CIRRELT

CIRRELT-2024-11

50 Years of Operations Research for Planning Consolidation-based Freight Transportation

Teodor Gabriel Crainic
Walter Rei

April 2024

Bureau de Montréal
Université de Montréal
C.P. 6128, succ. Centre-Ville
Montréal (Québec) H3C 3J7
Tél : 1-514-343-7575
Télécopie : 1-514-343-7121

Bureau de Québec
Université Laval,
2325, rue de la Terrasse
Pavillon Palasis-Prince, local 2415
Québec (Québec) G1V 0A6
Tél : 1-418-656-2073
Télécopie : 1-418-656-2624

50 Years of Operations Research for Planning Consolidation-based Freight Transportation

Teodor Gabriel Crainic*, Walter Rei

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) and School of Management, Université du Québec à Montréal

Abstract. Freight transportation is a vital sector for the global economy, essential for facilitating the trade upon which all societies depend and thrive. Carriers supply the transportation services that meet the demands expressed by shippers for the movement of goods between various locations within specific timelines. Consolidation is a widely spread strategy in transportation and logistics, which aims for increased operational and economic efficiency, by combining cargo of different shippers into the same vehicle for their complete or partial journeys. Consolidation-based carriers move a large and valuable part of the world trade over short, medium, long, and intercontinental distances. The focus of this paper is on consolidation-based freight transportation systems, and more specifically on the complex planning decision-making challenges that such systems entail. Operations Research has a rich history of fruitful interactions with decision-making in the planning and operation of freight transportation, leading to significant methodological developments. Over the past 50 years, research in this field has generated numerous Operations Research methodologies designed to tackle the complexities of solving difficult freight transportation problems and delivering high-quality solutions within the necessary decision-making time frames. We focus on the methodological developments addressing the tactical planning challenges of consolidation-based freight carriers, particularly the Service Network Design methodology that provides a general framework for optimizing carrier services and resource allocation. We structure our manuscript by exploring key dimensions in the inherent complexity of carrier-centric tactical planning and the O.R. innovations developed to address them. These dimensions include the explicit consideration of temporal aspects of system elements and dynamics; integrative decision-making, particularly in managing the resources carriers require to support their services and operations; and accounting for the uncertainty that directly impacts the planning of consolidation-based freight transportation systems. We should again emphasize that our objective is not to provide a comprehensive review of the literature. Instead, we offer insights into the main developments over the last 50 years and identify promising research avenues.

Keywords: Freight transportation, consolidation, tactical planning, service network design.

Acknowledgements. While working on the project, the first author held the UQAM Chair on Intelligent Logistics and Transportation Systems Planning, and was Adjunct Professor, Department of Computer Science and Operations Research, Université de Montréal. We gratefully acknowledge the financial support provided by the Natural Sciences and Engineering Council of Canada (NSERC), through its Discovery grant program. We also gratefully acknowledge the support of Fonds de recherche du Québec through their infrastructure grants.

---

* Corresponding author: teodorgabriel.crainic@cirrelt.net

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2024

1 Introduction

Freight transportation is a vital sector for the global economy, essential for facilitating the trade upon which all societies depend and thrive. It supports the diverse needs and necessities of people and organizations across the globe, operating within complex systems where various entities interact. On one hand, there are *carriers* who supply the transportation services that meet the demands expressed by *shippers*. These shippers represent a broad spectrum of organizational types, ranging from producers to distributors of goods and everything in between. They require the movement of goods between various locations within specific timelines, with carriers enabling this critical logistic function.

In characterizing carriers, a distinction is commonly made between the *dedicated* and the *consolidation-based* organizations. The former category involves situations where individual shipper requests are assigned to full loading units, with the carrier responsible for managing the entire journey of the load. This arrangement is typically suited for shippers who require exclusive use of a vehicle or container due to the volume or specific nature of the goods. Conversely, consolidation-based carriers facilitate the assignment of multiple shipper requests to the same loading units, effectively sharing capacity. This approach allows consolidated loads to be transported, with requests potentially sharing several loading units throughout their journey. Consolidation serves here as a general logistic strategy that benefits both shippers and carriers. Shippers, whose requests might not require a full loading unit or for whom it may not be cost-effective to pay for a dedicated transport service, can still benefit from favourable rates offered by carriers, who in turn can use their resources more efficiently (e.g., fill up their loading units with multiple shipper requests) to meet the overall demand. The focus of this paper is on consolidation-based freight transportation systems, and more specifically on the planning problems that such systems entail.

Given the wide array of organizations and stakeholders interacting within consolidation-based freight transportation systems, the planning and management of these systems present numerous complex decision-making challenges. Efficient navigation of these challenges necessitates a diverse array of tools and technologies to support decision-makers in their planning responsibilities. The field of *Operations Research* (*O.R.*) has a rich history of fruitful interactions with decision-making in the planning and operation of freight transportation, leading to significant methodological developments. Over the past 50 years, research in this field has generated numerous O.R. methodologies designed to tackle the complexities of solving difficult freight transportation problems and delivering high-quality solutions within the necessary decision-making time frames. A comprehensive review of all these developments would be quite challenging to accomplish within a single scientific paper. Instead, our aim in this manuscript is to provide a general overview of a particularly successful exchange between methodological developments in O.R. and an impactful decision-making process in freight transportation.

Our focus is on the supply side of consolidation-based freight transportation systems, governed by the activities of carriers. Specifically, carriers are tasked with planning the transportation and terminal resources and services necessary to meet shippers' demands. Among these decisions, tactical planning is crucial as it involves designing a set of scheduled services that align resource allocation with anticipated shipping needs over a medium to long-term horizon. Importantly, carrier tactical plans are established to accommodate demands of shippers that are repeatable and predictable, yet may still involve a degree of uncertainty. These tactical plans form the operational backbone for carriers, informing not only day-to-day operations but also strategic decisions, such as significant investments in infrastructure modifications, to enhance service efficiency and capacity. Consequently, we further focus our overview on the O.R. methodological developments made to solve the tactical planning problems pertinent to consolidation-based freight transportation systems. The primary methodology employed to address these challenges is Service Network Design (SND), which provides a general framework for optimizing carrier services and resource allocation.

We structure our manuscript by exploring key dimensions in the inherent complexity of carrier-centric tactical planning and the O.R. innovations developed to address them. These dimensions include the explicit consideration of temporal aspects of system elements and dynamics; integrative decision-making, particularly in managing the resources carriers require to support their services and operations; and accounting for the uncertainty that directly impacts the planning of consolidation-based freight transportation systems. We should again emphasize that our objective is not to provide a comprehensive review of the literature. Instead, we offer insights into the main developments over the last 50 years and identify promising research avenues.

The remainder of this paper is organized as follows. In Section 2, we review the key aspects that define the supply side of freight transportation systems. Additionally, we also provide a high-level classification of freight carriers and a comprehensive overview of the planning dynamics they employ, as well as the general O.R. methods utilized. Section 3 details the general SND models, presenting the basic formulations developed over the years and providing a historical perspective that highlights major inflection points in methodological developments. This section leads into a discussion of the main contributions made towards addressing some of the key dimensions in carrier-centric tactical planning. Section 4 specifically discusses the integration of time and related requirements such as managing delays and schedules within SND methods for freight transportation planning. Section 5 reviews how resource management has been explicitly considered in SND methods. Section 6 is dedicated to examining the main SND methods designed to address the effects of uncertainty on the tactical planning processes performed by carriers. Finally, we conclude in Section 7 with some perspectives on future research questions expected to shape the application of SND methods in the tactical planning of freight transportation systems.

2 The Supply Facet of Freight Transportation

As already mentioned, freight transportation is a complex activity field with numerous stakeholders with various interests and objectives. This complexity is not lower when the supply side is considered only. We therefore start with a high-level taxonomy of freight carriers, identifying the characteristics of consolidation and consolidation-based carriers and systems, which are within the scope of this paper (Section 2.1). Section 2.2 then recalls the basic definitions, system elements, and fundamental operations of consolidation carriers. Lastly, Section 2.3 synthesizes the main planning levels, decisions, and associated principal to O.R. methodologies for those same carriers.

2.1 Carriers & Systems

The taxonomy, illustrated in Figure 1, is far from exhaustive. We do not, for example, discuss in any detail (but mention it when required) the freight characteristics and requirements in terms of special care or equipment, e.g., dangerous goods, frozen or fresh produce, furniture, etc. Moreover, it is noteworthy that the vast majority of the O.R. developments related to freight transportation address issues associated to *for-hire* carriers, which offer transportation services to shippers. Consequently, we focus this article on this component of the field, and do not address explicitly the case of private carriers, i.e., of organizations owning and managing a fleet of vehicles for their own needs. It should be noted, however, that many developments proposed for the for-hire case may be applied to the private one as well.

Four dimensions are identified in Figure 1 addressing how customer demand is loaded and moved, the geographical scope of operations, the transportation mode, and the structure of the system and associated decision making.

The first dimension, *Load treatment*, indicates both the commercial shipper-carrier relation regarding how the carrier capacity is allocated to the shipper request, and the subsequent loading and movement activities. *Dedicated* refers to the case when a loading unit, e.g., a container, vehicle (truck or trailer for full-truckload motor carriers, vessel in maritime tramp shipping, etc.) or convoy (e.g., unit trains carrying minerals, cereals or petroleum products) is dedicated to a unique shipper demand, which pays for the complete journey. Once loaded, the corresponding shipment travels untouched until its final destination (even when the vehicle changes mode or carrier en route, e.g., a sealed container being transferred from ship to rail or truck).

Consolidation is the process of combining several shipments, of different shippers and potentially with different origins and destinations, for loading into the same vehicle or container for their complete or partial journeys. Consolidation is a widely spread strategy in transportation and logistics, born of two phenomena. On the shipper side, the volume or value of most shipments is too low to justify paying the tariffs associated with a direct, dedicated transport. On the carrier side, the same shipment characteristics do not permit

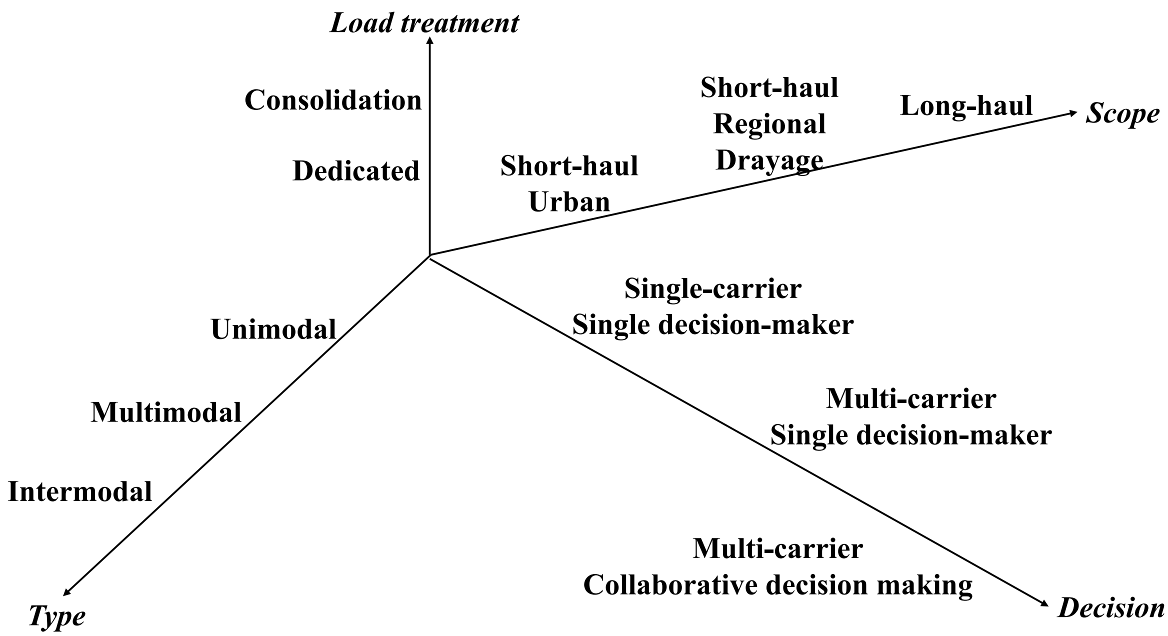

Figure 1: High-level classification of freight carriers

to offer profitable direct service with reasonable service quality (e.g., not waiting beyond customer willingness for vehicles to fill up with other demands).

Consolidation thus aims to reduce the unit shipment cost and the journey time, benefiting all parties involved. Railways, Less-than-Truckload (LTL) motor carriers, shipping companies moving containers on oceans, seas, rivers, and canals, postal services and express couriers are prime examples of consolidation-based carriers moving a large and valuable part of the world trade over short, medium, long, and intercontinental distances.

To achieve those goals, consolidation-based carriers are organized into so-called *hub-and-spoke networks*. The nodes of such a network are, for the most part, facilities handling freight and vehicles. Most facilities are of the *local/regional terminal* type, where most of the shipper demand from the surrounding regions is brought in to be transported by the carrier, and where the demand flows terminate their trips before being distributed to their final destinations. Rail stations, LTL regional terminals, most deep-sea and river/canal ports belong to this type. The *hubs* make up the second type. One finds in this category LTL breakbulks, major classification/blocking railroad yards, and major maritime ports for intermodal (container-based) traffic. While these terminals act as regional terminals for their hinterlands, their main role is to *consolidate* the flows in and out of their associated regional terminals for efficient long-haul transportation and economies of scale. It is noteworthy that, for same transportation modes, rail in particular, hubs are also handling the flows of individual power units and vehicles (railcars), the latter undergoing sorting (*classification*, in rail terms) and grouping (*blocking*) for efficient movement in *convoys* (trains). Road (hauling two or three trailers) and barge trains are also to be found. Several other vehicle- and power-unit-related activities are also performed in hubs, maintenance and repairs, in particular.

In a hub-and-spoke network, carriers thus first move low-volume loads available at a regional terminal to a hub, by so-called feeder services. At hubs, loads are sorted and consolidated into larger shipments, which are routed to other hubs by high-frequency, high-capacity services. Individual loads may go through more than one intermediary hub before reaching the final regional-terminal destination, being transferred from one service to another or undergoing re-classification and re-consolidation. It is noteworthy that, 1) high-frequency / capacity services may be run between a hub and a regional terminal or between two regional terminals when the level or value of demand justifies it; 2) more than one service, of possibly different modes, may be operated between the nodes of hub-and-spoke networks; a vehicle or convoy moving out of or towards a hub may visit more than one regional terminal on its route.

We focus on consolidation-based carriers in this article.

The major distinction along the *scope* dimension differentiates between *long-haul* and *short-haul* transportation. A second refinement categorizes short-haul operations, distinguishing between urban and periurban activities on one side, and regional activities on the other. The latter category includes drayage, which is a crucial practical issue in planning and managing the pickup, delivery, and transportation of both empty and loaded containers between ports and shipper locations within their immediate regions (e.g., Lai et al., 2013). The vast literature on vehicle routing addressing short-haul issues (see, e.g., Cordeau et al., 2007; Golden et al., 2008; Laporte, 2009; Toth and Vigo, 2002, 2014; Mor and Speranza, 2022) is beyond the scope of this article, which addresses long-haul transportation issues.

The *Type* dimension refers to the *mode* of transportation used by the carrier or multi-stakeholder system. It is worth mentioning that, "mode" is a general term, which may refer to different concepts/definitions according to the topic at hand. Thus, for example, a very general definition refers to fundamental "elements of nature", that is, land, water, air, and space transport. At the other end of the precision spectrum, a mode may refer to a particular combination of infrastructure, motorization, and even ownership, which may be part of a detailed analysis of a given transportation system, e.g., the combination of rail-track gauge (narrow, metric, or imperial), traction type (diesel or electric), and authority (state or country) used to model the rail transportation for evaluation and planning analyses either as a stand-alone system (e.g. Crainic et al., 1990b) or as part of a national strategic study (e.g., Crainic et al., 1990a; Guélat et al., 1990; Crainic and Florian, 2008).

The widely accepted, "classical" definition of modes is based on a high-level combination of elements such as nature, infrastructure, and vehicle/traction. One thus finds

- *Full-Truckload* (*TL*) and *Less-than-Truckload* (*LTL*) trucking on roads, particularly in inter-urban settings, with an increasingly larger array of road-based modes in cities, e.g., people-driven or autonomous electric or hydrogen (or mixed) powered vans, cargo bikes, and robots;

- Pipeline for moving liquids and gazes at various levels of viscosity are not a consolidation mode, hence, we do not discuss related issues in this article;
- *Rail*, sometimes more finely defined as *unit trains*, usually dedicated to bulk goods, e.g., grain, minerals, petroleum products, *general trains*, performing consolidation-based rail transportation, and *intermodal*, dedicated to moving containerized cargo (more on this "mode" in a moment);
- *Maritime navigation*, which may be further separated according to its geographical scope, deep-sea / ocean and coastal navigation, or cargo type, e.g., bulk, Roll-on / Roll-off (or RoRo, when cargo may be moved by its own means), Lift-on / Lift-off (LoLo when quay cranes do the loading and unloading work), and containerized cargo; Transportation performed by large ocean-going vessels on regular routes following regular schedules is identified as *liner shipping*;
- *River & canal navigation*, performed by barges, which may be further characterized by the type of cargo and operations, similar to maritime modes;
- *Air*, which until recently concerned air cargo only, with differentiation between dedicated transport and consolidation-based regular routes; one witnesses, however, a significant increase in drone-based transportation, mostly within cities for now, but not only, and the emergence of large non-rigid airships (aka, blimps) proposed currently for transporting heavy cargo into / out of difficult-to-access sites (e.g., following a natural or man-made disaster).

Unimodal, or single-mode transportation is then defined when a single mode is used from the origin to the destination of the cargo journey. Unimodal consolidation-based transportation is performed by LTL motor carriers (mostly between cities; a number of synthesis and review papers appeared over the years documenting the advancement of O.R.-related contributions, including Roy and Delorme, 1989; Bakir et al., 2021); railroads (e.g., Crainic, 1988; Cordeau et al., 1998; Crainic, 2009; Chouman and Crainic, 2021); liner-shipping maritime (e.g., Ronen, 1983; Christiansen et al., 2004, 2007, 2021) and river-barge companies (e.g., Bilegan et al., 2022); air-cargo (e.g., Feng et al., 2015; Srinivasan et al., 2023). Many journeys are unimodal and they are performed mostly using trucks. Most freight movements, however, involve two or more modes, freight being transferred from one mode to the next at *intermodal* terminals.

It is noteworthy that the difference between unimodal and multimodal transport very often depends upon the particular problem studied. Hence, for example, planning the long-haul operations of a LTL motor carrier clearly involves unimodal transport. In contrast, the system can be considered multimodal when planning simultaneously long-haul and feeder services, which bring cargo to, or distribute cargo from, a terminal in its immediate region. Similarly, from the point of view of the shipper, the journey is unimode when the same carrier picks it up at origin and delivers it at destination, independently of how the journey is performed (SteadieSeifi et al., 2014).

Intermodal transportation is generally defined as a chain of transportation services, most often of different modes but not necessarily so (e.g., transfer between two planes, trucks, or trains of different types or belonging to different companies), moving cargo packaged in such a way that it is not touched when transferred. A classical example of intercontinental intermodal transportation (Crainic and Kim, 2007) concerns loaded containers leaving a shipper's facility by truck to go either directly to port or to a rail yard from where a train delivers them to port. A ship then moves the containers from this initial port to a port on another continent, from where they are delivered to their final destinations by a single or a combination of "land" transportation means, road, rail, coastal or river navigation. Several intermodal terminals are part of this chain: the initial and final seaport container terminals, where containers are transferred between the liner navigation and land transportation modes, as well as in-land terminals (rail yards, river ports, etc.) providing transfer facilities between the land modes.

In its most general accepted definition, cargo is thus assumed to be packaged in containers, and many assimilate intermodal and container-based transportation. While this definition covers a good part of relevant transportation, a more general definition specifies that cargo has to be loaded into a "loading unit", which may be a container, but could also be a trailer, a swap body, and even boxes grouped on a pallet and tightly wrapped. Moreover, North American railroads introduced the term intermodal rail to designate their specific operations of handling and moving containers, often managed by particular administrative divisions of the railroad. A number of survey and synthesis papers address intermodal transport and planning, including Crainic and Kim (2007); Macharis and Bontekoning (2004); Bektaş and Crainic (2008).

Finally, the *decision* dimension reflects how the decision process is structured within the organization, or the group of organizations, supplying the services.

Most O.R. contributions in the last 50 years target the single-carrier case with its management team (or teams) making up the *single decision-maker* planing and managing the system to achieve the economic and service objectives of the organization. Increasingly, however, one witnesses studies addressing different organizational and decision-making settings, involving more than one stakeholder, more than one carrier, in particular.

