective applications in RFT. A connected vehicle uses GPS, communication technologies, and sensors to communicate with devices within the vehicle, other vehicles, and transportation networks. The vehicle needs to allocate its and other vehicle's positions, predict their movements, and plan movements accordingly (Dimokas et al. 2021). The potential of connected vehicles extends beyond providing personalized services to drivers, offering an excellent opportunity to improve traffic safety and transportation systems' sustainability by automatically sharing relevant information with nearby vehicles and transportation networks (Neilson et al. 2019).

The digital twin approach replicating a real road transport network operation through data and simulation also opens a new perspective toward fleets' decarbonization planning and forecasting costs. Such an approach can alleviate the risks of costly mistakes in planning and initial piloting during a fleet transitioning phase and further operational optimization and infrastructure planning (Vehviläinen et al. 2022). VTT Smart eFleet toolbox recreates the transport network by utilizing validated vehicle models and reconstruction of the road infrastructure. The road infrastructure includes practicalities (such as road curvature, elevation, and traffic),



**Fig. 1** Challenges in using Big Data in road freight modeling

positioning and capacity of the charging/refueling infrastructure, and environmental considerations such as seasonal temperature. The approach has been used in planning and evaluating various fleet deployment scenarios, including techno-economical evaluations, roadmaps for fleet decarbonization, and planning future policy measures. For example, Pihlatie et al. (2021) prescribed possible pathways for the growth of zero-emission transport in Finland, including the freight and logistics segment. In supporting the planning of governmental policy measures to speed up company adoption of vehicles with low-carbon energy sources, the analysis utilizes VTT Smart eFleet toolbox to calculate the total cost of ownership (TCO) of heavy-duty commercial vehicles for 15 years. Further, the digital twin simulation coupled with a system dynamics model allows to create scenarios for future demand of vehicles with different propulsion technology. When estimating TCO estimates, it takes into account uncertainties arising from the development of refueling infrastructure.

### Challenges in Using Big Data in Freight Modeling

This section discusses several critical challenges of applying Big Data in RFT research, analysis, and logistics operations management. These challenges could potentially hinder the access and broader utilization of Big Data in this sector. It has been argued that data sharing can improve freight communication, operation efficiency, system resilience, and technological advancement, and reduce costs and greenhouse emissions (Westervelt et al., 2022). The challenges are broken into organizational, technical, legal, and political challenges, discussed in greater detail below, and illustrated in Fig. 1.

### Organizational

Senior managers' lack of interest in data and data-driven decision-making is one of the main organizational challenges. Top managers who champion data analytics have seen the most success in their data collection programs and conduct analyses based on demand and performance data. Lack of support from top managers can make investments and hiring decisions to help gain efficiencies in Big Data-based projects and ICT difficult. Having the right technical expertise is essential to ensure project success after the proper procedures and equipment are in place. Personnel working in transport applications can benefit from advanced training in Big Data methods. Otherwise, errors can occur when decisions are based on poorly executed analyses by non-technical experts that fail to follow the essential due diligence described in the Technical Challenges ("Technical") (Sánchez-Martínez and Munizaga 2016).

### Technical

Data analytics must be robust, accessible, and provide interpretable results to provide meaningful opportunities and solutions to the industry. Credible data analysis and modeling tools are needed to utilize Big Data to harvest, clean effectively, and consolidate the data into a useful format for extracting insights. Poor technical analysis can affect the quality, comprehensiveness, collection, and analysis of Big Data (Mehmood et al. 2017). Technical standards are necessary for coordinating, managing, and governing the data (Neilson et al. 2019).

Technical challenges can be categorized based on the institution implementing Big Data applications: "preventable challenges" and "external challenges" (Kaplan and



Mikes 2012). Preventable challenges are internal risks that should be actively controlled and avoided. Preventable challenges can include data sharing, ownership, data cleaning, and meeting compliance standards (Rigby 2011). In contrast, external challenges are beyond the practitioners' control, such as the volume of unstructured data or whether there is a sufficient pool of skilled data scientists to enable Big Data processing capabilities (Mehmood et al. 2017). See below for these risks and challenges in greater detail.

## Preventable Technical Challenges

### Data Collection

Due to the frequent movement of vehicles of all types, data collected in freight may be inaccurate, incomplete, or unreliable in particular locations or at certain times (Alho et al. 2018; Ma et al. 2011). One possible way to tackle the challenge is by investing in new data collection technologies and improving data collection capability. With the development of IoT, new sensor techniques will continue to improve data collection and quality. Data capturing automation that minimizes manual data entry is also essential to quality improvement (Zhu et al. 2018).

### Storage

The data volume has jumped from terabyte to petabyte level, and the growth in data storage capacity lags data growth. Especially in transport, various data types, e.g., GPS measurements, are continuously produced from the various sensors. Traditional data storage infrastructure and database tools have been unable to cope with the increasingly large and complex mass data. Therefore, designing a data storage architecture that enables the proposed analyses has become a key challenge (Zhu et al. 2018).

Different data types, such as digital images, videos, statistical tables, time series, and social media data, have different formats (e.g., JSON, XML, PDF). This variety of formats presents a challenge in data storage as one type of database may not be optimal for storing and managing all kinds of data. A conventional relational database is not effective when retrieving or searching unstructured and heterogeneous data such as signal data generated by a sensor, free text data collected through social media, image data generated by satellite, and video data streamed by closed-circuit television. In addition, certain data types may require a specific type of database. Movement data, such as GPS locations and time stamps, can be stored and managed efficiently using parallel and distributed moving object databases.

Using a single type of data storage can limit the system's ability to meet different user expectations. Some data storage platforms are designed for large batches of data processing,

whereas others are optimized for real-time data processing. If a Big Data system only supports one type of data storage, the system will not be able to satisfy different user requirements for data processing. The system can meet the need for timely data processing using various platforms, such as vehicle tracking using in-memory data storage. In contrast, batch processing can be performed efficiently using HBase and Hadoop<sup>1</sup> (Neilson et al. 2019).

Various data exchange concepts for freight and logistics have been recently developed as part of EU-funded projects, such as (AEOLIX 2022; FENIX 2022). Their applicability is still limited to the extent of case-specific piloting activities that eventually need to be scaled up to a universal solution. Another example of EU-wide data exchange is the federated data spaces, such as Gaia-X (Gaia-X, 2022). This initiative aims to bring together different companies, research institutions, industry associations, and government administrations to build digital infrastructure and data spaces to serve various industries and co-create Europe's future digital infrastructure. The concept includes the development of a software framework for control and governance that implements a standard set of policies and rules that apply to any existing cloud/edge technology stack, with the final goal of achieving transparency, controllability, portability, and interoperability across data and services.

