Hybrid simulation and optimization approach for green intermodal transportation problem with travel time uncertainty

Original Citation:
Hrusovsky, Martin and Demir, Emrah and Jammernegg, Werner and van Woensel, Tom (2018)
Hybrid simulation and optimization approach for green intermodal transportation problem with travel time uncertainty.
Flexible Services and Manufacturing Journal, 30 (3).
pp. 486-516. ISSN 1936-6582
This version is available at: https://epub.wu.ac.at/5701/
Available in ePubWU: August 2017

License: Creative Commons: Attribution 4.0 International (CC BY 4.0)
ePubWU, the institutional repository of the WU Vienna University of Economics and Business, is provided by the University Library and the IT-Services. The aim is to enable open access to the scholarly output of the WU.
This document is the publisher-created published version. It is a verbatim copy of the publisher version.
Hybrid simulation and optimization approach for green intermodal transportation problem with travel time uncertainty

Martin Hrušovský1 · Emrah Demir2 · Werner Jammernegg1 · Tom Van Woensel3

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract The increasing volumes of road transportation contribute to congestion on road, which leads to delays and other negative impacts on the reliability of transportation. Moreover, transportation is one of the main contributors to the growth of carbon dioxide equivalent emissions, where the impact of road transportation is significant. Therefore, governmental organizations and private commercial companies are looking for greener transportation solutions to eliminate the negative externalities of road transportation. In this paper, we present a novel solution framework to support the operational-level decisions for intermodal transportation networks using a combination of an optimization model and simulation. The simulation model includes stochastic elements in form of uncertain travel times, whereas the optimization model represents a deterministic and linear multi-commodity service network design formulation. The intermodal transportation plan can be optimized according to different objectives, including costs, time and CO2 emissions. The proposed approach is successfully implemented to real-life scenarios where differences in transportation plans for alternative objectives are presented. The solutions for transportation networks with up to 250 services and 20 orders

Emrah Demir
demire@cardiff.ac.uk

Martin Hrušovský
martin.hrusovský@wu.ac.at

Werner Jammernegg
werner.jammernegg@wu.ac.at

Tom Van Woensel
t.v.woensel@tue.nl

1 WU Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria
2 Panalpina Centre for Manufacturing and Logistics Research, Cardiff Business School, Cardiff University, Cardiff CF10 3EU, UK
3 Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands

Published online: 10 December 2016
show that the approach is capable of delivering reliable solutions and identifying possible disruptions and alternatives for adapting the unreliable transportation plans.

**Keywords** Intermodal transportation · CO₂-equivalent · Travel time uncertainty · Simulation · Optimization

1 Introduction

The growing volumes of freight due to globalization lead to increased volumes of transportation on the limited transportation networks which results in delays and disruptions due to congestion, accidents and other unexpected events. This is especially true for road transportation which has been traditionally the preferred transportation option and still has the major share on the modal split in Europe (Commission European 2012; Eurostat 2014). Moreover, road transportation is one of the main contributors to the CO₂ emissions from transportation that are responsible for climate changes (European Commission 2014). Therefore, companies are searching for alternative transportation solutions that would minimize the negative impact of transportation.

One of the alternatives is the use of various transportation modes that offers flexibility and efficiency. The number of transportation alternatives can be increased by using different transportation modes and combining them in multimodal transportation chains. Multimodal transportation not only promotes advantages of each transportation mode but also eliminates their disadvantages. Intermodal freight transportation is a specialization of multimodal transportation. It consecutively uses various modes while moving the freight within a loading unit. Since a loading unit is often a standardized container, intermodal transportation is also called as containerized transportation. Intermodal transportation offers numerous advantages in addition to the noted flexibility offered by multimodal transportation. For example, standard sizes, faster transshipments, and reduced packaging expenses are essential benefits for shippers with large volumes (Jennings and Holcomb 1996).

Along with increasing trade volumes, also the research within the intermodal transportation has picked up in the last decades. Several review studies have already identified the importance of operations research (OR) for intermodal transportation (see e.g., Macharis and Bontekoning 2004; SteadieSeifi et al. 2014). One of the identified areas of need for OR solutions is the reliability and accuracy of the transportation networks under uncertainty. Rising awareness of the importance of low levels of tied-up capital (and thus inventories) leads to an increasing need for dependable on-time deliveries. Therefore, the reliability of transportation plans is progressively considered as a key performance indicator for logistics service providers and freight forwarders. A direct consequence of this changed operating environment is increasing awareness for the new developments in OR literature, such as modeling travel time uncertainties (see e.g., Noland and Small 1995; Gendreau et al. 1996; Kok et al. 2010).

Owing to tremendous achievements in both fields of OR and computer science, optimization models have been adapted such that uncertainties can be taken into
account. Their main purpose is to account for variabilities beforehand in order to develop transportation plans which are more accurate and also more reliable to external influences. Improved collection of real-time traffic flow information over the last decade, which enables the identification of traffic flow distributions, builds the data basis for such approaches (Cho et al. 2006).

On an operational level, within the transportation sector especially the Vehicle Routing Problem (VRP) and its multitude of extensions play a pivotal role in this context (see e.g., Laporte 1992; Eksioglu et al. 2009). While VRP’s literature offers extensive methods for mostly unscheduled road transportation, other transportation activities in intermodal transportation networks follow a fixed schedule. Especially train and flight connections as well as carriages on waterways, but also long-haul transports on the road come to mind. In such cases, service network design (SND) provides intriguing possibilities for the reproduction of transportation flows on more than one mode.