Most of these cases may be described as (more or less) integrated *Many-to-One-to-Many* (*M1M*) systems (Taherkhani et al., 2022; Bruni et al., 2024). Briefly (following Taherkhani et al., 2022), an M1M system involves shippers on one side, making *shipper-demand requests* for cost and time-efficient transportation for their loads, and carriers on the other side, which make *carrier-capacity offers* for transportation and warehousing space, while requesting profitable loads. The "One" decision maker in the middle (also sometimes named *Intelligent Decision Support Platform - IDSP*) plans and optimizes operations and resource-utilization to profitably and simultaneously satisfy the needs of both shippers and carriers. Shipper-demand requests and the carrier-capacity offers are made available to the system at different time periods. Hence, the IDSP receives time-dependent requests from both stakeholders and optimizes in time and space the selection of shipper-demand requests, carrier-capacity offers, shipment-to-carrier assignments, and shipment itineraries through the consolidation of loads of different shippers into the same vehicles and synchronization of activities.

On such M1M setting concerns the so-called *intermediary*, e.g., a logistics-service provider, also known as a xParty Logistics (xPL) firm, where the value of x indicates the range of activities a shipper or carrier may outsource to the xPL, from 1 indicating a classical carrier, to 3 and 4 (level 5 is currently being discussed/proposed) standing for outsourcing increasingly larger aspects of a shipper's logistics planning activities (e.g., Aguezzoul, 2014; Dufour et al., 2018; Giusti et al., 2019; Premkumar et al., 2021; Chen et al., 2021, for 3-4PL system reviews). We classify xPL as multi-carrier, single decision-maker systems.

Several transportation systems emerged rather recently proposing innovative "new" organizational and decision-making concepts aimed to enhance long-haul and urban freight transportation, that is, reduce the externalities and nuisances, more generally, the environmental footprint associated to the transportation of freight, while sustaining the social and economic development of the organizations, cities, and regions concerned. The best known, studied, and deployed in one form or another among these are City Logistics (Crainic et al., 2021c, 2024; Marcucci et al., 2024), Physical Internet (Montreuil, 2011; Ballot et al., 2014; Crainic and Montreuil, 2016; Crainic et al., 2023), and Synchromodality (Ambra et al., 2019, 2021; Giusti et al., 2019). All these systems display multi-stakeholder M1M characteristics and are based on consolidation. We classify them as single-carrier, single decision-maker systems when either a single major player makes most important decisions (e.g., a major express courier firm allocating the customers and routes to small cargo-bike firms within specified urban areas) or an arm-length company is set up by the collaborating stakeholders to safely exchange information and operate the IDSP platform. Most studies in the literature address this class of problems.

More complex collaborative structures may exist, however, defining particular rules to share resources, activities, risks, and costs/benefits, and involving specific collaborative decision-making mechanisms accounting for those rules. Such cases impose additional complexity to the decision processes and the O.R. decision-support models and methods, making up the third category of the decision dimension of the taxonomy.

2.2 Consolidation

Carriers operate on infrastructure networks. The nodes of these networks are the terminals where freight and vehicles are handled. Terminals may belong to the carriers, e.g., most rail yards and LTL terminals belong to railroads and motor carriers, respectively, or not, the carrier paying for their utilization, e.g., most maritime ports and airports. Terminals are connected by physical, e.g., roads and rail tracks, or conceptual, e.g., maritime, river, and air links, "infrastructure" paths. Similarly to terminals, the infrastructure may be carrier-owned, e.g., most North American rail, public, as most road, water, and air networks, or be of the rent/pay-for-passage type either in open or restricted sharing, e.g., toll roads and most European rail networks, respectively.

A shipper demand for transportation involves requesting the movement of a specific quantity and type of freight between two terminals. We use the generic term "freight type" to recall that each demand is for a particular product, with specific physical and transportation characteristics and requirements, including weight, size, fragility, risk (e.g., dangerous goods), as well as handling and product-vehicle adequacy rules. Demand is also characterized by economic and service elements. The latter generally involves the time allocated to the delivery of freight to destination. The former takes the form of the fee, *tariff*, shippers pay for the transportation of their goods, which is conditioned by a combination of freight type, distance, service requirements, and commercial understanding (e.g., long-term contracts offering discounts on regularity and volume). Penalties, for late deliveries and damage, for example, may be part of the commercial deal. The term *multi-commodity demand* is generally used to represent this diversity in the literature (Crainic et al., 2021a) and this paper.

The carrier answers this demand by operating vehicles, possibly organized in convoys, e.g., rail, road, and barge trains, according to a set of services defined between the terminals of its network. It is noteworthy that "service" is a generic term, used to designate both the customer-service, that is, what is sold to the shipper in terms of the transportation of its freight between designated terminals and according to specified conditions and tariffs, and the transportation service, i.e., the way the consolidation-carrier operations are organized. A *service* in the latter sense, and as it is used in the literature and this article, is characterized by origin and destination terminals, a route in the physical network with, possibly, a set of intermediate stops to take and deliver freight, as well as operational and economic attributes, including type of vehicle and traction, speed, capacity, costs and frequency or schedule. Both material, e.g., vehicles and traction (power) units, and human, crews within terminals and operating vehicles and convoys, are required to support the operations of the planned services.

In most cases, the customer service does not correspond to a unique transportation service. Indeed, the demand volume, value, or both, are not sufficiently high to justify an economically and service-quality efficient direct and rapid shipment, neither for the carrier, nor for the shipper. Hence, carriers will move demand through so-called *hub-and-spoke* physical and service networks, which provide the means to take advantage of the economies of scale of consolidation-based transportation.

Two major types of terminals are encountered in such networks. *Local/regional terminals* make up the largest set. The demand from its corresponding region is gathered at a local terminal to be consolidated and then moved by the carrier. Symmetrically, goods with destinations within the region are brought to the local terminal to be separated and then delivered to their consignees. Rail stations, LTL regional terminals, most deep-sea and river/canal ports, the larger part of postal and express-courier service centers belong to this type. The *hubs* make up the second category and form the core of consolidation-based transportation through the consolidation/de-consolidation of the flows in and out of their associated regional terminals for efficient long-haul transportation and economies of scale. LTL breakbulks, major classification/blocking railroad yards, major maritime ports for intermodal container-based traffic, and postal / express-courier sorting centers belong to this category.

Carriers thus first move low-volume loads available at a regional terminal to a hub, through what is known as *feeder* services. At hubs, loads are sorted (*classified* is the term used in a number of settings, freight railroads in particular) and consolidated into larger flows, which are routed to other hubs by high-frequency, high-capacity services (such services may be operated between a hub and a regional terminal or between two regional terminals when the demand level or value justifies it). Loads may go through more than one intermediary hub before reaching the regional-terminal destination, being transferred from one service to another or undergoing reclassification and re-consolidation. Notice that, more than one service, of possibly different modes, may be operated between consolidation and regional terminals.

It is noteworthy that railroads generally implement a more complex double consolidation policy, as cars are grouped into blocks, which are then grouped into trains. Thus, loaded and empty cars, with possibly different origins and destinations, at their origin terminal are classified and grouped into particular blocks. The *block* is then moved as a *single unit* by a series of trains, from its origin terminal to its destination terminal, where it is taken apart, its cars being either delivered to their final consignees or reclassified for inclusion into new blocks for the next part of their trips. Hence, train services are made up of blocks and, when appropriate, it is blocks that are picked up and delivered at intermediate stops, where they are *transferred* (*switched*) from one train to another. This classification and blocking operation contributes significantly to the economy-of-scale provided by rail transportation.

Carriers operating consolidation-based systems must plan those services to profitably satisfy their shipper customers. Defining the related issues, goals, and challenges, together with the O.R. methods developed to address them is the matter of the following sections.

2.3 Planning and Operations Research Methods

Consolidation-based freight transportation involves complex systems, operations, shipper-carrier relations, and decision-making, resulting in carriers engaging into a rather extensive set of planning activities prior to executing the resulting plans. One may thus describe *planning* as the preparation of operations in anticipation of a future situation, whereas *execution* is the acting, including updating/adjusting, of the pre-established plans. Planning activities involve an evaluation/prediction/evaluation of the future status of the system and its environment at the moment, and for the duration of, the corresponding planning horizon. With respect to the planning of the supply side we address, these predictions concern the demand, the status and behaviour of the various system (e.g., infrastructure, equipment, manpower) and environment (e.g., political, social, economic) components. These predictions also yield the input of the O.R. models and methods arrayed to support the carrier's planning and execution decisions.

The quality of the plans and the system's performance then depends on the prediction accuracy, the performance of the O.R. models and methods, and how the uncertainty inherent to predictions is accounted for. *Deterministic* formulations are developed when the degree of confidence regarding the prediction accuracy is deemed adequate for the contemplated planning horizon and decision-making process. The largest part of the literature concerns this case and it is summarized in Sections 3 - 5. *Stochastic* is the generic term when predicted uncertainty and its consequences are explicitly accounted for. Section 6 is dedicated to this case.

The planning activities consolidation-based carriers undertake may be broadly classified into three levels, complemented by the execution level. Figure 2 schematically displays these levels, together with their respective main planning horizon, scope, interconnections, and associated core O.R. methodologies.

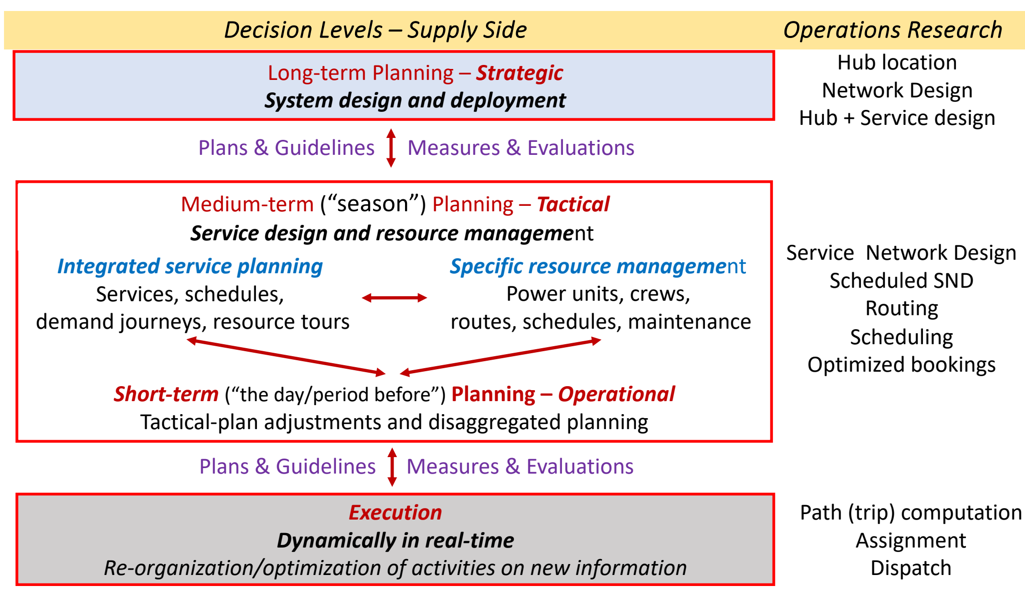

Figure 2: Decision & planning levels and related O.R. methodology

Strategic planning addresses long-term decisions on market deployment, system design, operation strategies, and acquisition of major resources. It concerns long planning horizons and involves rather high-level management. Discussing strategic-planning issues and O.R. methods is beyond the scope of this paper. Two observations only. First, *hub location* appears as core methodological instrument for the strategic design of the hub-and-spoke network structure characterizing consolidation-based carriers (see, e.g., O'Kelly, 1986; Aykin, 1995a,b; Cambell, 1994b; O'Kelly and Miller, 1994; Campbell et al., 2002; Alumur and Kara, 2008; Farahani et al., 2013; Alumur et al., 2021; Contreras, 2021). Second, tactical-planning models and methods (Sections 3 - 5) may be used as policy and performance-evaluation tools for strategic scenarios, either optimizing with appropriate abstraction/aggregation of operational details, demand, and costs, or simulating the scenario with appropriate approximation of carrier and shipper characteristics. Moreover, generalized service network design models may be built to address strategic-level issues such as determining the number, locations, and characteristics of terminals to build, use, or rent, the construction/enhancement of infrastructure, the type and dimension of the vehicle fleets, etc.

Medium to short-term planning, also identified as *tactical* and *operational*, respectively, has received the largest share of attention, producing the largest number of O.R. contributions. Tactical planning is performed by what is usually identified as medium-high level of management, for a medium-term planning horizon, also called *season* in the literature, which may extend from a few weeks (e.g., LTL motor carriers) to six or more (but twelve at maximum) months (e.g., railroads and liner shipping).

Tactical planning aims to build a transportation plan and schedule, for services and the resources required to execute them, to mitigate the possible drawbacks of consolidation (e.g., increased delays and costs due to terminal activities), satisfy the regular (repetitive) customer demand and service-quality requirements, and operate profitably and efficiently. Two major types of problem/decision-making settings are identified in Figure 2, one focusing on the more or less integrated planning of the service network, the second addressing the management of particular resources such as, power units, vehicles, and crews.

Service planning decides the selection and scheduling of services, the transfer and consolidation activities in terminals (as well as the convoy makeup and dismantling when relevant), the assignment and high-level management of resources supporting the selected services, and the demand itineraries to move the corresponding freight through the resulting service network. Hence, each *itinerary* is defined by the sequence of services used and the operations to be performed (e.g., transfer or re-classification and consolidation) at intermediary terminals. The goal is cost-efficient operation, together with timely and reliable delivery of demand according to customer specifications and the service-quality targets of the carrier. These are difficult problems, service network design being the methodology of choice in most cases (Sections 3 - 6).

The length of the planning horizon, the season, depends on the climate conditions (winter in the northern hemisphere may significantly impact transportation) and the intensity, type, and geographical distribution of the regular demand the carrier aims to satisfy. The *regular demand* generally corresponds to shippers that are strongly believed (due to, e.g., actual contracts, relations with trusted customers, and market knowledge by sales and customer-relation personnel) to bring business on a regular, that is, *repetitive*, basis for the coming season. In terms of total volume, regular demand is expected to make up a good part, e.g., 75% - 80%, of the peak demand to be serviced on a "normal" operating day. In terms of consistency, demand, and, hence, service, is expected to be repetitive according to a certain pattern, e.g., every day or every week. The plan is thus built for a given time duration, called *schedule length*, and is to be applied repetitively for the duration of the season.

Services need resources, human and material, and these are generally expensive and require particular utilization conditions. Work rules and schedules concern the human resources, while maintenance, inspection, routing and scheduling are relevant issues for material assets such as power units and vehicles. Traditionally, research focused on issues related to the management of specific resources, given a tactical plan and service network, and this class of topics is still of interest. Increasingly, however, resource management is integrated to the overall tactical-planning process. We discuss these issues in Section 5.

Operational planning might be performed by the same managers that built the tactical plans, but it is generally within the scope of decision-makers "close" to operations. The goal is to implement the plan given the observed conditions, in terms, e.g., of demand (the expected volume or not at the planned time or not) and service operations (mechanical issues, delays, etc.). The adjustment may be performed at iteratively and repeatedly various intervals, every week and day, for example. Although treated separately in most surveys and articles, we included tactical and operational planning decisions within the same box in Figure 2 to emphasize that the same methodology may be often used for both, albeit with different degrees of aggregation and time frames (e.g. Crainic et al., 2009, 2021c, for planning City Logistic systems).

The last class of decision-making structure addresses the (more or less) real-time management of activities and execution of operations, hopefully following the plan, in a dynamic environment. While O.R. is involved at this level as well (see, e.g., Crainic et al., 2024, in the context of City Logistics), the review of such issues and O.R. contributions is beyond the scope of this paper.

3 Service Network Design

Service Network Design (*SND*) is the core O.R. methodology supporting the planning of consolidation-based freight transportation carriers and systems. Used principally within tactical planning, i.e., yielding transportation plans covering regular activities for medium-to-short planning horizons, it may also be part of the scenario-evaluation phase of strategic planning for those organizations.

SND is part of the larger and important combinatorial-optimization family of *Network Design* (*ND*) problems and O.R. methodology (Magnanti and Wong, 1984; Balakrishnan et al., 1997). One my say that SND is ND applied to transportation planning (Crainic et al., 2021b). One may also claim that SND and ND are part of a virtuous circle of methodological innovation and enhancement targeting scientific development and performance-enhancing applications. Indeed, one observes that, many of the methodological developments contributed to network design have been inspired by transportation planning and, more particularly, SND-problem characteristics and challenges. One also observes that, fundamentally any SND formulation may be desribed as a ND model by suitably (re-)defining the service network. Hence, the many important results obtained in the last 50 years in understanding and addressing network-design problems may be mutatis mutandis applied to SND, as illustrated in this paper.

This section is dedicated to the presentation of the fundamental SND problem setting and formulations, which have been, and continue to be the topic of a large part of the contributions to literature and practice. This general formulation is also the methodological basis on which more refined models, addressing more complex applications, are built, as discussed in Sections 4 - 6.

We initiate the presentation with the basic notation and formulations, which provides the means to discuss the intimate relationships between SND and ND problems and models (Section 3.1). Similarly to all the other sections that follow, we complete the section with our perspective of the historical development of this field (Section 3.2).

3.1 Basic Notation and General SND Formulation

The basic notation underlying all the formulations discussed in the paper is presented first. We then proceed with the fundamental SND formulation, followed by some important variants.

3.1.1 Notation

Let $\mathcal{G}^{\text{PH}} = (\mathcal{N}^{\text{PH}}, \mathcal{A}^{\text{PH}})$ represent the physical network on which the carrier operates, with the nodes $\mathcal{N}^{\text{PH}}$ standing for the terminals where freight, services, and resources are handled, while the arcs of the set $\mathcal{A}^{\text{PH}}$ represent the physical or conceptual connections (paths, generally, travelled without freight-handling stops) among those. Physical nodes and arcs may have restricted capacities in terms of numbers of vehicles/ convoys and volumes of freight they may accommodate. Moreover, physical links could be mode-specific. Yet, to lighten the presentation, and without loss of generality, we do not introduce these characteristics in this article.

The multi-commodity demand is represented by the set of origin-destination demands $\mathcal{K}$ to be moved between the nodes of $\mathcal{G}^{\text{PH}}$. Each commodity $k \in \mathcal{K}$ is characterized by a quantity/volume of d^k units of freight to be moved from its origin $O(k) \in \mathcal{N}^{\text{PH}}$ to its destination $D(k) \in \mathcal{N}^{\text{PH}}$.

Let Σ represent the set of *potential* services, out of which one builds, at the tactical-planning level, the service network to be used during the next season. According to the particular application (e.g., short planning horizons or adjusting plans to new information), Σ may contain only the services which may be added to the network, or which could be modified (e.g., in the capacity they may carry or the schedule they follow), to adjust the tactical plan to updated information. In all cases, a service $\sigma \in \Sigma$ is defined by its origin $O(\sigma)$, destination $D(\sigma)$, and a route in the physical network made up of a set of facilities (if any) where the service stops to load and unload cargo. In the basic problem setting, services are direct, or *single-leg*, i.e., they move directly from the origin to the destination without any intermediary stops. We discuss the case of multi-leg services at the end of the present subsection. It is noteworthy that, irrespective of the service route definition, $\mathcal{A}$ may contain multiple arcs that have the same origin and destination, but differ in some attributes, e.g., mode and ownership, i.e., carrier-owned/managed or third-party carrier, each with specific cost and capacity features; partially overlapping service routes may also occur when multi-leg services are contemplated.

A so-called *fixed cost* f_σ is associated with service σ, and is incurred to set up and operate it, that is, in planning terms, select and include it in the service network for the next planning horizon. One defines c^k_σ, when the cost of transportation is commodity specific.

The service *capacity* u_σ represents the total demand the service may load and haul. The capacity may be measured in volume, tonnage, length (particularly for railroads), or number of units (e.g., containers for intermodal navigation and rail). More than one capacity measure may be relevant in any given problem setting. Moreover, particular capacities for particular products may also be imposed. To simplify the presentation, and if not otherwise indicated, we continue with a single capacity restriction in this article, which we identify by the generic term "volume". Even in this case, though, some applications involve particular commodity-specific capacities, noted u^k_σ in this paper, which complement the global capacity u_σ.