### Dimensionality

High dimensionality combined with a large sample size from freight transport creates high computational cost and algorithmic instability. It introduces unique computational and statistical challenges, including scalability and storage bottleneck, noise accumulation, spurious correlation, incidental endogeneity, and measurement errors. These challenges require a new computational and statistical paradigm (Fan et al. 2014). The applications include transport data preprocessing, state recognition, real-time control, dynamic route guidance, and real-time scheduling. The data processing system must be able to process more complicated and increasingly expanding data (Zhu et al. 2018). Moreover, the system's timeliness, reliability, and robustness must be tested thoroughly, as transport problems can be hard to predict (Neilson et al. 2019).

### Heterogeneity and Standardization

The massive samples in Big Data are typically aggregated from multiple sources, e.g., public and governmental

<sup>1</sup> HBase and Hadoop are software that facilitate using a network of many computers to solve problems involving massive amounts of data and computation.

organizations (Hee et al. 2018), at different times using various technologies. Such massive samples create issues of heterogeneity, experimental variations, and statistical biases and require developing more adaptive and robust procedures (Fan et al. 2014).

Other challenges include the lack of hardware and software standards, dealing with unusual events when analyzing average transportation performance, and recognizing statistical biases when interpreting data (Hee et al. 2018). Storing identified trips and whole schedules in a standard format presents an opportunity similar to those used for public transportation, such as general transit feed specification (GTFS) or network timetable exchange (NeTEx). Both formats could likely be adjusted for storing freight transportation data and could make data analytics more accessible through libraries and tools that typically emerge once a data format is standardized and widespread.

Another application that might benefit from standardization is the high-level data on freight flows. Bronzini and Singuluri (2009) proposed a scheme that investigates the feasibility of implementing a freight data exchange network with their project's partners to provide timely access to higher quality freight data. The scheme includes *highway performance monitoring system* data, commodity flow data, emissions data, vehicle classification count, and speed data.

## Data Privacy

In the era of Big Data, privacy is the most challenging and concerning problem (Zhu et al. 2018). Worries and legal constraints about privacy surveillance and data misuse challenges hinder data publication and use (Mehmood et al. 2017), especially in the case of data and origins and destinations of trips or fare transaction data. Big Data projects can capture large amounts of personal or commercially sensitive data. Companies are rightly concerned about personal data collection and usage (Hee et al. 2018; Sánchez-Martínez and Munizaga 2016). Personal privacy may be compromised during data transmission, storage, and use. The perceived legitimacy of institutions collecting and analyzing personal data is important to secure personal participation. In freight transport, logistics and other freight-related companies are reluctant to reveal their private information to rivals and publish information about their partner's activities that might risk future business.

The distributed nature of data storage and processing platforms, such as Hadoop and Cassandra, results in different security concerns (Neilson et al. 2019). Safeguarding Big Data is a complex task that requires appropriate authentication, access control, and encryption measures. Challenges in freight Big Data security currently being researched include the secure use of cloud-based services for devices

and networks, identifying internal and external threats, and the standardization of data security.

There is tension between the security and accessibility of data. To prevent unauthorized disclosure of personal private information, governments with related institutions (e.g., General Data Protection Regulation (GDPR) in the EU) could develop complete data privacy laws which include what data can be published, the scope of the data publishing, and use, the basic principles of data distribution, and data availability (Zhu et al. 2018). Care must be taken to safeguard this information and make it available only in anonymized or aggregated forms to prevent its misuse (Sánchez-Martínez and Munizaga 2016).

Randomization or generalization are common approaches to anonymize datasets. Randomization is the process that adds certain imprecision to a certain degree to the original data, shifts the values of attributes in a table so that artificial linking occurs, or groups attributes of an individual with others. Other approach is to generate synthetic data by randomly sampling from the "real" data such that a data set with similar properties is obtained but does not contain the actual observations (thus protecting privacy and business interests). Another approach is to share a limited sample of the data set that allows analyzers, e.g., private companies and researchers, to develop and validate models that may be later generally applied to the full dataset within companies. However, even with such methods, the results could contain sensitive information; thus, robust data management plans and iterative approaches to ensure data safety are needed.

## Validation

Inferring information from Big Data could be hard to validate. For instance, unlike travel diaries in which visited locations and trip purposes are reported, passively generated datasets such as GPS data often lack ground-truth data such as the stop locations of each truck to be validated against, as discussed in "Predictive applications". Even though a high level of accuracy is observed at the aggregate level, the inferred results at the individual level can have a significant number of errors (Chen et al. 2016).

The challenge of managing data collection from multiple sources also represents an opportunity to manage and deal with uncertainty in transportation data. For example, when there is conflicting information from GPS or roadside sensors, it may be possible to resolve it by checking the same information available on video or smartphone data (Zhu et al. 2018).

There is a trade-off between maintaining data quality, processing data as quickly as possible, and validating the data/results. For instance, data reliability is important, because it determines accuracy and precision. GPS signal interference from high buildings, a malfunction of the GPS devices,

or human errors (e.g., drivers forgetting to log the end of a trip) can lead to inaccuracies in location and time data. Such inaccuracies in data need to be resolved or minimized by validating them before performing other analytics on the data (Neilson et al. 2019).

## External Technical Challenges

### Data Ownership and Accessibility

One of the critical challenges is ownership, which connects to the commercial sensitivity of the data (Hee et al. 2018; Sánchez-Martínez and Munizaga 2016). Operators or agencies may hesitate to share data because competitors might use it. Also, when a private data collector owns data, data are not publicly available, mainly due to the data's monetary value. In most cases, data analyzers, e.g., private companies, researchers, and other governmental institutions, must sign a non-disclosure agreement that might inhibit the full use and publication of the research (Alho et al. 2018; Ma et al. 2011).

A government agency often manages data collected from traffic sensors and signals. The one in charge of freight transportation is usually in a different division of the same agency. Communication, cooperation, data storing, data formats, and data sharing between agencies in charge of signals and traffic and those in charge of freight transport planning are uncommon, making it hard to utilize data effectively.

