A Survey of Finished Vehicle Distribution and Related Problems from an Optimization Perspective

Abstract

In this survey paper, we critique the optimization studies on the distribution of finished vehicles from automobile manufacturers to dealers in the past three decades and propose promising prospective research to minimize the gap between industrial practice and academic research. First, we identify major decision makers involved in automobile distribution, summarize service models, and briefly describe the automobile shipping practice by transportation mode. After defining the automobile shipping optimization problem at the operational level, we present the automobile distribution problem taxonomy by classifying existing studies by the level of decision making, mode of transportation, and type of optimization decisions. Each subcategory of studies is reviewed in detail through comparisons by objective function, constraints, formulation, solution algorithm, and test instances. We conclude this survey paper by summarizing major review outcomes and proposing potential research directions. This survey will stimulate interested transportation researchers to conduct further research to keep pace with the rapid evolution of the automobile shipping practice.

Key words: Automotive supply chains; automobile distribution; integer programming; heuristics; integrated loading and routing; survey.

1. INTRODUCTION

1.1. Significance of automobile distribution

The automotive industry is one of the major industries in the world (Nieuwenhuis and Wells 2015). In some developed countries, such as Germany and Japan, the share of the automotive industry in the Gross Domestic Product (GDP) can exceed 5% (Saberi 2018). Over 10% of all manufacturing jobs in the Europe Union are in automobile production (Volling et al. 2013). Data compiled by the United States Bureau of Labor Statistics (USBLS) indicate that around 4.4 million people are directly involved in automobile manufacturing, wholesaling, and retailing in the US (USBLS 2021). According to the International Organization of Motor Vehicle Manufacturers (abbreviated as OICA in French), the total sales volume of passenger cars and commercial vehicles in 2019 is approximately 91 million (OICA 2020a), among which 46% are manufactured in China, the US, and Japan (OICA 2020b). In the US, on average 17 million automobiles are sold in each of the last five years.

As a very important component of the automotive industry, automobile distribution deals with the shipping of finished automobiles from hundreds of manufacturing plants across several different continents to tens of thousands of dealers across the US. The distribution of finished automobiles is thus a multi-billion-dollar industry considering the size of the U.S. automotive industry. In addition to transporting new cars, the shipping of preowned cars also accounts for a significant portion of the automobile transportation industry for the following reasons. First, as in the case of new cars, a lot of used cars are also exported from affluent countries such as Japan and South Korea to developing countries (Wang and Yeo 2016, Nieuwenhuis et al. 2007). Second, when people migrate, they ship rather than drive their own vehicles to a new place. According to the American Community Survey (ACS) conducted by the US Census Bureau, for various reasons (e.g., education, employment or lengthy vacations) around 7 million people each year relocate to another state (U.S Census Bureau 2018). Third, consumers may also purchase second-hand vehicles online (e.g., online auctions) and have them delivered to their locations. Therefore, the demand for automobile distribution has been consistently high.

1.2. Motivation and objective of this survey

Despite the long history of the automotive industry and clear significance of automobile distribution, there is a major gap between the academic research and industrial practice. Research lags behind practice by at least a few decades: it was not until the 1990s when optimization models for automobile shipping first appeared in the literature, although the mass production of automobiles started in the 1900s. At the operational level, Agbegha et al. (1998) first addressed the auto-carrier loading aspect of the highway-based automobile shipping. At the strategic level, Miller et al. (1996) developed a mixed integer program to optimize the mode selection and rail terminal location problems for an international automobile manufacturer. Due to the high competitiveness in the automobile shipping industry, operations research has been increasingly used to help automobile manufacturers and carriers make better decisions to reduce distribution costs and increase service quality. Over 50 optimization studies involving various automobile shipping decisions at different levels have been published since the 1990s. Nonetheless, an important literature gap remains: no systematic surveys of the academic literature on the automobile distribution problem have been conducted. Thus, this survey intends to contribute to the literature by bridging this gap.

Given that the focus of this survey is on automobile distribution, we do not review in detail the existing work that is primarily on transshipment terminal or port management. Interested readers are referred to Mattfeld and Kopfer (2003) and Mattfeld (2006), among others. Mattfeld (2006) studied some important issues related to the terminal operations in finished vehicle supply chains, such as vehicle transshipment planning, storage space allocation, and personnel scheduling. Mattfeld and Kopfer (2003) developed an integrated model for manpower planning and inventory management for a large vehicle port, the Bremerhaven hub in Germany.

Like other distribution problems in supply chain management, automobile distribution is a very complex decision making problem, because (1) many decision makers with different objectives are involved; (2) interrelated optimization decisions at various levels (i.e., operational, tactical, and strategic) are made; (3) different transportation modes (such as rail and highway) are employed to transport automobiles, which exhibit clear mode-specific features; (4) the problem size in practice tends to be very large; and (5) the primary factors and key considerations in automobile shipping vary significantly with region, implying a high diversity across the world. As indicated in a seminal paper Agbegha et al. (1998), one unique feature of the automobile shipping problem stems from the physical configuration of auto-carrying equipment. Highway-based automobile distribution can be used as an example to illustrate it. For economic reasons, on-road auto-carrier trailers are specially designed to allow the compact storage of automobiles on multiple levels of an auto-carrier while not violating any auto-carrier size regulations. A major operational challenge due to this special trailer configuration is that automobiles can only be loaded and unloaded through a single exit of an auto-carrier, which complicates loading and unloading operations for an auto-carrier. Unfortunately, the loading and unloading decisions, which will eventually determine how automobiles are stored in an auto-carrier, are interrelated with other decisions in automobile distribution, such as assigning automobiles to auto-carriers and routing auto-carriers. Therefore, the specialized auto-carrying equipment involved in automobile distribution clearly distinguishes the automobile distribution problem from other distribution problems.

Although studies since the 1990s have presented significant methodological advances and algorithmic improvements, additional research is practically necessary to keep pace with the rapid evolution of the automobile shipping industry. Therefore, we first provide a comprehensive and in-depth critique of the existing automobile shipping optimization studies spanning all modes of transportation across multiple decision-making levels to identify critical research gaps. Then, we propose prospective research directions for consideration by the transportation research community such that those gaps can be filled through future research. In addition, we discuss new research opportunities in automobile shipping considering the potential impact of emerging technologies, such as vehicle automation, artificial intelligence, and blockchain.

The rest of this paper is structured as follows. Section 2 presents the literature search procedure. We define major decision makers, summarize service models, introduce automobile shipping practice by mode in Section 3. An umbrella term named automobile shipping optimization problem is defined and a classification of this problem is then presented in Section 4. The next four sections (Sections 5 to 8) thus critique four subcategories of existing studies. In Section 9, we analyze major research gaps, propose plausible approaches to addressing such gaps, and analyze the impact of emerging technologies. In Section 10, concluding remarks are presented.

2. RESEARCH METHODOLOGY

In this survey paper, we intend to broadly cover those decision-making problems concerned with moving automobiles from one place to another. In the literature, quite a few terms with varying scopes are used, such as auto-carrier transportation (Tadei et al. 2002), automobile pickup and delivery (Wang et al. 2018), outbound logistics of automobiles (Wang and Chen 2020), and automobile distribution (Hei et al. 2014). All such related problems fall within the scope of this survey.

To ensure most relevant studies are covered in this survey, the following procedure is used to conduct the literature search systematically.

- First, we searched Web of Science and Google Scholar with keywords including “automobile shipping”, “automobile transportation”, “automobile distribution”, “automobile pickup and delivery”, and “auto-carrier loading and routing”. For those papers found in this step, we examined whether their main research methods were optimization-based and thus excluded purely qualitative research.
- Second, for each selected paper from the first step, we scanned its reference list to include those new studies that were not identified in the first step; we also used Google Scholar to scan relevant studies that referenced a paper identified from the first step.
- Third, we iterated the above steps until the pool of identified studies for review from the literature converged.

Although the above systematic approach should cover most of automobile shipping optimization studies written in English, we note that relevant studies published in other languages or in those journals not indexed by Web of Science or Google Scholar are not covered in this review. The final list of studies has more than 50 papers from mainstream journals in the field of operations research, such as Transportation Science, Transportation Research Parts B & E, Computers & Operations Research, Computers & Industrial Engineering, European Journal of Operational Research, Annals of Operations Research, and Management Science. The time period ranges from the 1990s to the present.

As mentioned earlier in this section, due to this paper’s focus on existing studies that use optimization and operations research as main research methods, existing qualitative studies are out of the scope of our survey. Two examples of such qualitative studies are Holweg and Miemczyk (2003) who studied the strategic implications of automotive logistics operations and Chandra et al. (2016b) who reviewed the management practices of outbound automotive logistics in India.

3. OVERVIEW OF AUTOMOBILE SHIPPING PROBLEM

There exists a wide array of problem variants of automobile shipping, depending on several practical factors, such as decision makers involved, transportation modes employed, and automobile carrying equipment used. In this section, we first define decision makers involved in automobile shipping; then, we describe three service models by characterizing how the decision makers interact with each other; finally, we introduce various transportation modes and on-road auto-carriers.

3.1. Decision makers

There are quite a few parties involved in the automobile shipping industry, such as customers, lead aggregators, brokers, and carriers. Those parties work collaboratively to ensure the arrival of automobiles at the destination. Definitions of such decision makers are given to facilitate further discussions.

Customers are the individuals (e.g., a person relocating to a different place) or corporates (e.g., an automobile manufacturer) who need to ship automobiles from an origin to a destination. Individual customers usually move their preowned vehicles from the current location to a new location they relocate to. Usually, they are both shippers and receivers. Unlike individuals, business customers usually ship finished vehicles from plants or warehouses to dealer locations. In this case, shippers differ from receivers.

Carriers are individuals or trucking companies which are engaged in the professional conveyance of automobiles. Most carriers do not communicate directly with shippers. The communication is usually through a *broker*, who acts as a facilitator for both shippers and carriers. On one side, brokers work with a network of carriers and verify the carriers’ licenses, insurances, and credentials; on the other side, brokers obtain shipping information either from shippers directly or from a *lead aggregator* that maintains a load board. With the information of available automobiles for shipping collected from individual shippers, the load board is usually accessible by the brokers with a paid subscription and is a critical tool for the brokers to find the best carrier or carriers to meet the need of one or a group of shippers. It should be noted that it is not very common for brokers to own any automobile-carrying (auto-carrying) equipment. Only in certain cases, an automobile transportation company or a Third-Party Logistics (3PL) company may act as a broker while it also owns a fleet of auto-carriers. The automobile shipments received by those broker-carriers may be fulfilled by their own dedicated fleets or get outsourced to other carriers.

Clearly, if a broker-carrier has its fleet and receives shipping orders directly from a shipper such as an automobile manufacturer, all decisions involved in the shipping process are made by itself, which thus justifies a centralized decision-making framework. If a broker does not own and operate any fleets, it may only make those decisions regarding how to assign shipping orders to carriers, while some other decisions (such as loading automobiles to auto-carriers) could be made by carriers. Depending on the context, routing decisions may be made by a broker or an individual carrier. Section 3.2 better characterizes the behavioral interactions among the above decision makers.

3.2. Service models

Depending on how decision makers interact, there are three service models, as illustrated in Figure 1.

In service model 1, the shipper, which is usually an automobile manufacturer, directly works with a carrier without involving any brokers. It seems that this service model is studied the most in the literature among the three service models. Tadei et al. (2002), Dell'Amico et al. (2015), Cordeau et al. (2015), Venkatachalam and Sundar (2016), and Wensing (2018) considered this service model. An instance of this service model is described as follows: automobile dealers located in different regions submit orders for new automobiles to an automobile manufacturer, which then sends newly assembled automobiles (i.e., BTO, build to order) to a staging area; all the vehicles including those stocked (i.e., BTS, build to stock) in the staging area are released to the carrier that has contracted with the manufacturer for delivery to the dealers.

In service model 2, automobiles are transported by multiple independent carriers that are selected by a broker without auto-carrying equipment, given the order information (such as vehicle time and shipping timeframe) received from a shipper. This service model is studied in Wang et al. (2018), and Bonassa et al. (2019), among others. In the problem studied by Wang et al. (2018), a 3PL provider, which does not own any auto-carrying equipment, connects with a network of trucking companies to deliver the finished automobiles from a manufacturer's outbound warehouses to various dealer locations. The shipping optimization decisions are made by the 3PL company.

Service model 3 involves a broker-carrier who has its private fleet and may also outsource its shipping orders to other carriers, depending on whether the orders it receives from multiple shippers (potentially through lead aggregators) exceed the capacity of its own fleet. In the first two service models, shippers are usually business customers (such as automobile manufacturers), whose shipping demand is usually large (e.g., dozens of automobiles per order) and predictable. In contrast, the shippers in service model 3 are usually individual customers who relocate to new places for various reasons (new employment or retirement). Individual shippers' demand is relatively small (one or two cars per order) and volatile. This difference also implies that shippers in service model 3 are more diverse than those in the other service models. Shippers in service model 3 submit the pickup and drop off locations, a timeframe, and other necessary vehicle details (such as make, model, and year) to the broker-carrier. After receiving all the shipment orders, the broker-carrier determines what orders can be served in-house and what orders must be outsourced to other carriers, based on inputs from shippers and carriers. Service model 3 has not been studied in the literature. Liu et al. (2016) mentioned the possibility of outsourcing remaining orders to other carriers if the fleet capacity of the logistics company in charge was not enough to cover all orders, although this was not explicitly incorporated into their optimization model.