3.1.2 SND formulation

The goal of a SND model is to select, out of the set of potential services Σ, the service network required to profitably and efficiently satisfy the multi-commodity demand $\mathcal{K}$. Hence, the model is built on a potential service network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, with the set of nodes $\mathcal{N} = \mathcal{N}^{\text{PH}}$, and where the set of arcs $\mathcal{A}$ are defined out of the set of services Σ according to the particular problem setting. In the *basic case*, with single-leg services in a *deterministic* (perfect knowledge of all parameter values, including demand, for the duration of the planning horizon) and *static* (time-related characteristics not explicitly considered, being assumed not to change significantly during the planning horizon), $\mathcal{A} = \Sigma$, with $a(\sigma)$ and $\sigma(a)$ notationally connecting the arc $a \in \mathcal{A}$ and its defining service $\sigma \in \Sigma$ (obviously, $u_a = u_{\sigma(a)}$, $u^k_a = u^k_{\sigma(a)}$, $c_a = c_{\sigma(a)}$, and $c^k_a = c^k_{\sigma(a)}$). Availability of the required resources is also assumed in the basic case; hence, their utilization is not explicitly represented in the formulation.

Two types of decision variables are to be found in all SND formulations, *selection* or *design* and *utilization* or *flow*. For the basic formulation, one thus has

Design variables $y_\sigma, \sigma \in \Sigma$, indicating whether service σ is selected as part of the service network; In its simple, and most often encountered in the literature form, design variables are binary, i.e., $y_\sigma = 1$ if service σ is selected, and 0, otherwise; When the service may be operated several times, modelling either several simultaneous or spread over the time duration considered (see Section 4), the design variables take non-negative integer values, that is, $y_\sigma \in \mathbb{Z}_+$;

Flow variables $x_a^k \geq 0, a \in \mathcal{A}, k \in \mathcal{K}$, modelling the volume (or percentage of the total volume) of commodity k moved on arc a (i.e., on the service defining the arc).

As already mentioned, tactical planning aims for profitable service and operations. In many cases, including the basic problem setting, it is assumed that demand is already known or may be predicted with sufficient confidence. Consequently, the estimated revenue is also assumed known and the goal becomes to minimize the total system cost to provide the service to address that demand. Let y and x stand for the vectors of design and flow decision variables, respectively, and let $\Phi(y,x)$ represent the total cost of operating the system, given the service network y and the distribution of the demand flows on that network, x. The specific form of $\Phi(y,x)$ is related to the particular application. In the literature relative to planning consolidation-based freight carriers, one generally defines this total cost as the sum of the fixed cost associated with selecting the service network $\phi(y)$ and the variable cost $\varphi(y,x)$ of using the service network to transport the demand. Formally, then, the basic SND seeks to

$$\min \quad \Phi(y,x) = \phi(y) + \varphi(y,x) \tag{1}$$

$$\text{s.t.} \quad \sum_{a\in\mathcal{A}_\eta^+} x_a^k - \sum_{a\in\mathcal{A}_\eta^-} x_a^k = \begin{cases} d^k, & \text{if } \eta = O(k), \\ -d^k, & \text{if } \eta = D(k), \\ 0, & \text{otherwise,} \end{cases} \quad \eta \in \mathcal{N}, k \in \mathcal{K}, \tag{2}$$

$$\sum_{k\in\mathcal{K}} x_a^k \leq u_a y_{\sigma(a)}, \quad a \in \mathcal{A}, \tag{3}$$

$$x_a^k \leq u_a^k y_{\sigma(a)}, \quad a \in \mathcal{A}, k \in \mathcal{K}, \tag{4}$$

$$y_\sigma \in \mathbb{Z}_+, \quad \sigma \in \Sigma, \tag{5}$$

$$x_a^k \geq 0, \quad a \in \mathcal{A}, k \in \mathcal{K}, \tag{6}$$

$$(y,x) \in \Psi, \tag{7}$$

where $\mathcal{A}_\eta^+ = \{(\eta,\eta') \in \mathcal{A}\}$ and $\mathcal{A}_\eta^- = \{(\eta',\eta) \in \mathcal{A}\}$ define the sets of incoming and outgoing arcs, for node $\eta \in \mathcal{N}$, respectively.

The value of the total system cost, (1), is restricted by several constraints representing the setting and dynamics of the problem. Equations (2) are *flow-balance*, also called *flow-conservation* constraints, ensuring that all the demand of each commodity leaves its origin (the first case), arrives at its destination (the second case), and departs any other locations at which it arrives (the third case). The *linking-capacit*y constraints (3) and (4) limit the total amount of demand and the volume of each individual commodity, respectively, which travels on arc $a \in \mathcal{A}$ to the capacity provided on that arc by the corresponding service, if selected. Constraints (5) - (6) define the domains of the decision variables, while constraints (7) stand for problem-setting-particular relations, such as, budget limits or topology restrictions on the numbers of incoming and outgoing arcs (the in- and out-degree of the node, respectively).

The objective function is written in general form, accommodating a broad range of cost-computing situations, including non-linearities due to the representation of, e.g., congestion in terminals or on the physical network and penalties for non compliance with service-quality targets or capacity restrictions (e.g. Crainic et al., 1984; Crainic and Rousseau, 1986), and non-continuities in the fixed cost of multi-departure services (e.g. Croxton et al., 2003b,a, 2007) In many such cases, it is assumed that the non-linear or non-continuous characteristics are distinguished for each service and demand individually, or to the individual fixed and variable costs, which yields formulations (8) and (9), respectively

$$\sum_{\sigma \in \Sigma} \phi_\sigma(y) + \sum_{k \in \mathcal{K}} \varphi_k(y, x) \tag{8}$$

$$\sum_{\sigma \in \Sigma} \phi_\sigma(y) y_\sigma + \sum_{k \in \mathcal{K}} \sum_{a \in \mathcal{A}} \varphi_{ak}(y, x) x_a^k \tag{9}$$

The most frequently used system-cost representation, however, assumes linear functions and takes the form (10). In this case, minimizing the total cost of the system is computed as the sum of the fixed costs associated with selecting services and the variable costs corresponding to using the resulting service network to transport the demand.

$$\sum_{\sigma \in \Sigma} f_\sigma y_\sigma + \sum_{k \in \mathcal{K}} \sum_{a \in \mathcal{A}} c_a^k x_a^k \tag{10}$$

This basic formulation, which appears in various forms, particularly related to the modelling of the objective function, in many studies targeting freight-transportation planning issues, emphasizes the network-design nature of the SND formulations. Indeed, any SND model may be cast as a ND formulation on an appropriately-defined network. Thus, formulation (1) - (6) is the same as the general network design model presented in the seminal Magnanti and Wong (1984) paper. Moreover, the linear-cost SND, that is the objective function (10) subject to constraints (2) - (7), is the same as that of the multi-commodity, fixed-cost, capacitated network design problem (see Crainic et al., 2021a, for a state-of-the-art presentation).

3.1.3 SND variants

We complete this presentation with variants of the SND problem setting and their corresponding formulations. These variants are related, respectively, to the capacity-feasibility issue, the service route, the explicit definition of the demand itineraries, the modelling of several interconnected designs, and the requirement to move, or not, all of a demand volume together.

Service capacity feasibility. Considering capacity limitations and including them explicitly in O.R. formulations may raise feasibility issues. These are disturbing, particularly since, in practice, there is "always" a feasible solution, even if quite costly, e.g., by calling on ad-hoc capacity provided by additional vehicles or outsourcing part of the demand transportation, or delaying transport and paying penalties. We return to this discussion in Sections 5 and 6, but the simplest approach is to include *dummy arcs* between the origin and destination of each commodity, with no capacity restrictions and appropriately high cost, The associated *slack-flow variables* then capture the volume of each demand unfulfilled by the capacity of the selected services, while taking care of the feasibility issue.

Let $a^k = (O(k), D(k))$ be the dummy arc associated to commodity $k \in \mathcal{K}$, with ς^k the corresponding slack-flow variable. The flow conservation constraints (2) then become

$$\sum_{a \in \mathcal{A}^+_\eta} x^k_a - \sum_{a \in \mathcal{A}^-_\eta} x^k_a = \begin{cases} d^k - \varsigma^k, & \text{if } \eta = O(k), \\ -d^k + \varsigma^k, & \text{if } \eta = D(k), \quad \eta \in \mathcal{N}, k \in \mathcal{K}, \\ 0, & \text{otherwise}, \end{cases} \tag{11}$$

and the term $\sum_{k \in \mathcal{K}} c_{a^k} \varsigma^k$ is added to the objective function. (These modifications are implicit when the set of dummy arcs a^k is included in $\mathcal{A}$.)

Multi-leg services. In many applications, a service $\sigma \in \Sigma$ stops at a number of different terminals along its route, between the origin $O(\sigma)$ and destination $O(\sigma)$ terminals, to load and unload freight, detach and attach vehicles from/to the convoy, change traction units or crews, and undergo various safety and maintenance verification. The service route is then described by the sequences of $n(\sigma)$ stops (origin, intermediary, and destination terminals), and service legs connecting them. A single-leg, direct, service has $n(\sigma) = 2$.

Following Crainic (2024b), let $\mathcal{N}^{\text{PH}}(\sigma) = \{\eta_i(\sigma) \mid i = 0, \ldots, n(\sigma), O(\sigma) = \eta_0, D(\sigma) = \eta_{n(\sigma)}\}$ be the stop sequence of service $\sigma \in \Sigma$. Then, the *service leg* $l_i(\sigma) = (\eta_{i-1}, \eta_i)$ is defined as the sub-path connecting the consecutive terminals η_{i-1}, $\eta_i \in \mathcal{N}^{\text{PH}}(\sigma)$ of the route of service σ, with $\mathcal{L}(\sigma) = \{l_i(\sigma), i = 1, \ldots, n(\sigma)\}, \sigma \in \Sigma$.

Multi-leg services yield several arcs in the potential service network G, an arc for each service in the service network, i.e., $\mathcal{A} = \mathcal{L} = \bigcup_{\sigma \in \Sigma} \mathcal{L}(\sigma)$. Let $a(l_i(\sigma))$ and $l_i(\sigma)(a)$ stand for the link defined by leg $i \in \mathcal{L}(\sigma)$ of service $\sigma \in \Sigma$ and the service leg defining arc $a \in \mathcal{A}$, respectively. Then $u_a = u_{l_i(\sigma)(a)}$ and $u^k_a = u^k_{l_i(\sigma)(a)}$ in constraints (3) - (4) of the SND formulation.

Path-flow SND formulation. OD demand shipments travel on itineraries, that is, they follow paths in $\mathcal{G} = (\mathcal{N}, \mathcal{A})$. Hence, flow variables and the SND model can be defined in terms of paths/itineraries.

Let Π^k identify the set of possible itineraries of commodity $k \in \mathcal{K}$ on the potential service network (all potential services and terminal operations on the flow), with $\mathcal{A}^k_\pi$ holding the sequence of arcs $a \in \mathcal{A}$ (services in Σ^k_π) making up the itinerary $\pi \in \Pi^k$, and $\delta^{\pi k}_a$ being the definitional Kronecker delta (i.e., $\delta^{\pi k}_a = 1$ when arc $a \in \mathcal{A}^k_\pi$, 0, otherwise).

Let $c^k_\pi = \sum_{a \in \mathcal{A}^k_\pi} c^k_a$ stand for the cost of itinerary $\pi \in \Pi^k$ of commodity $k \in \mathcal{K}$, and define the *itinerary-flow* decision variable h^k_π as the amount of commodity k moved on its itinerary $\pi \in \Pi^k$. The basic path-based SND formulation then becomes

$$\min \sum_{\sigma \in \Sigma} f_\sigma y_\sigma + \sum_{k \in \mathcal{K}} \sum_{\pi \in \Pi^k} c^k_\pi h^k_\pi \tag{12}$$

$$\text{s.t.} \sum_{\pi \in \Pi^k} h^k_\pi = d^k, \qquad k \in \mathcal{K}, \tag{13}$$

$$\sum_{k \in \mathcal{K}} \sum_{\pi \in \Pi^k} \delta^{\pi k}_a h^k_\pi \leq u_a y_{\sigma(a)}, \qquad a \in \mathcal{A}, \tag{14}$$

$$\sum_{\pi \in \Pi^k} \delta^{\pi k}_a h^k_\pi \leq u^k_a y_{\sigma(a)}, \qquad a \in \mathcal{A}, k \in \mathcal{K}, \tag{15}$$

$$y_\sigma \in \{0, 1\}, \qquad \sigma \in \Sigma, \tag{16}$$

$$h^k_\pi \geq 0, \qquad \pi \in \Pi^k, k \in \mathcal{K}. \tag{17}$$

It is well known that the basic arc and path-based formulations are equivalent, that is, they yield the same service network and objective-function value, with $x^k_a = \sum_{\pi \in \Pi^k} \delta^{\pi k}_a h^k_\pi, a \in \mathcal{A}, k \in \mathcal{K}$. The notation and model modifications for the multi-leg-service case are straightforward, and the previous note holds.

Multi-layer SND. Most SND models address problem settings considering a single level of consolidation of shipments into a vehicle or container. Some modes of transportation, e.g., rail, ocean liners, and intermodal barges, involve multiple levels of consolidation, however, as a vehicle may transport many containers, vehicles may be grouped into so-called *blocks* or *convoys* (rail and barge trains), or both. The corresponding SND formulations must, therefore, include these multiple layers of consolidation decisions. Similar modelling requirements arise when resource-management concerns are part of tactical planning and the SND formulation, requiring design layers for services and resource operations (Section 5), as well as in applications to motor-carrier platooning for long-haul movements (Albinski et al., 2020) and autonomous vehicles that can only travel autonomously in certain geographic regions (Scherr et al., 2019; Ammann et al., 2024).

It is noteworthy that, these design decisions display the particular characteristic of an arc in a given decision layer being defined with respect to a set of arcs, often making up a path or a cycle, in another decision layer. To illustrate, consider the case of railroad tactical planning where each potential block (group of cars handled together as a unit) is defined in the block layer, in terms of the path of service arcs that will transport it, if selected in the service layer (Zhu et al., 2014; Chouman and Crainic, 2021). Such interwoven definitions imply several connectivity relations and requirements in terms of both design (arc or node selection) and flow-distribution decisions, raising challenging network-design modelling and algorithmic issues.

Crainic et al. (2022) and Crainic (2024a) synthesize the literature on *Multi-layer Network Design* (*MLND*), proposing a general framework and taxonomy for the field. They also explore complex problem settings with applications to various consolidation-based freight transportation planning issues. We refer to some of these developments in the appropriate sections of this paper. Here, we present the basic multicommodity, fixed-cost, capacitated MLND formulation only.

Let $\mathcal{L}$ be the set of layers of multi-layer network $\mathcal{G} = (\mathcal{N}_l, \mathcal{A}_l) = \bigcup_{l \in \mathcal{L}} \{\mathcal{G}_l = (\mathcal{N}_l, \mathcal{A}_l)\}$, where $\mathcal{G}_l$ is the network on layer $l \in \mathcal{L}$, with $\mathcal{N}_l$ and $\mathcal{A}_l$ the corresponding sets of nodes and arcs. Let $l, l' \in \mathcal{L}$ be a couple of *(supporting, supported)* layers of $\mathcal{G}$ coupled by an arc definition specifying how an arc in *supported* layer l' is related to a subset of *supporting* arcs in layer l (e.g., the supporting arcs form the path defining the supported arc). For simplicity of presentation, we assume that all arcs in $\mathcal{G}$ are design arcs, and that a single set of OD demands $\mathcal{K}$ is defined on a given layer (notice that, the flows on that layer are projected on the layers associated with it through the arc definitions).

The rest of the notion defined for the previous SND models applies in this case as well (adjusted with the appropriate layer index), including for the decision variables $y_{al} = 1$ if arc $a \in \mathcal{A}_l$ of layer l is selected, 0, otherwise; and x^k_{al} indicating the quantity of demand $k \in \mathcal{K}$ assigned to arc a of layer l. The basic MLND formulation then becomes

$$\min \sum_{l \in \mathcal{L}} \left\{ \sum_{a \in \mathcal{A}_l} f_{al} y_{al} + \sum_{k \in \mathcal{K}} \sum_{a \in \mathcal{A}_l} c^k_{al} x^k_{al} \right\} \tag{18}$$

$$\text{s.t.} \sum_{a \in \mathcal{A}^+_{\eta l}} x^k_{al} - \sum_{a \in \mathcal{A}^-_{\eta l}} x^k_{al} = w^k_\eta, \qquad \eta \in \mathcal{N}_l, k \in \mathcal{K}, l \in \mathcal{L}, \tag{19}$$

$$\sum_{k \in \mathcal{K}} x^k_{al} \leq u_{al} y_{al}, \qquad a \in \mathcal{A}_l, \tag{20}$$

$$(\mathbf{y}, \mathbf{x}) \in (\mathcal{Y}, \mathcal{X})_{ll'}, \qquad (l, l') \in \mathcal{C}, l \in \mathcal{L}, \tag{21}$$

$$y_{al} = 1 \in \{0,1\}, x^k_{al} geq 0, \qquad a \in \mathcal{A}_l, k \in \mathcal{K}. \tag{22}$$

The formulation follows the standard network-design pattern. Note that, the flow-conservation constraints need enforcing for the demand-defining layer only in most cases, while Relations (21) stand for the sets of constraints corresponding to the design, flow, or attribute connectivity requirements proper to the multi-layer network design application at hand (Crainic, 2024a).

Split/no-plit demand policy. With respect to the handling of demand flows, two major problem classes are encountered in freight transportation, defined by the possibility to *split* the flow of any particular OD demand among several itineraries, or the obligation to follow a *no-split* policy and use a single itinerary.

The previous formulations address the former case. The definitions of the flow variables have to be modified when a no-split policy is to be modelled. Thus, for the arc-based (path-based) formulation, x_a^k (h_π^k) equals 1 if commodity $k \in \mathcal{K}$ travels on the arc $a \in \mathcal{A}$ (on its path $\pi \in \Pi^k$), and 0, otherwise.

The arc (path) formulation is then modified by multiplying x_a^k (h_π^k) by d^k in the objective function (1) ((12)) and constraints (2) - (4) ((13) - (15)), and changing the domain restrictions of the flow variables (6) to $x_a^k \in \{0, 1\}$ ((17) to $h_\pi^k \in \{0, 1\}$).

3.2 Historical Perspective

There is a broad and extensive literature on the Service Network Design problem and applications. General surveys and syntheses may be found in Crainic and Hewitt (2021); Wieberneit (2008); Crainic (2000, 2003); Crainic and Laporte (1997); Crainic and Roy (1988), while applications to particular fields are synthesized in, e.g., Assad (1980b); Crainic (1988); Cordeau et al. (1998); Newman et al. (2002); Ahuja et al. (2005a); Crainic (2009); Yaghini and Akhavan (2012); Chouman and Crainic (2021); Crainic (2009) for railroads, Bakir et al. (2021) for LTL motor carriers, Ronen (1983, 1993); Christiansen et al. (2020, 2021, 2007, 2004) for maritime transportation, Macharis and Bontekoning (2004); Bektaş and Crainic (2008); Crainic and Kim (2007); Archetti et al. (2022) for multi and intermodal transportation, and Bektaş et al. (2017); Crainic et al. (2024, 2021c) for City Logistics.

Early studies focus mostly on railroad and maritime shipping planning, the former making up the largest set of contributions. This focus is explained, most probably, by the high costs of infrastructure and equipment, the complexity of terminal (e.g., classification, consolidation/blocking, train formation/makeup, loading and unloading) and line/network activities and decisions, and the double consolidation of cars into blocks and of blocks into trains characterizing railroads with the consequent complexity.

Historically, it is noteworthy that before the 1980s, most contributions addressed operational-level issues, e.g., scheduling of operations in terminals and on the lines (accounting, in particular, for priorities when meeting or overtaking other trains), and allocation of vehicles and power units to re-balance supply and demand among the terminals of the carrier (see Section 5). A significant body of work also targeted the organization of the terminals, rail yards and container ports, and the understanding, as well as the modelling, of the delays dues to congested resources in the terminals and on railroad lines. Simulation and queuing methods were the methodologies of choice in those early studies. Discussing them in any depth goes beyond the scope and limits of this paper. We mention them, however, as they provided the basis for the development of tactical-level representations of those phenomena as, mostly convex, function approximations of the average delays, given the capacity of the facility and the traffic intensity aiming to use it (e.g., Crainic and Gendreau, 1986; Powell, 1986b; Powell and Humblet, 1986; Crainic and Quérin, 1988), for inclusion in non-linear SND formulations (e.g. Crainic et al., 1984; Crainic and Rousseau, 1986; Bektaş et al., 2010).

When system-wide planning issues started to be addressed, the contributions targeted single problems or combinations of a limited number of issues only, by applying variants of the basic formulations described in this section.

Early contributions include the pioneering train formation (also called service selection and makeup) model of Assad (1980a) for a line network, which took the form of a piece-wise SND formulation. Several contributions followed focusing on the service selection and the demand-itinerary building as defined in the basic formulations of this section, the main differences being the demand (OD car flows or blocks), the service types (e.g., regular and express), the modelling of costs (linear or non-linear), and the solution method proposed. Exact solution methods proved difficult to address realistically-sized instances (Marín and Salmerón, 1996a), and meta-heuristics were generally proposed (e.g. Martinelli and Teng, 1996; Marín and Salmerón, 1996b; Yaghini et al., 2014).