### Data Quality

Data quality refers to accuracy, completeness, reliability, and consistency. Poor data quality due to sensor failures or mistakes in manual data entry can add a burdensome prerequisite to data cleaning before helpful analysis can be carried out (Sánchez-Martínez and Munizaga 2016). Data quality can be compromised because of sensor malfunctioning or interferences, resulting in missing or conflicting data (Fan et al. 2014). There is a trade-off between opening up data quickly and making high-quality data available, creating one more challenge for open data (Zhu et al. 2018).

Freight data inherently contains a lot of noise and uncertainty. Big Data analysis has the potential to unveil hidden patterns associated with small subpopulations and weak commonalities across the whole population. But high dimensionality can also bring problems, including noise accumulation and spurious correlations. Spurious correlations refer to the fact that many uncorrelated random variables may have high sample correlations in high dimensions, causing false scientific discoveries and wrong statistical inferences. These problems can be preventable using the right tools and rigorous validations (“Preventable technical challenges”).

## Representativeness

A critical issue that has not been sufficiently addressed in Big Data applications is the representativeness of the data and the conclusions drawn from these data without careful caveats (Chen et al. 2016). For example, using GPS measurements from selected companies to provide policy advice on the whole freight movement could be misleading. Practical solutions include automatic data capturing and utilizing AI to verify the data. According to the application, samples should also be chosen more carefully to avoid biases. Lastly, transportation departments could expand data management processes to ensure accurate data (Zhu et al. 2018).

## Legal

The main legal challenge is related to privacy. To ensure data privacy, the EU initiated a data protection agreement called GDPR between states to harmonize data privacy laws across Europe (GDPR 2021). This agreement replaces current data protection laws, changing how EU organizations collect, share, and use their data to protect users' privacy (Richard Thomas 2018). However, it forbids the usage for purposes other than explicitly mentioned in the agreement and sharing data with third parties unless additional individual data-access agreements and data management plans are to be negotiated and signed by all involved parties. This requirement to protect all parties from liability also forms a significant obstacle in data collaboration and exchange.

## Political

Most data are gathered by private companies or by governmental institutions. It is hard to share data with third parties due to privacy issues, including other private companies, governmental institutions, or researchers. The environment for sharing data is still in its infancy, which requires political oversight to ensure the entities involved in data sharing that their data will be safe and contribute meaningfully toward sustainable freight transport ecosystems and the economy. An example was a push in 2009 by the US federal institutions to share selected and aggregated data in open data repositories where everyone could access and utilize them. Among many other advantages, these data sources are available to companies and startups in their work to generate economic revenue and evaluate governmental policies (Peled 2011).

Clear and transparent key performance indicators and good data analysis can aid sound decision-making and identify potentially counter-intuitive or even contradictory policy measures. Matowicki and Pribyl (2020) illustrated in a case study that an algorithm for connected and autonomous vehicles that aims to shorten the length of queues in intersections

significantly increases CO<sub>2</sub> emissions. The algorithm distributes traffic into different lanes to shorten the length of the queues at intersections. This change, however, breaks up the existing platoons: some vehicles slow down to move to another lane, and the remaining vehicles speed up to fill in the gaps in the platoon. Therefore, a decision-maker must prioritize and be transparent about the aims, followed by careful monitoring of the real-life efficiency of the implemented measures.

### Future Opportunities

This section identifies a few selected areas for further research. The need to handle the previous discussed challenges, e.g., preventable technical challenges, is an obvious next step to better implement Big Data applications in the freight sector. Many of the challenges require expertise from different domains, such as IT experts. The main focus is to discuss the need to support new trends in modeling and infrastructure development, expanding the potential in data analytics and utilization, achieving high data quality, and some further challenges that need to be addressed.

### Supporting the Need for New Trends in Modeling and Infrastructure Development

Common freight objectives are optimizing goods and service deliveries, real-time routing, and efficient vehicles and fuel supply infrastructures, e.g., refueling and energy delivery. A fundamental part of these systems is trip or tour-chain identification for fleet planning and management. Optimizing goods includes adaptive routing or continuous rerouting given known and planned traffic conditions, e.g., congestion, waiting times at ports and airports, and intermodal terminals.

Big Data promises to move beyond traditional OD matrices that describe aggregated behaviors. It captures individual data for analysis of overall patterns and the heterogeneity of behaviors, activities, and energy. Such a shift requires new modeling frameworks in traffic, demand, and energy, utilizing data from individual collection delivery points, consolidation–distribution centers, vehicles, energy infrastructure, and delivery points.

A significant trend in the freight sector is the focus on sustainability that requires a new way of planning infrastructure for the different alternative fuel technologies, e.g., electric charging, according to their different requirements, limitations, and new operation structures. Further alternative fuel vehicle requirements are energy management, especially for electric vehicles.

The increasing emphasis on environmental performance will require systems and data for environmental monitoring, reporting, and verification of the transport system. The

environmental impact of Big Data computing resources—energy demand, materials demand, and production—must also be measured.

### Expanding the Potential in Data Analytics and Utilization

Given the complex nature of transport networks, it will not be possible to identify single optimal routing solutions for the changes in demand over time. Therefore, adaptive decision-making as an ongoing learning problem may improve road freight performance, which will require the application of machine learning or deep learning methods to analyze how routing patterns can be changed to improve road freight efficiency, allowing for changes in traffic and demand models. One option for investigation could be the combination of agent-based modeling methods with AI.

While the use of AI in human–machine interactions has already been studied for many years, it is considered that the development of decision-making systems based on explainable AI will be one of the main future challenges in research. Despite the growing attractiveness of AI to analyze complex freight-related data, the number of real-life applications is still limited. Currently, application cases for demand forecasting, logistics, and vehicle routing witness a wave of solving certain problems by applying AI. Two main reasons limit current AI applications being more widely used in the freight sector: the complexity of many AI solutions and the "black-box" nature of AI tools. The blind trust in classical "black-box" concepts will not be enough to explain the outcomes of complex algorithms (Mugurusi and Oluka 2021) without the rationale from a human expert perspective. Yet, comprehensibility are centric characteristics of the applied solutions to reduce the distrust and aversion toward classical approaches in critical applications (Asatiani et al. 2020). Woschank et al., (2020) stated that soon, only human–machine hybrid systems will succeed in being applied in practice due to the skeptical nature of humans not accepting systems without rational explanation.