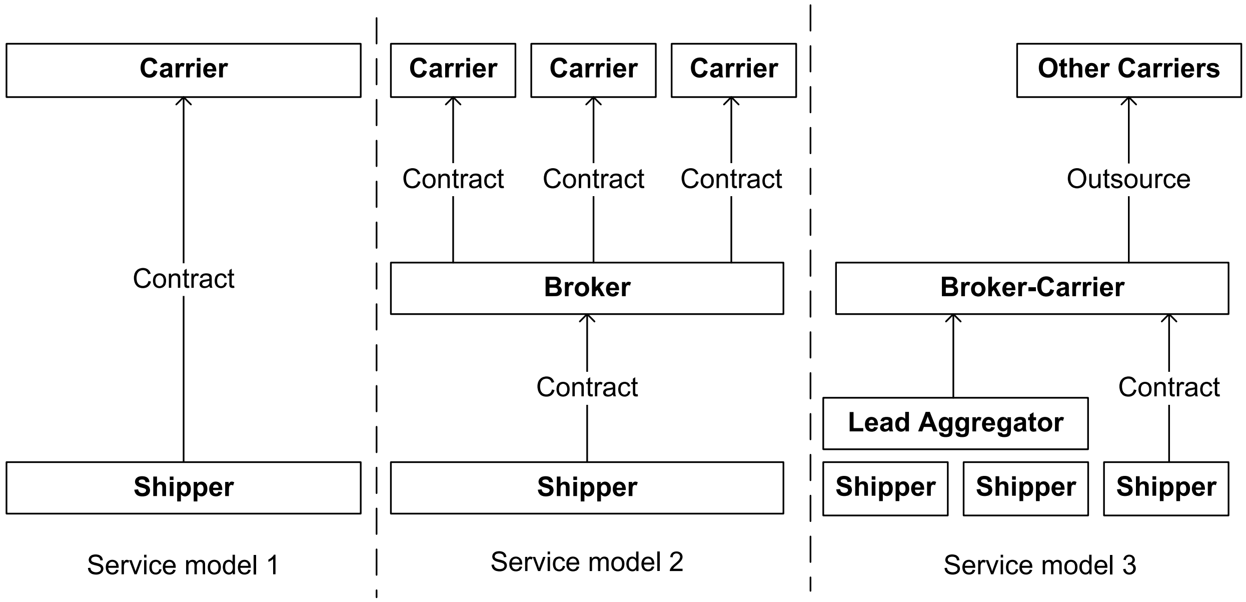
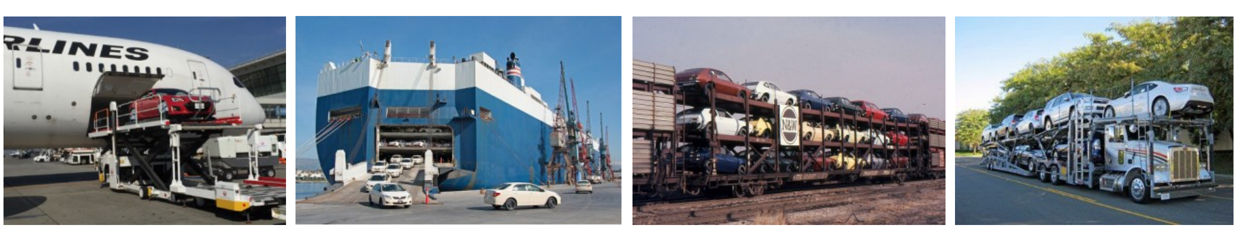


Figure 1: Service models for automobile shipping

3.3. Transportation modes

As shown in Figure 2, all five major transportation modes except pipelines are employed in practice for transporting automobiles. Shipping an automobile by air is not only feasible, but also has an incomparable speed advantage, which is ideal for shipping antique or very high-end vehicles (such as Formula one cars) overseas rapidly. As it takes at least a few thousand U.S. dollars to ship a car overseas, it is not a viable mode for large-scale distribution of automobiles.

Maritime transportation is the primary way of intercontinental transportation of automobiles. Leading automobile exporters such as Germany, Japan, and the U.S. rely on ocean to ship automobiles across continents or regions, for instance, from East Asia to North America or from Mexico to Northeastern U.S. Shipping by sea has the lowest transportation cost per car when the shipping volume is relatively large and the shipping is over a very large distance, while the clear drawback is its very slow speed. It usually takes weeks to cross the Pacific, which implies significant lead times (Katcoff 2011). Roll-on/roll-off or Ro-Ro ships (Kang et al. 2012) are used to carry automobiles because vehicles can be driven on and off the vessel on the vehicle's own wheels through ramps. A car-carrying vessel can carry thousands of automobiles, which would be infeasible by any other modes.



(a) Car carrying aircraft (b) Ro-Ro ship (c) Autorack (d) Roadway auto-carrier

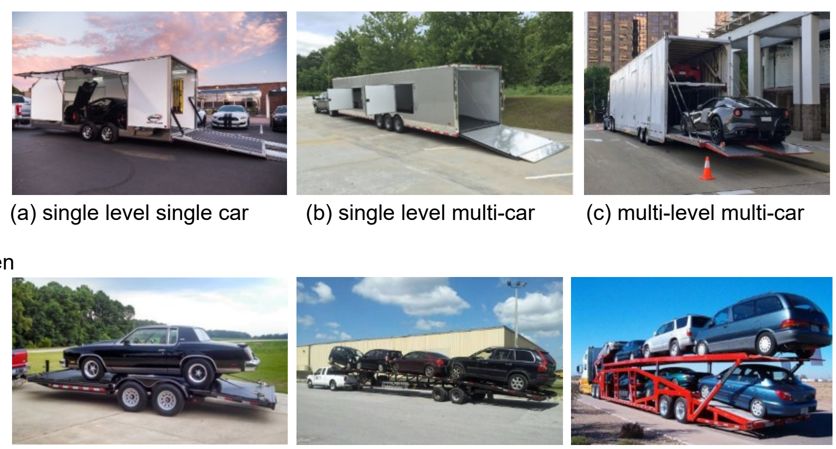
Figure 2: Automobile transportation modes

Railway is the primary long-haul mode for transporting vehicles, due to its cost-effectiveness for longer trips. According to Jin et al. (2010), around 70% of finished vehicles in the U.S. are shipped by railway. The special piece of rail rolling stock called autorack is used for carrying automobiles, which has two or three levels. Smaller vehicles such as sedans can be loaded onto tri-level railcars, while taller vehicles such as passenger trucks are loaded onto bi-level railracks (Katcoff 2011). Nowadays more and more autoracks are partially or fully enclosed to prevent automobiles from being damaged by inclement weather, falling rocks, being stolen, and other threats. In the U.S., travellers along the East Coast may take their automobiles while they ride a train, thus avoiding long-distance driving. Such a service is known as Auto Train in the U.S. or Motorail in Europe. In this case, passengers are carried in passenger rail cars while their automobiles are loaded into autoracks.

Highway offers the most flexible automobile transportation option, as it is the only mode that can provide door-to-door services. It has a relatively high delivery speed and medium transportation cost. Therefore, highway is the primary mode for short-haul automobile transportation. Automobile transportation by highway can be terminal-to-terminal or door-to-door, using a special type of truck with a tractor and trailer called car hauler, auto-carrier, or automobile transporter. Figure 3 shows some common auto-carriers.

Auto-carriers come in numerous sizes, types, and shapes. Like autoracks, roadway auto-carriers can be open or enclosed. Open auto-carriers are very widely used and can transport almost all types of cars, ranging from new cars to salvage ones. Enclosed transportation offers an extra layer of protection at a higher cost and can be used for transporting high-end vehicles. Auto-carriers may have one or two levels, with the carrying capacity (as measured by the number of available slots) ranging from one to more than ten. Although a single-car hauler, or better known as a tow truck, can be highly flexible to operate and can navigate narrow roads, it is cost effective for only short distances. A high-capacity carrier is more cost efficient due to economies of scale. Nonetheless, the maximum number of automobiles that can be carried by an auto-carrier is regulated to avoid severe damage to the pavement. The US Department of Transportation has regulations on the dimensions of auto-carriers, for instance, the overall length is up to 75 feet (U.S. Department of Transportation 2004).

Enclosed



(a) single level single car (b) single level multi-car (c) multi-level multi-car

Open

(d) single level single car (e) single level multi-car (f) multi-level multi-car

Figure 3: Different types of auto-carriers

Due to the special physical configuration, auto-carriers in general have only one access point from which automobiles are loaded and unloaded. For the maximum capacity utilization, automobiles of different dimensions should be loaded onto an auto-carrier in a certain sequence to avoid unnecessary and costly operations, such as *reloading*: a vehicle that is not yet destinated must be unloaded and loaded back to allow other vehicles to be moved. Aghbegha et al. (1998) pointed out that there is a unique combinatorial nature associated with auto-carrier loading and unloading that stems from the special physical characteristics of roadway auto-carriers. An auto-carrier has a set of slots of different dimensions, which are used to hold automobiles of various types and sizes. The number of possible ways to sequence automobiles for loading and assign them to slots increases exponentially with the number of slots in an auto-carrier.

While the above analyses are separated by individual shipping modes, it should be noted that automobiles are often transported using a combination of waterway, railway, and highway. For instance, imported cars from Japan can first be transported to a port of entry (such as Long Beach, California) by ocean; then, they are distributed to major warehouses throughout the U.S. by rail; last, they are delivered to dealer locations by highway.

4. PROBLEM CLASSIFICATION AND REVIEW OUTLINE

To make the review systematic, we start by defining the automobile shipping optimization problem at the operational level in Section 4.1; then, we briefly discuss related optimization problems at the strategic and tactical levels in Section 4.2. Finally, in Section 4.3, all relevant automobile shipping studies are classified based on the level of decision-making, primary transportation mode, and decision variables involved.

4.1. Automobile shipping optimization problem

The Automobile Shipping Optimization Problem (ASOP) is a class of operational optimization problems that involve (1) shippers that have automobiles to be shipped from one or more origins to one or more destinations, (2) carriers that operate auto-carrying equipment for transporting automobiles, and (3) coordinators that optimize various planning and operational decisions to achieve an objective, subject to the requirements and restrictions from both shippers and carriers. Each automobile to be shipped is associated with a pickup location, a drop-off location, a vehicle type, and possibly some other information. Carriers may be homogeneous and based in a central depot, or heterogeneous with a mixed fleet. As discussed in Section 3.2., there are multiple service models. Depending on which service model is followed, there may be an overlap between different roles, such as a hybrid of carrier and broker.

The underlying planning and operational decisions vary significantly with the mode employed to transport vehicles. Thus, we divide all the automobile shipping optimization studies into two broad categories: highway-based, and multimodal. As most studies in the literature involve highway transportation only, we further divide such studies based on what types of decisions are involved. In the first paper on roadway-based automobile transportation optimization problem, Aghbegha et al. (1998) described how various heuristics were used in the 1990s to solve the automobile shipping optimization problem. They pointed out that a commonly used heuristic in practice is a sequential approach, which is illustrated in Figure 4 and described as follows:

First, as not all the vehicles in the staging areas are to be delivered on a day, priorities are given to those vehicles that have been ordered earliest by dealerships. Dealers are then grouped so that the ordered vehicles from the dealers in the same group are delivered together by the same auto-carrier.

Next, as the vehicles to be delivered by an auto-carrier are finalized, a routing heuristic is used to determine the route and sequence to visit dealers.

Third, given a set of automobiles to be delivered and the optimized auto-carrier route, a detailed loading plan is determined to minimize reloading and avoiding damages. When an auto-carrier is en-route, vehicles that are not destined may be rearranged (such as temporarily unloaded and loaded back) in order to facilitate the delivery of some other vehicles.

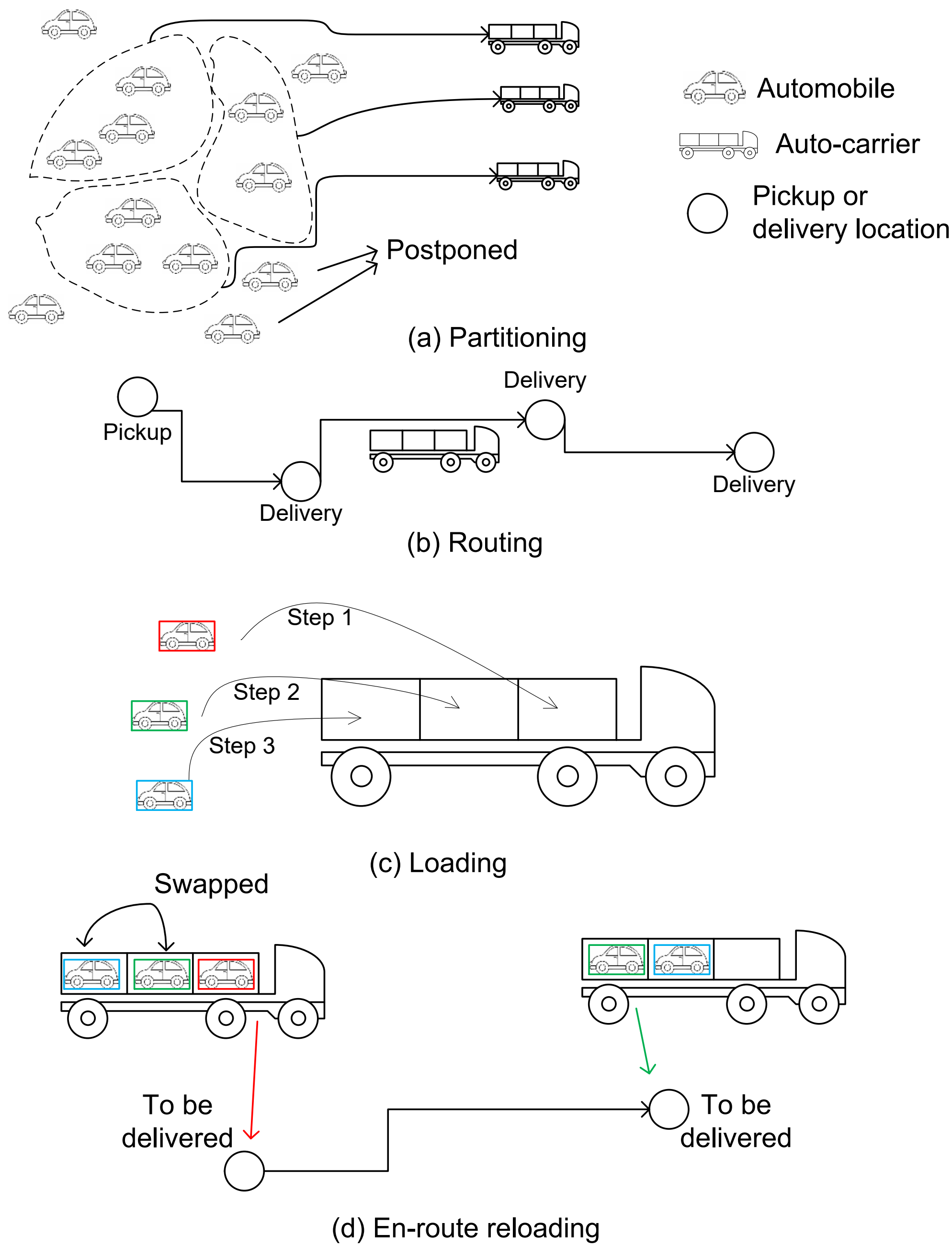


Figure 4: Types of decisions

As reviewed later, there are many existing studies that have adopted such a sequential optimization approach. There are also several existing studies that have focused on loading/unloading decisions only while assuming that auto-carrier routes are given. Although such a sequential approach is relatively simple to implement, it has a major drawback - it does not consider the interrelations between decisions made at different steps (routing and loading). For example, an "optimized" auto-carrier route without consideration of its associated loading implications may incur very expensive reloading operations. In the case of transporting very high-end cars, to avoid potential damage during reloading, a strict Last In, First Out (LIFO) policy may be adopted during loading; however, the resulting auto-carrier route may have significant detours, which is cost prohibitive. As indicated later in this survey (Section 9.2), although some existing studies have attempted to integrate routing and loading/unloading decisions, no existing study has developed a fully integrated optimization model, which is also a major motivation for this survey.

4.2. Tactical and strategic problems related to ASOP

The ASOP defined in Sections 4.1 is at the operational level, and involve operational decisions that are made relatively frequently, such as daily or twice per week. Like in other distribution problems, there are also tactical and strategic decisions related to automobile shipping. Such decisions are made infrequently due to their mid-term or long-term effects. An example of tactical decision is inventory control. For instance, Cachon & Olivares (2010) and Cachon et al. (2019) studied the relation between automobile sales and inventory with empirical data from auto dealerships in the US. An example of strategic decision is the location of vehicle distribution centers. For instance, Eskigun et al. (2005) optimized the automobile distribution network considering a few practical factors including lead time.

At the strategic and tactical levels, there are two types of optimization decisions that are most relevant to automobile shipping and thus studied extensively in the automobile distribution literature. Those decisions are: (i) design of automobile shipping networks and selection of proper shipping modes for automobile distribution, and (ii) management of empty autoracks including making repositioning decisions in railway-based automobile shipping (Sherali and Maguire 2000). Given the importance and relevance of these topics, we will review each of these problem classes in detail, along with the operational-level problems discussed in Section 4.1.