A parallel line of research is dedicated to the railcar-blocking problem. Blocking addresses the design of the block network, i.e., the selection of the blocks to move the classified (sorted and consolidated into blocks) railcars for all or part of their journeys and the construction of the demand itineraries in terms of block paths. The goal is to minimize the total costs composed of classification costs in yards and the transportation costs on blocks. The service selection is to be performed after, in order to efficiently move the selected blocks (e.g., the block-to-train assignment model of Jha et al., 2008). The models proposed in the literature take the form of the SND formulations presented in this section, the differences among contributions coming from the modelling of the costs of the system, side constraints (e.g., limits on the number of blocks one can build in each yard) and the solution method.

Bodin et al. (1980) proposed one of the first such models, a non-linear (piece-wise) SND formulation with side constraints, blocking delays being dependent on the number of cars assigned to each block. Newton et al. (1998) and Barnhart et al. (2000) also formulate the problem as a network design model, arcs representing candidate blocks among classification yards. No fixed costs are associated to blocks, the yard block-building limit being enforced through budget constraints. A path-formulation and a branch-and-price algorithm (Barnhart et al., 1998) are proposed in the first paper, while the second presents a dual-based Lagrangian relaxation decomposing the problem into easier-to-address subproblems (a continuous multicommodity flow problem and an integer block formulation that selects blocks satisfying yard capacity constraints). Metaheuristics for the arc or path-based formulations are proposed by, e.g., Ahuja et al. (2007) and Yaghini et al. (2011, 2012).

The 80s have seen the emergence of models aiming for the integration of system-wide tactical planning issues, through applications to railroads and motor-carriers. Crainic et al. (1984) presents what is probably the first rail service network design model addressing simultaneously the selection of services and their frequencies, car classification and blocking, train makeup, and freight routing. It is a non-linear path-based SND formulation, minimizing (9) subject to (13) - (17), which accounts for congestion and accumulation-delay phenomena in yards and on rail tracks, service-quality targets (e.g., market-specific delivery times), and trade-offs between operating and time-related costs. The non-linear generalized objective function combines operating and time-related costs for services and demand flows, as well as penalty costs for non compliance with service targets and the capacity limitations of terminals and services. The duration of terminal activities is modelled through convex approximations of average (and standard deviation) delays derived from queuing models of congestion in terminals. A similar approach is used for inter-terminal travel times as train services, with various priorities, are captive of the infrastructure. Block fixed costs are not included. They are rather approximated through the accumulation-delay costs and the limits on yard-specific block dimensions. The model also integrates the distribution of empty cars through one or several origin-destination demand matrices (generated through demand-distribution models from the surplus and shortage levels at yards, which were derived from the loaded-car demand). These matrices become commodities to be handled simultaneously with all other OD commodities in the problem. A heuristic solution method was used to address realistically-sized problem instances derived from the case of a large North-American railroad.

Crainic and Rousseau (1986) generalize the model for the tactical planning of any consolidation-based multicommodity multimode freight transportation systems, and propose a heuristic decomposition-based algorithm integrating discrete service-frequency modifications, network flow optimization, and column-generation principles. Bektaş et al. (2010) later proposed more refined solution methods integrating Lagrangean-based relaxation and decomposition algorithms. The authors show that, first, non-linearities may be handled efficiently through decomposition and, second, that the relaxation of the flow constraints, which yields an arc decomposition, has computationally better convergence properties than the dualization of the capacity constraints. These results are very encouraging for this demanding but important research topic.

Further important contributions to the basic SND methodology were made in the LTL planning field, simultaneously with the pioneering integrated SND methodology for railroad planning. On the one hand, Roy and Delorme (1989) generalize the methodology of Crainic et al. (1984) and Crainic and Rousseau (1986) to the LTL problem. The non-linear path-based formulation simultaneously selects services, itineraries, and balances the empty vehicles (trailers) between terminals. The objective function minimizes the total transportation and consolidation costs, with penalties for non respecting vehicle capacity and service targets. Data from large Canadian LTL carriers are used to show the interest of the methodology and to study the trade-offs between operation costs and time-related service targets in terms of solution structure and productivity (see also Crainic and Roy, 1988; Roy and Crainic, 1992) A different approach, motivated mainly by the structure of the U.S. LTL motor-carrier industry, was initiated by Powell and Sheffi (1983), followed by an important set of contributions. The model follows the basic, linear-cost SND model, but takes advantage of the structure of the studied LTL networks (each regional terminal is linked to a very limited number of hubs and services are single leg) and includes the inter-terminal balancing of empty vehicles. The LTL system structure is further exploited by introducing tree-based flow routing (Powell and Koskosidis, 1992). An add/drop heuristic is used to address the formulation (at each iteration one adds or drops services to/from the service network; see also Powell, 1986a; Farvolden and Powell, 1994).

4 SSND - Accounting for Time and Schedules

Events, decisions, and activities occur in time. How they are included in the problem settings and descriptions determines in a large part how time, and those time attributes, are represented in the Operations Research models proposed to assist decision making, as well as the type of O.R. model and solution method proposed.

Recall that, the length of the tactical planning horizon, the so-called season, is determined by the homogeneity of that time duration, in terms of regular demand and, hence, activities, in a stable environment. The season length varies with the carrier type, but may also vary with the climate, e.g., between a dry and a rainy season. As indicated previously, the length of the tactical plan is not equal to the planning horizon. It is, in fact, much shorter, determined by the repetition pattern of the regular demand. The activities planned for this schedule length are then repeatedly executed for the duration of the tactical planning horizon.

The problem settings and formulations of Section 3 do not appear to fit into this discussion, nor to integrate time. But, they do. The basic formulations address situations where there is no variation in demand for the duration of the schedule length. Either the schedule length is short and one assumes that everything "happens simultaneously", or the demand arrivals and service departures are assumed to be uniformly distributed over the schedule length for longer time spans, e.g., the 1440 minutes of the week (e.g., Crainic et al., 1984). Because the time attributes of the various problem elements are not explicitly included, those SND models are qualified as *static*.

The problem settings and models involving the *service frequency* are also qualified as static. The frequency of a service is defined as the number of times that "same" service is run during the schedule length (alternatively, the associated number of vehicle departures), and is formulated as $y_\sigma \in \mathbb{Z}_+, \sigma \in \Sigma$ in constraints (5). The operation hypothesis in such cases is that departures are equally spread out over the schedule length. The models summarized at the end of the previous section, including Crainic and Rousseau (1986); Crainic et al. (1984); Roy and Delorme (1989), and Powell and Sheffi (1983), belong to this category, the former three including service frequency as decisions, the latter extracting the service frequencies out of the optimization results.

Addressing in a more refined way the variations of demand in time and the scheduling of the services selected to answer the transportation requests marks an important step in tactical-planning definition and the development of service network design methodology. Various time attributes are explicitly considered in such *time-dependent*, some authors prefer the term *time-sensitive*, problem settings. Hence, demand $k \in \mathcal{K}$ is further characterized by an *availability time* $\alpha(k)$ at origin $O(k)$ and a due date $\beta(k)$ at destination $D(k)$. Similarly, services are characterized by a *schedule* indicating the departure and arrival times, $\alpha(\eta_i)$ and $\beta(\eta_i), i = 1, \dots, n(\sigma)$, respectively, at each of the terminals $\mathcal{N}^{\text{PH}}(\sigma)$ on its route. Services are further characterized by a total duration $\tau(\sigma)$, that includes the time spent in terminals and the moving time associated to each leg $\tau(l_i(\sigma))$. Schedules may be strict, as for most European and Canadian railroads and regular containership liners, or more of an "indicative" nature, the schedule being eventually modified to account for particular events (e.g., an important customer for which one stops a direct LTL service) or how much freight is already loaded.

The time-related characteristics of demand and services are generally incorporated into SND models by explicitly attaching them to a *service time-space network* $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ capturing the dynamics of the system for the duration of the schedule length. The service legs provide the arcs specifying the potential movements through space and time of the corresponding vehicles and convoys, while itineraries perform the same role for the transportation of demand, from it origin terminal at its availability moment in time, to its destination terminal at a time compatible with its due date. When formulated on such a network, the SND is often referred to as a *Scheduled Service Network Design* (*SSND*) model.

A time-space network is often built by partitioning the schedule length into non-overlapping periods of time, wherein all activities at terminals during a period will be modelled as occurring at the same time. This approach is also known under the term *time discretization*. The granularity of the partition and the definition of each period are normally governed by the characteristics of the system elements and operation practice. Most applications in the literature, however, implement the classical approach, first introduced by Ford and Fulkerson (1958), according to which all periods are of the same length and apply to all the nodes of the network. We follow this approach to introduce the topic, and discuss alternatives to conclude the section.

Let T be the schedule length. Let $\{0, 1, \dots, |\mathcal{T}| - 1, |\mathcal{T}| = T\}$ be the set of time instants t that partition the schedule length into $|\mathcal{T}|$ periods of equal length, and $\mathcal{T}$ the ordered set of those periods. Period t is then defined as $(t - 1, t)$, all events taking place during this period being assigned to time instant t. The node set of the time-space network, may then be defined as $\mathcal{N} = \{\eta_t, \eta \in \mathcal{N}^{\text{PH}},\ t = 0, \dots, |\mathcal{T}|\}$, including copies of all the terminals in the physical network at all the time periods (instants) defined.

The arc set $\mathcal{A}$ consists of two types of arcs. The first encompasses the *moving* arcs, that is, the service legs according to their schedules. Specifically, the moving arc $a \in \mathcal{A}$, standing for service leg $l_i(\sigma) = (\eta_{i-1}, \eta_i), i = 0, \ldots, n(\sigma), \sigma \in \Sigma$, is defined as $a = (\eta_{\alpha(\eta_{(i-1)})}, \eta_{\beta(\eta_i)})$, representing the departure of the service leg from its origin terminal η_{i-1} at time instant $\alpha(\eta_{(i-1)})$ and arriving at its destination terminal η_i at time $\beta(\eta_i)$.

The second arc type, often referred to as a *holding* arc, represents the possibility to wait, or to "hold", for one period goods or resources at a terminal. It is thus of the form $a = (\eta_t, \eta_{n+1}), \eta \in \mathcal{N}^{\text{PH}},\ t = 0, \ldots, |\mathcal{T}|$. There are no fixed costs associated with holding arcs. A number of capacity and unit cost parameters may be defined, however, to represent the handling or warehousing capabilities of the terminal and costs of this handling or of keeping resources idle.

It is worth recalling that 1) the SSND aims for a repeatable tactical plan answering the regularity of demand over the planning horizon, and 2) not all demands and services have their initial and terminal instants within the schedule length; some may start during the previous application of the plan and terminate currently; others, start during the current application of the plan, but terminate during the following one. These issues are addressed by having the moving and holding arcs modelling service legs and terminal activities that would end during the next application of the plan *wrap-around*, which corresponds to having the respective arcs end after the beginning of the current schedule length. To avoid the apparent paradox of an activity terminating before it starts, the time computations are performed *modulo*(T) (see, e.g., Crainic and Hewitt, 2021, for details).

The SSND formulations then take the form of the corresponding basic SND models of Section 3 applied to the time-space network $\mathcal{G}$ as defined above. The SSND considers the same two sets of decision variables, selecting scheduled services (with frequencies, in some cases, representing the number of simultaneous departures of the service) and building itineraries in the service network for demand-flow distribution. The SSND constraint sets are analog to those of the SND, but they are written to account for the time attributes of the system, e.g., for the flow-balance equations defined according to the availability instant and the delivery due date.

The differences and challenges come from applications to particular problem settings and the dimensions of those problems and of the corresponding formulations.

Focusing on railway planning, Haghani (1989) presents a model which attempts to combine train routing and scheduling, make-up, as well as empty car distribution on a time-space network with fixed travel times and pre-specified traffic rules. A heuristic is used to address a somewhat simplified version of the model and illustrate the interest of integrated planning.

Zhu et al. (2014) propose a multi-layer time-space SSND model, which appears to be the first comprehensive formulation to select the train services and schedules to operate for a given schedule length, the car classification policies, the blocks to build in each terminal, with their routes within the service network, the train makeup, and the demand itineraries using these services and blocks. The authors also introduce a matheuristic solution methodology combining slope scaling, a dynamic block-generation mechanism, long-term memory-based perturbation strategies, and an ellipsoidal search, i.e., a new intensification mechanism to thoroughly explore very large neighborhoods of elite solutions in an efficient way using information from the history of the search. Experimental results show that the proposed solution method is efficient and robust, yielding high-quality solutions for realistically-sized problem instances.

Express-package deliveries often involve more than one mode, typically road and air (or, recently, high-speed trains) services. Early efforts to design such service networks focused on the specifics of the transportation network and potential service routes to build reduced time-space networks (Kim et al., 1999; Barnhart et al., 2002). While studying the same problem, Armacost et al. (2002) proposed a very interesting modelling approach based on defining "new", composite design decision variables for aircraft operations that encode the selection of multiple aircraft services. The authors then showed that, with an appropriate set of constraints, the resulting model provides sufficient capacity to transport all demand, even though demand flows are not explicitly modelled.

Time-expanded SSND models for LTL service network design do not appear in the literature until the 2000s. Thus, Jarrah et al. (2009) propose a SSND formulation on a time-space network for the LTL service-design and flow planning problem with explicit service commitments (measured in delivery days). The authors leverage LTL-carrier policies (e.g., single-path per shipment) and the previous work by Powell and Koskosidis (1992) to formulate the problem using the in-tree structure of the problem, and generating in-trees through a column-generation procedure within a meta-heuristic scheme. The in-tree structure was also exploited in Erera et al. (2013a) in the context of a matheuristic scheme which at each iteration chooses a destination terminal and then solves an integer program to route freight destined for that terminal, holding fixed the routes for freight destined for other terminals. Continuing this line of research, Lindsey et al. (2016) proposes a meta-heuristic with neighbourhoods defined by adjusting many in-trees, and associated time-space demand itineraries, simultaneously.

About the same time at the turning of the millennium, a number of interesting SSND developments were proposed targeting City Logistics. The first modelling framework for the short to medium-term planning of two-tier City Logistics systems was proposed by Crainic et al. (2009). The problem setting considers different vehicle types and specific product-to-vehicle assignment rules at each tier, time-dependent origin-destination demand, and the need to schedule and synchronize first-tier multi-modal services and second-tier road-based multi-tour work assignments. The modelling framework combines in a path-based formulation, a SSND model for the first tier, and a scheduled, multi-depot, multi-tour, heterogeneous vehicle routing with time windows formulation for the second tier. Demand itineraries linked the two tiers and provide the synchronization environment.

The authors observed that, for mid-term tactical planning and the evaluation of long-term strategic alternatives, the second-tier routing problem could be approximated and added to the first-tier formulation through appropriately-defined service costs on links connecting intermediate facilities (the satellites) and customer-zone. Fontaine et al. (2021) capitalized on this observation, while addressing a richer problem setting than in previous literature. The authors additionally integrate inbound and outbound demands, consider passenger-transportation modes (buses, tramways, etc.), multiple compartment vehicles, multiple satellite capacity measures (expressed in terms of freight volume and numbers of first and second-tier vehicles), several heterogeneous limited-size fleets of particular transportation modes, and decisions on the assignment of customers to consolidation centers and satellites. The problem is formulated as an arc-based SSND model and an efficient Benders decomposition algorithm is developed to solve it.

The representation of time and the time-space formulations raise a number of issues and challenges and, while quite a volume of work has been already dedicated to these issues, this is still a very broad and rich research area.

One may notice, first, that the schedule-length partition does not need to be the same at all terminals. Indeed, the relative importance of the terminal with respect to the overall work load may point to the need to have a fine granularity for high-utilization terminals, hubs, for example. Many services operate, start, stop, or terminate, at such terminals over most of the schedule length duration, and many demand flows need to be handled. In contrast, several smaller regional terminals may have to operate at less intensive levels and at certain time moments only. Mini-time-space networks may then be used to appropriately model each terminal in time, these mini-networks being then connected through the services arriving and departing at the node during the interval. Moreover, a continuous-time representation may be used for the activities modelled by the arcs of a mini-network. While mentioned in the literature (Crainic and Hewitt, 2021), we are not aware of many actual applications (Pedersen and Crainic, 2007).

An interesting observation is that schedule-length partitioning is not necessarily required when the schedules of potential services are rather strict. One may rather use the arrival and departure times of services at terminals to create the time instants of the time-space network. Then, the physical nodes are duplicated at relevant time instants only (that is, only when the event takes place at the terminal), and the availability time of each demand at its origin terminal is "projected" on the first time instant following arrival (Morganti et al., 2020; Kienzle et al., 2024). One still models on a time-space network in such a representation, but determining an appropriate granularity of partition appears less of an issue.

Different approaches may be called upon when service schedules are not so strict (or, when time-related uncertainty is considered) and one is more interested in the number of departures within a given time interval to ensure timely delivery of demands at destinations. Modelling the schedule length and associated events using so-called continuous time is such an approach. The timing of events related to service, and itinerary, operations then become part of the decision variables, and a significant amount of constraints governing arrivals, departures, and activity synchronization at terminals has to be included (e.g., Lange et al., 2024). More research is required in this area.

A coarse-granularity-based approach may be used when the schedule length is of a limited duration (e.g., a working day as encountered in many LTL applications). The periods, generally of unequal length, then may be tailored to fit the arrival pattern of demand and the desired pattern of departures to ensure on-time arrival at the next stops of out-going services (e.g., three periods, morning, early afternoon, late afternoon / early evening). In such a case, the time attributes of each service would indicate the coarse time of the day when it is supposed to leave or arrive at the terminal, and a frequency-based formulation determining the number of departures during that period.

In many cases, however, time discretization is the preferred methodology for SSND and the issue of the partition granularity must be addressed. A fine granularity yields short time periods and provides the means to build a detailed representation of time and time-related activities. But, it results in very large time-space networks and very high solution-method challenges. A coarser granularity alleviates partially this problem, but may result in a poorer representation of decisions and operations in time (Boland et al., 2019) Consequently, an important body of work has focused on determining the appropriate granularity for SSND time-space networks, particularly in the context of LTL applications.

Introduced by Boland et al. (2017), the *Dynamic Discretization Discovery* (*DDD*) algorithmic strategy starts with a coarse granularity and iteratively refines it. The initial so-called *partially time-expanded network* is formulated in such a way that the corresponding SSND is a relaxation of an SSND formulated on a time-space network derived from the finest possible granularity in managerial terms (also called "complete enumeration", e.g., the minute). Then, at each iteration of the DDD, the partially time-expanded SSND is solved and the solution is examined to see if it can be converted to an optimal solution to the complete-enumeration SSND, in which case, the method stops, Otherwise, the current partially time-expanded network is refined and the algorithm continues.

Contributions targeting DDD enhancements and generalizations followed. Hewitt (2019) proposes speed-up techniques for DDD when used to address LTL-inspired SSND instances. Marshall et al. (2021) propose a variant of DDD based on a differently-formed partially time-expanded network. DDD has been adapted to other SSND-related problems, as well. Medina et al. (2019); He et al. (2023) propose adaptations of the algorithm to SSND problems that also determine local delivery routes. Escobar-Vargas and Crainic (2024) adapt the DDD concept to address a mixed-integer formulation combining discretized and continuous time representations for a two-echelon location-routing problem with time-dependent demand and synchronization requirements.

5 SSND-RM - Accounting for Resource Management

Carriers need physical and human resources in order to operate their planned service network. We focus on the resources required "in the field", that is, manning services and performing terminal operations on services and freight. Physical resources, the term "asset" is also found in the literature, refer mainly to the vehicles (e.g., trucks, ships, barges, airplanes, cargo bikes, and drones), traction/power units (e.g., road tractors and locomotives), and loading/hauling units (e.g., trailers, railcars/wagons, and containers) required to move freight. Terminal physical resources include cranes of various types and yard locomotives. We use the term *crew* to refer to a group of people required to operate these resources and services.

Assigning human resources to services provides the means to execute them (except when automated vehicles are providing the complete service, which is still very rare). On the other hand, assigning material resources to services define their characteristics. Power units determine how much can be hauled in terms of combined load and vehicle weight on each type of physical-network arc. Similarly, the loading/hauling units used determine the capacity of the service. Consequently, the resource management and resource-to-service assignment issues are very important for the definition, optimization, and performance of the carrier and its service network.