Additionally, there are further generic issues with new data. One challenge is summarizing continuous data to reveal short-term and long-term dynamics patterns, such as demand and traffic. Another challenge is the development of algorithms or AI methods for identifying critical data points/regions, e.g., from continuous monitoring by video cameras. An important area for development is cross-checking and validating new data sets by combining different data sources. One primary application is tour generation by combining vehicle GPS data with infrastructure and location data from geoinformation services to infer vehicle activities and locations (loading, unloading, cross-loading, consolidation), rest and refueling stops, and waiting times. For demand modeling or planning, the tours need to include vehicle parking data for planning loading, unloading, consolidation, rest,

and refueling stops. Finally, methods must be developed to understand the representativeness or generalizability of patterns found in Big Data.

Mobile application usage data mining is a promising data source as mobile apps data output is constantly growing. Models, algorithms, and AI could be implemented using mobile data to identify freight vehicles, e.g., light trucks, from other vehicle types (Iwan et al. 2014; Keimer et al. 2018). Data collected by a mobile application can be used to enhance the energy and demand modeling, for example, by identifying adoption patterns or dependence on other factors (in time and space) that characterize different categories or groups of vehicles of large populations. Behavior analysis can provide a deeper understanding of behaviors and preferences beyond standard analytics. Similarly, location-based social networks can be data mined (Dietz et al. 2020).

### Achieving High Data Quality

Achieving high data quality includes identifying missing data points or inconsistencies in a data set between data points and accounting for these in the pattern analysis. As noted in the challenges section (i.e., “Technical”), a significant issue with Big Data is the development of analysis structures for ensuring high-quality data and understanding the range of uncertainty of the results. Procedures for data validation, both within and for data sets combining different sources, must be developed. Data sources, e.g., GIS and geoinformation location attributes (e.g., types of locations or names of companies visited), are often incomplete, error-prone, and with inhomogeneous levels of aggregation or categorization. Few studies have tried to address the problem of missing data for long durations other than applying conventional heuristics such as filtering or removing data based on simple rules. Improving this requires extensive domain knowledge of the applications and methods to account for the missing information (Nawaz et al. 2020).

Another aspect of improving data quality is the ‘cleaning’ aspect—identifying spurious data points from measurement and reporting device malfunctions. Methods for combining different tracking data sources (i.e., data fusion), such as GPS, aerial, and infrastructure tracking data, need to be developed. For example, combining cargo tracking (parcels, pallets, containers, bulk) and vehicle or transport system tracking data can be useful for logistics management.

### Addressing Further Organizational Challenges

All that is technically possible is not always applicable in all firms, especially in logistics, where there is a structure of a few global, well-resourced firms and many small transport operators with minimal technical resources. Therefore, management systems need to be developed to enable Big Data

systems and analysis incorporation into logistics firms' decision-making process that covers all organizations or firms in the supply chain. A part of this issue is that the use of Big Data requires the development of user interfaces to present it in an understandable way to users with lower technical backgrounds compared to Big Data experts. Also, the insights need to be understandable, robust, and transparent regarding their limitations, transferability, and uncertainty. That would encourage decision-makers to adopt Big Data analytics in their decision-making process to overcome many of their related organizational challenges.

## Discussion and Conclusion

In this paper, we identify different Big Data types used in RFT that differ according to their collecting method, sampling format, resolution, and information collected. Transport-related data from local transport entities are not new to RFT modeling and logistics management. What is new is the combination with other data sources in the Big Data applications. We base our analysis on published academic literature. The implication can be a bias toward academic research perspectives emphasizing the opportunities for research and innovation but perhaps with insufficient emphasis on firm-level challenges such as competition and individualization. Therefore the readers should take our review with a grain of salt, recognizing that the discussions for the potential trade-offs are outside of the scope of this study. Still, our review gives an overview of the opportunities and challenges related to the Big Data and RFT.

Emerging data types have not been fully utilized in freight applications to reach their potential. Noticeably, GPS measurement is the most important Big Data source for RFT applications, especially in the private sector that tracks individual vehicles to improve logistics and energy performance efficiencies. The high spatial and temporal resolutions collected at the vehicle level enable GPS measurements to be used in various data analytics applications and scopes of transport modeling, e.g., freight demand and energy. These details could enhance or even replace traditional models (e.g., the four-step model) with tour-based analysis in RFT applications. The availability of such data types is crucial to many recent prescriptive applications, e.g., connected vehicles. Other Big Data types, such as data from analog and communication devices, are not collected for transport purposes, and they require conversion process to obtain freight-related information. However, these data types are essential to provide additional complementary information to individual vehicle data, such as trip purposes and cargo details. Noticeably, there is not enough literature discussing the use of such data in RFT applications compared to applications utilizing both transport-related data from local

transport entities and vehicle tracking devices. An explanation for that could be the indecisiveness of current methods and algorithms, e.g., AI, in extracting transport-related information from such data types, e.g., collecting land use information from satellite images.

Having the right technical expertise is essential to ensure project success after the proper procedures and equipment are in place. Personnel working in transport applications can benefit from advanced training in Big Data methods. Otherwise, errors can occur when decisions are based on poorly executed analyses by non-technical experts that fail to follow the essential due diligence described in the Technical Challenges.

The privacy challenge is a critical challenge and concern hindering sharing and accessing Big Data. Acquiring the data is difficult without both legally and technically assuring the data provider of the confidentiality of data users. Various challenges hinder access and utilization of Big Data for freight transport applications, including organizational, technical, legal, and political. Organizational challenges are mainly attributed to the lack of support from top managers at their institutions, making it difficult for investments and hiring in Big Data-based projects. Challenges in the organization, data privacy, data quality, and its validation are related to both data and analytics and could be considered significant. These challenges impact the results of RFT applications and still require efforts and experience to overcome their impact. Although it is a technical challenge, data privacy is also a legal issue for data users or seekers. Before sharing the data with third parties, companies must implement many legal contracts and safeguards. One approach would be anonymizing the data either through randomization or generalization. By adding imprecision to a certain degree to the original data, shifting the values of attributes in a table so that artificial linking occurs or grouping an individual with others. Anonymizing Big Data could overcome the privacy issue, but that takes away many important data details, such as taking away the start time and locations of GPS measurements. While such safeguards and legal contracts are important in protecting privacy, it also prevents wider data use that could create broader societal values. Organizations and policymakers have a significant role in overcoming the privacy issue via developing complete data privacy laws, including what data can be published, the scope of the data publishing and use, and the basic principles of data distribution and availability.