It should be noted that although there are multiple decision-making levels, existing studies on automobile shipping across various decision levels are not uniformly distributed, with most of them at the operational level. This is understandable because automobile distribution problems at the operational level are dramatically different from other distribution problems because of the unique auto-carrying equipment involved, as discussed in Section 1.2. As we climb up the decision hierarchy, fewer and fewer operational features are present, and the problem uniqueness diminishes. For instance, auto-carrier reloading issue is prominent at the operational level, while it certainly becomes trivial at the network design level.

4.3. Taxonomy and review outline

We classify all relevant studies by the decision-making level and the primary transportation mode involved. As most of the studies are on operational-level decisions of the automobile shipping, we review such studies with priority in this survey. This explains why we review such studies before studies on higher-level decision-making problems, although strategic-level decisions usually precede lower-level decisions in the decision process. As already explained in Section 4.1, we divide all the operational optimization studies into two broad categories: highway-based, and multimodal. We further divide all those roadway-based operational models into two groups, depending on whether they consider routing optimization decisions explicitly. Studies on tactical and strategic level decisions, including mode selection and network design, and empty autorack management, are reviewed separately. Figure 5 shows how the reviews of these different classes of problems are organized in Sections 5 to 8.

It should be noted that operational-level optimization problems involving different transportation modes differ substantially due to the mode-specific characteristics involved such as loading platform configuration, carrier capacity, and routing complexity. Thus, it does not make much sense to examine those modeling differences among different modes in a single section, as there is very little commonality. For instance, a highway-based auto-carrier carries around ten automobiles and visits multiple dealer locations. Reloading operations (which are essential) directly contribute to the optimization complexity. On the contrary, a Ro-Ro ship carries thousands of automobiles and has a single destination. When automobiles are shipped by Ro-Ro ships, reloading becomes irrelevant. Clearly, the resulting optimization models are quite different, which further means that they should not be reviewed together. This explains why in each of the following four sections, one specific group of studies is reviewed as illustrated further in Figure 5.

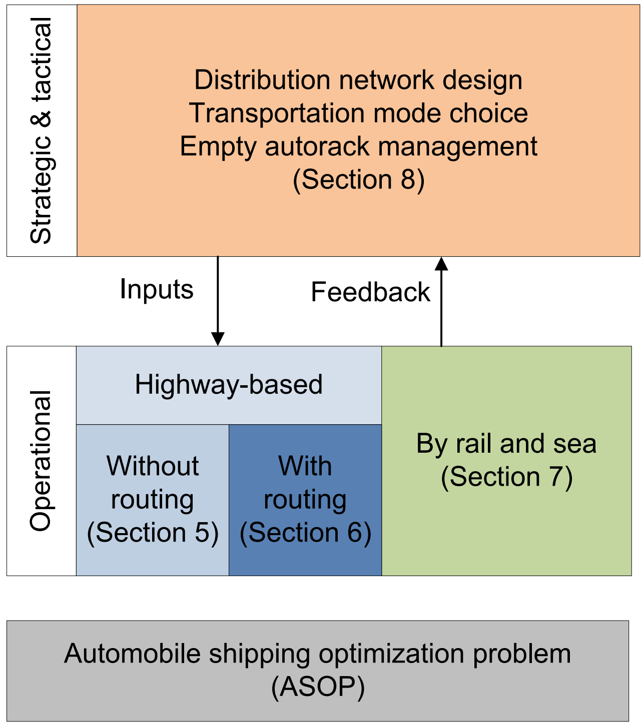


Figure 5: Classification scheme and coverages of problem classes in Sections 5 to 8

5. LOADING AND RELOADING OPTIMIZATIONS

As the existing roadway-based automobile shipping optimization models without involving any routing implications are primarily focused on loading and unloading optimizations, we review such studies in this section. Loading is the final step in the sequential decision-making process described in Section 4.1, which is to assign a given set of automobiles to various slots of an auto-carrier whose route is already fixed. Given the configuration of an auto-carrier, reloading is usually needed when some automobiles that are not to be delivered at a location prevent the unloading of destinated automobiles. Such automobiles blocking the unloading path of destinated automobiles must be temporarily unloaded and loaded back either to the original slot or another available slot, after those destinated automobiles are unloaded and delivered.

5.1. Formulations and solution algorithms

Agbegha was the first to study the auto-carrier loading problem (ACLP), according to Billing et al. (2018) and Wensing (2018), among others. Agbegha et al. (1998), which was based on Agbegha's doctoral dissertation (Agbegha 1992), focused only on loading decisions, although it was acknowledged that loading decisions were interrelated with other optimization decisions (such as routing) in automobile shipping. In Agbegha et al. (1998), a so-called loading network was proposed for an auto-carrier trailer, where each node represents a slot and an arc from one node to another indicates the unloading precedence. Depending on the configuration of an auto-carrier trailer, certain loading constraints were considered, such as single car constraint (a slot can accept only a few types of automobiles), and pairwise constraint (some automobiles cannot be assigned to adjacent slots). The optimization objective was to minimize the reloading cost, for a given auto-carrier route, which specifies the sequence for visiting auto dealers. The loading optimization problem was formulated as a geometric assignment problem, which was solved by a heuristic algorithm involving a branch-and-bound procedure.

Lin (2010) followed Agbegha et al. (1998) in adopting the loading network and formulated a quadratic assignment problem. A key assumption made in Agbegha et al. (1998), which requires a reloaded automobile to be loaded back to the original slot, was relaxed in Lin (2010) by allowing a reloaded automobile to move to any feasible available slot whenever reloading is needed. The quadratic assignment problem was directly solved with a nonlinear programming solver. Lin (2010) claimed that (1) the solution was exact; (2) better solutions were found than Agbegha et al. (1998).

Chen (2016a) followed this line of research and pointed out that the number of reloads was incorrectly computed in Agbegha et al. (1998) if empty slots were present. After fixing this issue and making other changes, Chen (2016a) formulated the loading optimization problem as an integer programming problem. Computational results demonstrated that the new heuristic developed in Chen (2016a) outperformed the heuristic of Agbegha et al. (1998).

In the three ACLP studies reviewed above (i.e., Agbegha et al. 1998, Lin 2010, Chen 2016a), the loading optimization problem was solved at an auto-carrier depot, before the auto-carrier was dispatched. Chen (2016b) studied a sequential auto-carrier loading problem (SACLP) that was solved along the route at each dealer location, assuming the solution to the ACLP was given as an input. Chen (2016b) formulated SACLP as a series of binary integer programs, but proposed a heuristic algorithm to solve the problem. Computational results showed that the solution of SACLP along the auto-carrier route helped reduce the reloading cost as compared with solving ACLP once at the auto-carrier depot.

Table 1 provides a quick comparison of those four studies that are focused on loading and reloading decisions. First, the simplifying assumption made by Agbegha et al. (1998) on reloading an automobile to its original slot was still made in Chen (2016a). Second, the majority of loading optimization problems were solved by various heuristics, with the exception of Lin (2010) who directly solved their quadratic assignment problem with a solver.

Table 1: Summary of existing studies on loading and reloading problems

Study	Formulation	Back to original slot once reloaded	Solution method
Agbegha et al. (1998)	Geometric assignment problem	Yes	Heuristic based on branch-and-bound
Lin (2010)	Quadratic assignment problem	Not required	Solved directly by a nonlinear programming solver
Chen (2016a)	Integer programs (IP)	Yes	Heuristic
Chen (2016b)	A series of binary IPs	Not required	Heuristic

5.2. Test Instances

In the pioneering work Agbegha et al. (1998), four sets of ACLP test instances (named #1, #2, #3, and #4) were developed based on the auto-carrier loading practice in the U.S. Each set consisted of 15 instances (numbered from 1 to 15). In each set of the test instances, there were a unique auto-carrier type and four types of automobiles, such as cars, vans, and long vans. Based on the configuration of an auto-carrier and types of automobiles, loading constraints were also generated. Each automobile in the set of automobiles assigned to the auto-carrier (i.e., a load) had a destination (which was a dealer location), although multiple automobiles may go to the same dealer location. Each instance in a test instance set also had a unique sequence of dealer locations (i.e., route). Agbegha et al. (1998) reported their optimization results for all 60 instances, which included the computational time and the number of reloads in the solution for each instance.

Results showed that for some test instances, feasible solutions were not found using the heuristic they developed.

Lin (2010) used the same 60 instances in Agbegha et al. (1998) and reported that his formulation yielded better solutions for a dozen of instances. However, it should be noted that Lin (2010) did not solve the same problem as Agbegha et al. (1998), because in Lin (2010) it was no longer a requirement to reload an automobile to its original slot. Lin (2010) also compared their solution time with Agbegha et al. (1998), despite the clear difference in computing power. The computation time for an instance ranged from less than a second to nearly one minute.

Chen (2016a) adopted 9 instances from 3 sets (#1, #2, and #4) defined in Agbegha et al. (1998). He reported a higher solution efficiency than Agbegha et al. (1998), as all instances were solved within 21 seconds. Chen (2016a) also reported that feasible solutions were found for some unsolvable instances in Agbegha et al. (1998). As in Lin (2010), the significant difference in computing power was not considered when solution times were compared. Chen (2016b) tested their method on 5 instances, part of which were from Agbegha et al. (1998). The rest of them were generated by Chen (2016b). Solution times were not reported in Chen (2016b).

Table 2 provides an overview of the test instances used in those loading and reloading studies. It clearly shows that the majority of test instances used for testing loading optimization algorithms were from Agbegha et al. (1998). All the test instances are available in these papers. In terms of solution efficiency, it takes at least a few seconds to find an optimized loading plan for a given route covering about five dealer locations.

Table 2: Overview of test instances in existing loading and reloading studies

Study	# of instances	# of auto types	# of carrier types	Data source	Instance availability
Agbegha et al. (1998)	60	4	4	Original	Yes
Lin (2010)	60	4	4	Agbegha et al. (1998)	Yes
Chen (2016a)	9	4	3	Agbegha et al. (1998)	Yes
Chen (2016b)	5	4	2	Mixed	Yes

6. OPTIMIZATION OF ROUTING AND RELATED DECISIONS

In the loading and reloading optimization studies reviewed in Section 5, the auto-carrier route is considered given and no routing related costs are considered. On one hand, routing has a direct impact on loading; on the other hand, routing decisions are clearly affected by partitioning decisions (which automobiles are selected and delivered by which auto-carrier), as described in the sequential decision-making approach in Section 4.1. Therefore, it is very challenging to optimize the routing decisions considering their interrelations with other decisions. Routing has also attracted much more attention than loading in the literature (e.g., Tadei et al. 2002, Dell'Amico et al. 2015, and Wang et al. 2018). In this section, we focus on roadway-based automobile shipping optimization models involving routing and other related decisions.

6.1. Evolution of auto-carrier routing studies considering loading

We first review each relevant study in detail chronologically to show the evolution of those optimization models. Readers may skip Section 6.1 and proceed to Sections 6.2 to 6.5 for syntheses of our review findings.

Tadei et al. (2002) considered three aspects of automobile shipping, namely loading, vehicle selection, and routing, in a simplified manner. First, regarding loading, an auto-carrier was modelled as a single loading plane where automobiles were loaded horizontally. Each automobile to be loaded has its equivalent length, depending on the vehicle class. If the total equivalent length of the automobiles to be loaded did not exceed the capacity threshold of an auto-carrier, the loading plan was considered feasible. Clearly, loading solutions generated this way may not even be feasible considering a realistic configuration of an auto-carrier, such as discrete slots of various sizes in Agbegha et al. (1998). Second, although the routing aspect was involved in the problem considered by Tadei et al. (2002), detailed auto-carrier routes were not generated explicitly. Instead, they simply divided geographical regions into clusters and required that only automobiles with destinations in the same cluster can be delivered together by an auto-carrier. This replaced routing decisions by clustering of auto dealers located in a region. Tadei et al. (2002) developed a three-step heuristic based on an integer programming formulation to solve the integrated optimization problem. The equivalent length concept was later adopted in Li and Zhang (2016) and Wensing (2018), among others.

Miller (2003) also tried to optimize auto-carrier routing while considering loading complications in his master's thesis. His work made a few key simplifying assumptions, including not considering automobile sizes so that any car can be loaded to any slot on an auto-carrier, and not considering distances between dealer locations. The auto-carrier modeling in Miller (2003) is more realistic than Tadei et al. (2002), although Miller (2003) assumed that an auto-carrier had two flat levels on which automobiles were loaded and unloaded straight. Construction and improvement heuristics were designed to solve the optimization problem in Miller (2003).

Dell'Amico et al. (2015) claimed that they were the first to generate detailed solutions of both the loading and routing components of the automobile shipping problem simultaneously. Their algorithm was an iterated local search procedure centered around the routing decisions, coupled with a fairly sophisticated procedure used as a subroutine to check whether a given route was feasible considering several loading constraints, such as weight and precedence. Although this work marked a milestone in the auto-carrier routing optimization, it could be improved. For instance, the notion of equivalent length used in Tadei et al. (2002) was kept, which prevented the consideration of detailed loading plans, as in Agbegha et al. (1998). Dell'Amico et al. (2015) strictly imposed the last-in-first-out (LIFO) policy, which implied no reloading was considered at all. Cordeau et al. (2015) extended Dell'Amico et al. (2015) by considering a multi-day planning problem with a rolling horizon approach.

Hu et al. (2015) approached the loading constraints differently by modeling each automobile as a three-dimensional irregular shape. The nature of loading in their study was thus to pack irregular items to an auto-carrier. They also introduced a learning procedure to discover loading patterns that mapped a combination of automobiles to an auto-carrier. They developed an evolutionary algorithm to solve a vehicle routing problem with packing constraints. Similarly, Liu et al. (2016) considered the three-dimensional geometric contour of finished automobiles when automobiles were assigned to auto-carriers. Routing decisions were not studied, because it was assumed that an auto-carrier followed a fixed route. Nonetheless, Liu et al. (2016) did include routing costs, which was determined only by the fixed route of an auto-carrier.

In an unpublished working paper by Venkatachalam and Sundar (2016), a simple heuristic was used to determine the auto-carrier route; a branch-and-price algorithm was developed to solve the subsequent loading problem for the given route. In the loading subproblem, precise automobile-to-slot assignment decisions were considered, subject to various loading constraints about automobile height, weight, and length. Although the loading aspect considered in this work was relatively sophisticated, loading was not taken into account when routing decisions were made. Thus, the overall solution approach was essentially a sequential approach. The full integration of routing and loading in the branch and price framework was explicitly mentioned as prospective research.

Wang et al. (2018) studied how to optimize the assignment of automobiles to auto-carriers and the route of each auto-carrier so as to maximize the total value of the assigned automobiles minus the total transportation cost. Although Wang et al. (2018) considered some loading constraints, such as the downward compatibility between cars to be loaded and slots on an auto-carrier, they did not consider precise loading and unloading plans. A column generation based heuristic was used to solve the resulting pickup and delivery problem.