Different modeling approaches of SND offer solutions for intermodal transportation planning on the tactical as well as on the operational level. While SND on the tactical level is especially concerned with the frequency a service should be operated with per time frame (Frequency Service Network Design) and the itineraries of the offered services in order to properly serve the demand, SND problems on the operational level deal with the selection of available services for specific transports. On each level of planning, SND offers advantages for the consolidation of transports as well as the consideration of multiple modes (Crainic 2003). Moreover, it offers methodological possibilities which enable the representation of transshipment as well as the consolidation of containers. There is a rich literature on both intermodal and SND problems in the past decades (see e.g., Crainic and Rousseau 1986; Crainic 2000; Wieberneit 2008).

In order to include uncertainties into SND formulations, SND methodologies can be divided into static and dynamic problems. While for static problems all of the characteristics of possible services are fixed, dynamic problems allow for modifications of the values of at least one variable. The research for dynamic SND problems is still in its early days, though, which leads to a lack of applications to as well as the development of new methods for service network environments. Most of the limited publications in this domain are dealing with demand uncertainty (see e.g., Ukkusuri et al. 2007; Lium et al. 2009; Crainic et al. 2011) while only a minority takes travel time uncertainties into account (see e.g., Demir et al. 2016). Input from practitioners, though, suggests that travel time uncertainty - next to demand uncertainty - is one of the two most important sources of variability to consider when trying to make accurate transportation plans.

Whereas considering uncertainty in conventional methods such as linear or dynamic programming might lead to problems due to high number of possible scenarios, simulation can be used in this context as it can imitate the real behavior of systems by providing test bed for various experiments (Kelton and Law 2000). It is being widely used to analyze complex dynamic and stochastic situations in supply chain and logistics management (see e.g., Wahle et al. 2002; Davidsson et al. 2005; Mes et al. 2007; Reis 2014; Holmgren et al. 2014). In order to investigate the dynamic behavior of such systems, the literature on simulation and its integration
with standard optimization methods is recently getting more attention in OR literature. In her PhD thesis, Preusser (2008) studied a combined simulation and optimization approach to improve the supply chain by simultaneously optimizing a large number of possible transportation decisions. In another study, Song et al. (2013) proposed a simulation-based approach for sustainable transportation optimization by searching for optimal combination of transportation planning and operations strategies that minimize generalized costs of multimodal traveling.

An extensive literature is available for combination of optimization and simulation in hybrid approaches in Supply Chain Management (SCM) context (see e.g., Almeder et al. 2009; Acar et al. 2010; Nikolopoulou and Ierapetritou 2012; Sahay and Ierapetritou 2013), especially for production and distribution (see e.g., Bilgen and Çelebi 2013; Sel and Bilgen 2014; Safaei et al. 2010), location decisions, and perishable goods (see e.g., Keizer et al. 2015). However, the application to transportation area is limited. To the best of our knowledge, there is no study to date that proposes hybrid simulation and optimization approach in the domain of intermodal freight transportation. The aim of simulation is usually to adjust results from deterministic optimization or to estimate some parameters (e.g., inventory, costs). This is different from our approach, where simulation is used to evaluate the reliability of the created deterministic plans under stochastic conditions.

The aim of this paper is to combine the advantages of both simulation and optimization methods by considering a dynamic and stochastic environment in the context of intermodal transportation and to offer solutions which do not only minimize costs but also reduce the negative impact of transportation activities. The contributions of this paper are the following:

1. We propose a green approach for the movement of containers between a point of origin and a specified destination combining economic (costs, time) and environmental (carbon dioxide equivalent (CO2e) emissions) objectives. The emissions for each service and transshipment are modeled using specific models considering several factors such as distance, speed, payload, etc.
2. To the best of our knowledge, we propose the first hybrid simulation and optimization approach for intermodal transportation. A mixed integer linear program being small enough to provide optimal deterministic solution within a reasonable computational time is proposed. These plans together with the transportation network including different transportation modes and transshipment locations are then used in the stochastic simulation model. The whole process is repeated until a feasible and reliable transportation plan for each order is found. In this way, the size of the problem instances can be increased and the time needed for finding the optimal solution can be reduced in comparison to the traditional stochastic linear programming models.
3. The proposed methodology takes into account specific characteristics of each transportation service (e.g., capacity, schedules, costs) and considers uncertainties connected with travel times. The simulated travel times have a three-point distribution for each service, representing the uncongested, congested and disrupted travel time with their respective probabilities. Extensive
computational experiments using real-life data have been performed to show the benefits of using a hybrid simulation and optimization approach.

The paper is structured as follows. Section 2 introduces the intermodal transportation problem whereas Sect. 3 details the solution methodology. Section 4 discusses the case study and other extensive computational results obtained with the proposed approach. Conclusions and managerial insights are stated in Sect. 5.

2 Problem description

The intermodal transportation chain consists of a number of transportation services served by different transportation modes that connect intermodal terminals where transshipment has to be handled. There might exist multiple connections between two terminals representing different services. These services need to be coordinated in order to ensure smooth flow of goods in containers through the network from their origin to the destination within time windows specified by the customer. Typically, there exist various alternative routes within the network between the planned origin and destination of a container and the aim is to find the optimal route which fulfills the criteria set by the decision maker (e.g., minimization of costs, time or emissions).