With the availability of Big Data, there are abundant new opportunities to create more individual-level data analysis, extracting new insights not only into the aggregate patterns but also into the heterogeneity of behaviors and activities. Moreover, opportunities for new data sources and methods, e.g., agent-based modeling and optimization, can deliver more reliable information for decision-making. To

increase data quality efforts, procedures for data validation are expected to utilize and combine different data sources. Such procedures would identify missing or inconsistencies in the pattern analysis.

**Author Contributions** The authors confirm their contribution to the paper as follows: study conception and design: WS, SY, FS, JK, YT, and PP; analysis, and interpretation of results: WS, SY, FS., and JK; draft manuscript preparation: WS, SY, and FS discussing the challenges: all authors; all authors reviewed the text and approved the final version of the manuscript.

**Funding** Open access funding provided by Chalmers University of Technology. This work is supported by the EU STORM project funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 101006700.

**Data Availability** Not applicable.

## Declarations

**Conflict of Interest** On behalf of all authors, the corresponding author states that the authors have no competing interests as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- ACEA and T&E, 2021. Zero-emission trucks: Industry and environmentalists call for binding targets for infrastructure.
- AEOLIX, 2022. AEOLIX [WWW Document]. URL <https://aeolix.eu> Accessed 4 Nov 2022).
- Alho AR, You L, Lu F, Cheah L, Zhao F, Ben-Akiva M (2018) Next-generation freight vehicle surveys: Supplementing truck GPS tracking with a driver activity survey. *IEEE Conf Intell Transp Syst.* 22:2974–2979. <https://doi.org/10.1109/ITSC.2018.8569747>
- Arias MB, Bae S (2016) Electric vehicle charging demand forecasting model based on big data technologies. *Appl Energy* 183:327–339. <https://doi.org/10.1016/j.apenergy.2016.08.080>
- Asatiani A, Malo P, Nagbøl PR, Penttinen E, Rinta-Kahila T, Salovaara A (2020) Challenges of explaining the behavior of black-box AI systems. *MIS Q Exec* 19:259–278
- Basso R, Kulcsár B, Sanchez-Diaz I (2021) Electric vehicle routing problem with machine learning for energy prediction. *Transp Res Part B Method* 145:24–55. <https://doi.org/10.1016/j.trb.2020.12.007>