As pointed out in Section 2.3, one finds two major categories of resource-management problem settings and research, one focusing on particular resources, the second on the integration of resource-management concerns into system planning and SND modelling. Presenting a comprehensive review of the long history of research targeting the former category, and of the associated rich corpus of literature, is beyond the scope of this article. We briefly survey a number of milestones in the following. For detailed surveys and syntheses of the literature, the interested reader may turn to Dejax and Crainic (1987) for contributions until the end of the 80's, Cordeau et al. (1998); Piu and Speranza (2014) for rail, as well as to the general articles and chapters mentioned in Section 3.2.

The beginnings of O.R. work on the management of resources may be traced to the problem of repositioning resources. Indeed; trade and, hence, traffic is unbalanced, both in volume and the nature of the freight moved. Consequently, once at destination, the vehicles that brought freight in are not required to move freight out, but they, or similar resources, are needed at other terminals (including the one from where the initial movement initiated). At a network-wide level, one thus observes surpluses of certain resources at a number of terminals and deficits of the same resources at others. Resources then need to be repositioned or balanced.

Starting with the pioneering work on empty cars and containers (Bomberault and White, 1966; White, 1968; White and Bomberault, 1969; White, 1972), as well as on locomotive management (Florian et al., 1976), most of this literature and developments address operational planning issues, e.g., distribution and routing. Network optimization is the methodology of choice in this field, evolving from the initial transportation problem models to the contemporary mixed-integer time-space multicommodity formulations integrating various practical rules and constraints, including Jordan and Turnquist (1983); Crainic et al. (1993); Joborn et al. (2004); Powell and Topaloglu (2005); Ahuja et al. (2005b); Vaidyanathan et al. (2008b,a); Bouzaïene-Ayari et al. (2016); Piu et al. (2015); Ortiz-Astorquiza et al. (2021); Miranda et al. (2020). The scheduling of crews required to man a given service network has not received the same level of attention. We signal the contributions by Crainic and Roy (1992) and Erera et al. (2013b) for LTL driver scheduling, and Balakrishnan et al. (2016) for railroad train crews. We also signal the important research effort and related body of contributions on Approximate Dynamic Programming Powell (2011, 2021) and applications to fleet management (e.g. Powell and Carvalho, 1997, 1998; Godfrey and Powell, 2002a,b; Powell et al., 2007; Bouzaïene-Ayari et al., 2016).

Turning now to the second category, the integration of resource-management concerns into system planning and SND modelling, one notices that many of the early service network design models do not consider resource availability. The formulations aim only to provide services with sufficient capacity to transport the demand shipments. It was assumed that a resource-centric operational model, of the first category discussed above, would then be used to provide the resources required to support the selected plan.

Early acknowledgement of the possible shortcomings of such strategies, yielded SND formulations that accounted for the need to reposition empty vehicles. Initial contributions model the repositioning of railcars or tractor-trailer units as sets of OD demands to be distributed simultaneously with the "regular'" commodities (Crainic et al., 1984; Crainic and Rousseau, 1986; Crainic and Roy, 1988).

Similar concerns while addressing LTL and express-delivery planning, yield the inclusion of explicit decision variables capturing the number of empty ground vehicles to move between terminal pairs to balance the numbers of inbound and outbound vehicles at each terminal (e.g., Powell, 1986a; Smilowitz et al., 2003; Jarrah et al., 2009; Erera et al., 2013a). The same approach may also be found when costly vehicles, e.g., planes, vessels, and locomotives, are part of the system and of the definition of services, and their flow has to be balanced at each terminal and time period (e.g. Armacost et al., 2002; Lai and Lo, 2004; Pedersen and Crainic, 2007; Pedersen et al., 2009; Bilegan et al., 2022).

This strategy is generally identified as *design-balanced* SND / SSND. It aims to ensure that the number of services that arrive at a node equals the number of services that depart, by adding the set of node-degree constraints (23) to the formulations introduced previously.

$$\sum_{a \in \mathcal{A}_\eta^+} y_a - \sum_{a \in \mathcal{A}_\eta^-} y_a = 0, \quad \forall \eta \in \mathcal{N}. \tag{23}$$

This approach may be extended. Thus, the service-design variables may be multiplied by an appropriate factor when more than one unit of resource is associated to the service, which ensures the balance of the respective resource at the nodes of the network. Moreover, more than one type of resource may be considered by instantiating a set of constraints similar to (23). Note, however, that additional constraints may be required to govern the inter-resource relations. To illustrate, consider several types of container-carrying railcars, of different lengths, which can be used simultaneously. Then, one must ensure that the railcar combinations planed to be blocked together or to move together on the same service do not exceed the block or the train maximum permitted length (Kienzle et al., 2024).

The design-balancing requirement brings an additional challenge to addressing SND formulations as, for example, rounding up a fractional solution to the linear programming relaxation is no longer guaranteed to yield a feasible solution to the original problem. Heuristic solution methods are therefore generally proposed (see, e.g., Pedersen et al., 2009; Vu et al., 2013; Chouman and Crainic, 2015).

Yet, it also induces a structure to the problem and solutions. Specifically, that a design can be decomposed into cycles. In other words, that resources move according to cycles on the (potential) service network, limited in time by the need to return to their home base for inspection and maintenance. Andersen et al. (2009b,a) thus extend SND formulations to include resource cycles, cyclic schedules, and the coordination/synchronization of several railroads and navigation services at particular junction points. The authors also show that cycle-based formulations provide more modelling flexibility and computational efficiency. Andersen et al. (2011) exploit this structure in a branch-and-price-based scheme for the problem wherein vehicles flow on cycles and commodities flow on paths, with both cycles and paths generated dynamically via column generation.

Let $\Theta = \{\theta\}$ stand for the set of feasible cycles the units of the resource considered may perform, f_θ the "fixed" cost of selecting and operating the resource cycle $\theta \in \Theta$, and δ_θ^σ the cycle-to-service assignment indicator, where $\delta_\theta^\sigma = 1$ if the resource performing cycle $\theta \in \Theta$ may support service $\sigma \in \Sigma$, and 0 otherwise. Define the binary decision variable $y_\theta = 1$, if cycle $\theta \in \Theta$ is selected, and 0 otherwise. The basic SSND with single resource cyclic management then becomes

$$\min \quad \sum_{\sigma \in \Sigma} f_\sigma y_\sigma + \sum_{\theta \in \Theta} f_\theta y_\theta + \sum_{k \in \mathcal{K}} \sum_{a \in \mathcal{A}} c_a^k x_a^k \tag{24}$$

subject to constraints (2) - (7) enriched with

$$y_\sigma \leq \sum_{\theta \in \Theta} \delta_\theta^\sigma y_\theta, \quad \forall \sigma \in \Sigma, \tag{25}$$

$$y_\theta \in \mathbb{Z}_+, \quad \forall \theta \in \Theta, \tag{26}$$

where (24) minimizes the selection and operation costs of services and resources, plus the cost of moving the demand flows, while constraints (25) link the selection of services and the resources required to operate them.

It is noteworthy that models incorporating design-balance type of constraints only acknowledge that resources are needed to support services and transport shipments and, thus, may have to move empty to be positioned for future activities. They do not recognize, however, that resources need to periodically return to their specific home-base terminal, and are not easily extended to account for other considerations such as the size of the fleet, for example. Such issues are addressed in models explicitly incorporating the *resource management* considerations, as illustrated by the SSND with cyclic resource operations shown above ((24) - (26)).

Crainic et al. (2014b, 2018); Hewitt et al. (2019); Crainic and Hewitt (2021) propose more general *Scheduled Service Network Design with Resource Acquisition and Management*, *SSND-RAM*, formulations, which consider several types of resources and integrate outsourcing servicing certain markets, resource acquisition, allocation, and re-allocation decisions. Moreover, Hewitt et al. (2019) explicitly recognizes uncertainty in demand volume when tactical decisions are taken relative to the design of the service network and the resource allocation (see Section 6). The solution methods proposed in these papers combine dynamic cycle-generation schemes and mechanisms to choose cycles and services, as well as to move demand flows given the resulting capacity.

6 Uncertainty

The tactical plans developed by carriers are formulated well in advance of their implementation. As noted earlier, the aim of this planning process is to establish a series of services for a designated schedule length, which are intended for repeated application over a specified horizon to meet the regular transportation demands of a set of shippers. While the demand from these shippers is expected to have a repeatable nature, the specific details of their requests may change randomly each time they are made. Additionally, unexpected events can further influence the execution of scheduled services, altering the planned transportation supply. Consequently, carriers that engage in tactical planning invariably encounter a degree of uncertainty within the overall informational context that they face.

When employing SND models to address carrier-centric tactical planning problems, it is important to first identify the main sources of uncertainty present in the informational context. On the demand side, using the basic linear-cost SND model minimizing (10) subject to (2) - (6) as an illustrative example, it is frequently observed that commodities $k \in \mathcal{K}$, associated with requests from regular shippers, exhibit randomly varying volumes or quantities (i.e., $d(k), k \in \mathcal{K}$). These fluctuations are often attributed to the random nature of demands, which affect shipper requests throughout the horizon over which the tactical plan is repeatedly applied. On the supply side, the characteristics of the potential services set Σ can also undergo random changes. Specifically, for a given service $\sigma \in \Sigma$, while the fixed cost f_σ is often assumed to be known (reflecting the overall setup cost committed upfront to include the service in the transportation supply over the considered horizon), the variable cost $c^k_{\sigma(a)}, a \in \mathcal{A}$, incurred for using the service to repeatedly transport demand commodities can be subject to random changes. Such variations can arise from, for example, fluctuating delays in performing the required transportation operations. Furthermore, each service $\sigma \in \Sigma$, when selected, is expected to supply a given capacity u_σ enabling demands to be loaded and hauled. However, disruptive events, such as unavailabilities in the resources supporting the given service (e.g., vehicle breakdowns), may randomly change this capacity (effectively reducing u_σ) in a given application of the tactical plan.

In this paper, we consider the form of uncertainty known as *randomness*, as described in Klibi et al. (2010), which is associated with the regular operations of a carrier. Specifically, randomness phenomena are encountered when a carrier performs its business as usual activities. It is thus expected to both occur in normal planning settings and be more readily quantifiable (past data being often available to help formulate these random variations). Therefore, in this section, it is assumed that the type of uncertainty that the carriers face can be accurately represented through well-defined random variables.

When planning amidst such uncertainty, a primary challenge arises when initially established tactical plans, including specific services selected and anticipated commodity itineraries, become infeasible or financially burdensome to implement due to random changes on either the demand or the supply sides. Indeed, this uncertainty can render any well-conceived tactical plan, developed through the use of forecasts, ineffective, necessitating significant revisions often associated with considerable additional costs. The critical question, then, is how to effectively address this challenge during the initial tactical planning phase performed by carriers. In this section, we explore the methodological developments made over the years specifically aimed at tackling this issue. Our focus is on discussing the stochastic optimization methods proposed for solving service network design problems relevant to freight transportation planning. These methods represent the main efforts dedicated to addressing the tactical planning problems of carriers under conditions of uncertainty.

The remainder of this section is divided as follows: using the basic linear-cost SND model as an illustrative example, we first recall the two general modelling paradigms used in applying stochastic optimization (Section 6.1), and then review the main methodological advancements that have been made in formulating and solving stochastic scheduled service network design models (Section 6.2). The latter subsection provides an historical overview of the methodological progress in this area.

6.1 Modelling Paradigms

There are two general modelling paradigms when applying stochastic optimization. The first involves imposing a set of probabilistic constraints, known as *chance constraints*, which serve to limit the likelihood of solutions becoming infeasible at execution time, due to the uncertainty present in the model's parameters. The second paradigm explicitly formulates the model's decisions in accordance with the information revelation process inherent to the problem. Thus, the formulation explicitly differentiates between decisions made under uncertainty, referred to as *a priori decisions*, and those made after the uncertainty is partially or fully revealed, referred to as *recourse decisions*. Models resulting from the application of this paradigm are termed *stochastic models with recourse*. In this section, we present these two paradigms as they are applied to the basic linear-cost SND model. This discussion aims to lay the groundwork for clarifying the methodological contributions that are then examined in the subsequent subsection.

Before detailing any stochastic SND model, it is essential first to formulate the uncertainty considered. Stochastic optimization is applied under the general assumption that the uncertainty affecting the model's parameters can be accurately modelled using random distributions. Therefore, let us first define the following probability space $(\Omega, \mathcal{F}, \mathbb{P})$ associated with a random experiment that reveals the considered stochastic parameters. The set Ω contains the possible outcomes $\omega \in \Omega$ of the random experiment. As for $\mathcal{F}$, it defines the set of all events. Lastly, let $\mathbb{P}$ be the measure that assigns probabilities to the possible outcomes of the random experiment.

The two modelling paradigms are illustrated through the more general case concerning the informational uncertainty inherent in SND. We thus assume that the following parameters are stochastic: 1) the commodity volumes $d^k, k \in \mathcal{K}$; 2) the service variable costs $c^k_{\sigma(a)}, k \in \mathcal{K}, \sigma \in \Sigma, a \in \mathcal{A}$; and 3) the service capacities $u_a, \sigma \in \Sigma$. For simplicity, we assume that only a general capacity applies to the total amount of commodity flow that can be transported through the selected arcs. Consequently, we exclude the constraints (4) from our presentation, although all models can be readily extended to incorporate them. For notation purposes, vectors $\mathbf{d}, \mathbf{c}, \mathbf{u}$ contain the random variables for these parameters and vector ξ is defined as the concatenation of the previous three vectors (i.e., $\xi^\top = [\mathbf{d}^\top, \mathbf{c}^\top, \mathbf{u}^\top]$). Additionally, $\mathbf{d}(\omega), \mathbf{c}(\omega), \mathbf{u}(\omega), \omega \in \Omega$, are the vectors of parameter values corresponding to the possible realizations of the random event.

Chance constraints. To properly introduce the chance constraints formulation, one must consider the potential impediments that prevent a carrier from directly implementing a tactical plan in light of random changes occurring within the informational context. Let $\overline{y}_{\sigma(a)}, a \in \mathcal{A}$ and $\overline{x}^k_a, a \in \mathcal{A}, k \in \mathcal{K}$, represent a specific feasible solution obtained from the deterministic model: Minimize (10) subject to (2), (3), (5), and (6), which utilizes forecasts for the uncertain parameters. This tactical plan, illustrated by the solution vector $(\overline{y}, \overline{x})$, outlines both the services to be operated (i.e., the design decisions $\overline{y}$) and the itineraries chosen for the various commodities (i.e., the flow decisions $\overline{x}$). However, due to the uncertainties discussed earlier, it may not always be feasible to execute the tactical plan without alterations. In particular, for some realizations $\omega \in \Omega$, given the observed volumes of commodities $\mathbf{d}(\omega)$ and service capacities $\mathbf{u}(\omega)$, the chosen services $\overline{y}$ and planned itineraries $\overline{x}$ might fail to satisfy the demands as they emerge from the shippers or be viable given the capacities observed for the selected services.

Among the significant concerns for carriers, two scenarios stand out prominently. First is the situation where the designed transportation service supply and planned commodity itineraries are inadequate to cover the total demand issued by the shippers. As highlighted earlier, such circumstances often compel carriers to resort to ad-hoc capacity, incurring substantially higher costs or penalties that detrimentally affect their financial outcomes. Second, there's the risk that selected services may not support the intended commodity flow to be transported due to unforeseen reductions in available capacities, undermining the feasibility of the planned itineraries. Therefore, formulating tactical plans that limit the risk of observing these scenarios can be highly beneficial for carriers. Chance constraints are instrumental in embedding these probabilistic considerations within the tactical planning framework, offering a method to preemptively address potential execution challenges.

As detailed earlier, dummy arcs $a^k = (O(k), D(k))$ for each commodity $k \in \mathcal{K}$ can be effectively used to model the utilization of ad-hoc capacity for transporting portions of commodity volumes. Given the stochastic nature of the planning context (where both shipper demands and service capacities can randomly vary), one can deduce that the dependency on ad-hoc capacity fluctuates with each specific realization of the observed random experiment. Consequently, we define $\varsigma^k(\omega)$, for each $\omega \in \Omega$ and $k \in \mathcal{K}$, as the *realization-specific slack-flow variables.* These variables precisely capture the volumes of commodities transported using ad-hoc capacity under varying circumstances. Thus, probabilistic constraints imposed on the tactical plans can be formulated by setting limits on the flows assigned to these variables.

Similarly, the feasibility of linking-capacity constraints under the explicit consideration of uncertainty is another critical factor. Specifically, for a given realization $\omega \in \Omega$, the observed capacities of the selected services might prove inadequate to support the planned commodity flows they should carry. To address these limitations, additional capacity on specific services would be necessary to balance the shortfall. Therefore, $\varrho_{\sigma(a)}(\omega)$ is introduced for each $\omega \in \Omega$ and $\sigma(a) \in \Sigma$, signifying the *supplemental service capacity needed on a realization-specific basis* to implement the planned itineraries. Consequently, probabilistic constraints for the tactical plans can be established based on the required extra capacity to compensate for the observed shortfalls.

Following a similar modelling approach as the one presented in Hewitt et al. (2021), let us first define function $F_i^k(x)$ as the net flow for commodity $k \in \mathcal{K}$ at node $i \in \mathcal{N}$, i.e., $F_i^k(x) = \sum_{a \in \mathcal{A}_\eta^+} x_a^k - \sum_{a \in \mathcal{A}_\eta^-} x_a^k$. A chance constraint that limits the total amount observed on the slack-flow variables can then be defined as follows:

$$\mathbb{P}\left(\omega \in \Omega \mid \exists \varsigma^k(\omega) \geq 0, k \in \mathcal{K} : \quad \begin{array}{c} \sum\limits_{k \in \mathcal{K}} \varsigma^k(\omega) \geq \alpha_{\mathcal{K}}, \\ F_{O(k)}^k(x) = d^k(\omega) - \varsigma^k(\omega), \\ F_{D(k)}^k(x) = -d^k(\omega) + \varsigma^k(\omega), \end{array}\right) \leq 1 - \beta_{\mathcal{K}}. \tag{27}$$

Consequently, the chance constraint (27) limits the probability of observing realizations $\omega \in \Omega$ where the total commodity volumes that are transported via the dummy arcs exceeds the threshold $\alpha_{\mathcal{K}}$ (i.e., $\sum_{k \in \mathcal{K}} \varsigma^k(\omega) \geq \alpha_{\mathcal{K}}$) to at most $1 - \beta_{\mathcal{K}}$, where $\beta_{\mathcal{K}}$ delineates the requisite level of reliability for the planned itineraries. Such conditions could also be established on a per-commodity basis for each $k \in \mathcal{K}$. In such a case, a commodity-specific threshold α_k, for $k \in \mathcal{K}$, would dictate the limit beyond which the flow assigned to $\varsigma^k(\omega)$ is considered as an unacceptable outcome $\omega \in \Omega$ by the carrier.

The same modelling approach can be employed to define a chance constraint that limits the total amount of significant shortfalls in the capacities of the selected services:

$$\mathbb{P}\left(\omega \in \Omega \mid \exists \varrho_{\sigma(a)}(\omega) \geq 0, \sigma(a) \in \Sigma : \begin{array}{c} \sum\limits_{\sigma(a)\in\Sigma} \varrho_{\sigma(a)}(\omega) \geq \alpha_\Sigma, \\ \sum\limits_{k\in\mathcal{K}} x_a^k \leq u_a(\omega) y_{\sigma(a)} + \varrho_{\sigma(a)}(\omega), \end{array}\right) \leq 1-\beta_\Sigma. \quad (28)$$

Here, the parameters α_Σ and β_Σ serve to, respectively, specify the maximum allowable threshold for the aggregate additional service capacity required and the desired level of reliability for the designed service network. Furthermore, similar to the previous case, the probabilistic constraint (28) can also be disaggregated to apply a specific threshold $\alpha_{\sigma(a)}$ for each service $\sigma(a) \in \Sigma$, thus providing more granular control over the service-specific reliability requirements.

Stochastic models with recourse. The alternative stochastic modelling paradigm that has been adopted to address uncertainty in SND utilizes stochastic models with recourse. These models are designed to encapsulate the decision-making dynamics that are integral to the tactical planning processes performed by carriers. In this paradigm, the stochastic formulation comprehensively incorporates both tactical decisions for designing the transportation supply and operational-level decisions for the repeated application of the tactical plan in a randomly varying environment.

It is important to emphasize that all these decisions occur within a context of randomly varying informational conditions. As outlined previously, the tactical plan is established in advance, introducing uncertainty in the decision-making process due to the lack of precise information. Conversely, operational decisions are made in a setting where the actual values of stochastic parameters become known, aiming to adaptively meet shippers' demands based on real-time information. This setting inherently encompasses multiple *stages* - specific moments when stochastic parameters become known, allowing decisions to be made in reaction to this newly acquired information. In a general sense, the tactical decisions are thus made during the first stage, often referred to as *a priori* decisions, in the face of complete uncertainty. Meanwhile, operational decisions would naturally be executed in subsequent stages, known as *recourse* decisions, as the stochastic parameters become progressively observed.