One of the characteristics of intermodal transportation networks are fixed departure times of services which are running according to planned schedules. This is especially true for rail or waterway services whereas truck services are usually more flexible as they do not have fixed time slots when they can use the available infrastructure. This constraint further increases the complexity of the intermodal transportation problem since the fixed departure times have to be considered when coordinating the individual services in a transportation plan. Whereas schedules can be easily incorporated into planning if only deterministic travel times under ideal conditions are considered, they might lead to disruptions of the network when delays occur and the goods are delivered to the terminal only after the next planned service has already left. Therefore, the reliability of the plans plays an important role as it can be also seen in Fig. 1. Here the objective is to move the containers from node A to node F which can be done using either route A–B–C–F or route A–D–E–F. The lower part of the figure gives a Gantt chart of departure and travel times of individual services whereby the deterministic travel times (without any disruptions) are represented by the black, green and orange bars. In this setting, the upper route A–B–C–F seems to be the optimal one. However, if also possible travel time uncertainties and delays due to disruptions based on e.g., historical data and depicted by red bars are taken into account, it is visible that the waterway service B–C might lead to delays resulting in disruptions since the goods on the vessel will miss the departure of the following train service C–F. Therefore, it might be beneficial to use reliable route A–D–E–F where the delays seem to be shorter so that the plan can be followed also in case of disruptions.

The travel time uncertainties incorporating possible common delays (e.g., congestion) can be represented in form of travel time distribution which can be
modeled in different ways. One possibility is to define a discrete distribution with three time points representing early, on-time and delayed arrival as it is done in Noland and Polak (2002). However, since the intermodal services are operated according to schedules and usually do not arrive earlier, we adapted this approach and we propose a three-point distribution consisting of the uncongested travel time representing the ideal state without any delays, the congested travel time accounting for possible small delays (e.g., due to congestion) and disrupted travel time covering longer delays due to e.g., an accident which is blocking the transportation link. In order to compare this approach with a continuous time distribution, a shifted exponential distribution is also used for travel time modeling (see e.g., Noland and Small 1995) in the computational study in Sect. 4.

In the proposed approach, each travel time realization has a certain probability of occurrence. This travel time distribution is different for each service which results in a high number of possible travel time combinations within the whole intermodal transportation network. As the possibilities of capturing all these combinations are limited for the traditional optimization models (see Sect. 3), we propose a combination of an optimization model and simulation model. In this combination, the optimization model is used to compute the optimal transportation plans under the deterministic uncongested travel times which are then evaluated for their reliability under stochastic travel time distributions in the simulation model. In this way it is possible to obtain the transportation plans quickly and evaluate them based on a higher number of travel time scenarios. Before this solution approach is presented in Sect. 3, the modeling of services and emissions from transportation in our approach is shortly described in the following sections.

2.1 Modeling transportation services

In our study, the term service is defined as a scheduled transportation with a specific vehicle. A service, thereby, is specified by its scheduled start as well as travel time

Fig. 1 An example of the intermodal transportation problem
(or service time) and route. While the start times are constant, travel times can take on values of varying degrees dependent on the traffic state and infrastructure utilization. As there can be multiple scheduled transports between terminals, the network has to be able to represent multiple services between terminals. The focus of this research is not on the tactical scheduling of services, but on their operational sequencing to routing plans which fulfill the demand and provide reliable solutions in respect to the uncertainties considered.

Supplemental to the scheduled start time as well as the service time itself, each service is characterized by its free capacity in terms of loading units (or containers) as well as its approximated costs and consumption of CO$_2$e emissions when using one of these open container slots. The services, thereby, have varying specifications which can be modeled as a problem with heterogeneous fleet.

The demand orders to be fulfilled are specified by the amount of containers demanded, origin and destination node as well as the earliest departure and latest arrival time of the delivery. In order to increase the efficiency, the scope of the network is reduced beforehand. In doing so, the network is carefully downsized by service options definitely inappropriate for the fulfillment of the demand orders.

2.2 Modeling carbon dioxide equivalent emissions

As carbon dioxide is the dominant man-made greenhouse gas (GHG), the impacts of other gases can also be calculated based on carbon dioxide equivalent (CO$_2$e) emissions (Demir et al. 2015). CO$_2$e emissions cause atmospheric changes and climate disruptions which are harmful to the natural and built environments, and pose health risks (Dekker et al. 2012; Demir et al. 2014; Bektas et al. 2016).

Despite the fact that transportation sector is one of the biggest contributors of CO$_2$e emissions, a survey performed by Demir et al. (2013) showed that calculation of emissions is only slowly becoming part of the transportation plans. Even when emissions are taken into account in planning software, they are only reported as an additional factor and they are not used as an optimization objective. Usually only costs are taken into account for optimization and in case of multiple objectives costs are combined with service, distance, time, etc. This development might be caused by several reasons which make the calculation of emissions challenging. The possible reasons are discussed below.

- The amount of emissions is dependent on the energy needed for moving a vehicle coming either from diesel fuel or electricity consumption. Although the energy consumption can be easily measured after the transport has been conducted, calculation of energy consumption before the start of the transport is problematic as it is dependent on a number of factors which are not always known. These factors include the characteristics of the vehicle (e.g., weight, air and rolling resistance), route and driving characteristics (e.g., gradient, speed, number of stops), and the amount of goods transported (Eichlseder et al. 2009). In order to estimate the emissions, a number of different models requiring detailed inputs have been developed as shown by Demir et al. (2011) and Demir et al. (2014). Besides these detailed microscopic models, emission calculators
(macroscopic models) based on real-world measurements and recommended values for a typical vehicle are also available (see e.g., Boulter and McCrae 2009; IFEU 2011). However, each of these models and calculators is based on oversimplified assumptions which lead to discrepancies between estimated and actual emissions.