- Blackburn R, Lurz K, Priebe B, Göb R, Darkow I-L (2015) A predictive analytics approach for demand forecasting in the process industry. *Int Trans Oper Res* 22:407–428. <https://doi.org/10.1111/itor.12122>
- Borgi T, Zoghliani N, Abed M, Naceur MS (2017) big data for operational efficiency of transport and logistics: a review, in 2017. *IEEE Inter Conf Adv Logis Trans.* 22:113–120. <https://doi.org/10.1109/ICAdLT.2017.8547029>
- Bronzini, M.S., Singuluri, S., 2009. Scoping Study for a Freight Data Exchange Network.
- Bucher JD, Bradley TH (2018) Modeling operating modes, energy consumptions, and infrastructure requirements of fuel cell plug in hybrid electric vehicles using longitudinal geographical transportation data. *Int J Hydrogen Energy* 43:12420–12427. <https://doi.org/10.1016/j.ijhydene.2018.04.159>
- Chankaew N, Sumalee A, Treerapot S, Threepak T, Ho HW, Lam WHK (2018) Freight traffic analytics from national truck GPS data in Thailand. *Transp Res Procedia* 34:123–130. <https://doi.org/10.1016/j.trpro.2018.11.023>
- Chase CW (2013) Using big data to enhance demand-driven forecasting and planning. *J Bus Forecast.* <https://doi.org/10.1002/9781118691861>
- Chen YH (2020) Intelligent algorithms for cold chain logistics distribution optimization based on big data cloud computing analysis. *J Cloud Comput.* <https://doi.org/10.1186/s13677-020-00174-x>
- Chen C, Ma J, Susilo Y, Liu Y, Wang M (2016) The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transp Res Part C Emerg Technol* 68:285–299. <https://doi.org/10.1016/j.trc.2016.04.005>
- Cho W, Choi E (2017) DTG big data analysis for fuel consumption estimation. *J Inf Process Syst* 13:285–304. <https://doi.org/10.3745/JIPS.04.0031>
- Christopher Frey H, Kim K (2006) Comparison of real-world fuel use and emissions for dump trucks fueled with B20 biodiesel versus petroleum diesel. *Transp Res Rec* 11:110–117. <https://doi.org/10.1177/0361198106198700112>
- Chung S-H (2021) Applications of smart technologies in logistics and transport: A review. *Transp Res Part E Logist Transp Rev.* 153:102455. <https://doi.org/10.1016/j.tre.2021.102455>
- Cicconi P, Landi D, Germani M (2016) A virtual modelling of a hybrid road tractor for freight delivery. *ASME Int. Mech Eng Congr Expo Proc* 12:1–8. <https://doi.org/10.1115/IMECE201668013>
- Costa C, Chatzimilioudis G, Zeinalipour-Yazti D, Mokbel MF (2017) Towards real-time road traffic analytics using Telco Big Data. *ACM Int Conf Proc Ser Part.* <https://doi.org/10.1145/3129292.3129296>
- Dias D, Antunes AP, Tchepel O (2019) Modelling of emissions and energy use from biofuel fuelled vehicles at urban scale. *Sustain.* <https://doi.org/10.3390/su11102902>
- Dietz LW, Sen A, Roy R, Wörndl W (2020) Mining trips from location-based social networks for clustering travelers and destinations. *Inf Technol Tour* 22:131–166. <https://doi.org/10.1007/s40558-020-00170-6>
- Dimokas N, Margaritis D, Gaetani M, Koprubasi K, Bekiaris E (2020) A big data application for low emission heavy duty vehicles. *Transp Telecommun* 21(265):274. <https://doi.org/10.2478/tjt-2020-0021>
- Dimokas N, Margaritis D, Gaetani M, Favenza A (2021) A Cloud-Based Big Data Architecture for an Intelligent Green Truck, *Advances in Intelligent Systems and Computing.* Springer Inter Publish. 54:1076
- Du W, Murgovski N, Ju F, Gao J, Zhao S (2022) Stochastic model predictive energy management of electric trucks in connected traffic. *IEEE Trans Veh Technol.* <https://doi.org/10.1109/TVT.2022.3225161>
- Duan M, Qi G, Guan W, Guo R (2020) Comprehending and analyzing multiday trip-chaining patterns of freight vehicles using a multiscale method with prolonged trajectory data. *J Transp Eng Part A Syst* 146:04020070. <https://doi.org/10.1061/jtepbs.0000392>
- Fan J, Han F, Liu H (2014) Challenges of Big Data analysis. *Natl Sci Rev* 1:293–314. <https://doi.org/10.1093/nsr/nwt032>
- FENIX, 2022. FENIX [WWW Document]. URL <https://fenix-network.eu/> Accessed 4 Nov 2022.
- Figueiras P, Gonçalves D, Costa R, Guerreiro G, Georgakis P, Jardim-Gonçalves R (2019) Novel Big Data-supported dynamic toll charging system: Impact assessment on Portugal’s shadow-toll highways. *Comput Ind Eng* 135:476–491. <https://doi.org/10.1016/j.cie.2019.06.043>
- Forrest K, Mac Kinnon M, Tarroja B, Samuelson S (2020) Estimating the technical feasibility of fuel cell and battery electric vehicles for the medium and heavy duty sectors in California. *Appl Energy* 276:115439. <https://doi.org/10.1016/j.apenergy.2020.115439>
- Franzen J, Stecken J, Pfaff R, Kühlenkötter B (2019) Using the digital shadow for a prescriptive optimization of maintenance and operation bt - advances in production, logistics and traffic. In: Clausen U, Langkau S, Kreuz F (eds) Springer International Publishing. Cham
- Gaia-X, 2022. Gaia-X [WWW Document]. URL <https://www.gaia-x.eu/> Accessed 4 Nov 2022).
- Gao Z, Lin Z, Franzese O (2017) Energy consumption and cost savings of truck electrification for heavy-duty vehicle applications. *Transp Res Rec* 2628:99–109. <https://doi.org/10.3141/2628-11>
- GDPR, 2021. General Data Protection Regulation [WWW Document]. URL <https://gdpr-info.eu/> Accessed 28 Jun 2021
- Ge, M., Friedrich, J., 2020. 4 Charts Explain Greenhouse Gas Emissions by Countries and Sectors [WWW Document]. URL <https://www.wri.org/insights/4-charts-explain-greenhouse-gas-emissions-countries-and-sectors> Accessed 29 Jun 2021
- Gingerich K, Maoh H, Anderson W (2016) Classifying the purpose of stopped truck events: An application of entropy to GPS data. *Transp Res Part C Emerg Technol* 64:17–27. <https://doi.org/10.1016/j.trc.2016.01.002>
- Giusti R, Manerba D, Bruno G, Tadei R (2019) Synchromodal logistics: An overview of critical success factors, enabling technologies, and open research issues. *Transp Res Part E Logist Transp Rev* 129:92–110. <https://doi.org/10.1016/j.tre.2019.07.009>
- Hee K, Mushtaq N, Özmen H, Rosselli M, Zicari RV, Hong M, Akkerar R, Roizard S, Russotto R, Teoh T (2018) Leveraging big data for managing transport operations: Understanding and mapping big data in transport sector. Belgium, Brussels
- Hess S, Quddus M, Rieser-Schüssler N, Daly A (2015) Developing advanced route choice models for heavy goods vehicles using GPS data. *Transp Res Part E Logist Transp Rev* 77:29–44. <https://doi.org/10.1016/j.tre.2015.01.010>
- Iwan S, Małecki K, Stalmach D (2014) Utilization of Mobile Applications for the Improvement of Traffic Management Systems. In: Mikulski J (ed) *Telematics - Support for Transport.* Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 48–58
- Jensen CS, Kligys A, Pedersen TB, Timko I (2004) Multidimensional data modeling for location-based services. *VLDB J* 13:1–21. <https://doi.org/10.1007/s00778-003-0091-3>
- Joubert JW, Axhausen KW (2013) A complex network approach to understand commercial vehicle movement. *Transportation (amst)* 40:729–750. <https://doi.org/10.1007/s11116-012-9439-0>
- Joubert JW, Meintjes S (2015) Repeatability & reproducibility: Implications of using GPS data for freight activity chains. *Transp Res Part B Methodol* 76:81–92. <https://doi.org/10.1016/j.trb.2015.03.007>