Wensing (2018) adopted the sequential decision-making framework in Agbegha et al. (1998) without considering the loading optimization. Like Tadei et al. (2002), Wensing (2018) considered the length equivalents for both the automobile and auto-carrier. The notion of loading pattern was also used to model the loading feasibility. The focus of this study was to compare two scenarios, both over a 10-day period: (i) all arriving shipment orders must be accepted and shipped on the day of their arrival, and (ii) the transport company can freely accept or reject an arriving order. Two scenarios were evaluated in terms of measures such as overall profit and driver productivity. A greedy heuristic based on the manual planning experience and a sophisticated heuristic consisting of construction, perturbation and combination subroutines were compared through numerical studies.

Billing et al. (2018) used a three-phase heuristic to solve an integrated auto-carrier transportation problem, where selection of automobiles, assignment of automobiles to auto-carriers, and routing of auto-carriers were jointly considered, while the precise loading was not involved. A unique contribution was that they explored the potential of using probabilistic information in reducing various costs.

Bonassa et al. (2019) studied a dynamic version of the partitioning problem where automobiles were selected and assigned to auto-carriers, considering various loading restrictions. In this study, it was assumed that the dealers served by an auto-carrier were located in a cluster, which made routing trivial. Also, the actual route length was not part of the objective function. So, routing decisions were not optimized. Loading was not optimized either because it was assumed that precise loading plans were up to individual auto-carrier operators after automobiles were assigned to an auto-carrier.

Juárez Pérez et al. (2019) developed a two-phase heuristic for the auto-carrier transportation problem. First, they used an insertion heuristic to generate an auto-carrier route; then they checked whether it was feasible for all the assigned automobiles to be loaded to the auto-carrier without violating the automobile height constraint or the LIFO policy.

Wang and Chen (2020) considered a similar problem to the one studied in Wang et al. (2018), but significantly simplified routing decisions by approximating the actual transportation cost when the involved dealers were located within a small radius of the auto-carrier depot, as the authors believed that the precise optimization of routing was not worthwhile. As in Wang et al. (2018), Wang and Chen (2020) considered an urgency level of each automobile, a full load requirement, and a downward loading compatibility constraint. They also developed a column generation based heuristic to solve their problem.

6.2. Optimization objective and cost modeling

Among the several optimization objectives considered in the literature, two most common choices were the maximization of total profit (revenue minus cost) and minimization of total cost. Table 3 presents a comparison of all those studies reviewed in Section 6.1 by the cost type being considered. As the choice of optimization objective is often related to the auto-shipping practice in a region or country, the applicable country is also listed in Table 3. Routing cost is the travel cost/time related to a few factors, such as the actual travel distance of auto-carriers, the number of dealership visits, and the number of auto-carriers used. If the precise routing is not considered, the routing cost may be approximated (Wang and Chen 2020) or even considered fixed (Liu et al. 2016). Loading cost is usually the number of reload operations. Revenues represent incomes or rewards for selecting and shipping a subset of automobiles when not all automobiles can be shipped by available auto-carrier capacity. For instance, some priority measure defined for each automobile (“criticality index” in Tadei et al. (2002) and “urgency level” in Wang et al. (2018)) is multiplied by a coefficient to represent the reward or income for shipping this automobile. It can be seen from Table 3 that routing is considered in almost all the studies, in an exact or approximate manner. While all studies reviewed in Section 6 involved the loading aspect, very few of them have considered loading cost explicitly. In those studies, loading feasibility was checked in some simple or sophisticated way, but the quality of loading decisions (if considered at all) was not quantified.

An illustrative example can be used to demonstrate that checking the loading feasibility only without measuring the quality of loading plans may yield inferior solutions. Suppose there are two candidate auto-carrier routes associated with the same set of automobiles to be delivered, whose routing costs are 10 and 8, respectively. While for both routes, the loading constraints are satisfied, the minimum numbers of reload operations for the two routes are 1 and 6, respectively. Suppose the cost of each reloading operation is equivalent to one unit of routing cost. Clearly, in this example, without considering the routing impact on reloading, the second route would be viewed as “optimal”; however, after considering the number of reloads, the first route should be more desirable.

Unlike most other studies which did not involve the payments to individual auto-carriers, Bonassa et al. (2019) carefully analyzed the cost structure of shipping an automobile from the customers’ perspective based on the practice in Brazil. The cost of delivering a car was based on the distance between a depot and its destination, not the actual route length. Their optimization problem was defined from the perspective of a broker (a third-party logistics company), which outsourced shipments to independent truckers. The objective was thus to minimize the total shipping payments made from a broker to truckers.

Table 3: Comparison by cost types considered

Study	Routing cost	Loading cost	Revenues	Country
Tadei et al. (2002)	Yes	No	Yes	NA
Miller (2003)	Yes	Yes	No	NA
Dell’Amico et al. (2015)	Yes	No	No	Italy
Cordeau et al. (2015)	Yes	No	No	Italy
Hu et al. (2015)	Yes	No	No	China
Liu et al. (2016)	Yes	No	Yes	China
Venkatachalam and Sundar (2016)	Yes	Yes	No	NA
Li and Zhang (2016)	Yes	No	No	NA
Wang et al. (2018)	Yes	No	Yes	China
Wensing (2018)	Yes	No	Yes	Germany
Billing et al. (2018)	Yes	No	No	Germany
Bonassa et al. (2019)	Yes	No	No	Brazil
Juárez Pérez et al. (2019)	Yes	No	No	Mexico
Wang and Chen (2020)	Yes	No	Yes	China

6.3. Constraints

Many constraints on loading, reloading, and routing were considered in the literature, mainly depending on how the physical configuration of an auto-carrier was modelled. Tadei et al. (2002) treated an auto-carrier as a single loading plane with a fixed length. Thus, a single capacity constraint was used to model all the loading complications. Dell'Amico et al. (2015) extended Tadei et al. (2002) by considering multiple loading platforms instead of only one, while the concept of equivalent length was kept. Dell'Amico et al. (2015) argued that modern European auto-carriers were very flexible and could not be modelled using slots (see Agbegha et al. (1998) for the U.S. auto-carrier configuration). Similar to Agbegha et al. (1998) and based on the practice in China, Wang et al. (2018) modelled an auto-carrier as a set of a fixed number of slots. Such slots had different dimensions and could hold automobiles of different sizes. Wang et al. (2018) also considered the so-called downward compatibility, which prevented a small slot from holding a large automobile while allowing a large slot to hold automobiles of any size. Unlike the above studies, Hu et al. (2015) and Liu et al. (2016) considered a three-dimensional capacity of an auto-carrier so that loading was essentially a three-dimensional bin packing problem. Liu et al. (2016) further considered the possibility of lifting the front of an automobile to form an angle with the loading platform in order to reduce the length requirement (which would allow more automobiles to be loaded within the fixed platform length). Understandably, this placement of automobiles at an angle reduced the length requirement but would require more vertical space. Although the choice of auto-carrier capacity modeling approach was based on the local context, it is noted that certain approaches prevented the consideration of important operations in loading. For instance, reloading cannot be further considered in the physical models assumed by Tadei et al. (2002), Hu et al. (2015), and Liu et al. (2016).

Table 4 shows a comparison of the relevant studies by the constraints considered. Most studies have considered the detailed dimensions or types/classes of automobiles, and some of them further considered the automobile weight. The LIFO loading requirement (Dell'Amico et al. 2015) was critical because it directly related auto-carrier routes to loading plans. Dell'Amico et al. (2015) imposed the LIFO loading to avoid complex and expensive reloading operations; however, this strict requirement unnecessarily increased the auto-carrier travel distance, or the number of auto-carriers needed. The impact of this LIFO loading restriction can be seen in the following example. There are two possible auto-carrier routes, namely R1: [A+, B+, B-, A-] and R2: [A+, B+, A-, B-], where "+" means pickup and "-" means delivery. In terms of routing, the second route R2 has a significantly lower cost than R1. However, if the LIFO loading restriction is considered, R2 is no longer feasible, leaving R1 as the only choice. Clearly, if the LIFO loading restriction is lifted in some cases, a significantly better solution may be found. Therefore, Venkatachalam and Sundar (2016) did not adopt this strict LIFO loading requirement and imposed a limit on the number of reloads instead, which represented a better compromise between two interrelated decisions of loading and routing.

Although split deliveries (a delivery is split among multiple vehicle visits to a customer, Dror and Trudeau 1989) may not be preferred by customers due to the inconvenience, Dell'Amico et al. (2015) explored the benefits of allowing split deliveries by conducting a sensitivity analysis of the penalty for split deliveries. They found that as the penalty for split delivery increased, the number of dealership visits significantly dropped, while both the number of required auto-carriers and auto-carrier kilometers increased. Therefore, a good balance between the dealers' preferences and the carriers' costs should be carefully selected.

Time window constraints were rarely considered in the existing studies. In most cases, some simple priority measures or urgency levels were considered, such as in Tadei et al. (2002). As all the studies were for the delivery problem of finished vehicles, the starting point of all auto-carriers was the same, which was called depot in Dell'Amico et al. (2015), vehicle distribution center in Hu et al. (2015), or new car storage yard in Juárez Pérez et al. (2019). It was also assumed that all finished vehicles were loaded all together at the same location. Miller (2003) pointed out that routing would be much more complex when vehicles were picked up at multiple locations or simultaneous pickups and deliveries could occur at one location, such as in the case of picking up and delivering used vehicles for individual customers. Wang et al. (2018) made a simplifying assumption on the order of pickup and delivery by restricting all pickups to occur before any delivery.

While Table 4 shows major constraints, not all constraints that were considered in the literature are included in this table. Also, a detailed review shows that researchers may have modelled some constraints quite differently. For instance, a full-load constraint was enforced in Wang et al. (2018) and Wang and Chen (2020), while in Bonassa et al. (2019), the number of automobiles per auto-carrier ranged from 8 to 11. For another example, in most studies the supply of auto-carriers was assumed to be limited. Nonetheless, Billing (2018) and Bonassa et al. (2019) assumed an unlimited supply of auto-carriers, which implied that all orders available on one day could be shipped on the same day. Some additional side constraints were also considered, such as carrier-city compatibility constraint (Wang et al. 2018) that required an auto-carrier to visit a limited set of dealers, and the infrastructure-carrier compatibility constraint (Wensing 2018) that prevented a fully loaded auto-carrier from traversing a specific narrow road segment or bridge.

Note that "NA" in Table 4 means not applicable. For instance, Tadei et al. (2002) did not study precise routing or loading optimization, and hence whether LIFO was required or not and whether there were time windows become irrelevant in their problem.

Table 4: Comparison by constraints considered

Study	Auto-carrier Capacity	Automobile dimension	Weight limit	LIFO required	Reload	Split delivery	Time windows
Tadei et al. (2002)	One level, equivalent length	Yes	No	NA	NA	NA	NA
Miller (2003)	Two levels, ten slots	No	No	NA	Yes	NA	No
Dell'A. et al. (2014)	Multi-platform, equivalent length	Yes	Yes	Yes	No	Allowed	No
Cordeau et al. (2015)	Multiple levels, equivalent length	Yes	Yes	Yes	No	Allowed	Yes
Hu et al. (2015)	Three-dimensional space	Yes	Yes	NA	NA	Prohibited	Yes
Liu et al. (2016)	Three-dimensional space	Yes	Yes	NA	NA	Allowed	NA
Venkatachalam and Sundar (2016)	Multiple heterogeneous slots	Yes	Yes	No	Yes	Allowed	No
Li and Zhang (2016)	Two levels, equivalent length	Yes	No	NA	NA	NA	No
Wang et al. (2018)	Multiple heterogeneous slots	Yes	No	NA	NA	NA	No
Wensing (2018)	As in Hu et al. (2015)	Yes	No	NA	NA	NA	No
Billing et al. (2018)	As in Hu et al. (2015)	Yes	No	NA	NA	Prohibited	No
Bonassa et al. (2019)	Equivalent length	Yes	No	NA	NA	NA	NA
Juárez Pérez et al. (2019)	Multiple levels	Yes	No	Yes	No	NA	Yes
Wang and Chen (2020)	Multiple heterogeneous slots	Yes	No	NA	NA	NA	No

6.4. Solution algorithms

The auto-carrier routing problem with loading considerations is clearly NP-hard, as Tadei et al. (2002) showed that the simplified version of the problem they studied was already strongly NP-hard. Therefore, almost all existing studies developed heuristics to solve their problems. The heuristic algorithms used in the literature can be classified into the following types: sequential heuristics, iterated heuristics, math programming (MP) based heuristics, and other heuristics.

Sequential heuristics can be used when multiple types of decisions are optimized sequentially in separate steps. They include the three-step heuristic in Tadei et al. (2002), the three-phase heuristic in Billing et al. (2018), and the two-phase heuristics in Venkatachalam and Sundar (2016) and Juárez Pérez et al. (2019). In the first step, Tadei et al. (2002) simplified the auto-carrier capacity constraints and developed an integer program to assign auto-carriers to a subregion of the whole geographic area. For each auto-carrier in a subregion, initial feasible solutions were constructed in Step 2 and improved in Step 3. Similarly, in Phase 1 of their algorithm, Billing et al. (2018) selected orders for delivery on a day based on the relative distance between dealers. In Phase 2, they used the routing heuristic of Clark and Wright (1964) to construct vehicle routes; in Phase 3, vehicle routes were improved through some ad hoc strategies. Although such sequential heuristics were relatively easy to implement and may work well in some cases, a primary drawback is that possible effect of decisions made in a later stage on earlier decisions was ignored. For instance, without fully exploring the routing possibilities, it is suboptimal to determine whether an auto-carrier should be assigned to a specific region or whether an automobile should be selected for delivery. Venkatachalam and Sundar (2016) used a heuristic to solve the routing subproblem of the overall problem first. Given an auto-carrier route, the loading problem was then solved by a column generation based branch-and-bound algorithm. Juárez Pérez et al. (2019) used an insertion heuristic to generate routes first, followed by another heuristic to generate feasible loading plans for the given routes.

Unlike sequential heuristics, iterated heuristics repeatedly evaluate the quality of earlier decisions so as to improve the current decisions. Miller (2003) proposed a construction heuristic to generate an initial solution and improvement heuristics such as k-opt or 2-opt to improve the initial solution. Dell'Amico et al. (2015) started with some initial solutions generated by a greedy heuristic. Then, in each iteration, they perturbed the current solution to obtain an updated solution, which triggered the next iteration. The loop was iterated until a time limit was reached. Under this iterated solution framework, Dell'Amico et al. (2015) designed their own greedy heuristic, perturbation method, and local search operators. For instance, they considered the following local search operators: intra-route move, one-to-one dealer swap, auto-carrier interchange, and route addition, among others, to obtain a better auto-carrier route. To ensure the loading feasibility of an auto-carrier route, a procedure for checking loading feasibility was invoked throughout this iterated heuristic. Cordeau et al. (2015) generalized the iterated heuristic of Dell'Amico et al. (2015) to a more general problem. Wensing (2018) proposed two heuristics: a greedy construction type of heuristic based on the manual planning technique, and an iterated heuristic consisting of construction, local search, perturbation and recombination procedures.

While most papers formulated their problems mathematically as MIP or IP formulations, only a few of them actually designed algorithms based on these formulations. A column generation based heuristic was developed in Wang et al. (2018) based on a set packing type of formulation for their problem. In this heuristic, they used a two-step heuristic to generate initial auto-carrier routes for each auto-carrier as an initial set of columns, and solved the problem heuristically by decomposing the formulation into a master problem and a subproblem. The LP relaxation of the master problem was solved to generate dual variable values, which were then used in the subproblem. The subproblem was solved to generate new auto-carrier routes, which were added to the master problem. A similar column generation based heuristic was also used in a later study Wang and Chen (2020). Liu et al. (2016) formulated a loading problem as an integer program and designed a branch-and-bound algorithm with a greedy search based on oscillation analysis to solve it.