- According to the GHG protocol, emissions can be divided into three scopes. These include emissions from resources owned by a company [e.g., production (Scope 1)], indirect emissions from purchased energy (Scope 2), and all other emissions including also other stages of supply chain [e.g., suppliers, transportation, distribution (Scope 3)] (Toffel and Sice 2011; Hoen et al. 2014). The emissions caused by transportation activities can either be calculated as emissions from fuel consumption directly in the vehicle (tank-to-wheel, TTW) or can also include emissions from production of the fuel (well-to-wheel, WTW). Inclusion of emissions from fuel production is especially important in cases where electric vehicles are involved since emissions from electricity consumption are equal to zero (Kranke et al. 2011).

- The monetary value of CO$_2$e emissions is still unclear. Since the long-term effects of emissions on climate change and the amount of released emissions cannot be easily predicted, the estimation of emission costs is based on a number of assumptions including different discount rates for future events and risk attitude of the decision makers. The social costs of emissions are estimated to be between 0 EUR to 700 EUR per ton of emissions depending on the model (Anthoff et al. 2011; Nordhaus 2011). Therefore the monetary value of emissions cannot be easily compared to transportation costs.

In case of road transportation, the Passenger car and Heavy duty vehicle Emission Model (PHEM) developed by TU Graz is an important basis for the Handbook on Emission Factors for Road Traffic (HBEFA) that offers fuel consumption factors for vehicles with different engine types driving on different road categories. These factors are based on real-world measurements simulating predefined driving cycles and provide the fuel consumption for an empty as well as fully loaded vehicle (Eichlseder et al. 2009). There is a linear relationship between the load factor and the fuel consumption of the vehicle and therefore the fuel consumption can be calculated easily for different payloads. In addition to the road category and load factor, the gradient also plays an important role. According to Knörr et al. (2011), the influence of gradient on fuel consumption in road transport is 5–10% which means that the emissions calculated for flat countries have to be multiplied by 1.05 for hilly and by 1.1 for mountainous countries.

In case of rail transportation, emissions are also dependent on a number of factors, such as rolling and aerodynamic resistance, and speed as well as gradient (Boulter and McCrae 2009). However, as the comparison with real-world values shows, energy consumption of a train can be related to its gross weight in tons (Knörr et al. 2011). The fuel consumption function can be also calculated for hilly countries and therefore has to be multiplied by 0.9 for flat countries and by 1.1 for mountainous countries (Knörr et al. 2011). In order to calculate the amount of emissions from the energy consumption, the result has to be multiplied by the
specific emission factor. Here it has to be differentiated between the diesel train causing emissions by burning diesel and the electric train where the emissions are caused only by energy production and not by energy consumption (Kranke et al. 2011).

Since the energy consumption does not produce any emissions in case of electric trains, the well-to-wheel (WTW) emission scope has to be chosen in order to make emissions from different transportation modes comparable. This scope includes emissions from energy consumption and energy production (Kranke et al. 2011). The emission factor used for diesel is therefore 3.24 kg CO$_2$e emissions per liter of diesel and for electric energy specific emission factors are taken for each country as proposed by Knörr et al. (2011) and DSLV (2013).

In case of inland waterway transportation (IWT), the available emission calculators usually use a fixed average emission factor per ton kilometre (tkm) (see e.g., Knörr et al. 2011; IFEU 2011) or use the same methodology for IWT and sea transportation (NTM 2008). These methods are employed due to the fact that a detailed calculation of emissions from IWT requires a high amount of input data which is not always available. However, the results are inaccurate and do not show the real performance of IWT. Therefore, in our research, the model ARTEMIS developed by Boulter and McCrae (2009) is applied. The input data for this model consists of vessel characteristics (e.g., type, length, breadth, draught, engine power), route properties (e.g., waterway depth and width, distance, vessel speed and direction) and cargo characteristics (e.g., type and weight). All of these input parameters are either proposed by the model itself or can be manually set for a specific vessel which allows a more accurate calculation of emissions.

For the terminals an energy consumption factor of 4.4 kWh/transshipment multiplied by the respective emission factor of the country is taken (Knörr et al. 2011). In order to express the emissions in terms of costs, a value of 70 EUR per ton of CO$_2$e emissions was chosen as recommended by the German Federal Environment Agency (PLANCO 2007).

We note that the amount of emissions released per twenty-foot container (TEU) strongly depends on the capacity utilization of the vehicle. Whereas the relationship between the load factor and the amount of emissions is linear in case of trucks, the emission functions for trains and inland vessels have an exponential character. Hence, in order to reduce the complexity, the emissions per TEU were calculated assuming the utilization of 80% for trains (PLANCO 2007) and 90% for inland vessels (via donau 2007).

3 The development of the hybrid approach

This section presents the methodological framework for the solution of green intermodal transportation problem which was originally defined by Demir et al. (2016). The objective here is to find optimal and reliable transportation plans for a set of transportation orders that are represented by their demand, origin and destination nodes as well as earliest release and due times. While the release time
represents a hard time constraint to work with, the due time is represented as a soft constraint, leading to penalty costs for late deliveries.