While stochastic multi-stage formulations adeptly encapsulate the dynamics of operational decision-making, they considerably augment the complexity of the optimization model, especially as the model expands with additional stages. This complexity is particularly pronounced when incorporating tactical planning phases characterized by essential discrete decisions. Furthermore, when solving tactical planning problems, detailed operational specifics may not be essential for accurately establishing tactical plans; a good approximation can often provide sufficient guidance. Consequently, the predominant modelling approach for SND, especially focusing on carrier-centric tactical planning, employs a two-stage formulation.

How each stage is defined should reflect the specific planning problem under consideration. Typically, the first stage involves design decisions that specify the services to be operated, thereby establishing the carrier's transportation supply. These decisions are made well before the actual utilization of the services, requiring them to comprehensively address the full spectrum of anticipated uncertainty.

When it comes to defining potential recourse actions in the second stage, a broad range of options can be established, offering varying degrees of flexibility in adjusting the overall plan. On one end of the spectrum is the simple recourse option, which involves imposing a penalty proportional to the extent of the plan's infeasibility. This approach essentially adopts a general *observe-and-pay* strategy for defining recourse actions. Alternatively, more complex recourse actions can be formulated at the network level. A notable strategy within this context is the establishment of commodity itineraries as second-stage decisions, effectively treating the flow decisions as recourse actions. This is the strategy taken in the illustrative model presented below and has been shown to be effective in a variety of tactical planning transportation problems. Recourse actions can also extend to structural adjustments in the service network initially designed during the first stage. However, adopting such a strategy introduces greater complexity, as it would necessitate making discrete decisions as part of the recourse actions.

In all cases, this general modelling approach presupposes that all uncertainties are clarified in a single stage, meaning all stochastic parameters are assumed to be observed simultaneously. Therefore, the second stage offers an approximation of the carrier's real operations to meet shipper demands, guided by the tactical plan established in advance. Despite its approximate nature, this stage of the formulation provides nonetheless valuable insights, estimating operational costs to significantly influence the tactical plan's design. Solving the two-stage stochastic model thus yields a comprehensive tactical plan. The plan not only outlines the setup of the service network by the carrier but also specifies cost-efficient adjustments to accommodate observed random changes.

Using again the modelling approach presented in Hewitt et al. (2021), in the stochastic model with the flow-defined recourse strategy, the variables y are designated as the first-stage design decisions, while the flow variables are set as $x_a^k(\omega)$ for $a \in \mathcal{A}$, $k \in \mathcal{K}$, and $\omega \in \Omega$. This definition indicates that the flow decisions, which also delineate the commodity itineraries, act as recourse actions undertaken in response to specific realizations of the stochastic event. Assuming the system cost is linearly defined as in equation (10), the model formulation is as follows:

$$\min \quad \sum_{\sigma \in \Sigma} f_\sigma y_\sigma + \mathbb{E}_\xi \Big[Q\big(y, \xi(\omega)\big) \Big] \tag{29}$$

$$\text{s.t.} \qquad y_\sigma \in \mathbb{Z}_+, \sigma \in \Sigma, \tag{30}$$

where,

$$Q\big(y,\xi(\omega)\big) = \min \sum_{k\in\mathcal{K}}\sum_{a\in\mathcal{A}} c_a^k(\omega)x_a^k(\omega) \tag{31}$$

$$\text{s.t.}\sum_{a\in\mathcal{A}_\eta^+} x_a^k(\omega) - \sum_{a\in\mathcal{A}_\eta^-} x_a^k(\omega) = \begin{cases} d^k(\omega), & \text{if } i = O(k), \\ -d^k(\omega), & \text{if } i = D(k), i \in \mathcal{N}, k \in \mathcal{K}, \\ 0, & \text{otherwise,} \end{cases} \tag{32}$$

$$\sum_{k\in\mathcal{K}} x_a^k(\omega) \leq u_a(\omega)y_{\sigma(a)}, a \in \mathcal{A}, \tag{33}$$

$$x_a^k(\omega) \geq 0, a \in \mathcal{A}, k \in \mathcal{K}. \tag{34}$$

Model (29) - (30) addresses the first-stage decisions, which seek to design the service network under complete uncertainty. The objective function (29) thus combines a deterministic component - the total fixed costs for the selected services - with a second term that assesses the anticipated future costs of adjusting / adapting the tactical plan, given by the design decisions made, to revealed information (i.e., the recourse cost function). This involves computing the expected value of the cost function $Q\big(y,\xi(\omega)\big)$, which evaluates the total cost incurred to move the demand flows using the service network defined by y, upon observing the stochastic parameter values associated with realization $\omega \in \Omega$, i.e., the value vector $\xi(\omega)$. Consequently, function $Q\big(y,\xi(\omega)\big)$ is characterized as the optimal solution to the minimum cost multi-commodity flow model (31)-(34) configured for the realized commodity volumes $\mathbf{d}(\omega)$, service variable costs $\mathbf{c}(\omega)$, and service capacities $\mathbf{u}(\omega)$. As highlighted earlier, in this context, the flow variables fulfill the role of recourse actions and represent the decisions made in the second stage.

In the next subsection, we explore some of the principal innovations proposed for solving stochastic scheduled service network design problems. To ensure clarity in our discussion, we directly reference the stochastic modelling paradigms previously outlined when presenting these innovations.

6.2 Stochastic SSND Methods

Methodological advancements in stochastic SSND have emerged relatively recently. By far, the most studied variant of the problem involves the case where the shipper demand volumes are assumed stochastic. One of the first contribution to this field is by Lium et al. (2007), where the authors investigate the impact of accounting for randomly fluctuating demands on the configuration of service networks when using two-stage stochastic models with recourse to solve the problems. Specifically, the study examines the effects of integrating prevailing demand correlations and provides numerical evidence that different correlation setups can markedly affect the design of service networks. Subsequent research by the same authors in Lium et al. (2009) further explores the impact of demand uncertainty on SSND. This follow-up study delineates the structural differences between networks devised via two-stage stochastic models with recourse and those from deterministic approaches, highlighting the distinct advantages of the former. Notably, this research demonstrates that optimal networks derived from stochastic formulations exhibit unique characteristics not present in networks derived from deterministic models. Specifically, when the uncertainty is explicitly considered, the designed networks tend to feature multiple alternative paths to connect commodity origin-destination node pairs, enhancing flexibility to accommodate fluctuating commodity flows. Consolidation, in the form of alternative paths sharing common arcs, is also observed as a strategy to minimize the overall costs of the solutions obtained.

While deterministic approximations might not incorporate efficiency-inducing structures in the networks they generate, it is important to recognize that deterministic models can still provide significant insights for applying SSND methods in transportation service planning amidst demand uncertainty. This perspective is supported by Wang et al. (2019), where the authors show that elements of a service network designed using a deterministic model can effectively inform decision-makers on efficiently *upgrading* the network (i.e., applying local modifications to the network's structure) for enhanced performance when used under uncertain conditions. Such comparative analyses of stochastic and deterministic optimization approaches in SSND offer a clear perspective on the decisional impacts of uncertainty within the tactical planning processes performed by carriers.

Over the years, contributions to the development of stochastic SSND methods, particularly concerning uncertain demands, have been primarily directed towards two main areas: methodological advancements, which include algorithmic developments for efficient problem-solving, and the incorporation of various carrier-centric tactical planning dimensions within the developed methods. Regarding the first area, one of the earliest attempts to introduce a meta-heuristic solution approach tailored to two-stage SSND models under stochastic demands is presented by Hoff et al. (2010). The authors propose a heuristic algorithm that is based on the principles of the variable neighbourhood search strategy. A key contribution of their work is the development of neighbourhood structures designed to identify zones within the service network where the anticipated reliance on ad-hoc capacity, which incurs significantly higher costs, is particularly pronounced. Identifying these specific network zones is crucial for the neighbourhood search meta-heuristic to then pinpoint moves that improve the current solution.

Another heuristic search method recognized as an efficient solution strategy for stochastic SSND is the progressive hedging meta-heuristic. Originally developed by Crainic et al. (2011) for general stochastic network design models, specifically those involving stochastic demands, this approach has been effectively adapted for different SSND variants under uncertainty. The method employs scenario decomposition, breaking down the two-stage stochastic formulations according to the scenarios that represent potential realizations of the uncertain demands. It then solves these scenario-specific subproblems iteratively, updating their formulations (i.e., through fixed cost adjustments) to gradually seek a consensus among all scenarios, aiming for convergence towards a unified service network solution. Enhanced versions of this solution strategy are detailed in Crainic et al. (2014a) and Jiang et al. (2021), where the introduction of scenario clustering approaches to refine the decomposition strategy (resulting in multi-scenario subproblems) is shown to significantly improve the efficiency of the consensus-seeking process to achieve a single service network.

In the second area of contributions, studies on SSND methods for the variant with stochastic demands have aimed to incorporate specific carrier tactical planning dimensions into the stochastic optimization models, thereby increasing their applicability and relevance to real-world scenarios. In Bai et al. (2014), the authors introduce a two-stage stochastic SND model where rerouting options serve as recourse actions in the second stage, offering added flexibility to manage fluctuating shipper demands. Carriers often favour rerouting strategies over dependence on ad-hoc capacity - often an outsourced solution - due to the enhanced control over the services they offer to shippers and potential for cost savings.

Actually, determining the optimal balance between rerouting or adjusting planned services and the utilization of ad-hoc capacity has emerged as a critical question in this domain over the years. It is clearly shown in Wang and Wallace (2016) that explicitly considering freight transportation spot markets when designing a service network for tactical planning in the presence of stochastic demands, where a carrier can either outsource part of the shipper demand or sell excess capacity on the planned services, can lead to notable gains in the plan's efficiency. The trade-offs between dedicating efforts to service adjustments versus relying on ad-hoc capacity to meet fluctuating demand volumes are further explored in Crainic et al. (2016), specifically within the context of planning two-tiered city logistics systems. The study introduces various two-stage stochastic optimization models for tactical planning in these systems, employing alternative recourse strategies. These strategies either utilize ad-hoc city freighters for time-dependent urban distribution operations or implement localized changes to service dispatches to efficiently execute the tactical plans. In Müller et al. (2021a), the integration of scheduled service adjustments as part of the recourse actions for accommodating demand variability is also explored within a two-stage stochastic framework for tactical planning in an intermodal transportation system. The study effectively emphasizes how employing such a strategy can significantly enhance operational flexibility and the system's overall efficiency. More recently, in Liu et al. (2023), a two-stage stochastic formulation is introduced to determine vehicle allocation decisions between 1) fixed services (routes determined in the first stage for repeated use over the planning horizon), and 2) flexible services (routes that are decided in the second stage as part of the recourse actions). This study reinforces the critical importance of finding an optimal balance between advanced planning for cost reduction and incorporating adaptability into the plans at the time of execution.

As indicated previously, the allocation and utilization of transportation resources (such as vehicles, capacity units, and drivers) form a fundamental component of the tactical planning performed by carriers. This process is also significantly influenced by the random demand variations that can be observed during operational execution. In Dong et al. (2015), the authors explore service capacity planning in a shipping network under uncertain demand conditions, introducing a two-stage stochastic model for the problem. The first stage pre-assigns specific containers to services, setting up the planned capacities, while the second stage focuses on routing these containers based on actual shipper demands as they become known. To address the complexities of this model, the authors proposed a solution algorithm designed using progressive hedging techniques.

When a carrier directly manages the resources necessary to perform the planned services, circumstances may arise that necessitate addressing uncertainties related to the transportation capacities these resources provide. In Hasany and Shafahi (2017), the authors introduce a two-stage stochastic model to tackle the railroad blocking problem - a crucial aspect of tactical planning for railroad carriers - that accounts for the variability in shipper demands, flow capacities, and total transit times of shipments. The first stage involves selecting blocks, thereby defining a set of potential paths through the rail network for the shipments that are to be determined in the subsequent stage. Given the potential variability in shipment sizes (i.e., demands), flow throughput at stations, and transit times, the model aims to identify a set of blocking paths that minimize the overall costs, encompassing both the operating costs of the selected blocks and the expected transit time costs of the shipments.

The explicit consideration of resource acquisition (such as vehicles) within the context of SSND amid demand uncertainty is addressed in Hewitt et al. (2019). Here, a two-stage framework is introduced, integrating both the acquisition (or contracting) and allocation of resources, alongside the selection of scheduled services as part of the first-stage tactical plan. Subsequently, the second stage executes these selected services to meet realized demands, with service dispatches being conducted using either owned or outsourced resources. The obtained stochastic model is then solved using a column-generation based matheuristic.

The management of resources in the context of formulating SSND under uncertainty can further enable the adoption of specific distribution strategies that leverage the characteristics of the available resources. In Scherr et al. (2022), the authors propose a two-stage stochastic formulation that develops a tactical service plan for an urban transportation system where different vehicle types are present. This plan includes fleet-sizing decisions and the selection and scheduling of services (performed using the acquired vehicles) in the first stage to address stochastic requests, which are specified in the second stage. These requests arise from a series of automated vehicles performing distribution operations within the considered urban transportation system. In this particular context, platooning services are planned, allowing a human-driven vehicle to lead a series of automated vehicles through roads not accessible to them, to service the randomly varying demands within the system.

Although the majority of contributions to stochastic SSND methods have focused on cases involving stochastic demands, there has also been notable effort directed towards exploring the implications of stochastic service times as well. This line of inquiry examines the impacts of variability in service durations on the design and operational efficiency of service networks, recognizing that such stochasticity can significantly affect the reliability and responsiveness of the transportation services provided by carriers. For instance, in Demir et al. (2016), the authors extend their investigation to include not just stochastic demands but also the uncertainty in travel times, specifically within the planning context for services operated over intermodal freight transportation systems. The proposed stochastic model integrates recourse actions and probabilistic constraints to tackle these uncertainties effectively. Recourse actions come into play when actual demands exceed forecasts, allowing for the possibility to reroute excess demands through planned services with available capacity, or to accommodate them via ad-hoc capacity, represented by road vehicles presumed to be always at hand for such situations. Moreover, to account for the possibility that random variations in travel times might hinder the timely arrival of planned services for transfer operations between transportation modes, a probabilistic constraint is applied. This constraint ensures the tactical plan's reliability by limiting the probability of observing such delays, thereby enhancing the overall robustness and efficiency of the service network. Building on this framework, Zhao et al. (2018) further explore SSND within an intermodal transportation system context, where, alongside uncertainties in travel times and demands, the variability of transfer times at terminals was also explicitly considered. The use of probabilistic constraints in this study aims to ensure that the two-stage formulation proposed yields a dependable and efficient set of transportation services, accounting for the full spectrum of operational uncertainties.

Finally, an alternative approach to manage travel time uncertainty is proposed in Lanza et al. (2021), which sets quality targets for deliveries by applying penalties for any delay observed. This method utilizes a two-stage formulation: in the first stage, services are selected and scheduled, and itineraries for commodities are established. The second stage addresses the penalties incurred during the plan execution under specific travel time realizations. By optimizing a tactical plan that minimizes both the total operational costs and the expected total penalties, this approach systematically designs service networks that are cost-efficient and effectively mitigates the negative impacts of travel time variability on delivery schedules. To address this stochastic model, the authors designed a meta-heuristic based on the progressive hedging strategy.

The use of penalties for delivery delays was similarly adopted by Müller et al. (2021b) to address stochastic transit times in the integrated planning of services and vehicle routes within an intermodal system. This approach presents a two-stage stochastic model that combines the scheduling of pick-up and delivery vehicle routes with the planning of services that enable freight to be transported between transshipment terminals across an extensive network, taking into account the variability of transit times. In the first stage, the model outlines both the services and the itineraries, which here include vehicle routes for distributing orders. The subsequent stage then focuses on determining the timing decisions, specifically the start times of the chosen services and the adjustments required for delivery schedules (i.e., possibly delays) in response to actual transit times observed. Considering the complexity of the resulting stochastic model, the authors propose an iterated local search algorithm to address it.

7 Conclusions and Perspectives

In this paper, we aimed to spotlight significant scientific contributions made over the last 50 years in the development of O.R. methodologies to address crucial freight transportation challenges. As highlighted, freight transportation activities are conducted over complex systems that can involve various transportation modes, moving loads by possibly different carriers over various distances. Our focus was specifically on methods that have been successfully applied to the tactical planning of consolidation-based freight transportation systems, with an emphasis on carrier-centric problem settings. Consolidation-based carriers design the transportation service supply to meet the demand expressed by their customers, i.e., the shippers requiring freight to be transported between various locations. To ensure efficiency and profitability, the planning of such systems necessitates dedicated O.R. methods to support planning at all decision levels, tactical, short-to-medium term planning, in particular. Our overview of the field clearly shows that Scheduled Service Network Design is an essential O.R. methodology in assisting to conduct the necessary planning.

As our review indicates, contributions to SSND have evolved, developing efficient tools to address several critical aspects of the tactical planning performed by carriers. These aspects include: 1) managing time requirements affecting both commodity itineraries and transportation services, such as delays and schedules, for which efficient methods were proposed to integrate time dimensions into the optimization models developed; 2) managing resources that support carrier services (e.g., vehicles, drivers, capacity units), where additional combinatorial characteristics were added to the SSND models to properly account for the allocation and use of resources; and 3) comprehensively understanding and addressing the impacts of various sources of uncertainty on the planning processes, illustrating that deterministic approximations cannot substitute for specialized models that explicitly account for uncertainty to produce efficient plans. The methodological development in SSND has matured into a field that exemplifies how innovations in discrete optimization can be effectively applied to tackle a broad range of practical transportation planning challenges. Looking towards the future, we see new and interesting directions emerging that, in our view, represent the natural next steps in this field of scientific inquiry.

As an overarching trend, transportation markets have become significantly competitive, with large carriers exerting ever-increasing pressure on smaller ones by covering larger market segments and maintaining low fares. Simultaneously, governments worldwide are enforcing stricter limits on greenhouse gas (GHG) emissions to mitigate environmental impacts, directly influencing the transportation sector. This new regulatory environment is prompting carriers to revise their overall plans and to adopt more environmentally friendly transportation modes. All in all, these pressures are significantly shaping the evolution and future direction of freight transportation systems.

Broadly, there is an undeniable trend toward strategic partnerships among carriers of all sizes as a means to enhance overall profitability. Additionally, general concepts such as the Physical Internet, Synchromodality, and City Logistics are increasingly being promoted as strategies for improving efficiency in the planning and operation of freight transportation systems. These concepts all seek to leverage economic opportunities through the mutualization of services and resources among partner organizations to produce more integrated and interconnected transportation systems. They also advocate for the use of specialized containers, along with common protocols, interfaces, and data formats, to streamline operations across systems for all stakeholders involved, including both carriers and shippers. The advantages of such standardization are to enable easy integration (enhancing consolidation efforts and synchronization of services), as well as to improve and facilitate handling, transfer, and storage operations. Overall, these emerging systems emphasize the optimization of the system's performance to achieve higher levels of financial efficiency for all stakeholders (maximizing profits for carriers and minimizing costs for shippers) and to promote environmental sustainability (reducing waste and minimizing the environmental footprint of transportation operations).

With such initiatives being advanced, new classes of SSND problems are emerging and undergoing study. In particular, Multi-Stakeholder and Multi-Layered SSND problems, which involve planning services across interconnected and layered systems, are gaining attention. These problems typically include multiple stakeholders (e.g., carriers) who mutualize and coordinate their resources to support and execute the services that constitute the overall transportation supply. Addressing these challenges successfully requires further significant contributions to be made towards: defining accurate and meaningful optimization models that properly formulate the complex stakeholder interactions and transportation operations occurring within these new complex systems and developing innovative solution methodologies.

Firstly, planning services over multi-layered systems presents significant discrete optimization challenges. Such systems often encompass diverse transportation operations being conducted over varying distances, raging from intercontinental to urban last-mile delivery, and incorporate different transportation modes. Modern modes now often include environmentally friendly options such as cargo bikes, autonomous electric vehicles, and even drones (which are being considered more and more to perform distribution operations). Each layer and mode thus introduces unique complexities that must be addressed effectively. Furthermore, the possible use of varied infrastructure and the inter-layer connectivity requirements, which govern the timing (i.e., synchronizing services conducted across multiple layers) and capacity (i.e., defining the service capabilities of each layer based on its supporting layers), contribute to the complexity of these problems (combining location decisions, capacity planning and network design in a single optimization model). This complexity results in significantly challenging discrete optimization models that are non-linear, large-scale, and involve elements of uncertainty to be solved.