Bonassa et al. (2019) formulated their problem as an integer program and solved it directly by a commercial solver.

The heuristics used in a couple of other papers do not seem to fall within any type of heuristics reviewed above. Hu et al. (2015) developed an evolutionary algorithm. A sequence of dealers represented an auto-carrier route, as the auto-carrier began and ended a route at the same depot. This sequence was also coded as a chromosome in the evolutionary algorithm. Selection, crossover and mutation operators were also used to generate better routes. Li and Zhang (2016) combined simple heuristics with integer programming. They first used Dijkstra's algorithm to calculate the shortest paths from the distribution center to all customers. Then, according to the demand of the customers on each shortest path, they used an integer linear programming model to calculate the number of cars unloaded at each point of the shortest path. Finally, a heuristic was used to figure out the loading plan and routing strategy of each transporter traveling along the shortest path.

Table 5 presents a comparison of those studies by formulation and solution method. Although in some studies, dynamic or stochastic information was involved, most studies deal with static and deterministic problems. While the underlying problems can all be formulated as IP or MIP formulations, few studies utilized such formulations to develop heuristics. It is known that math programming based heuristics generally have a more robust performance than rule or local search based heuristics. As Agbegha et al. (1998) pointed out, since automobile transportation was a major industry, even one percent of cost savings could represent millions of dollars. Therefore, there is a clear need to develop better optimization methods, especially IP / MIP based heuristics, to yield more cost savings.

Table 5: Comparison by formulation and solution method

Study	Formulation	Solution Method
Tadei et al. (2002)	Mixed integer program (MIP)	Sequential heuristics
Miller (2003)	NA	Iterated heuristics
Dell'Amico et al. (2015)	Integer program (IP)	Iterated heuristics
Cordeau et al. (2015)	NA	Iterated heuristics
Hu et al. (2015)	MIP	Other heuristics
Liu et al. (2016)	IP	MP-based heuristics
Venkatachalam and Sundar (2016)	MIP	Sequential heuristics
Li and Zhang (2016)	IP	Other heuristics
Wang et al. (2018)	IP	MP-based heuristics
Wensing (2018)	NA	Iterated heuristics
Billing et al. (2018)	MIP	Sequential heuristics
Bonassa et al. (2019)	MIP	MIP solved directly by a commercial solver
Juárez Pérez et al. (2019)	IP	Sequential heuristics
Wang and Chen (2020)	IP	MP-based heuristics

6.5. Test Instances

Randomly generated test instances as well as real-world instances were used to test the solution methods reviewed in Section 6.4. Table 6 compares the existing studies by the test instances used. They are divided into three groups by the overall size of test instances (small, medium, and large), in terms of the number of auto-carriers available, the number of automobiles to be shipped, and the number of dealers involved. Very few of those test instances were made publicly available with the exception of the ones used by Dell'Amico et al. (2015). Not all studies have described clearly how those test instances were generated, with Billing et al. (2018) as one of the exceptions. We thus review in detail the instances used in Dell'Amico et al. (2015) and Billing et al. (2018).

Dell'Amico et al. (2015) considered in total 23 test instances, one for each working day. For all instances, two types of auto-carriers with various carrying capacities were used. The first type of auto-carrier had four loading platforms, while the second type had two platforms. The platform dimension, payload and other relevant parameters were also provided for both types of auto-carriers. There were 723 automobile models, which were consolidated into 14 loading classes for the first type of auto-carrier and 8 loading classes for the second type of auto-carrier. Each order had detailed information about the vehicle to be shipped and its destination dealer. In each instance, the shortest distance between each pair of locations (auto-carrier depot or dealer location) derived from some GIS software was also provided. The test instances can be found at http://www.or.unimore.it/site/home/online-resources/auto-carrier-transportation-problem/instances.html

Billing et al. (2018) considered all the cities with a population larger than 30,000 in a German state called North Rhine-Westphalia. Each city was assumed to have at least one auto dealer, while for cities with a much larger population, there were more dealers proportionally. The travel distance between any pair of locations was a linear function of the geographical distance. The probability for each dealer to make an order was given and on average 50 dealers made at least one order. The order size (number of automobiles) was uniformly distributed on the range 1 to 3. The automobile type was uniformly sampled from small, medium and large. The delivery deadline was also uniformly sampled from 0, 1, 2, and 3 days into the future. As Billing et al. (2018) solved a multi-period planning problem, the length of the planning horizon was 30 days.

Table 6: Comparison of test instances

Size	Study	Instance type	# of instances	# of auto-carriers	# of automobiles	# of dealers
Small	Miller (2003)	Synthetic	12	3	30	15-30
	Li and Zhang (2016)	Synthetic	1	10	~100	9
	Liu et al. (2016)	Real-world	18	1-18	3-20	-
	Wensing (2018)	Real-world	1	3	25	11
Medium	Venkatachalam and Sundar (2016)	Synthetic	72	15-100	100-600	5-25
	Wang et al. (2018)	Synthetic	27	5-15	100-250	5-50
		Real-world	7	1-15	68-241	28-49
	Billing et al. (2018)	Synthetic	20	-	-	159

	Bonassa et al. (2019)	Synthetic	30	unlimited	8-23	-
		Real-world	4	-	535-863	7-78
	Wang and Chen (2020)	Synthetic	24	5-40	100-500	10-80
Large	Tadei et al. (2002)	Synthetic	12	100-400	2,000-4,000	400-800
		Real-world	3	146-160	1,610-3,569	553-762
	Dell'Amico et al. (2015)	Real-world	23	30-117	272-1,139	96-251
	Cordeau et al. (2015)	Same test instances as in Dell'Amico et al. (2015)				
	Hu et al. (2015)	Synthetic	12	-	-	50-420
	Juárez Pérez et al. (2019)	Synthetic	10	4-490	20-3,000	44
		Real-world	1	655	3,884	44

A review of the test instances used in the literature yields the following findings:

First, the test instances were used to test algorithms that were not developed for the same loading and routing problem. In fact, as reviewed in Sections 6.2 and 6.3, the problems studied can differ significantly in terms of both the objective function and the constraints involved. Thus, it is not recommended to directly compare sizes of the test instances (e.g., number of auto-carriers, number of dealer locations) without looking into the problem context first.

Second, unlike in the loading and reloading studies reviewed in Section 5, which all used the entirety or part of the test instances from Agbegha et al. (1998), in the studies on the integrated optimization problem, almost all the authors only used the instances developed by themselves or their co-authors. For instance, Cordeau et al. (2015) adopted the instances of Dell'Amico et al. (2015), while Dell'Amico was a co-author of Cordeau et al. (2015); Billing et al. (2018) partially adopted the instances of Wensing (2018), who co-authored Billing et al. (2018). The only test instances that are made available online are the ones used by Dell'Amico et al. (2015). This means that it is difficult to replicate the computational experiments in all those studies except Dell'Amico et al. (2015).

Third, the computational times reported in those studies were not comparable, because (1) different computing platforms (e.g., programming language, CPU, RAM) were used, and (2) the underlying optimization problem was not the same, as discussed above.

Although the various heuristics proposed in the existing studies may have worked well for their problems, it is not clear whether these algorithms will still be efficient for problems where routing and loading decisions are more closely integrated than in the existing problems. For instance, the column generation based heuristic takes a few hundred seconds to an hour to solve an instance in Wang et al. (2018), which did not involve loading optimization. The iterated local search heuristic of Dell'Amico et al. (2015) had a solution time limit of 1,500 seconds, among which 534 seconds on average were spent on checking only the loading feasibility for each test instance. As Dell'Amico et al. (2015) only checked the loading feasibility rather than generating a precise loading plan, the computation time will likely be much higher when loading optimization is further incorporated.

7. AUTO SHIPPING BY RAIL AND SEA

Despite the authors' extensive literature search, no optimization studies for automobile shipping by air have been found in the literature. In this section, we briefly review existing optimization studies related to automobile shipping by rail and sea.

7.1. Motorail load planning

As reviewed in Section 3.3, motorail transportation is the long-haul transportation of passengers with their personal vehicles on the same train. The loading aspect of motorail transportation is to assign many automobiles of various types and dimensions to the auto-carrying wagons, decks, and slots, respecting weight, height, width, and other technical constraints. As the train to be loaded has a fixed origin and destination, all automobiles to be loaded to a train must also have the same origin and destination, which means all automobiles are loaded (and unloaded) simultaneously, eliminating any reloading complexities during the train route.

Lutter and Werners (2015) developed the first loading optimization model for motorail transportation. In their problem, each auto-carrying wagon has two decks, on each of which there are a fixed number of positions/slots. Lutter and Werners (2015) proposed a 4-index integer programming formulation with the objective of maximizing the total number of loaded automobiles. With this loading optimization model, they further developed a decision support system to help determine whether an arriving automobile order should be accepted. Lutter (2016) later extended Lutter and Werners (2015) by proposing two tighter reformulations of the original 4-index formulation. The first 3-index reformulation was derived by converting the height constraint in Lutter and Werners (2015) into a knapsack-type constraint for each loading deck, which was shown to yield a stronger linear programming formulation. To overcome the symmetry in the 3-index reformulation, Lutter (2016) proposed another extended formulation, which was solved by a customized branch and price algorithm. Lutter (2016) demonstrated the performance of the proposed various formulations by solving 100 real-world test instances from a company named DB Fernverkehr AG in Germany. Computation results indicated that the last formulation solved by the branch and price algorithm had the best efficiency: 95% of the test instances could be solved to optimality in 90 seconds.

In contrast with the highway-based auto-carrier loading problem, the loading optimization problem for motorail transportation could be solved much more efficiently. The primary reason is that in motorail loading only assignment decisions are considered, while no reloading complications are involved.

7.2. Ro-Ro shipping and container shipping

When automobiles are shipped overseas, they are usually shipped via Ro-Ro vessels. It is also possible that some of them are shipped in containers via container ships.

Øvstebø et al. (2011a) formulated the Ro-Ro ship stowage problem for a given Ro-Ro ship whose sailing route was pre-determined. In this ship stowage problem, Øvstebø et al. (2011a) tried to determine what automobiles should be selected for transportation and how to assign them to a deck of a Ro-Ro ship and further to a lane on a deck, at each port of the sailing route. When vehicles were selected for loading, they were rolled on board the Ro-Ro ship; they were rolled off at the destination port. Each lane on the Ro-Ro ship could take automobiles in a LIFO fashion. Some penalty costs were incurred if automobiles in the back were rolled off before those in the front. Øvstebø et al. (2011a) developed a mixed integer program with the objective of maximizing the difference between total revenues and penalties due to the temporal rearrangements. A special stability constraint was incorporated to avoid placing too many automobiles on one side of a ship. They proposed two solution approaches: solving directly with the solver Xpress and using a customized heuristic. Through solving a large set of randomly generated instance, they concluded that neither solution approach dominated the other. In other words, the solver Xpress outperformed the proposed heuristic for some instances, not all. Øvstebø et al. (2011b) extended their study Øvstebø et al. (2011a) by further considering the routing and scheduling of a Ro-Ro ship. They developed a mixed integer program, which was solved by the solver Xpress directly or a tabu search algorithm. Computational results indicated that only the tabu search algorithm could solve realistically sized problem instances.

The stowage problem in Øvstebø et al. (2011a) was analogous to the auto-carrier loading problem studied by Agbegha et al. (1998). While there were many other mode-specific differences (such as quite different auto-carrying equipment), a clear difference was that in auto-carrier loading, stability was not given as much attention as in Ro-Ro ship stowage. In addition, in the routing and scheduling problem studied by Øvstebø et al. (2011b), only a single Ro-Ro ship was considered, while in the highway-based auto-carrier routing problem, such as those in Dell'Amico et al. (2015) and other papers reviewed in Section 6, multiple auto-carriers were considered.

Kang et al. (2012) developed an integer program to determine car allocations to each vessel and the voyage route with the objective of minimizing the total logistics cost. The integer program was solved by a genetic algorithm. Additional studies related to other aspects of Ro-Ro shipping, such as fleet deployment, inventory management (Chandra et al. 2016a) and speed optimization (Andersson et al. 2015), were not reviewed in detail here. This is because in most such general Ro-Ro shipping studies, very few specific considerations with respect to automobiles were included, as Ro-Ro ships are designed to carry any wheeled cargo, such as automobiles, trailers, railroad cars, farming equipment, and military equipment.

In contrast with shipping via Ro-Ro ships, not many studies have studied shipping automobiles in containers. Xuan (2014) studied the containerized automobile shipping and indicated all Tesla vehicles were shipped using containers. Xuan conducted cost-benefit analyses from automobile manufacturers and container lines and identified target markets for the China Shipping Group. Dael (2014) also proposed the stowage of automobiles in a container to take advantage of the excess capacity in a container vessel. Through an analysis, Dael (2014) tried to convince auto manufacturers that shipping automobiles in containers may be more cost-effective than shipping them on a pure car carrier. As a container can hold very few automobiles (e.g., no more than 4) and it is impossible to rearrange them until the container is delivered, automobile reloading complexities are likely to be avoided. Once automobiles are properly loaded into a container, the optimization problem is a general containership routing and scheduling problem, which was reviewed by Christiansen et al. (2013).

As compared with those automobile shipping optimization studies involving rail and water, it is relatively clear that highway-based automobile shipping has been studied the most extensively and thoroughly. Therefore, highway-based optimization models may be adapted to for the water-based and rail-based automobile distribution problems, despite various mode-specific characteristics.

8. STRATEGIC AND TACTICAL PLANNING STUDIES

All studies reviewed in Sections 5, 6, and 7 are at the operational level, as all those underlying optimization decisions are made every day or every few days. Those decisions at the tactical or strategic level are made much less frequently, such as every five years or once in a decade. Examples of such tactical and strategic decisions are locating rail terminals in a multimodal automobile distribution network and negotiating a contract with a major railway company. In this section, we review those automobile shipping optimization studies at a higher level than operational.

8.1. Mode selection and network design

The design of automobile distribution network is critical because the network structure eventually defines how efficient automobile distribution can become in the long run. As pioneers in this area, Miller et al. (1996) developed a mixed integer program to help an international automobile manufacturer to design the automobile distribution network in North America in the long run. Cars and light trucks that were manufactured in the U.S. or imported from about 100 ports of entry should be delivered to approximately 275 dealer regions throughout the U.S. As there were a few candidate transportation modes, such as truck, rail, and containerized intermodal rail, the first decision was to determine the optimal modal mix. As rail played an essential role in the intermodal distribution network, the second decision was to optimize the locations of rail terminals. The objective of the proposed mixed integer program was to minimize the total logistics cost of delivering automobiles. Miller et al. (1996) used this optimization model to compare a few different network scenarios by the delivery cost, mode share, and transit time.

Eskigun et al. (2005) also studied the automobile distribution network design problem considering lead-time related costs, due to the increasing interest of the automotive industry in reducing lead-time. They formulated a nonlinear programming model to minimize the sum of the fixed cost of opening distribution centers, transportation cost, and lead-time related cost. After introducing additional binary variables and constraints, the nonlinear program was linearized and solved with a Lagrangian heuristic. They generated and solved 56 problem instances based on the real-world data. Numerical analyses indicated that the proposed heuristic solved the formulation very well: achieving a less than 1 percent optimality gap in about 10 hours on average.