The plans can be optimized according to three different objectives - transportation costs, time and CO$_2$ emissions (in form of emission costs). These objectives can be considered either separately or in combination depending on the weight which is assigned to each objective before the optimization process is started. In this way the traditional economic objective of minimizing costs can be combined with reduction of environmental impact of transportation which contributes to more sustainable transportation plans. Moreover, the resulting transportation plans have to be reliable accounting for possible minor or major disruptions which are represented by stochastic travel times as described in Sect. 2.

In their original work, Demir et al. (2016) proposed a mixed-integer linear programming formulation including sample average approximation (SAA) method for solving the defined problem. In this approach, samples of travel time combinations for the transportation network are created in order to represent the travel time uncertainty. Based on these samples, the reliability of transportation plans is then evaluated directly during the optimization process. Although this allows to consider uncertainty within the optimization model, the presented results show that the additional complexity of the model caused by stochasticity is a limiting factor for the sample size and also the size of instances which can be solved to optimality. Besides that, the time objective is represented only by penalty costs for late deliveries which means that any route that does not violate the required delivery time can be chosen in case of time optimization without the possibility to evaluate the optimality of the plan.

In this paper we propose a new hybrid approach combining deterministic optimization model described in Sect. 3.1 with a simulation model presented in Sect. 3.2. In this approach, the optimization model is solved in the first step using deterministic (uncongested) travel times in order to obtain optimal transportation plans without consideration of uncertainty. In this deterministic model, the time objective consists of two parts: firstly, the total transportation time is minimized using the the costs for inventory in transit dependent on the total transportation time. Secondly, penalty costs for late delivery of goods to the destination are also included.

In the second step, the plans serve as an input for the simulation model which then checks the reliability of the solution under uncertain travel times. If the transportation plan is evaluated as reliable, it is fixed for further execution, otherwise the optimization model is used again in order to replace an unreliable plan by a new plan for the affected order. This process is repeated until reliable transportation plans are found for all orders. The detailed solution procedure is presented in Sect. 3.3.

The proposed approach has a number of advantages which contribute to an improved and more efficient solution of the problem:

- The inclusion of in-transit inventory costs enables better evaluation of plans which minimize the transportation time.
The division of the solution procedure into two steps reduces the complexity of the optimization model since only one travel time is considered for each service instead of a number of different scenarios. As a result, the deterministic (optimal) solution is obtained faster and the model can solve larger instances without memory problems of the solver.

The use of simulation for checking the reliability of transportation plans enables to consider a higher number of scenarios in each run which increases the statistical significance of the results. Although this has negative effect on the computational time for small instances, the time needed for each simulation run tends to be stable with the increasing size of the instances. This is in contrast with the exponential time increase in case of the SAA approach.

Since the simulation model can easily handle problems with higher complexity, further sources of uncertainty can be added (e.g., demand, vehicle capacity, etc.). Besides that, different probability distributions can be used for representing each source of uncertainty (Borshchev 2013).

Whereas the original approach using SAA only reports the optimal route without giving more information about alternative routes and causes of disruptions, the simulation model in the proposed approach is capable of showing where and how often the disruptions on the proposed route can usually occur. Moreover, it is possible to define additional decision rules for changing the route in situations where the original deterministic plan becomes infeasible in stochastic environment. In this way, additional decision criteria (e.g., costs of an alternative solution) can be taken into account so that also plans that appear to be unreliable can be considered as optimal if the additional costs in case of infeasibility are very low and the costs of the next best alternative are very high (see computational study in Sect. 4 as an example).

### 3.1 The optimization model

This section provides a linear mixed-integer mathematical formulation of the deterministic optimization model, which is used in the first step of the solution methodology to find an optimal plan for orders \( p \in \mathcal{P} \) defined by their demand \( d^p \), origin \( i \) and destination \( j \) nodes as well as earliest release \( r^p_{\text{release}} \) and due time \( r^p_{\text{duetime}} \). Moreover, \( \mathcal{P}^p(i,j) = \{ (p \in \mathcal{P}) | i \in \mathcal{N} \text{ and } j \in \mathcal{N} \} \) is a set of orders with origin \( i \) and destination node \( j \). The orders can be routed in a transportation network consisting of services \( s \in \mathcal{S} \) (scheduled transports) and nodes \( i, j \in \mathcal{N} \) (transshipment locations). Each service, since it is connected to a schedule and vehicle, is unique and connects transshipment locations \( i \) and \( j \). Therefore, \( \mathcal{S}(i,j,v,D^s_m) = \{ (s \in \mathcal{S}) | i \in \mathcal{N} \text{ and } j \in \mathcal{N} \text{ and } v \in \mathcal{V} \} \) is a set of services executed by vehicle \( v \) between origin \( i \) and destination node \( j \) within the starting time window bounded by \( T^s_{\text{min}} \) and \( T^s_{\text{max}} \). In addition to that, services are characterized by their scheduled departure time \( D^s \) and service time \( t^s \) as well as service slot price \( c^s \) and CO\(_2\)e emissions per container \( e^s \). Services on the road as well as transshipments are - due to their lack of scheduled supply in reality - assumed to be available when needed.
We first present sets, parameters and decision variables and then provide the mathematical formulation of the model (Tables 1, 2). This model extends the model introduced by Demir et al. (2016) by adding in-transit inventory costs to the original time-related cost component of the objective.