- Kaffash S, Nguyen AT, Zhu J (2021) Big data algorithms and applications in intelligent transportation system: A review and bibliometric analysis. *Int J Prod Econ* 231:107868
- Kaplan RS, Mikes A (2012) Managing risks: a new framework. *Harv Bus Rev* 90:21
- Katrakazas C, Antoniou C, Vazquez NS, Trochidis I, Arampatzis S (2019) Big data and emerging transportation challenges: Findings from the noesis project. *MT-ITS Int Conf Model Technol Intell Transp Syst*. <https://doi.org/10.1109/MTITS.2019.8883308>
- Kavianipour M, Fakhrrmoosavi F, Singh H, Ghamami M, Zockaie A, Ouyang Y, Jackson R (2021) Electric vehicle fast charging infrastructure planning in urban networks considering daily travel and charging behavior. *Transp Res Part D Transp Environ* 93:102769. <https://doi.org/10.1016/j.trd.2021.102769>
- Keimer A, Laurent-Brouty N, Farokhi F, Signargout H, Cvetkovic V, Bayen AM, Johansson KH (2018) Information patterns in the modeling and design of mobility management services. *Proc IEEE* 106:554–576. <https://doi.org/10.1109/JPROC.2018.2800001>
- Khan M, Machemehl R (2017) Analyzing tour chaining patterns of urban commercial vehicles. *Transp Res Part A Policy Pract* 102:84–97. <https://doi.org/10.1016/j.tra.2016.08.014>
- Kim TH, Kim SJ, Ok H (2017) A study on the cargo vehicle traffic patterns analysis using big data. *ACM Int Conf Proc Ser*. 102:55–59. <https://doi.org/10.1145/3149572.3149598>
- Kretzschmar J, Gebhardt K, Theiß C, Schau V (2016) Range Prediction Models for E-Vehicles in Urban Freight Logistics Based on Machine Learning. In: Tan Y, Shi Y (eds) *Data Mining and Big Data*. Springer International Publishing, Cham, pp 175–184
- Kumar A, Calzavara M, Velaga NR, Choudhary A, Shankar R (2019) Modelling and analysis of sustainable freight transportation. *Int J Prod Res* 57:6086–6089. <https://doi.org/10.1080/00207543.2019.1642689>
- Laranjeiro PF, Merchán D, Godoy LA, Giannotti M, Yoshizaki HTY, Winkenbach M, Cunha CB (2019) Using GPS data to explore speed patterns and temporal fluctuations in urban logistics: The case of São Paulo. *Brazil J Transp Geogr* 76:114–129. <https://doi.org/10.1016/j.jtrangeo.2019.03.003>
- Lárusdóttir EB, Ulfarsson GF (2015) Effect of driving behavior and vehicle characteristics on energy consumption of road vehicles running on alternative energy sources. *Int J Sustain Transp* 9:592–601. <https://doi.org/10.1080/15568318.2013.843737>
- Lee CKH (2017) A GA-based optimisation model for big data analytics supporting anticipatory shipping in Retail 4.0. *Int J Prod Res* 55:593–605. <https://doi.org/10.1080/00207543.2016.1221162>
- Lv Y, Duan Y, Kang W, Li Z, Wang FY (2015) Traffic Flow Prediction with Big Data: A Deep Learning Approach. *IEEE Trans Intell Transp Syst* 16:865–873. <https://doi.org/10.1109/TITS.2014.2345663>
- Ma X, McCormack ED, Wang Y (2011) Processing commercial global positioning system data to develop a web-based truck performance measures program. *Transp Res Rec* 77:92–100. <https://doi.org/10.3141/2246-12>
- Matowicki, M., Pribyl, O., 2020. The need for balanced policies integrating autonomous vehicles in cities, in: 2020 Smart City Symposium Prague. <https://doi.org/10.1109/SCSP49987.2020.9134030>
- Mehmood R, Meriton R, Graham G, Hennelly P, Kumar M (2017) Exploring the influence of big data on city transport operations: a Markovian approach. *Int J Oper Prod Manag* 37:75–104. <https://doi.org/10.1108/IJOPM-03-2015-0179>
- Mesgarpour M, Dickinson I (2014) Enhancing the value of commercial vehicle telematics data through analytics and optimisation techniques. *Arch Transp Syst Telemat* 7:27–30
- Milne D, Watling D (2019) Big data and understanding change in the context of planning transport systems. *J Transp Geogr* 76:235–244. <https://doi.org/10.1016/j.jtrangeo.2017.11.004>
- Misra S, Bera S (2018) Introduction to big data analytics. *Smart Grid Technol*. 76:38–48. <https://doi.org/10.1017/9781108566506.005>
- Mou J (2020) Intersection traffic control based on multi-objective optimization. *IEEE Access* 8:61615–61620. <https://doi.org/10.1109/ACCESS.2020.2983422>
- Mugurusi G, Oluka PN (2021) Towards explainable artificial intelligence (xai) in supply chain management: a typology and research agenda. In: Din M (ed) *IFIP advances in information and communication technology*. Springer International Publishing, Newyork
- Munizaga MA (2019) Big data and transport. *A Res Agenda Transp Policy* 10:00032
- Naumov V, Szarata A, Vasiutina H (2022) Simulating a macrosystem of cargo deliveries by road transport based on big data volumes: a case study of poland. *Energies*. <https://doi.org/10.3390/en15145111>
- Nawaz A, Huang Z, Wang S, Akbar A, AlSalman H, Gumaei A (2020) GPS Trajectory Completion Using End-to-End Bidirectional Convolutional Recurrent Encoder-Decoder Architecture with Attention Mechanism. *Sensors*. <https://doi.org/10.3390/s20185143>
- Neilson A, Indratno D, B., Tjandra, S., (2019) Systematic review of the literature on big data in the transportation domain: concepts and applications. *Big Data Res* 17:35–44. <https://doi.org/10.1016/j.bdr.2019.03.001>
- Nguyen T, Zhou L, Spiegler V, Ieromonachou P, Lin Y (2018) Big data analytics in supply chain management: A state-of-the-art literature review. *Comput Oper Res* 98:254–264. <https://doi.org/10.1016/j.cor.2017.07.004>
- Nicolaides D, Madhusudhanan AK, Na X, Miles J, Cebon D (2019) Technoeconomic analysis of charging and heating options for an electric bus service in London. *IEEE Trans Transp Electrif* 5:769–781. <https://doi.org/10.1109/TTE.2019.2934356>
- Parmar Y, Natarajan S, Sobha G (2019) DeepRange: deep-learning-based object detection and ranging in autonomous driving. *IET Intell Transp Syst* 13:1256–1264
- Peled A (2011) When transparency and collaboration collide: The USA open data program. *J Am Soc Inf Sci Technol* 62:2085–2094. <https://doi.org/10.1002/asi.21622>
- Perrotta F, Parry T, Neves LC, Buckland T, Benbow E, Mesgarpour M (2019) Verification of the HDM-4 fuel consumption model using a Big data approach: A UK case study. *Transp Res Part D Transp Environ* 67:109–118. <https://doi.org/10.1016/j.trd.2018.11.001>
- Phadke AA, Khandekar A, Abhyankar N, Wooley D, Rajagopal D (2021) Why regional and long-haul trucks are primed for electrification now, energy technology area. Lawrence Berkeley National Laboratory, Los Angeles
- Pihlatie, M., Laurikko, J., Naumanen, M., Wiman, H., Rökman, J., Pettinen, R., Paakkinen, M., Hajduk, P., Rahkola, P., Laukkanen, M., Sahari, A., 2021. Kaupallisten ajoneuvojen rooli liikenteen ilmastopolitiikassa Kaupallisten ajoneuvojen rooli liikenteen ilmastopolitiikassa. Karoliina
- Plötz, P., Speth, D., 2021. Truck Stop Locations in Europe Final report. Karlsruhe.
- Plötz P, Wachsmuth J, Gnann T, Neuner F, Speth D, Link S (2021) Net-zero-carbon Transport in Europe until 2050 Targets. *Technol Policies Long-Term Strat*. 31:765
- Rigby DK (2011) The future of shopping. *Harv Bus Rev* 21:443
- Romano Alho A, Sakai T, Chua MH, Jeong K, Jing P, Ben-Akiva M (2019) Exploring algorithms for revealing freight vehicle tours, tour-types, and tour-chain-types from gps vehicle traces and stop