Along this research line, Jin et al. (2010) optimized the mode choice (truck or rail) for each plant-dealer pair for a U.S. manufacturer based on delivery cost. Trucking companies charged a fixed cost per automobile and a variable cost based on mileage, while a railcar's cost followed a piecewise linear function of the shipping volume, reflecting the quantity discounts or scale economies. An integer program was developed for this mode choice problem, which was solved directly by a commercial solver. Results from solving randomly generated test instances showed that the integer program could be solved in a few seconds. Jin et al. (2010) also simulated the multi-round contract negotiation process between the auto manufacturer and a rail company.

Hei et al. (2014) considered the optimal distribution of automobiles from one distribution center to other locations on a multimodal freight network. Three modes were considered, namely direct trucking, road-rail, and road-water. While trucking was the most flexible (i.e., can be dispatched almost at any time), both trains and ships had weekly schedules. Hei et al. (2014) built an integer program to determine what mode and route should be selected for those vehicles heading to each destination. In the numerical studies, vehicles were to be distributed from Shanghai to other 17 locations in China. The multimodal network consisted of 17 trucking routes, 18 road-rail intermodal routes, and 8 road-water routes. In the numerical studies, a commercial solver was used to directly solve the integer program, as the solution time was within 1 minute. Meng et al. (2015) expanded the sensitivity analyses of train and ship capacities in Hei et al. (2014) by considering the uncertain automobile shipment volume. As the auto manufacturer must purchase train and ship capacities in advance for a fixed time period (e.g., 1 year), Meng et al. (2015) developed an optimization model to determine the optimal amount of capacity to procure from rail and shipping companies. The problem was complicated by uncertain delivery demand, which fluctuated over time. Therefore, a two-stage stochastic program was formulated to minimize the expected delivery cost, which was solved by a sample average approximation method. They also presented a case study based on the network used by the Shanghai Automobile Industry Corporation in China.

Wang et al. (2016) used a fuzzy Delphi method to evaluate the multimodal auto transportation network design to ship used cars from South Korea to Centra Asia, based on the following five factors: cost, reliability, transportation capability, total time, and security. Chandra et al. (2020) studied the viability of modal shift from in-land transportation to short-sea shipping at a strategic level for auto shipping in India.

They developed a mixed integer program to select the route alternative, ship type and share of shipment volumes by coastal shipping. One origin (Chennai) and a set of destinations in India were considered in the case study. Depending on the availability of coastal shipping and cost parameters, they considered 22 configurations. The formulation was solved directly by a commercial solver with computation time ranging from a few minutes to more than two hours. In a follow-up study (Dong et al. 2020), the tradeoff between economic costs and environmental impacts of the distribution network design was furthered considered.

In summary, quite a few studies have addressed the design of a multimodal automobile distribution network, although the optimization decisions considered in such studies differed. In those studies, a centralized decision-making framework was adopted for the auto manufacturer. However, in reality, trucking companies, rail companies and other carriers were also involved. In those existing studies, behavioral responses from other parties involved were not properly modelled. Therefore, a game-theoretic approach may be considered to fully characterized the behavioral interactions between the manufacturer and other parties in the design of automobile distribution network.

8.2. Empty autorack management

Sherali and his collaborators (Sherali and Tuncbilek 1997, Sherali and Suharko 1998, Sherali and Lunday 2011) have studied various problems in the fleet management of autorack (railcars used for shipping automobiles) for about three decades. After automobiles were delivered, the empty autoracks should be distributed to load other automobiles. Under a naive approach, empty autoracks were returned to their points-of-origin, known as reverse routing. A more efficient approach was based on the pooling of empty autoracks as a common resource among multiple shippers and carriers. Sherali and Tuncbilek (1997) studied how many autoracks an automobile manufacturer should acquire in any given year, i.e., fleet sizing problem. Sherali and Suharko (1998) focused on a tactical planning problem of repositioning empty autoracks to where they needed to be reloaded. Sherali and Lunday (2011) focused on the potential equity issues in the pooling agreement and proposed a few alternative schemes to apportion autoracks to automobile manufacturers. Although the empty autorack repositioning problems studied by Sherali were motivated by the rail-based auto shipping practice, the general class of empty fleet (not necessarily autorack fleet) management problems has been extensively studied in the literature. Interested readers are directed to Dejax and Crainic (1987), Spieckermann and Voß (1995), and Gorman (2015), among others.

9. RESEARCH GAPS AND PROSPECTIVE OPPORTUNITIES

In this section, we identify important literature gaps and propose strategies to address them. We also evaluate the impacts of disruptions and emerging technologies.

9.1. Hybrid service model

Three service models are summarized in Section 3.2. The review findings indicate that all the existing automobile optimization studies assume the first two service models, because they focus on the shipping of finished vehicles. Service model 3 is a hybrid one, as shippers may be individuals rather than auto manufacturers; the broker may have its own fleet that can be used to deliver orders, but may also outsource some orders to other independent carriers. The consideration of service model 3 has a few important implications for the automobile distribution optimization, which are discussed as follows.

9.1.1 Pickup and drop-off precedence

In the case of finished vehicle distribution, all automobiles for shipping are often loaded simultaneously at a common location and delivered at their respective dealer locations. Route design usually involves delivery destinations only and is essentially about sequencing of dealer locations. Conversely, individual shippers generally have not only different drop-off locations, but also different pickup locations, which means route design will need to consider the sequencing of both pickup and drop-off locations. It is quite possible that not all pickups should occur prior to any deliveries, unlike what is assumed in most existing studies reviewed in Section 6. Clearly, when multiple individual shippers are served by an auto-carrier, the routing complexity is much higher than that for an auto manufacturer that has all vehicles ready for pickup at a single location, even though the number of automobiles assigned to an auto-carrier remains the same.

9.1.2 Auto-carrier capacity

Because in service models 1 and 2, there is only one pickup location where the auto-carrier has the highest load, which means if the auto-carrier capacity constraint is not violated at the beginning of a route, the capacity constraint is met at any other locations. Service model 3 is more general, because not all pickups occur at the same location, and some locations may even allow for simultaneous pickups and drop-offs. In other words, the cumulative load in an auto-carrier route (measured by the total number of automobiles, or the sum of their equivalent lengths, depending on how an auto-carrier is modelled) may exceed the auto-carrier capacity (measured by either the number of slots or an equivalent length limit). Thus, from a capacity constraint point of view, automobile shipping optimization in the context of service model 3 becomes much more complex.

9.1.3 Pricing of auto shipments in customer quotes

In finished vehicle distribution, the automobile manufacturer usually has a long-term contract with a carrier, which is renewed annually or every five years. A manufacturer usually negotiates with rail carriers every five years (Jin et al. 2010), because the competition in the rail sector is relatively low and locations of rail assets (e.g., tracks, intermodal terminals) are fixed (Katcoff 2011). The contract with a highway carrier is mostly renewed annually. Unlike corporate shippers, individual shippers do not sign any long-term contract with a broker or carrier; instead, they seek a customized and one-time quote, usually from a broker, as shown in Figure 6. A shipper may try to obtain quotes from multiple brokers. When serving individual shippers, a broker needs to optimize the quote for an individual shipper, by taking into account what other automobile shipping orders it receives, and the available capacity of the auto-carrier fleet owned by itself or its contractors. Clearly, the optimization of a quote is interrelated with other operational decisions, such as automobile assignment, auto-carrier loading, and routing. We also note that quote optimization is largely about pricing, which has a direct impact on whether a shipper will accept a quote or not, and hence directly affects a broker's profitability.

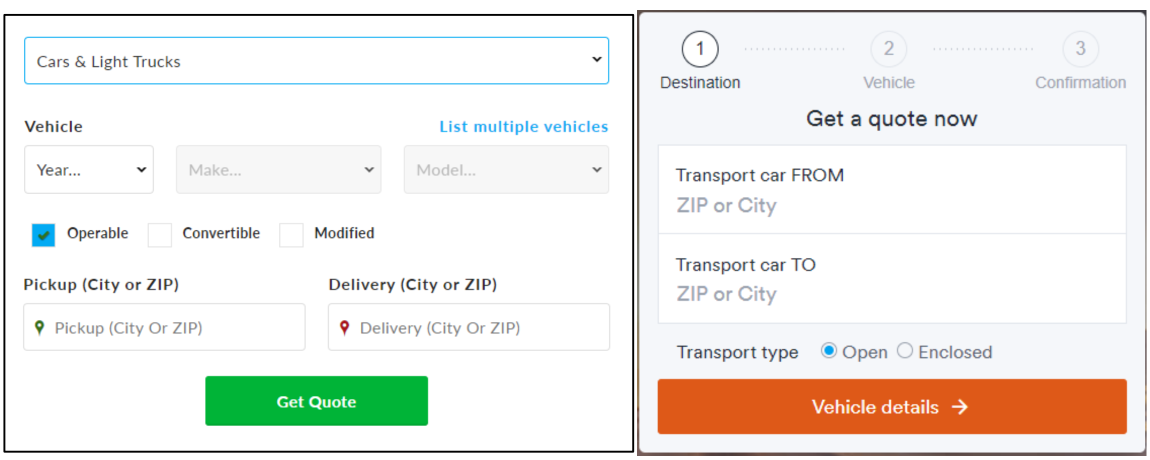


Figure 6: Getting a quote from two anonymous automobile shipping brokers

9.1.4 Consideration of partially filled auto-carriers

In service models 1 and 2, all auto-carriers are assumed to be empty at the vehicle depot. Such an empty auto-carrier does not have a predetermined destination, because its destination depends on the assigned automobile shipments to it. As the distribution is usually one-way, meaning auto dealers do not ship automobiles to auto manufacturers, and hence those auto-carriers are assumed to return to the auto-carrier depot empty (Liu et al. 2016). In contrast to other two service models, in service model 3, individual truckers (also called owner-operators) may make capacity of their trucks (auto-carriers) partially available to a broker, because their trucks are partially filled already (through another source, such as load board or a different broker) and must follow a preliminary itinerary (e.g., arrive at a prespecified location by a due date). Those owner-operators accept further automobile shipment orders allocated by the broker, on the condition that their predetermined schedules are not violated. Therefore, in service model 3, it is necessary to consider both those auto-carriers with full capacity ready and no predetermined travel plans and other partially filled auto-carriers that must follow a predetermined travel plan. Moreover, in service models 1 and 2, it might be reasonable to assume auto-carriers return to their depots carrying no automobiles on a return trip, while in service model 3, such empty backhaul miles can probably be avoided through optimized matching of loads and auto-carriers.

9.2. Full integration of loading and routing optimization

9.2.1 Shortcoming of the sequential heuristic

For ease of discussion, we first present a general mathematical program for the integrated auto-carrier loading and routing optimization problem. We use x and y to represent routing and loading decisions, respectively. The formulation is written as follows:

$$max \quad f(x) + h(y) \tag{1}$$

$$s.t. \quad g(x) \geq 0 \tag{2}$$

$$p(x, y) \geq 0 \tag{3}$$

$$x \in Z^m \tag{4}$$

$$y \in Z^n \tag{5}$$

The objective (1) is to minimize the sum of routing cost $f(x)$ and loading cost $g(x)$. Constraint (2) is a routing constraint. Constraint (3) models the interrelation between routing and loading.

Although quite a few studies (e.g., Dell'A. et al. 2014) have considered loading and routing aspects of automobile shipping and attempted to solve the integrated problem (1-5), none of the existing studies have managed to develop an optimization model that produces simultaneously precise routing plans and precise loading plans. A majority of existing studies focused on routing optimization, while considering loading feasibility only. In these existing studies, even though the problems were already simplified by not considering precise loading plan, the simplified optimization problems were not solved optimally, as various heuristics were used that did not provide any solution quality guarantee.

The idea of the widely adopted sequential heuristic (described in Section 4.1) can be formally stated as follows. First, the routing subproblem (6-8) is solved to find the optimal routing plan x^*.

$$max\, f(x) \tag{6}$$

$$s.t. \quad g(x) \geq 0 \tag{7}$$

$$x \in Z^m \tag{8}$$

Then, the loading subproblem (9-11) is solved to obtain the best possible loading plan given the "optimal" route x^*.

$$max\, h(y) \tag{9}$$

$$s.t. \quad p(x^*, y) \geq 0 \tag{10}$$

$$y \in Z^n \tag{11}$$

The above-mentioned sequential heuristic has a major drawback: it does not consider the interrelations between decisions made at different steps (routing and loading, i.e. x and y variables). For example, an "optimized" auto-carrier route without consideration of its associated loading implications may incur very expensive reloading operations. In the case of transporting very high-end cars, to avoid potential damage during reloading, a strict Last In, First Out (LIFO) policy may have to be adopted during loading; however, the resulting auto-carrier route may have significant detours, which can result in an excessively high routing cost. In a representative study, Venkatachalam and Sundar (2016) run their branch and price algorithm to solve the loading problem for each auto-carrier route. As the number of possible auto-carrier routes is very large and for each given auto-carrier route there are many possible loading plans, this sequential strategy (routing first, loading second) used in Venkatachalam and Sundar (2016) does not work efficiently, which may be the reason this sequential strategy was not implemented in their computational experiments. As of now, there are no studies that have explored full integration of loading and routing decisions. Therefore, research on full integration of loading and routing optimization is needed to fill this critical gap in the literature.

9.2.2 A novel framework for the integrated optimization

It is nontrivial to solve the fully integrated problem from the perspective of optimization. The full integration requires the real-time interaction between routing optimization and loading optimization, and can have a significant advantage in solution efficiency. For instance, when routing decisions being optimized by some insertion-base heuristic, if the effect of an insertion on loading is considered simultaneously, many routes may have been eliminated in the first place due to their enormous loading complexities (implying high loading costs). The advantage of the full integration strategy over the routing first, loading second strategy can also be illustrated with a simple example shown in Figure 7.

Figure 7(a) shows three of the many possible routing plans for a given auto-carrier and a predetermined set of automobiles, as well as all feasible loading plans corresponding to each of these three routes. The underlined numbers in Figure 7 are the routing or loading costs. If a sequential approach is used (routing first, loading second), then each route is constructed completely first, followed by constructing all possible loading plans for the route. Conversely, if an integrated approach is used (route construction and loading plan generation are considered simultaneously), then the mutual impact of route construction and loading plan generation is considered simultaneously and timely before a complete route or loading plan is built. In the process of building route 3 (before it is completely built), it is found that part of the route is identical to some segment of route 2 (which was already built earlier along with all its loading plans), and based on this known information, it is calculated that (i) if constructed fully, route 3 would cost at least 7, and (ii) any loading plan for route 3 would cost at least 2. Thus, any solution combining route 3 and a feasible loading plan would cost at least 9. Therefore, the partially constructed route 3 can be discarded immediately because it would never yield a better solution than the solution combination of route 1 and loading plan 1a (which has a total cost of 8).