The objective function (1) of the mathematical model minimizes a weighted sum of the total costs. The weights enable the reflection of individual preferences regarding direct transportation ($\omega_1$), time-related ($\omega_2$) and CO2e emissions-related ($\omega_3$) costs. The direct transportation costs consist of transportation costs per container and service $c^s$, which include the fixed transportation costs per service allocated to one container as well as the direct transportation costs per container and transshipment costs per container ($c_j$). The time-related costs are represented by in-transit inventory costs for the total time spent since the release of goods at the origin until the arrival of the order to the destination. In addition to that, charges for delayed deliveries ($c_{pen}^p$) are also included in time-related costs. As the third objective the CO2e emissions-related costs per kg ($c_{emi}$) for the emissions consumed per container serviced ($e^s$) and transshipped ($e_j$) are also included. The transportation plans can be optimized for single objectives [e.g., (1, 0, 0) for transportation

Table 1 Sets and parameters used in the model

<table>
<thead>
<tr>
<th>Sets and parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{N}$</td>
<td>Set of all transshipment locations</td>
</tr>
<tr>
<td>$\mathcal{N}^+$</td>
<td>Set of start terminals of transportation orders</td>
</tr>
<tr>
<td>$\mathcal{N}^-$</td>
<td>Set of end terminals of transportation orders</td>
</tr>
<tr>
<td>$\mathcal{P}$</td>
<td>Set of transportation orders</td>
</tr>
<tr>
<td>$\mathcal{S}$</td>
<td>Set of transportation services</td>
</tr>
<tr>
<td>$\mathcal{V}$</td>
<td>Set of vehicles</td>
</tr>
<tr>
<td>$\gamma^{\text{release}}_p$</td>
<td>Earliest release time of order $p$</td>
</tr>
<tr>
<td>$\gamma^{\text{duetime}}_p$</td>
<td>Due time of order $p$</td>
</tr>
<tr>
<td>$c_j$</td>
<td>Transshipment costs per container in terminal $j$</td>
</tr>
<tr>
<td>$c^s$</td>
<td>Transportation costs of a service $s$</td>
</tr>
<tr>
<td>$c_{emi}$</td>
<td>Emissions-related costs per kg of CO2e emissions</td>
</tr>
<tr>
<td>$c_{pen}^p$</td>
<td>Penalty costs in case of late delivery of goods</td>
</tr>
<tr>
<td>$c_{inv}^p$</td>
<td>In-transit inventory costs per hour for order $p$</td>
</tr>
<tr>
<td>$\text{cap}^s$</td>
<td>Free capacities of services $s$</td>
</tr>
<tr>
<td>$d^p$</td>
<td>Demand (in containers) of order $p$</td>
</tr>
<tr>
<td>$e_j$</td>
<td>Emissions in kg per transshipment of container in terminal $j$</td>
</tr>
<tr>
<td>$e^s$</td>
<td>Emissions in kg per transportation of container on service $s$</td>
</tr>
<tr>
<td>$L$</td>
<td>Large (enough) number</td>
</tr>
<tr>
<td>$t_j$</td>
<td>Separate loading and unloading time at terminal $j$</td>
</tr>
<tr>
<td>$r^s$</td>
<td>Transportation time of service $s$</td>
</tr>
<tr>
<td>$T_{\text{min}}^s$</td>
<td>Start of the departure time window for service $s$</td>
</tr>
<tr>
<td>$T_{\text{max}}^s$</td>
<td>End of the departure time window for service $s$</td>
</tr>
<tr>
<td>$\omega_i$</td>
<td>Weight for the objective $i$</td>
</tr>
</tbody>
</table>
Table 2 Decision variables used in the model

<table>
<thead>
<tr>
<th>Decision variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_{delay}^p )</td>
<td>Delay of order ( p ) at destination node ( j )</td>
</tr>
<tr>
<td>( A^s )</td>
<td>Arrival time of service ( s ) at the associated destination node ( j )</td>
</tr>
<tr>
<td>( AD^p )</td>
<td>Arrival time of order ( p ) to its destination</td>
</tr>
<tr>
<td>( D^s )</td>
<td>Departure time of service ( s ) at the associated departure node ( i )</td>
</tr>
<tr>
<td>( delay_{qr}^p )</td>
<td>Delay between preceding service ( q ) and succeeding service ( r ) of order ( p )</td>
</tr>
<tr>
<td>( n_j )</td>
<td>The number of containers transshipped at terminal ( j )</td>
</tr>
<tr>
<td>( y^i, y^p )</td>
<td>A binary variable equal to 1 if service ( s ) is used (for order ( p ))</td>
</tr>
<tr>
<td>( x^p )</td>
<td>The number of containers of order ( p ) carried via service ( s )</td>
</tr>
<tr>
<td>( z_{qr}^p )</td>
<td>The number of containers of order ( p ) that have to be transshipped between preceding services ( q ) and succeeding service ( r )</td>
</tr>
</tbody>
</table>

Minimize

\[
\begin{align*}
\omega_1 \sum_{p \in P} \sum_{s \in S} x^p s + \omega_1 \sum_{j \in N} n_j c_j + \omega_2 \sum_{p \in P} c^p (AD^p - \Gamma_{release}^p) \\
+ \omega_2 \sum_{p \in P} \sum_{s \in S} d_{delay}^p c_{pen}^p + \\
\omega_3 c_{emi} \sum_{p \in P} \sum_{s \in S} x^p s + \omega_3 \sum_{j \in N} n_j c_j 
\end{align*}
\]