Secondly, it is crucial to accurately define the interactions and collaborations among stakeholders within these emerging systems. Additionally, since carriers operate as commercial entities, these collaborative systems must ensure profitability for each participating organization. To maintain efficiency and foster sustainable relationships, it is necessary to effectively balance competitive and cooperative dynamics, ensuring that all parties find the arrangements mutually beneficial. Toward this end, integrating *Revenue Management* considerations into the planning processes of such transportation systems is essential. In this context, the transportation supply of the system is defined through various service options, differentiated by price and service quality (e.g., time delivery guarantees), catering to different shipper demand segments interested in these options. Such considerations add further complexity to the discrete optimization models used for the system's planning.

Thirdly, the increased standardization in the new transportation systems also brings new combinatorial considerations to be included in the optimization models that are developed to solve carrier-centric tactical planning problems. Specifically, the use of modular specialized containers, which are designed for enhanced consolidation and simplified handling and transfer operations, naturally lead towards explicitly considering *Bin Packing* dimensions when solving SSND problems. Such dimensions not only directly bring more operational considerations in the tactical planning settings performed by carriers, they also lead towards the study of new discrete optimization model variants, which seek to design networks while imposing bin packing requirements on the network's flow.

Finally, in light of all the work accomplished to advance SSND methodology, there is still a lot of work that remains to be done to advance modelling and solution methods, and to use them in the broader context of decision support systems for practitioners. Towards this end, there are multiple questions that remain open. On top of the list there is how predictive analytics can be efficiently combined with SSND methods for better organizational integration? SSND methods, which are of the prescriptive type, to be effectively used by carriers require integration with the informational contexts in which the companies evolve (such contexts are directly used to quantify the parameters of SSND models, be them deterministic or stochastic). At the same time, such methods, to be used in practical settings, also need to be integrated in the broader decisional and organizational processes of carriers. The latter involving, for example, how tactical planning is best integrated with the operational processes (and even real-time decision-making) that are conducted by carriers to perform their transportation activities. Hence, the Service Network Design field of research will remain quite vibrant in the future.

Acknowledgments

While working on the project, the first author held the UQAM Chair on Intelligent Logistics and Transportation Systems Planning, and was Adjunct Professor, Department of Computer Science and Operations Research, Université de Montréal, while the second held the Canada Research Chair in Stochastic Optimization of Transport and Logistics Systems. We gratefully acknowledge the financial support provided by the Natural Sciences and Engineering Council of Canada (NSERC), through its Discovery grant program. We also gratefully acknowledge the support of Fonds de recherche du Québec through their infrastructure grants. Finally, during the preparation of this work the authors used ChatGPT-4, under its most strict privacy enforcing settings, for text edit purposes exclusively. After using this tool, the authors reviewed and further edited the text as needed and take full responsibility for the content of the publication.

References

A. Aguezzoul. Third-party logistics selection problem: A literature review on criteria and methods. *Omega*, 49:69–78, 2014.

R.K. Ahuja, C.B. Cunha, and G. Şahin. Network Models in Railroad Planning and Scheduling. In *Tutorials in Operations Research INFORMS 2005*, pages 54–101. INFORMS, 2005a. Published online: 14 Oct 2014; http://dx.doi.org/10.1287/educ.1053.0013.

R.K. Ahuja, J. Liu, J.B. Orlin, and L.A. Sharma, D.and Shughart. Solving Real-Life Locomotive-Scheduling Problems. *Transportation Science*, 39(4):503–517, 2005b.

R.K. Ahuja, K.C. Jha, and J. Liu. Solving Real-Life Railroad Blocking Problems. *Interfaces*, 37:404–419, 2007.

S. Albinski, T.G. Crainic, and S. Minner. The Day-before Truck Platooning Planning Problem and the Value of Autonomous Driving. Technical Report CIRRELT-2020-04, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et les transports, Université de Montréal, Montréal, QC, Canada, 2020.

S. Alumur and C. Kara. Network Hub Location Problems: The State of The Art. *European Journal of Operational Research*, 190(1):1–21, 2008.

S. Alumur, J.F. Cambell, Y. Contreras, B.Y. Kara, V. Marianov, and M.E. O'Kelly. Perspectives on Modeling Hub Location Problems. *European Journal of Operational Research*, 291(1):1–17, 2021.

T.K. Ambra, A. Caris, and C. Macharis. Towards Freight Transport System Unification: Reviewing and Combining the Advancements in the Physical Internet and Synchromodal Transport Research. *International Journal of Production Research*, 57(6): 1606–1623, 2019.

T.K. Ambra, C. Mommens, and C. Macharis. Intermodal and Synchromodal Freight Transport. In Vickerman, R. et al., editor, *International Encyclopedia of Transportation*, volume 3, pages 456–462. Elsevier U.K., 2021.

P. Ammann, S. Albinski, T.G. Crainic, and R. Kolisch. Joint Truck and Driver Routing and Scheduling for Semi-Autonomous Platoons. Technical Report CIRRELT, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et les transports, Université de Montréal, Montréal, QC, Canada, 2024.

J. Andersen, T.G. Crainic, and M. Christiansen. Service Network Design with Management and Coordination of Multiple Fleets. *European Journal of Operational Research*, 193(2):377–389, 2009a.

J. Andersen, T.G. Crainic, and M. Christiansen. Service Network Design with Asset Management: Formulations and Comparative Analyzes. *Transportation Research Part C: Emerging Technologies*, 17(2):197–207, 2009b.

J. Andersen, M. Christiansen, T.G. Crainic, and R. Grønhaug. Branch-and-Price for Service Network Design with Asset Management Constraints. *Transportation Science*, 46(1):33–49, 2011.

C. Archetti, L. Peirano, and M.G. Speranza. Optimization in Multimodal Freight Transportation Problems: A Survey. *European Journal of Operational Research*, 299:1–20, 2022.

A.P. Armacost, C. Barnhart, and K.A. Ware. Composite Variable Formulations for Express Shipment Service Network Design. *Transportation Science*, 36(1):1–20, 2002.

A.A. Assad. Modelling of Rail Networks: Toward a Routing/Makeup Model. *Transportation Research Part B: Methodological*, 14:101–114, 1980a.

A.A. Assad. Models for Rail Transportation. *Transportation Research Part A: Policy and Practice*, 14:205–220, 1980b.

T. Aykin. The Hub Location and Routing Problem. *European Journal of Operational Research*, 83(1):200–219, 1995a.

T. Aykin. Networking Policies for Hub-and-Spoke Systems with Application to the Air Transportation System. *Transportation Science*, 29(3):201–221, 1995b.

R. Bai, S. W. Wallace, J. Li, and A. Y.-L. Chong. Stochastic Service Network Design with Rerouting. *Transportation Research Part B: Methodological*, 60:50–65, 2014.

I. Bakir, A. Erera, and M.W.F. Savelsbergh. Motor Carrier Service Network Design. In T.G. Crainic, M. Gendreau, and B. Gendron, editors, *Network Design with Applications in Transportation and Logistics*, chapter 16, pages 427–467. Springer, Boston, 2021.

A. Balakrishnan, T.L. Magnanti, and P. Mirchandani. Network Design. In M. Dell'Amico, F. Maffioli, and S. Martello, editors, *Annotated Bibliographies in Combinatorial Optimization*, pages 311–334. John Wiley & Sons, New York, NY, 1997.

A. Balakrishnan, A. Kuo, and X. Si. Real-Time Decision Support for Crew Assignment in Double-Ended Districts for U.S. Freight Railways. *Transportation Science*, 50(4): 1139–1393, 2016.

E. Ballot, B. Montreuil, and R. D. Meller. *The Physical Internet: The Network of Logistics Networks*. La Documentation Française, Paris, France, 2014.

C. Barnhart, E.L. Johnson, G.L. Nemhauser, M.W.F. Savelsbergh, and P.H. Vance. Branch-and-Price: Column Generation for Solving Huge Integer Programs. *Operations Research*, 46(3):316–329, 1998.

C. Barnhart, H. Jin, and P.H. Vance. Railroad blocking: A network design application. *Operations Research*, 48(4):603–614, 2000.

C. Barnhart, N. Krishnan, D. Kim, and K. Ware. Network Design for Express Shipment Delivery. *Computational Optimization and Applications*, 21(3):239–262, 2002.

T. Bektaş and T.G. Crainic. A Brief Overview of Intermodal Transportation. In Taylor, G.D., editor, *Logistics Engineering Handbook*, chapter 28, pages 1–16. Taylor and Francis Group, Boca Raton, FL, USA, 2008.

T. Bektaş, M. Chouman, and T.G. Crainic. Lagrangean-based Decomposition Algorithms for Multicommodity Network Design with Penalized Constraints. *Networks*, 55(3):272–280, 2010.

T. Bektaş, T.G. Crainic, and T. Van Woensel. From managing urban freight to smart city logistics networks. In K. Gakis and P. Pardalos, editors, *Networks Design and Optimization for Smart Cities*, volume 8 of *Series on Computers and Operations Research*, pages 143–188. World Scientific Publishing, 2017.

I.C. Bilegan, T.G. Crainic, and Y. Wang. Scheduled Service Network Design with Revenue Management Considerations and an Intermodal Barge Transportation Illustration. *European Journal of Operational Research*, 300(1):164–177, 2022.

LD. Bodin, B.L. Golden, A.D. Schuster, and W. Romig. A Model for the Blocking of Trains. *Transportation Research Part B: Methodological*, 14(1):115–120, 1980.

N. Boland, M. Hewitt, L. Marshall, and M.W.F. Savelsbergh. The Continuous-Time Service Network Design Problem. *Operations Research*, 65(5):1303–1321, 2017.

N. Boland, M. Hewitt, L. Marshall, and M.W.F. Savelsbergh. The Price of Discretizing Time: A Study in Service Network Design. *EURO Journal on Transportation and Logistics*, 8(2):195–216, 2019.

A.M. Bomberault and W.W. White. Scheduling Empty Box Cars. Technical report, IBM New York Scientific Center, Hawthorne, N.Y., 1966.

B. Bouzaïene-Ayari, C. Cheng, S. Das, R. Fiorillo, and W.B. Powell. From Single Commodity to Multiattribute Models for Locomotive Optimization: A Comparison of Optimal Integer Programming and Approximate Dynamic Programming. *Transportation Science*, 50(2):366–389, 2016.

M.E. Bruni, T.G. Crainic, and G. Perboli. Bin packing methodologies for capacity planning in freight transportation and logistics. In T.G. Crainic, M. Gendreau, and F. A., editors, *Contributions to Combinatorial Optimization and Applications - A Tribute to Bernard Gendron*, chapter 6. Springer, 2024. forthcoming.

J.F. Cambell. A Survey of Network Hub Location Problems. *Studies in Locational Analysis*, 6:31–49, 1994b.

J.F. Campbell, A.T. Ernst, and M. Krishnamoorthy. Hub Location Problems. In Drezner, Z. and Hamacher, H., editors, *Facility Location: Application and Theory*, pages 373–407. Springer Verlag, Berlin, 2002.

Z.-S. Chen, X. Zhang, K. Govindan, X.-J. Wang, and K.-S. Chin. Third-party reverse logisticspprovider selection: A computational semantic analysis-based multi-perspective multi-attribute decision-making approach. *Expert Systems with Applications*, 166: 114051, 2021.

M. Chouman and T.G. Crainic. Cutting-Plane Matheuristic for Service Network Design with Design-Balanced Requirements. *Transportation Science*, 49(1):99–113, 2015.

M. Chouman and T.G. Crainic. Freight Railroad Service Network Design. In T.G. Crainic, M. Gendreau, and B. Gendron, editors, *Network Design with Applications in Transportation and Logistics*, chapter 13, pages 383–426. Springer, Boston, 2021.

M. Christiansen, K. Fagerholt, and D. Ronen. Ship Routing and Scheduling: Status and Perspectives. *Transportation Science*, 38(1):1–18, 2004.

M. Christiansen, K. Fagerholt, B. Nygreen, and D. Ronen. Maritime Transportation. In Barnhart, C. and Laporte, G., editors, *Transportation*, volume 14 of *Handbooks in Operations Research and Management Science*, pages 189–284. North-Holland, Amsterdam, 2007.

M. Christiansen, E. Helsen, D. Pisinger, D. Sacramento, and C. Vilhelmsen. Liner Shipping Network Design. *European Journal of Operational Research*, 286(1):1–20, 2020.

M. Christiansen, E. Helsen, D. Pisinger, D. Sacramento, and C. Vilhelmsen. Liner Shipping Network Design. In T.G. Crainic, M. Gendreau, and B. Gendron, editors, *Network Design with Applications in Transportation and Logistics*, chapter 15, pages 469–505. Springer, Boston, 2021.

I. Contreras. Hub Network Design. In T.G. Crainic, M. Gendreau, and B. Gendron, editors, *Network Design with Applications in Transportation and Logistics*, chapter 18, pages 567–598. Springer, Boston, 2021.

J.-F. Cordeau, G. Laporte, M.W.F. Savelsbergh, and D. Vigo. Vehicle Routing. In Barnhart, C. and Laporte, G., editors, *Transportation*, Handbooks in Operations Research and Management Science, pages 367–428. North-Holland, Amsterdam, 2007.

J.F. Cordeau, P. Toth, and D. Vigo. A Survey of Optimization Models for Train Routing and Scheduling. *Transportation Science*, 32(4):380–404, 1998.

T.G. Crainic. Rail Tactical Planning: Issues, Models and Tools. In L. Bianco and A. La Bella, editors, *Freight Transport Planning and Logistics*, pages 463–509. Springer-Verlag, Berlin, 1988.

T.G. Crainic. Network Design in Freight Transportation. *European Journal of Operational Research*, 122(2):272–288, 2000.

T.G. Crainic. Long-Haul Freight Transportation. In Hall, R.W., editor, *Handbook of Transportation Science*, pages 451–516. Kluwer Academic Publishers, Norwell, MA, second edition, 2003.

T.G. Crainic. Service Design Models for Rail Intermodal Transportation. In Bertazzi, L., Speranza, M.G., and van Nunen, J.A.E.E., editors, *Lecture Notes in Economics and Mathematical Systems*, volume 619, pages 53–67. Springer-Verlag, Berlin Heidelberg, 2009.

T.G. Crainic. Multi-layer network design for consolidation-based transportation planning. In T.G. Crainic, M. Gendreau, and F. A., editors, *Contributions to Combinatorial Optimization and Applications - A Tribute to Bernard Gendron*. Springer, 2024a. forthcoming.

T.G. Crainic. Consolidation-based Transportation Planning: The Service Network Design Methodology. In S. Parragh and T. Van Woensel, editors, *Research Handbook on Transport Modeling*. Edgar Elgar Publishing, 2024b. forthcoming.

T.G. Crainic and M. Florian. National Planning Models and Instruments. *INFOR*, 46 (4):81–90, 2008.

T.G. Crainic and M. Gendreau. Approximate Formulas for the Computation of Connection Delays under Capacity Restrictions in Rail Freight Transportation. In *Research for Tomorrow's Transport Requirements*, pages 1142–1155. Fourth World Conference on Transport Research, Vancouver, Canada, 1986. Vol. 2.

T.G. Crainic and M. Hewitt. Service Network Design. In T.G. Crainic, M. Gendreau, and B. Gendron, editors, *Network Design with Applications in Transportation and Logistics*, chapter 12, pages 347–382. Springer, Boston, 2021.

T.G. Crainic and K.H. Kim. Intermodal Transportation. In Barnhart, C. and Laporte, G., editors, *Transportation*, volume 14 of *Handbooks in Operations Research and Management Science*, chapter 8, pages 467–537. North-Holland, Amsterdam, 2007.

T.G. Crainic and G. Laporte. Planning Models for Freight Transportation. *European Journal of Operational Research*, 97(3):409–438, 1997.

T.G. Crainic and B. Montreuil. Physical Internet Enabled Hyperconnected City Logistics. In E. Taniguchi and R. G. Thompson, editors, *9th International Conference on City Logistics, June 2015, Tenerife, Spain*, volume 12 of *Transportation Research Procedia*, pages 383–398. Elsevier, 2016.

T.G. Crainic and L. Quérin. Approximation of Over-the-line Delay for Single Line - Single Speed Freight Rail Transportation. Publication 632, Centre de recherche sur les transports, (CRT) Université de Montréal, Montreal, Canada, 1988.

T.G. Crainic and J.-M. Rousseau. Multicommodity, Multimode Freight Transportation: A General Modeling and Algorithmic Framework for the Service Network Design Problem. *Transportation Research Part B: Methodological*, 20:225–242, 1986.

T.G. Crainic and J. Roy. O.R. Tools for Tactical Freight Transportation Planning. *European Journal of Operational Research*, 33(3):290–297, 1988.

T.G. Crainic and J. Roy. Design of Regular Intercity Driver Routes for the LTL Motor Carrier Industry. *Transportation Science*, 26(4):280–295, 1992.

T.G. Crainic, J.-A. Ferland, and J.-M. Rousseau. A Tactical Planning Model for Rail Freight Transportation. *Transportation Science*, 18(2):165–184, 1984.

T.G. Crainic, M. Florian, J. Guélat, and H. Spiess. Strategic Planning of Freight Transportation: STAN, An Interactive-Graphic System. *Transportation Research Record*, 1283:97–124, 1990a.

T.G. Crainic, M. Florian, and J.-E. Léal. A Model for the Strategic Planning of National Freight Transportation by Rail. *Transportation Science*, 24(1):1–24, 1990b.

T.G. Crainic, M. Gendreau, and P.J. Dejax. Dynamic Stochastic Models for the Allocation of Empty Containers. *Operations Research*, 41(1):102–126, 1993.

T.G. Crainic, N. Ricciardi, and G. Storchi. Models for Evaluating and Planning City Logistics Transportation Systems. *Transportation Science*, 43(4):432–454, 2009.

T.G. Crainic, X. Fu, M. Gendreau, W. Rei, and S.W. Wallace. Progressive Hedging-based Meta-heuristics for Stochastic Network Design. *Networks*, 58(2):114–124, 2011.

T.G. Crainic, M. Hewitt, and W. Rei. Scenario Grouping in a Progressive Hedging-based Meta-heuristics for Stochastic Network Design. *Computers & Operations Research*, 43 (2):90–99, 2014a.

T.G. Crainic, M. Hewitt, M. Toulouse, and D.M. Vu. Service Network Design with Resource Constraints. *Transportation Science*, 50(4):1380–1393, 2014b.

T.G. Crainic, F. Errico, W. Rei, and N. Ricciardi. Modeling Demand Uncertainty in Two-Tier City Logistics Tactical Planning. *Transportation Science*, 50(2):559–578, 2016.

T.G. Crainic, M. Hewitt, M. Toulouse, and D.M. Vu. Scheduled Service Network Design with Resource Acquisition and Management. *EURO Journal on Transportation and Logistics*, 7(3):277–309, 2018.

T.G. Crainic, M. Gendreau, and B. Gendron. Fixed-Charge Network Design Problems. In T.G. Crainic, M. Gendreau, and B. Gendron, editors, *Network Design with Applications in Transportation and Logistics*, chapter 2, pages 15–28. Springer, Boston, 2021a.

T.G. Crainic, M. Gendreau, and B. Gendron, editors. *Network Design with Applications in Transportation and Logistics*. Springer, Boston, 2021b.

T.G. Crainic, G. Perboli, and N. Ricciardi. City Logistics. In T.G. Crainic, M. Gendreau, and B. Gendron, editors, *Network Design with Applications in Transportation and Logistics*, chapter 16, pages 507–537. Springer, Boston, 2021c.

T.G. Crainic, B. Gendron, and M. Kazemzadeh. A Taxonomy of Multilayer Network Design and a Survey of Transportation and Telecommunication Applications. *European Journal of Operational Research*, 303(1):161–179, 2022.

T.G.. Crainic, K. W., and B. Montreuil. Hyperconnected city logistics: A conceptual framework. In E. Marcucci, V. Gatta, and M. Le Pira, editors, *Handbook on City Logistics and Urban Freight*, chapter 20, pages 398–421. Edward Elgar, Northampton, MA, U.S.A.,, 2023.

T.G.. Crainic, J. Gonzalez-Feliu, N. Ricciardi, F. Semet, and T. Van Woensel. Operations research for planning and managing city logistics systems. In E. Marcucci, V. Gatta, and M. Le Pira, editors, *Handbook on City Logistics and Urban Freight*, chapter 10, pages 190–223. Edward Elgar, Northampton, MA, U.S.A., 2024.

K.L. Croxton, B. Gendron, and T.L. Magnanti. Models and Methods for Merge-in-Transit Operations. *Transportation Science*, 37(1):1–22, 2003a.

K.L. Croxton, B. Gendron, and T.L. Magnanti. A Comparison of Mixed-integer Programming Models for Nonconvex Piecewise Linear Cost Minimization Problems. *Management Science*, 49(9):1268—1273, 2003b.