Efficient methods need to be developed to jointly optimize routing and loading decisions. One plausible approach is to improve the existing sequential approaches and make them more intelligent so that they can identify suboptimal routes and loading plans and discard them before they are completely constructed. The example illustrated in Figure 7(b) is based on such an approach. To further improve the loading optimization efficiency, one can also investigate whether data driven approaches (that are dramatically different from the existing loading optimization framework based on loading networks) can be applied for evaluating possible loading plans for a given route. For example, a machine learning algorithm may be trained on the known loading patterns that can be generated from the practice or simulation experiments. The trained machine learning model can thus be used to quickly evaluate the loading implications in the process of constructing routes. In other words, the trained learning algorithm can instantaneously predict the loading cost $\eta(x)$ for a given route x, thus eliminating the need to solve the loading subproblem (9-11) given routing decision x. We then can reformulate the integrated optimization model (1-5) into problem (12-14):

$$max\ f(x) + \eta(x) \tag{12}$$

$$s.t. \quad g(x) \geq 0 \tag{13}$$

$$x \in Z^m \tag{14}$$

If this learning algorithm can be successfully developed, inferior routes can thus be identified early on and hence computational time can be reduced.

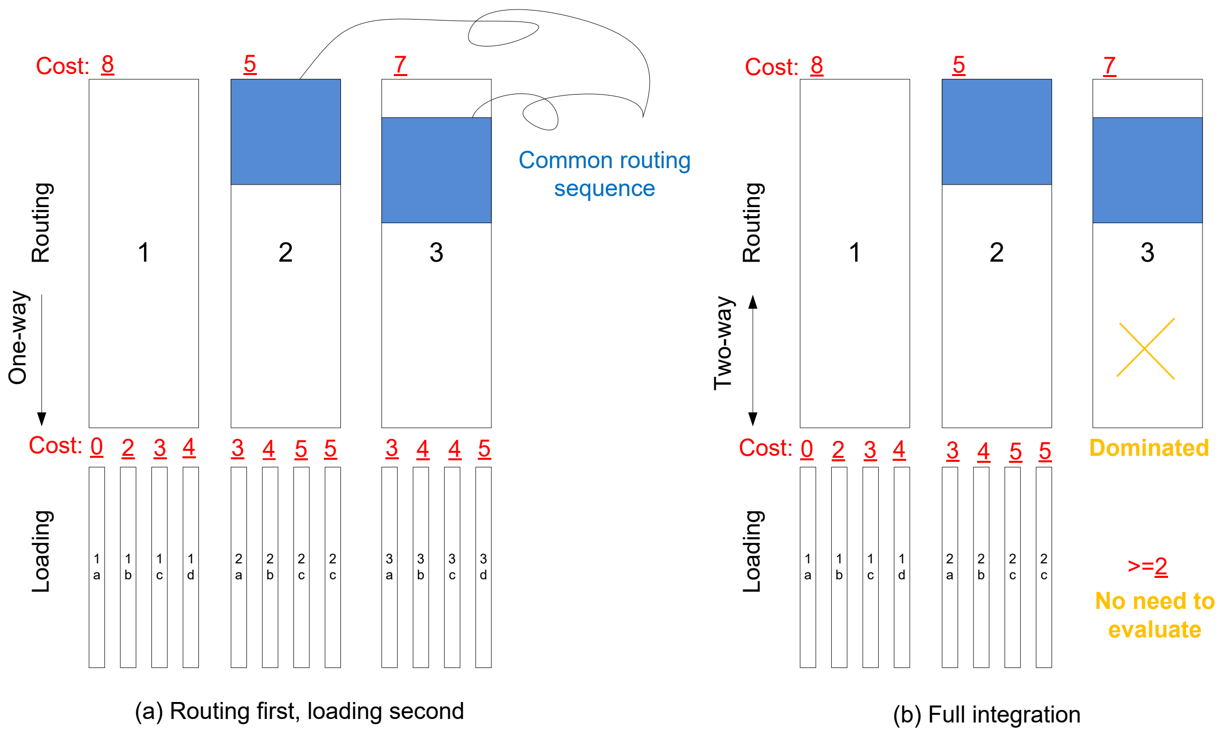


Figure 7: Advantage of the real-time interaction between routing and loading

Finally, we believe that math programming based approaches are ideal if we aim for robust, high quality solutions. One can investigate whether the column generation framework used by Wang et al. (2018) can be extended to solve fully integrated problems. The complexity under this methodology is mainly at the subproblem level. For a fully integrated problem, the subproblem will be much more challenging.

To benchmark the fully integrated optimization algorithms to be developed in the future, additional efforts are also needed to develop standard test instances that are not yet available in the existing literature.

9.3. Realistic routing constraints and other considerations

9.3.1 Time window constraint

Commonly encountered constraints in vehicle routing problems, such as time window constraints, have not been fully incorporated in the auto-carrier routing studies. Although it is widely known that the automobile shipments have their due dates, very few studies have tried to consider the time window requirement explicitly. As reviewed earlier, Tadei et al. (2002) considered a criticality index and Wang et al. (2018)

considered an urgency level as proxies. To make the relevant models more realistic, an explicit consideration of delivery time windows is necessary.

9.3.2 DOT rules

Another important practical issue to consider is that drivers of auto-carriers must follow Department of Transportation (DOT) rules, which specify working hour rules for the safety of long-haul drivers. To our knowledge, none of the existing studies in the automobile shipping optimization literature have incorporated DOT rules into their models. Obviously, consideration of DOT rules will make routing decisions much more challenging. In the vehicle routing literature, some studies (e.g. Xu et al. 2003, Rancourt et al. 2013, Alcaraz et al. 2019) have taken driver working hour rules into routing models. It would be interesting to see if some algorithmic ideas in such existing studies could be further incorporated.

9.3.3 Customized shipping option

The final set of practical issues that we think deserve research is related to individual automobile shipping options. There are two options for individual automobile shipping, namely terminal-to-terminal and door-to-door. In the case of terminal-to-terminal, the owner of a vehicle (an individual shipper) first drops off the vehicle at a terminal (such as a rail terminal or port) and later picks it up from a different terminal close to the shipper's destination. On the contrary, the door-to-door option saves the drop-off and pickup hassles for the shipper. The door-to-door option is thus more convenient and less time consuming for shippers. From the perspective of auto shipping, terminal-to-terminal shipping enables a carrier to achieve economies of scale by grouping and hauling multiple automobiles with the same origin and destination terminals. Therefore, terminal-to-terminal shipping is more affordable than door-to-door. For a broker that offers both terminal-to-terminal and door-to-door auto delivery services, there are two new routing issues to be considered. First, the broker needs to decide whether to treat terminal-to-terminal orders separately from door-to-door orders or consider them together by possibly mixing these two different types of orders in a shipment. The corresponding routing and loading optimization obviously will be very different. Second, if some auto shippers are flexible and accept both delivery options, which are associated with different shipping costs, then an additional decision that the broker needs to make is to determine which shipping option to deliver the car of each such auto shipper. Understandably, routing decisions are different under the two shipping options discussed here. As of now, there is no research that has addressed the issues discussed here.

9.4. Intermodal optimization models

At the strategic decision-making level, the intermodal automobile shipping has been studied, while there are no operational optimization studies that have considered the intermodal automobile shipping. Thus, there is a research need in this area. At the operational level, special attention is needed to model the transfer operations between different modes of transportation at intermodal terminals. For instance, at a rail-to-truck transfer terminal, autorack arrivals should be coordinated with auto-carrier departures to minimize the transfer time of automobiles at the terminal for timely delivery. In addition, considering that train arrivals may be subject to delays, real-time train holding strategies (Sun and Schonfeld 2016) may be developed to minimize the impact of random delays on the transfer coordination at terminals.

9.5. Stochastic optimization models

Strategic automobile distribution decisions are very infrequently made, which have a major impact on the automobile distribution in the next few decades. Such decisions must be carefully evaluated considering a wide range of future scenarios, instead of a single averages-based scenario. In the existing literature, very few strategic studies have considered uncertainties in automobile distribution. Clearly, the automobile shipping demand could be uncertain; the shipping cost parameters may also fluctuate as technologies (such as vehicle automation) evolve over time. Therefore, another promising research direction is to develop stochastic optimization models, which can be solved by approaches such as the sample average approximation technique (Kiya and Davoudpour 2012), among others.

9.6. Impact of natural and man-made disasters

Natural and man-made disasters can both create significant disruptions to supply chains and logistics systems and cause sudden shifts in demand patterns. Automobile distribution is no exception. In particular, as indicated by many experts (Ivanov 2020, Choi 2020, Choi 2021), the ongoing COVID-19 public health crisis has reshaped the norm of logistics operations. On one hand, the sales of new vehicles dropped in 2020 because car buyers were more cautious about getting a new car due to the uncertain economic condition. On the other hand, the demand for used cars increased because many people who relied on public transit for commuting decided to buy preowned cars due to concerns with virus transmission inside public transit vehicles (Rosenbaum 2020). The increased online trading of preowned cars have increased automobile shipping demand. Given such new changes, a promising direction for future research is to build a robust automobile distribution network that could withstand various changes ranging from changing consumer preferences to technologies innovations.

Another area that we think needs more research is disruption management in automobile distribution. When a part of the distribution network experiences a reduced capacity or is closed off completely due to emergencies caused by disasters, already dispatched auto-carriers must be re-routed quickly, and new routing and loading plans may have to be created for the auto-carriers soon to be dispatched. Similar actions may have to be taken in other situations such as auto-carrier breakdowns and last-minute change of time windows due to emergencies. There is a well-developed body of literature on vehicle routing problems with disruptions (e.g., Li et al. 2009, Ju et al. 2017, Eglese and Zambirinis 2018). It will be interesting to see if some of the existing algorithms from this literature can be extended to tackle automobile shipping problems under various disruption scenarios.

9.7. Opportunities due to emerging technologies

9.7.1 Impact of vehicle automation

With various policy and funding support from the US government (USDOT 2020), cities across the US have started hosting autonomous vehicle pilots, which implies autonomous vehicles (AVs) are likely to become a reality in the foreseeable future. When the self-driving technology is market ready, the current automobile distribution scheme will be disrupted. For instance, at present, driving new vehicles directly and individually from a warehouse to dealer locations is economically infeasible, considering the high cost associated with delivery drivers. Nonetheless, when drivers are no longer needed for operating a vehicle, it may become feasible to eliminate the expensive auto-carrier operations by dispatching each AV individually to the corresponding dealer location. More importantly, when auto-carriers can also self-drive, a new vehicle distribution scheme can bring significant benefits through the routing cost tradeoff between auto-carriers and finished automobiles. This distribution requires the collaboration of larger autonomous vehicles (i.e., auto-carriers) and smaller autonomous vehicles (i.e., finished automobiles). Although both auto-carriers and finished automobiles can operate autonomously, their routing costs clearly differ due to their various sizes, weights, and vehicle technologies. Therefore, the collaboration between auto-carriers and finished vehicles presents a promising solution, as it strikes a balance between routing auto-carriers exclusively and operating finished autonomous automobiles only. New optimization models and algorithms will be needed to support this novel distribution scheme.

9.7.2 Impact of artificial intelligence and machine learning

Major breakthroughs made in artificial intelligence can be used to advance the automobile distribution practice by marrying machine learning (ML) to integer programming (IP). There is a rapidly growing interest in using ML to help solve combinatorial optimization problems (Bengio et al. 2021). ML is valuable because it can be used to make accurate predictions on some key values in real time for given inputs when sufficient training data are presented. For instance, in solving a vehicle routing problem, a vehicle route can be provided as an input to an ML algorithm while an approximate routing cost is a possible output from this algorithm. When a ML algorithm is extensively trained under various problem settings, the learning algorithm could identify hidden relations between certain problem characteristics and solution patterns. Therefore, ML can complement IP by guiding the search in the solution space by, for example, avoiding some computationally expensive subproblems, or by deciding whether or not a specific solution approach should be used (Kruber et al. 2017). Clearly, there is a great need to build the methodological foundation for the integration of ML and optimization algorithms for the automobile distribution problem and develop a set of new ML enhanced algorithms for solving this problem.

9.7.3 Impact of blockchain

A blockchain is an expanding chain of linked data packages or blocks, each of which has a cryptographic hash of the previous block, a random number for validating the hash, a timestamp, and transaction data (Nofer et al. 2017). This design ensures the integrity of the entire chain. Representing a new way of organizing data, blockchain has attracted an increasing number of supply chain researchers to explore its impact on supply chain operations (Dutta 2020). As automobile distribution involves multiple parties, who frequently exchange their data, such as orders, contracts, and payments, blockchain has the potential to innovate the way automobile distribution related data are collected, stored, and managed. This will also have a major implication for real-time decision-making in automobile distribution, such as dynamic pricing of automobile shipments, and dynamic routing of auto-carriers.

10. CONCLUDING REMARKS

This paper contributes to the literature by presenting the first systematic review of the automobile distribution and related optimization problems. This survey covers three major modes of transportation employed for shipping automobiles (i.e., highway, railway, and waterway), and involves optimization decisions at operational, tactical, and strategic levels. Reflecting the reality of the existing literature, the bulk of the survey is on highway-based auto-carrier routing optimization with various levels of auto-carrier loading and unloading complexities.

Even though the automobile shipping optimization literature has rapidly evolved in the last few decades and significant advancements have been made in developing efficient and effective optimization algorithms, we conclude that several major gaps remain between the practical needs in the industry and solution capabilities of the existing methods. First, hybrid service models with additional real-world elements of automobile distribution should be considered. Second, the full integration of loading and routing optimization is a promising research direction. Third, the impacts of uncertainties, disruptions, and emerging technologies on automobile distribution should also be well incorporated into relevant optimization studies.

By critiquing the automobile distribution optimization literature and proposing potential strategies to address identified research gaps, we expect this survey to stimulate transportation researchers to develop advanced optimization models to keep pace with the rapid development of the automobile distribution practice.

REFERENCES

Agbegha, G. Y. (1992). An optimization approach to the auto-carrier problem, Doctoral dissertation, Case Western Reserve University, Cleveland, OH.

Agbegha, G. Y., Ballou, R. H., & Mathur, K. (1998). Optimizing auto-carrier loading. *Transportation Science*, 32(2), 174-188.

Alcaraz, J. J., Caballero-Arnaldos, L., & Vales-Alonso, J. (2019). Rich vehicle routing problem with last-mile outsourcing decisions. *Transportation Research Part E: Logistics and Transportation Review*, 129, 263-286.

Andersson, H., Fagerholt, K., & Hobbesland, K. (2015). Integrated maritime fleet deployment and speed optimization: Case study from RoRo shipping. *Computers & Operations Research*, 55, 233-240.

Billing, C., Jaehn, F., & Wensing, T. (2018). A multiperiod auto-carrier transportation problem with probabilistic future demands. *Journal of Business Economics*, 88(7-8), 1009-1028.

Bengio, Y., A. Lodi, & A. Prouvost. (2021). Machine learning for combinatorial optimization: a methodological tour d'horizon. *European Journal of Operational Research*, 290 (2), 405-421.

Bonassa, A. C., Cunha, C. B., & Isler, C. A. (2019). An exact formulation for the multi-period auto-carrier loading and transportation problem in Brazil. *Computers & Industrial Engineering*, 129, 144-155.

Cachon, G. P., & Olivares, M. (2010). Drivers of finished-goods inventory in the US automobile industry. *Management Science*, 56(1), 202-216.

Cachon, G. P., Gallino, S., & Olivares, M. (2019). Does adding inventory increase sales? Evidence of a scarcity effect in US automobile dealerships. *Management Science*, 65(4), 1469-1485.