Subject to:

\[
\begin{align*}
\sum_{s \in \delta(s \in S | n = i)} x^p = d^p \quad \forall n \in N | n = i, p \in P \\
\sum_{s \in \delta(s \in S | n = j)} x^p = d^p \quad \forall n \in N | n = j, p \in P \\
\sum_{s \in \delta(s \in S | n = i)} x^p - \sum_{s \in \delta(s \in S | n = j)} x^p = 0 \quad \forall n \in N | (n \neq i \quad n \neq j), p \in P \\
\sum_{p \in \gamma(p \in P)} x^p - y^s cap^s \leq 0 \quad \forall s \in \delta(s \in S) \\
x^p \leq y^p L \quad \forall s \in \delta(s \in S), p \in \gamma(p \in P) \\
x^p \geq y^p \quad \forall s \in \delta(s \in S), p \in \gamma(p \in P)
\end{align*}
\]
\[ y^s \leq \sum_{p \in \mathcal{P}, q \in \mathcal{P}} x^{sp} \quad \forall s \in \delta(s \in \mathcal{S}) \]  

\[ \sum_{p \in \mathcal{P}} \sum_{s \in \mathcal{S}} x^{sp} - 2 \sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{S}} \sum_{r \in \mathcal{S}} \delta(s \in \mathcal{S}|i = n|j = n) \delta(q \in \mathcal{S}|i = n) \delta(r \in \mathcal{S}|j = n) \]  

\[ \varepsilon^{opp} = n_n \quad \forall n \in \mathcal{N} \]  

\[ D^s + t^s - A^s \leq L(1 - y^s) \quad \forall s \in \delta(s \in \mathcal{S}) \]  

\[ A^q + t_{xq}^{opp} + t_{xq}^{opp} - 2t_{zq}^{opp} - \text{delay}^{opp} - D^s - L(1 - y^{opp}) \leq L(1 - y^{opp}) \]  

\[ \forall q \in \delta(s \in \mathcal{S}|j \in \mathcal{N}), \]  

\[ p \in \mathcal{P}, r \in \delta(s \in \mathcal{S}|i \in \mathcal{N}) \]  

\[ D^s - y^{opp} \Gamma_{\text{release}}^p \geq 0 \quad \forall p \in \mathcal{P}, s \in \delta(s \in \mathcal{S}|i \in \mathcal{N}^+) \]  

\[ A^s - \alpha_{\text{delay}}^p - \Gamma_{\text{due time}}^p \leq L(1 - y^{opp}) \quad \forall p \in \mathcal{P}, s \in \delta(s \in \mathcal{S}|i \in \mathcal{N}^-) \]  

\[ AD^p \leq A^s - L(1 - y^{opp}) \quad \forall p \in \mathcal{P}, s \in \delta(s \in \mathcal{S}|i \in \mathcal{N}^-) \]  

\[ T_{\text{min}}^s \leq D^s \leq T_{\text{max}}^s \quad \forall s \in \mathcal{S} \]  

\[ \text{delay}^{opp} \leq y^{opp} L \quad \forall q \in \mathcal{S}, r \in \mathcal{S}, p \in \mathcal{P} \]  

\[ \text{delay}^{opp} \leq y^{opp} L \quad \forall q \in \mathcal{S}, r \in \mathcal{S}, p \in \mathcal{P} \]  

\[ y^s, y^{opp} = \{0, 1\} \quad \forall s \in \mathcal{S}, q \in \mathcal{S}, r \in \mathcal{S}, p \in \mathcal{P} \]  

\[ \alpha_{\text{delay}}^p, x^{opp}, \varepsilon^{opp}, \text{delay}^{opp}, D^s, A^s, AD^p \geq 0 \quad \forall s \in \mathcal{S}, q \in \mathcal{S}, r \in \mathcal{S}, p \in \mathcal{P}. \]  

Constraints (2)–(4) handle the movement of containers. While constraints (2) and (3) focus on the origin and destination nodes, constraint (4) manages the transshipment. Demand, in that regard, is positive if more containers are planned to originate from a specific node than are destined for that node. Constraint (5) ensures that capacity limits of services are adhered to. Constraints (6)–(8) make sure that a service is only allowed to process any amount of containers when it is selected. While (9) tracks the transshipment necessary, constraints (10) and (11) ensure the timely sequencing of the services within the network. As seen in (10), each service has interrelated departure, service and arrival times. In addition to the synchronization at nodes in terms of loading units [(2)–(4)], constraint (11) takes care of the timely synchronization. It ensures the relation of sequential services at a transshipment location. This is necessary due to more or less fixed schedules of services, which permit services with earlier departure times than possible preceding services from following up on them. Constraints (12) and (13) provide the time frame for
each order to plan within. The lower limit (earliest pick-up time) is fixed while the upper limit (due date) can be bent, with penalties - if desired - allocated to late deliveries (\(d_p^d\)). Constraint (14) defines the arrival time of the order to the destination which is dependent on the arrival of the last service which the order is carried on. Constraint (15) gives the time window within which services can depart with \(T_{min}^s = T_{max}^s\) being valid for scheduled services. Constraints (16) and (17) ensure that the feasibility of two consecutive services is only checked if these services are designated to be used within the same routing plan. The domain of the decision variables is given in constraints (18) and (19).