- activity data. *J Big Data Anal Transp* 1:175–190. <https://doi.org/10.1007/s42421-019-00011-x>
- Rosero F, Fonseca N, López JM, Casanova J (2021) Effects of passenger load, road grade, and congestion level on real-world fuel consumption and emissions from compressed natural gas and diesel urban buses. *Appl Energy*. <https://doi.org/10.1016/j.apenergy.2020.116195>
- Ruan M, Lin JJ, Kawamura K (2012) Modeling urban commercial vehicle daily tour chaining. *Transp Res Part E Logist Transp Rev* 48:1169–1184. <https://doi.org/10.1016/j.tre.2012.06.003>
- Sánchez-Martínez GE, Munizaga M (2016) Workshop 5 report: harnessing big data. *Res Transp Econ* 59:236–241. <https://doi.org/10.1016/j.retrec.2016.10.008>
- Santos G (2017) Road transport and CO2 emissions: what are the challenges? *Transp Policy* 59:71–74. <https://doi.org/10.1016/j.tranpol.2017.06.007>
- Seyedan M, Mafakheri F (2020) Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. *J Big Data* 7:53. <https://doi.org/10.1186/s40537-020-00329-2>
- Shi Q, Abdel-Aty M (2015) Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transp Res Part C Emerg Technol* 58:380–394. <https://doi.org/10.1016/j.trc.2015.02.022>
- Singh A, Shukla N, Mishra N (2018) Social media data analytics to improve supply chain management in food industries. *Transp Res Part E Logist Transp Rev* 114:398–415. <https://doi.org/10.1016/j.tre.2017.05.008>
- Souza GC (2014) Supply Chain Analytics Bus Horiz 57:595–605. <https://doi.org/10.1016/j.bushor.2014.06.004>
- Speth D, Plötz P, Funke S, Vallarella E (2022) Public fast charging infrastructure for battery electric trucks – a model-based network for Germany. *Environ Res Infrastruct Sustain*. 2:998
- Thakur A, Pinjari AR, Zanjani AB, Short J, Mysore V, Tabatabaee SF (2015) Development of algorithms to convert large streams of truck GPS data into truck trips. *Transp Res Rec* 2529:66–73. <https://doi.org/10.3141/2529-07>
- Thomas R (2018) GDPR: how will the new data protection law affect the transport sector? Sect GDPR how will new data Prot. *Law Affect Transp*. 11:544
- Tiwari S, Wee HM, Daryanto Y (2018) Big data analytics in supply chain management between 2010 and 2016: Insights to industries. *Comput Ind Eng* 115:319–330. <https://doi.org/10.1016/j.cie.2017.11.017>
- Vehviläinen M, Rahkola P, Keränen J, Pippuri-Mäkeläinen J, Paakinen M, Pellinen J, Tammi K, Belahcen A (2022) Simulation-based comparative assessment of a multi-speed transmission for an e-retrofitted heavy-duty truck. *Energies*. <https://doi.org/10.3390/en15072407>
- Venkadavarahan M, Raj CT, Marisamynathan S (2020) Development of freight travel demand model with characteristics of vehicle tour activities. *Transp Res Interdiscip Perspect*. 8:100241. <https://doi.org/10.1016/j.trip.2020.100241>
- Verendel V, Yeh S (2019) Measuring Traffic in Cities Through a Large-Scale Online Platform. *J Big Data Anal Transp* 1:161–173. <https://doi.org/10.1007/s42421-019-00007-7>
- Wang Y, Sarkis J (2021) Emerging digitalisation technologies in freight transport and logistics: Current trends and future directions. *Transp Res Part E Logist Transp Rev*. 108:14229
- Wang J, Rakha HA (2018) Virginia Tech Comprehensive Powered-based Fuel Consumption Model: Modeling Compressed Natural Gas Buses. *IEEE Conf Intel Trans. Sys. Proc*. 176:1882–1887. <https://doi.org/10.1109/ITSC.2018.8569252>
- Wang G, Gunasekaran A, Ngai EWT, Papadopoulos T (2016) Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *Int J Prod Econ* 176:98–110. <https://doi.org/10.1016/j.ijpe.2016.03.014>
- Wei X, Ye M, Yuan L, Bi W, Lu W (2022) Analyzing the freight characteristics and carbon emission of construction waste hauling trucks: big data analytics of Hong Kong. *Int J Environ Res Public Health*. <https://doi.org/10.3390/ijerph19042318>
- Westervelt, M., Aland, R., Dupraz, I., 2022. Solving the Global Supply Chain Crisis with Data Sharing.
- Woschank M, Rauch E, Zsifkovits H (2020) A review of further directions for artificial intelligence, machine learning, and deep learning in smart logistics. *Sustain*. <https://doi.org/10.3390/su12093760>
- Xu Y, Gbologh FE, Lee DY, Liu H, Rodgers MO, Guensler RL (2015) Assessment of alternative fuel and powertrain transit bus options using real-world operations data: Life-cycle fuel and emissions modeling. *Appl Energy* 154:143–159. <https://doi.org/10.1016/j.apenergy.2015.04.112>
- Yang X, Sun Z, Ban XJ, Holguín-Veras J (2014) Urban freight delivery stop identification with GPS Data. *Transp Res Rec J Transp Res Board* 2411:55–61. <https://doi.org/10.3141/2411-07>
- Yuan Y, Xiong Z, Wang Q (2017) An incremental framework for video-based traffic sign detection, tracking, and recognition. *IEEE Trans Intell Transp Syst* 18:1918–1929. <https://doi.org/10.1109/TITS.2016.2614548>
- Yves R, El-Houssaine A, Nikolaos K, Desmet B (2018) Temporal big data for tactical sales forecasting in the tire industry. *Inform* 48:121–129. <https://doi.org/10.1287/inte.2017.0901>
- Zanjani AB, Pinjari AR, Kamali M, Thakur A, Short J, Mysore V, Tabatabaee SF (2015) Estimation of statewide origin-destination truck flows from large streams of GPS data: application for florida statewide model. *Transp Res Rec* 2494:87–96. <https://doi.org/10.3141/2494-10>
- Zhu L, Yu FR, Wang Y, Ning B, Tang T (2018) Big data analytics in intelligent transportation systems: a survey. *IEEE Trans Intell Transp Syst* 20:266–294. <https://doi.org/10.4018/978-1-5225-7609-9.ch009>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.