K.L. Croxton, B. Gendron, and T.L. Magnanti. Variable Disaggregation in Network Flow Problems with Picewise Linear Costs. *Operations Research*, 55(1):146–157, 2007.

P.J. Dejax and T.G. Crainic. A Review of Empty Flows and Fleet Management Models in Freight Transportation. *Transportation Science*, 21(4):227–247, 1987.

E. Demir, W. Burgholzer, M. Hrušovský, E. Arikan, W. Jammernegg, and T. V. Woensel. A Green Intermodal Service Network Design Problem with Travel Time Uncertainty. *Transportation Research Part B: Methodological*, 93:789–807, 2016.

J.-X. Dong, C.-Y. Lee, and D.-P. Song. Joint Service apacity Planning and Dynamic Container Routing in Shipping Network with Uncertain Demands. *Transportation Research Part B: Methodological*, 78:404–421, 2015.

É. Dufour, G. Laporte, J. Paquette, and M.-È. Rancourt. Logistics service network design for humanitarian response in east africa. *Omega*, 74:1–14, 2018.

A. Erera, M. Hewitt, S. M,.W.F., and Y. Zhang. Improved Load Plan Design Through Integer Programming Based Local Search. *Transportation Science*, 47(3):412–427, 2013a.

A,.L. Erera, M. Hewitt, S. M,.W.F., and Y. Zhang. Creating Schedules and Computing Operating Costs for LTL Load Plans. *Computers and Operations Research*, 40(3): 691–702, 2013b.

D. Escobar-Vargas and T.G. Crainic. Multi-attribute Two-Echelon Location Routing: Formulations and Dynbamic Discretization Discovery Approach. *European Journal of Operational Research*, 314(1):66–78, 2024.

M.E. Farahani, M. Hekmatfar, A.B. Arabani, and E. Nikbakhsh. Hub Location Problems: A Review of Models, Classification, Solution Techniques, and Applications. *Computers & Industrial Engineering*, 64(4):1096–1109, 2013.

J.M. Farvolden and W.B. Powell. Subgradient Methods for the Service Network Design Problem. *Transportation Science*, 28(3):256–272, 1994.

B. Feng, Y. Li, and Z.-J. Shen. Air cargo operations: Literature review and comparison with practices. *Transportation Research Part C: Emerging Technologies*, 56:263–280, 2015.

M. Florian, G. Bushell, J. Ferland, G. Guertin, and L. Nastansky. The Engine Scheduling Problem in a Railway Network. *INFOR*, 14:121–138, 1976.

P. Fontaine, T.G. Crainic, O. Jabali, and W. Rei. Scheduled Service Network Design with Resource Management for Two-tier Multimodal City Logistics. *European Journal of Operational Research*, 294(2):558–570, 2021.

L.R. Ford and D.R. Fulkerson. Constructing Maximal Dynamic Flows from Static Flows. *Operations Research*, 6(3):419–433, 1958.

R. Giusti, D. Manerba, and R. Tadei. Synchromodal Logistics: An Overview of Critical Success Factors, Enabling Technologies, and Open Research Issues. *Transportation Research Part E: Logistics and Transportation Review*, 129:92–110, 2019.

G.A. Godfrey and W.B. Powell. An Adaptive Dynamic Programming Algorithm for Dynamic Fleet Management I: Single Period Travel Times. *Transportation Science*, 36 (1):21–39, 2002a.

G.A. Godfrey and W.B. Powell. An Adaptive Dynamic Programming Algorithm for Dynamic Fleet Management II: Multiperiod Travel Times. *Transportation Science*, 36 (1):40–54, 2002b.

B. L. Golden, S. Raghavan, and E.A. Wasil, editors. *The Vehicle Routing Problem: Latest Advances and New Challenges*. Springer, New York, 2008.

J. Guélat, M. Florian, and T.G. Crainic. A Multimode Multiproduct Network Assignment Model for Strategic Planning of Freight Flows. *Transportation Science*, 24(1):25–39, 1990.

A.E. Haghani. Formulation and Solution of Combined Train Routing and Makeup, and Empty Car Distribution Model. *Transportation Research Part B: Methodological*, 23 (6):433–452, 1989.

R.M. Hasany and Y. Shafahi. Two-stage Stochastic Programming for the Railroad Blocking Problem with Uncertain Demand and Supply Resources. *Computers & Industrial Engineering*, 106:275–286, 2017.

Y. He, M. Hewitt, F. Lehuédé, N. Medina, and O. Péton. A Continuous-time Service Network Design and Vehicle Routing Problem. *Networks*, 83(2):300–323, 2023.

M. Hewitt. Enhanced Dynamic Discretization Discovery for the Continuous-Time Load Plan Design Problem. *Transportation Science*, 53(6):1731–1750, 2019.

M. Hewitt, T.G. Crainic, M. Nowak, and W. Rei. Scheduled Service Network Design with Resource Acquisition and Management under Uncertainty. *Transportation Research Part B: Methodological*, 128:324–343, 2019.

M. Hewitt, W. Rei, and S.W. Wallace. Stochastic Network Design. In T.G. Crainic, M. Gendreau, and B. Gendron, editors, *Network Design with Applications in Transportation and Logistics*, chapter 10, pages 283–315. Springer, Boston, 2021.

A. Hoff, A.-G. Lium, A. Lokketangen, and T.G. Crainic. A Metaheuristic for Stochastic Service Network Design. *Journal of Heuristics*, 16(1):653–679, 2010.

A.I. Jarrah, E. Johnson, and u. Neubert. Large-scale, Less-than-Truckload Service Network Design. *Operations Research*, 57(3):609–625, 2009.

K.C. Jha, R.K. Ahuja, and G. Şahin. New Approaches for Solving the Block-to-Train Assignment Problem. *Networks*, 51(1):48–62, 2008.

X. Jiang, R. Bai, S. W. Wallace, G. Kendall, and D. Landa-Silva. Soft Clustering-based Scenario Bundling for a Progressive Hedging Heuristic in Stochastic Service Network Design. *Computers and Operations Research*, 128:105182, 2021.

M. Joborn, T.G. Crainic, M. Gendreau, K. Holmberg, and J.T. Lundgren. Economies of Scale in Empty Freight Car Distribution in Scheduled Railways. *Transportation Science*, 38(2):459–464, 2004.

W.C. Jordan and M.A. Turnquist. A Stochastic Dynamic Network Model for Railroad Car Distribution. *Transportation Science*, 17:123–145, 1983.

J. Kienzle, T.G. Crainic, E. Frejinger, and S. Bisaillon. The Intermodal Railroad Blocking and Railcar Fleet Management Planning Problem. Technical Report CIRRELT-2023, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et les transports, Université de Montréal, Montréal, QC, Canada, 2024.

D. Kim, C. Barnhart, K. Ware, and G. Reinhardt. Multimodal Express Package Delivery: A Service Network Design Application. *Transportation Science*, 33(4):391–407, 1999.

W. Klibi, A. Martel, and A. Guitouni. The Design of Robust Value-Creating Supply Chain Networks: A Critical Review. *European Journal of Operational Research*, 203 (2):283–293, 2010.

M. Lai, T.G. Crainic, M. Di Francesco, and P. Zuddas. Heuristic Search for the Routing of Heterogeneous Trucks with Single and Double Container Loads. *Transportation Research Part E: Logistics and Transportation*, 56:108–118, 2013.

M.F. Lai and H.K. Lo. Ferry Service Network Design: Optimal Fleet Size, Routing and Scheduling. *Transportation Research Part A: Policy and Practice*, 38:305–328, 2004.

J. Lange, T.G. Crainic, N. Gschwind, and W. Rei. An Integrated Planning Approach for a Two-tier City Logistics System with Satellite Synchronization. Publication CIRRELT, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport, Université de Montréal, Montréal, QC, Canada, 2024.

G. Lanza, T.G. Crainic, W. Rei, and N. Ricciardi. Service Network Design Problem with Quality Targets and Stochastic Travel Times. *European Journal of Operational Research*, 288(1):30–46, 2021.

G. Laporte. Fifty Years of Vehicle Routing. *Transportationm Science*, 43(4):408–416, 2009.

K. Lindsey, A. Erera, and M.W.F. Savelsbergh. Improved Integer Programming-based Neighborhood Search for Less-than-Truckload Load Plan Design. *Transportation Science*, 50(4):1360–1379, 2016.

C. Liu, S. Lin, Z.-J. M. Shen, and J. Zhang. Stochastic Service Network Design: The Value of Fixed Routes. *Transportation Research Part E: Logistics and Transportation*, 174:103118, 2023.

A.-G. Lium, T.G. Crainic, and S.W. Wallace. Correlations in Stochastic Programming: A Case from Stochastic Service Network Design. *Asia-Pacific Journal of Operational Research*, 24(2):1–13, 2007.

A.-G. Lium, T.G. Crainic, and S.W. Wallace. A Study of Demand Stochasticity in Service Network Design. *Transportation Science*, 43(2):144–157, 2009.

C. Macharis and Y.M. Bontekoning. Opportunities for OR in Intermodal Freight Transport Research: A Review. *European Journal of Operational Research*, 153(2):400–416, 2004.

T.L. Magnanti and R.T. Wong. Network Design and Transportation Planning: Models and Algorithms. *Transportation Science*, 18(1):1–55, 1984.

E. Marcucci, V. Gatta, and M. Le Pira, editors. *Handbook on City Logistics and Urban Freight*. Edward Elgar, Northampton, MA, U.S.A., 2024.

A. Marín and J. Salmerón. Tactical planning of Rail Freight Networks. Part I: Exact and heuristic Methods. *European Journal of Operational Research*, 90:26–44, 1996a.

A. Marín and J. Salmerón. Tactical planning of Rail Freight Networks. Part II:Local Search Methods with Statistical Analysis. *European Journal of Operational Research*, 94:43–53, 1996b.

L. Marshall, N. Boland, M.W.F. Savelsbergh, and M. Hewitt. Interval-based Dynamic Discretization Discovery for Solving the Continuous-Time Service Network Design Problem. *Transportation Science*, 55(1):29–51, 2021.

D.R. Martinelli and H. Teng. Optimization of Railway Operations Using Neural Networks. *Transportation Research Part C: Emerging Technologies*, 4C(1):33–49, 1996.

N. Medina, M. Hewitt, F. Lehuédé, and O. Péton. Integrating Long-haul and Local Transportation Planning: Service Network Design and Routing Problem. *EURO Journal on Transportation and Logistics*, 8(2):119–145, 2019.

P. Miranda, J.-F. Cordeau, and E. Frejinger. A Time-Space Formulation for the Locomotive Routing Problem at the Canadian National Railways. Technical Report CIRRELT-2020-19, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et les transports, Université de Montréal, Montréal, QC, Canada, 2020.

B. Montreuil. Towards a Physical Internet: Meeting the Global Logistics Sustainability Grand Challenge. *Logistics Research*, 3(2-3):71–87, 2011.

A. Mor and M.G. Speranza. Vehicle Routing Problems over Time: A Survey. *Annals of Operations Research*, 314:255–275, 2022.

G. Morganti, T.G. Crainic, E. Frejinger, and N. Ricciardi. Block Planning for Intermodal Rail: Methodology and Case Study. *Transportation Research Procedia*, 47:19–26, 2020.

J. P. Müller, R. Elbert, and S. Emde. Intermodal Service Network Design with Stochastic Demand and Short-term Schedule Modifications. *Computers & Industrial Engineering*, 159:107514, 2021a.

J. P. Müller, R. Elbert, and S. Emde. Integrating vehicle routing into intermodal service network design with stochastic transit times. *EURO Journal on Transportation and Logistics*, 10:100046, 2021b.

A. M. Newman, L. K. Nozick, and C. A. Yano. Optimization in the rail industry. In P. M. Pardalos and M. G. C. Resende, editors, *Handbook of Applied Optimization*, pages 704–718. Oxford University Press, New York, NY, 2002.

H. N. Newton, C. Barnhart, and P.H. Vance. Constructing railroad blocking plans to minimize handling costs. *Transportation Science*, 32(4):330–345, 1998.

M.E. O'Kelly. The Location of Interacting Hub Facilities. *Transportation Science*, 20: 92–106, 1986.

M.E. O'Kelly and H.M. Miller. The Hub Design Problems: A Review and Synthesis. *The Journal of Transport Geography*, 2:31–40, 1994.

. Ortiz-Astorquiza, J.-F. Cordeau, and E. Frejinger. The Locomotive Assignment Problem with Distributed Power at the Canadian National Railway Company. *Transportation Science*, 55(2):510–531, 2021.

M.B. Pedersen and T.G. Crainic. Optimization of Intermodal Freight Service Schedules on Train Canals. Publication CIRRELT-2007-51, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport, Montréal, QC, Canada, 2007.

M.B. Pedersen, T.G. Crainic, and O.B.G. Madsen. Models and Tabu Search Metaheuristics for Service Network Design with Asset-Balance Requirements. *Transportation Science*, 43(2):158–177, 2009.

F. Piu and M.G. Speranza. The Locomotive Assignment Problem: A Survey on Optimization Models. *International Transactions in Operational Research*, 21(3):327–352, 2014.

F. Piu, V. Première Kumar, M. Bierlaire, and M.G. Speranza. Introducing a Preliminary Consists Selection in the Locomotive Assignment Problem. *Transportation Research Part E: Logistics and Transportation Review*, 82:214–237, 2015.

W.B. Powell. A Local Improvement Heuristic for the Design of Less-than-Truckload Motor Carrier Networks. *Transportation Science*, 20(4):246–357, 1986a.

W.B. Powell. Approximate, Closed Form Moment Formulas for Bulk Arrival, Bulk Service Queues. *Transportation Science*, 20(1):13–23, 1986b.

W.B. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. John Wiley and Sons, 2nd edition, 2011.

W.B. Powell. *Reinforcement Learning and Stochastic Optimization - A Unified Framework for Sequential Decisions*. Wiley-Interscience, Hoboken, N.J., 2021.

W.B. Powell and T.A. Carvalho. Dynamic Control of Multicommodity Fleet Management Problems. *European Journal of Operational Research*, 98:522–541, 1997.

W.B. Powell and T.A. Carvalho. Real-Time Optimization of Containers and Flatcars for Intermodal Operations. *Transportation Science*, 32(2):110–126, 1998.

W.B. Powell and P. Humblet. Queue Length and Waiting Time Transforms for Bulk Arrival, Bulk Service Queues with a General Control Strategy. *Operations Research*, 34:267–275, 1986.

W.B. Powell and Y.A. Koskosidis. Shipment Routing Algorithms with Tree Constraints. *Transportation Science*, 26(3):230–245, 1992.

W.B. Powell and Y. Sheffi. The Load Planning Problem of Motor Carriers: Problem Description and a Proposed Solution Approach with Predetermined Step Sizes. *Transportation Researc Part A: General*, 17(6):471–480, 1983.

W.B. Powell and H. Topaloglu. Fleet Management. In Wallace, S. and Ziemba, W., editors, *Applications of Stochastic Programming*, Math Programming Society - SIAM Series on Optimization, pages 185–216. SIAM, Philadelphia, PA, 2005.

W.B. Powell, B. Bouzaïene-Ayari, and H.P. Simaõ. Dynamic Models for Freight Transportation. In Barnhart, C. and Laporte, G., editors, *Transportation*, volume 14 of *Handbooks in Operations Research and Management Science*, pages 285–365. North-Holland, Amsterdam, 2007.

P. Premkumar, S. Gopinath, and A. Mateen. Trends in third party logistics–the past, the present & the future. *International Journal of Logistics Research and Applications*, 24(6):551–580, 2021.

D. Ronen. Cargo Ships Routing and Scheduling: Survey of Models and Problems. *European Journal of Operational Research*, 12(2):119–126, 1983.

D. Ronen. Ship Scheduling: The Last Decade. *European Journal of Operational Research*, 71(3):325–333, 1993.

J. Roy and T.G. Crainic. Improving Intercity Freight Routing with a Tactical Planning Model. *Interfaces*, 22(3):31–44, 1992.

J. Roy and L. Delorme. NETPLAN: A Network Optimization Model for Tactical Planning in the Less-than-Truckload Motor-Carrier Industry. *INFOR*, 27(1):22–35, 1989.

Y.O. Scherr, B.A.N. Saavedra, M. Hewitt, and D.C. Mattfeld. Service Network Design with Mixed Autonomous Fleets. *Transportation Research Part E: Logistics and Transportation Review*, 124:40–55, 2019.

Y.O. Scherr, M. Hewitt, and D.C. Mattfeld. Stochastic Service Network Design for a Platooning Service Provider. *Transportation Research Part C: Emerging Technologies*, 144:103912, 2022.

K.R. Smilowitz, A. Atamtürk, and C.F. Daganzo. Deferred Item and Vehicle Routing within Integrated Networks. *Transportation Research Part E: Logistics and Transportation*, 39:305–323, 2003.

D. Srinivasan, R.K. Amit, and W.-C. Chiang. Air Cargo Revenue Management: A State-of-the-Art Review. *International Journal of Revenue Management*, 13(3):125–215, 2023.

M. SteadieSeifi, N.P. N. P. Dellaert, W. Nuijten, T. van Woensel, and R. Raoufi. Multimodal Freight Rransportation Planning: A Literature Review. *European Journal of Operational Research*, 233(1):1–15, 2014.

G. Taherkhani, I.C. Bilegan, T.G. Crainic, M. Gendreau, and W. Rei. Tactical Capacity Planning in an Integrated Multi-stakeholder Freight Transportation System. *OMEGA*, 110, 2022. Article 102628.

P. Toth and D. Vigo, editors. *The Vehicle Routing Problem*. SIAM Monographs on Discrete Mathematics and Applications. SIAM, Philadelphia, PA, 2002.

P. Toth and D. Vigo, editors. *Vehicle Routing: Problems, Methods, and Applications - Second Edition*. MOS-SIAM Series on Optimization. SIAM, Philadelphia, PA, 2014.

B. Vaidyanathan, R.K. Ahuja, J. Liu, and L.A. Shughart. Real-Life Locomotive Planning: New Formulations and Computational Results. *Transportation Research Part B: Methodological*, 42(2):147–168, 2008a.

B. Vaidyanathan, R.K. Ahuja, and J.B. Orlin. The Locomotive Routing Problem. *Transportation Science*, 42(4):492–507, 2008b.

D.M. Vu, T.G. Crainic, and M. Toulouse. A Three-Stage Matheuristic for the Capacitated Multi-commodity Fixed-Cost Network Design with Design-Balance Constraints. *Journal of Heuristics*, 19:757–795, 2013.

X. Wang and S. W. Wallace. Stochastic scheduled service network design in the presence of a spot market for excess capacity. *EURO Journal on Transportation and Logistics*, 5:393–413, 2016.

X. Wang, T.G. Crainic, and S.W. Wallace. Stochastic Network Design for Planning Scheduled Transportation Services: The Value of Deterministic Solutions. *INFORMS Journal on Computing*, 31(1):153–170, 2019.

W.W. White. A Program for Empty Freight Car Allocation. Technical Report 360D.29.002, IBM Contributed Program Library, IBM Corporation, Program Information Department, Hawthorne, N.Y., 1968.

W.W. White. Dynamic Transshipment Networks: An Algorithm and its Application to the Distribution of Empty Containers. *Networks*, 2(3):211–236, 1972.

W.W. White and A.M. Bomberault. A Network Algorithm for Empty Freight Car Allocation. *IBM Systems Journal*, 8(2):147–171, 1969.

N. Wieberneit. Service network design for freight transportation: a review. *OR spectrum*, 30(1):77–112, 2008.

M. Yaghini and R. Akhavan. Multicommodity Network Design Problem in Rail Freight Transportation Planning. *Procedia Social and Behavioral Sciences*, 43:728–739, 2012.

M. Yaghini, M. Seyedabadi, and M.M. Khoshraftar. Solving Railroad Blocking Problem Using Ant Colony Optimization Algorithm. *Applied Mathematical Modelling*, 35(12): 5579–5591, 2011.

M. Yaghini, M. Seyedabadi, and M.M. Khoshraftar. A Population-based Algorithm for the Railroad Blocking Problem. *Journal of Industrial Engineering International*, 8(8): 1–11, 2012.

M. Yaghini, M. Momeni, and M. Sarmadi. Solving Train Formation Problem Using Simulated Annealing Algorithm in a Simplex Framework. *Journal of Advanced Transportation*, 48:402–416, 2014.

Y. Zhao, Q. Xue, Z. Cao, and X. Zhang. A Two-Stage Chance Constrained Approach with Application to Stochastic Intermodal Service Network Design Problems. *Journal of Advanced Transportation*, page 6051029, 2018.

E. Zhu, T.G.. Crainic, and M. Gendreau. Scheduled Service Network Design for Freight Rail Transportations. *Operations Research*, 62(2):383–400, 2014.