Chandra, S., Christiansen, M., & Fagerholt, K. (2016a). Combined fleet deployment and inventory management in roll-on/roll-off shipping. *Transportation Research Part E: Logistics and Transportation Review*, 92, 43-55.

Chandra, S., Christiansen, M., & Fagerholt, K. (2020). Analysing the modal shift from road-based to coastal shipping-based distribution–a case study of outbound automotive logistics in India. *Maritime Policy & Management*, 47(2), 273-286.

Chandra, S., Ghosh, D., & Srivastava, S. K. (2016b). Outbound logistics management practices in the automotive industry: An emerging economy perspective. *Decision*, 43(2), 145-165.

Chen, H. K. (2016a). A heuristic algorithm for the auto-carrier loading problem. *International Journal of Shipping and Transport Logistics*, 8(1), 21-50.

Chen, H. K. (2016b). Sequential auto carrier loading problem. *Transportation Research Record*, 2548(1), 53-61.

Chen, F., & Wang, Y. (2020). Downward compatible loading optimization with inter-set cost in automobile outbound logistics. *European Journal of Operational Research*, 287(1), 106-118.

Choi, T. M. (2020). Innovative "bring-service-near-your-home" operations under Corona-virus (COVID-19/SARS-CoV-2) outbreak: Can logistics become the Messiah?. *Transportation Research Part E: Logistics and Transportation Review*, 140, 101961.

Choi, T. M. (2021). Risk Analysis in Logistics Systems: A Research Agenda During and After the COVID-19 Pandemic. *Transportation Research Part E: Logistics and Transportation Review*, 145, 102190.

Christiansen, M., Fagerholt, K., Nygreen, B., & Ronen, D. (2013). Ship routing and scheduling in the new millennium. *European Journal of Operational Research*, 228(3), 467-483.

Clarke, G., & Wright, J. W. (1964). Scheduling of vehicles from a central depot to a number of delivery points. *Operations Research*, 12(4), 568-581.

Cordeau, J. F., Dell'Amico, M., Falavigna, S., & Iori, M. (2015). A rolling horizon algorithm for auto-carrier transportation. *Transportation Research Part B: Methodological*, 76, 68-80.

Dael, A. R. (2014). The carriage of automobiles in containers: an alternative method to address the excess capacity in the liner trade. *World Maritime University Dissertations*. 484. http://commons.wmu.se/all_dissertations/484.

Dejax, P. J., & Crainic, T. G. (1987). Survey paper—a review of empty flows and fleet management models in freight transportation. *Transportation Science*, 21(4), 227-248.

Dell'Amico, M., Falavigna, S., & Iori, M. (2015). Optimization of a real-world auto-carrier transportation problem. *Transportation Science*, 49(2), 402-419.

Dong, B., Christiansen, M., Fagerholt, K., & Chandra, S. (2020). Design of a sustainable maritime multi-modal distribution network–Case study from automotive logistics. Transportation Research Part E: Logistics and Transportation Review, 143, 102086.

Dror, M., & Trudeau, P. (1989). Savings by split delivery routing. *Transportation Science*, 23(2), 141-145.

Dutta, P., Choi, T. M., Somani, S., & Butala, R. (2020). Blockchain technology in supply chain operations: Applications, challenges and research opportunities. *Transportation Research Part E: Logistics and Transportation Review*, 142, 102067.

Eglese, R., & S. Zambirinis, (2018). Disruption management in vehicle routing and scheduling for road freight transport: A review. *Top*, 26(1), 1-17.

Eskigun, E., Uzsoy, R., Preckel, P. V., Beaujon, G., Krishnan, S., & Tew, J. D. (2005). Outbound supply chain network design with mode selection, lead times and capacitated vehicle distribution centers. *European Journal of Operational Research*, 165(1), 182-206.

Gorman, M. F. (2015). Empty railcar distribution. In Handbook of Operations Research Applications at Railroads (pp. 177-189). Springer, Boston, MA.

Hei, X., Meng, Q., Wang, S., & Mao, H. (2014). Optimal automobile distribution model in multimodal freight transportation networks. *Transportation Research Record*, 2410(1), 50-57.

Holweg, M., & Miemczyk, J. (2003). Delivering the '3-day car'—the strategic implications for automotive logistics operations. *Journal of Purchasing and Supply Management*, 9(2), 63-71.

Hu, Z. H., Zhao, Y., Tao, S., & Sheng, Z. H. (2015). Finished-vehicle transporter routing problem solved by loading pattern discovery. *Annals of Operations Research*, 234(1), 37-56.

Ivanov, D. (2020). Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. *Transportation Research Part E: Logistics and Transportation Review*, 136, 101922.

Jin, M., Eksioglu, S. D., Eksioglu, B., & Wang, H. (2010). Mode selection for automotive distribution with quantity discounts. *Networks and Spatial Economics*, 10(1), 1-13.

Ju, G. Zhou, T. Chen, (2017) Disruption management for vehicle routing problem with time-window changes, *International Journal of Shipping and Transport Logistics*, 9(1), 4-28.

Juárez Pérez, M. A., Pérez Loaiza, R. E., Flores, Q., Malaquias, P., Atriano Ponce, O., & Flores Peralta, C. (2019). A Heuristic Algorithm for the Routing and Scheduling Problem with Time Windows: A Case Study of the Automotive Industry in Mexico. *Algorithms*, 12(5), 111.

Kang, M. H., Choi, H. R., Kim, H. S., & Park, B. J. (2012). Development of a maritime transportation planning support system for car carriers based on genetic algorithm. *Applied Intelligence*, 36(3), 585-604.

Katcoff, E. (2011). From plant to dealer: Improving route optimization for outbound vehicle distribution at an automobile manufacturer, Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Kiya, F., & Davoudpour, H. (2012). Stochastic programming approach to re-designing a warehouse network under uncertainty. *Transportation Research Part E: Logistics and Transportation Review*, 48(5), 919-936.

Kruber, M., M.E. Lübbecke, & A. Parmentier. 2017. Learning when to use a decomposition. In Integration of AI and OR Techniques in Constraint Programming, edited by D. Salvagnin and M. Lombardi, Vol. 10335, 202–210. Springer, International Publishing.

Li, J. Q., P.B. Mirchandani, & D. Borenstein, (2009). A vehicle rescheduling problem with real-time vehicle reassignments and trip cancellations. *Transportation Research Part E: Logistics and Transportation Review*, 45, 419-433.

Li, Z., & Zhang, Y. (2016, July). Research on loading and routing problem of finished car logistics. In 2016 International Conference on Logistics, Informatics and Service Sciences (LISS) (pp. 1-6). IEEE.

Lin, C.H. (2010). An exact solving approach to the auto-carrier loading problem. *Journal of Society for Transportation and Traffic Studies (JSTS)*, Vol. 1, No. 1, 93-107.

Liu, J., Smith, A. E., & Qian, D. (2016). The vehicle loading problem with a heterogeneous transport fleet. *Computers & Industrial Engineering*, 97, 137-145.

Lutter, P., & Werners, B. (2015). Order acceptance for motorail transportation with uncertain parameters. *OR Spectrum*, 37(2), 431-456.

Lutter, P. (2016). Optimized load planning for motorail transportation. *Computers & Operations Research*, 68, 63-74.

Mattfeld, D. C. (2006). The management of transshipment terminals: decision support for terminal operations in finished vehicle supply chains (Vol. 34). Springer Science & Business Media.

Mattfeld, D. C., & Kopfer, H. (2003). Terminal operations management in vehicle transshipment. *Transportation Research Part A: Policy and Practice*, 37(5), 435-452.

Meng, Q., Hei, X., Wang, S., & Mao, H. (2015). Carrying capacity procurement of rail and shipping services for automobile delivery with uncertain demand. *Transportation Research Part E: Logistics and Transportation Review*, 82, 38-54.

Miller, B. M. (2003). Auto carrier transporter loading and unloading improvement. Master's thesis, Air Force Institute of Technology, Wright-Patterson AFB, OH.

Miller, T., Wise, D., & Clair, L. (1996). Transport network design and mode choice modeling for automobile distribution: a case study. *Location Science*, 4(1-2), 37-48.

Nieuwenhuis, P., Beresford, A., & Choi, K. Y. (2007). Shipping air?; Tracking and forecasting the shipments of new and used cars. *Journal of Maritime Research*, 4(3), 17-36.

Nieuwenhuis, P., & Wells, P. (Eds.). (2015). The global automotive industry. John Wiley & Sons. ISBN: 9781118802397, https://onlinelibrary.wiley.com/doi/book/10.1002/9781118802366

Nofer, M., Gomber, P., Hinz, O., & Schiereck, D. (2017). Blockchain. *Business & Information Systems Engineering*, 59(3), 183-187.

OICA, 2020a, International Organization of Motor Vehicle Manufacturers, http://www.oica.net/category/sales-statistics/, Accessed on October 1st, 2020

OICA, 2020b, International Organization of Motor Vehicle Manufacturers, http://www.oica.net/category/production-statistics/2019-statistics/, Accessed on October 1st, 2020

Øvstebø, B. O., Hvattum, L. M., & Fagerholt, K. (2011a). Optimization of stowage plans for RoRo ships. *Computers & Operations Research*, 38(10), 1425-1434.

Øvstebø, B. O., Hvattum, L. M., & Fagerholt, K. (2011b). Routing and scheduling of RoRo ships with stowage constraints. *Transportation Research Part C: Emerging Technologies*, 19(6), 1225-1242.

Rancourt, M.-E., J.-F. Cordeau, G. Laporte. (2013), Long-haul vehicle routing and scheduling with working hour rules. *Transportation Science*, 47(1), 81 – 107.

Rosenbaum, R. (2020). The used car boom is one of the hottest, and trickiest, coronavirus markets for consumers. https://www.cnbc.com/2020/10/15/used-car-boom-is-one-of-hottest-coronavirus-markets-for-consumers.html, Accessed on January 30th, 2021

Saberi, B. (2018). The role of the automobile industry in the economy of developed countries. *International Robotics & Automation Journal*, 4(3), 179-180.

Sherali, H. D., & Tuncbilek, C. H. (1997). Static and dynamic time-space strategic models and algorithms for multilevel rail-car fleet management. *Management Science*, 43(2), 235-250.

Sherali, H. D., & Suharko, A. B. (1998). A tactical decision support system for empty railcar management. *Transportation Science*, 32(4), 306-329.

Sherali, H. D., & Maguire, L. W. (2000). Determining rail fleet sizes for shipping automobiles. *Interfaces*, 30(6), 80-90.

Sherali, H. D., & Lunday, B. J. (2011). Equitable apportionment of railcars within a pooling agreement for shipping automobiles. *Transportation Research Part E: Logistics and Transportation Review*, 47(2), 263-283.

Spieckermann, S., & Voß, S. (1995). A case study in empty railcar distribution. *European Journal of Operational Research*, 87(3), 586-598.

Sun, Y., & Schonfeld, P. (2016). Holding decisions for correlated vehicle arrivals at intermodal freight transfer terminals. *Transportation Research Part B: Methodological*, 90, 218-240.

Tadei, R., Perboli, G., & Della Croce, F. (2002). A heuristic algorithm for the auto-carrier transportation problem. *Transportation Science*, 36(1), 55-62.

U.S. Bureau of Labor Statistics (2021), Automotive Industry: Employment, Earnings, and Hours, https://www.bls.gov/iag/tgs/iagauto.htm, Accessed on January 30th, 2021

U.S Census Bureau (2018). State-to-State migration flows. https://www.census.gov/data/tables/time-series/demo/geographic-mobility/state-to-state-migration.html, Accessed on October 1st, 2020

U.S. Department of Transportation (2004). Federal size regulations for commercial motor vehicles https://ops.fhwa.dot.gov/freight/publications/size_regs_final_rpt/ Accessed on October 1st, 2020

US Department of Transportation (USDOT) (2020). USDOT automated vehicles activities. https://www.transportation.gov/AV, Accessed on January 30th, 2021

Venkatachalam, S., & Sundar, K. (2016). Branch-and-price algorithm for an auto-carrier transportation problem, unpublished preprint, https://arxiv.org/abs/1605.09030

Volling, T., Matzke, A., Grunewald, M., & Spengler, T. S. (2013). Planning of capacities and orders in build-to-order automobile production: A review. *European Journal of Operational Research*, 224(2), 240-260.

Xu, H., Z.-L. Chen, S. Rajagopal, S. Arunapuram. (2003). Solving a practical pickup and delivery problem. *Transportation Science*, 37(3), 347 – 364.

Xuan, X. (2014). Evaluation of shipping finished automotive in multimodal containers: a marketing plan for shipping company. World Maritime University Dissertations. 470. http://commons.wmu.se/all_dissertations/470

Wang, Y., & Yeo, G. T. (2016). A Study on international multimodal transport networks from Korea to Central Asia: Focus on secondhand vehicles. *The Asian Journal of Shipping and logistics*, 32(1), 41-47.

Wang, Y., Chen, F., & Chen, Z. L. (2018). Pickup and delivery of automobiles from warehouses to dealers. *Transportation Research Part B: Methodological*, 117, 412-430.

Wensing, T. (2018). Social and environmental impact of advances in economically driven transport optimization: Case study in automobile distribution. In Social and Environmental Dimensions of Organizations and Supply Chains (pp. 195-210). Springer, Cham.

APPENDIX

Table 7 provides a quick comparison of 35 selected optimization studies that have been carefully reviewed in this paper due to their high relevance. The comparison is by the type of decisions involved (strategic or tactical (abbreviated as ST), operational without routing (O1), and operational with routing (O2)), and the transportation mode used for automobile distribution (highway (H), rail (R), water (W), intermodal (I)). See Sections 5 to 8 for detailed reviews of these studies.

Table 7: Overview of 35 selected studies that have been reviewed in detail

Authors & Year	Type	Mode
Agbegha et al. (1998)	O1	H
Billing et al. (2018)	O2	H
Bonassa et al. (2019)	O2	H
Chandra et al. (2020)	ST	I
Chen (2016a)	O1	H
Chen (2016b)	O1	H
Wang and Chen (2020)	O2	H
Cordeau et al. (2015)	O2	H
Dael (2014)	O1	W
Dell'Amico et al. (2015)	O2	H
Dong et al. (2020)	ST	I
Eskigun et al. (2005)	ST	I
Hei et al. (2014)	ST	I
Hu et al. (2015)	O2	H
Jin et al. (2010)	ST	I
Juárez Pérez et al. (2019)	O2	H
Kang et al. (2012)	O2	W
Li and Zhang (2016)	O2	H
Lin (2010)	O1	H
Liu et al. (2016)	O2	H
Lutter (2016)	O1	R
Lutter and Werners (2015)	O1	R
Meng et al. (2015)	ST	I
Miller (2003)	O2	H
Miller et al. (1996)	ST	I
Øvstebø et al. (2011a)	O1	W

Øvstebø et al. (2011b)	O2	W
Sherali and Lunday (2011)	ST	R
Sherali and Suharko (1998)	ST	R
Sherali and Tuncbilek (1997)	ST	R
Tadei et al. (2002)	O2	H
Venkatachalam and Sundar (2016)	O2	H
Wang et al. (2016)	ST	I
Wang et al. (2018)	O2	H
Wensing (2018)	O2	H
Xuan (2014)	O1	W