### 3.2 The simulation model

Simulation model represents the second step of the solution methodology where the travel time uncertainty is considered. In contrast to the simulation models described in Sect. 2, which are used to estimate a value of a certain variable, the objective of this simulation model is rather to evaluate the reliability of a transportation plan by checking its feasibility under different travel time scenarios. Since the simulation model does not optimize the plans but only checks them and chooses the pre-defined option in case of infeasibility, it allows to increase the significance of the results by running a high number of scenarios in a relatively short time.

The simulation is modeled using the Anylogic University 7.2.0 software (Borshchev 2013) which is a Java-based simulation tool that supports the most common simulation methodologies (e.g., discrete-event, agent-based and system dynamics) and their combinations. In order to represent the complex intermodal transportation network and the associated processes, a combination of agent-based and discrete-event simulation is used. Within this framework, the agents represent entities that can have different states depending on the current activity which the entity is executing or on a decision which has to be made. The discrete-event processes are mainly used to represent the connection of the container with a certain vehicle during the loading, transportation and unloading processes including also waiting and travel times. The agents used in the simulation are terminals, vehicles and containers.

The population of Terminal agents represents the terminals in the intermodal network which can serve either as origin, destination or transshipment point for an order. Each terminal can be uniquely identified by its ID number and its characteristics include the name, transshipment time, costs and \(CO_2\) emissions per transshipped container. Since a Geographic Information System (GIS) is also a part of Anylogic, each terminal can be visualized on a map with its exact geographic coordinates.

The population of Vehicles comprises all types of vehicles that can be operated on a transportation network. The agents of type Vehicle can be either trucks, trains or vessels. All of them are characterized by the ID of the service which they are operated on, origin and destination terminal with their specific IDs, distance, departure time, transportation costs and \(CO_2\) emissions per container as well as the capacity of each vehicle. The travel time uncertainty is represented by the three
possible travel time realizations with their correspondent probabilities that are also stored as parameters for each service. At the initialization of each service, one travel time realization is randomly chosen according to the probability distribution. Whereas trains and vessels have their departure times fixed according to a database, trucks are assumed to be available when they are needed. Therefore they are waiting in the terminal until an order is available for using them.

A *Container* is an agent representing an order or a group of containers from an order which are transported on the same route determined by the transportation plan resulting from the optimization model. The information required for simulating the route of the *Container* comprises the unique identification in form of order ID and group ID, the origin and destination of the order, release and due time as well as in-transit inventory costs, penalty costs for late delivery and total number of containers. The route is defined as a sequence of services which have to be used in order to arrive at the final destination.

The simulation process is displayed in Fig. 2. Each *Container* agent starts in its terminal of origin (Terminal A) where it is waiting for the first service of its plan (represented by a *Vehicle* agent). When the service arrives, *Container* is loaded onto the *Vehicle* and transported to the next planned terminal (Terminal B) whereby the travel time is chosen randomly from the possible states. After the arrival to the next terminal, the *Container* is unloaded and the *Vehicle* is discarded since it has reached its final destination. At this stage it has to be decided for the *Container* whether it has also arrived at the planned final destination or whether further transportation is required. If the transportation has to be continued, the current terminal is again considered as origin and the *Container* has to wait for the next service. If the next service has already left before the *Container* is ready for further transportation, the original transportation plan becomes infeasible and an alternative solution has to be chosen. In this case it is assumed that an extraordinary truck service is organized which carries the *Container* directly to the final destination. However, since this

![Fig. 2 An overview of the simulation model](https://example.com/fig2.png)
extraordinary truck is not planned in advance, the costs of this service are higher in comparison to a planned truck on the same route in order to reflect the need for organizing a truck service in very short time which usually causes higher costs in reality. For this a truck matrix consisting of all possible connections between two terminals within the network with their respective costs, travel times and \( CO_2e \) emissions is part of the simulation model. When the Container reaches its final destination, the total transportation costs, \( CO_2e \) emissions and arrival time are recorded to evaluate the costs of the transportation plan and the agent is discarded.

Within one simulation run, all services and orders are simulated using the randomly chosen travel times. In order to evaluate the reliability of a plan, the simulation runs are repeated a certain number of times using different travel time combinations. In each simulation run, the total costs (including transportation costs, in-transit inventory costs, penalty costs for late arrival and emission costs) and the feasibility of the plan are recorded for each order. After the experiment is finished, the results for each order are evaluated based on two criteria: the number of runs in which the plan was infeasible and the average additional costs of this infeasibility in comparison to the optimal deterministic solution. If the number of scenarios where the plan was not feasible and the average total costs exceed a certain threshold defined in advance, the plan is not reliable and the particular order is sent back to the optimization model to find a new and reliable route. The orders with reliable plans are saved for transportation execution and the capacity of services which are used by these orders is updated. For the plans which are not reliable, the sequence of services is stored and sent to the optimization model in order to avoid the choice of this plan repeatedly. For this, an additional constraint per each unreliable route is added to the mathematical model as stated below.

\[
\sum_{s \in \delta(x \in \delta)} x^{sk} \leq |K^k| - 1 \quad \forall k \in \mathcal{K},
\]

where \( K \) is the set of orders which have unreliable routes and \( |K^k| \) is the number of services used for order \( k \), which has at least one unreliable route.

### 3.3 Combination of optimization and simulation model

The solution procedure combining the two models presented before is shown in Fig. 3. Although each model is run on a different specialized software, both have access to the same database consisting of the characteristics of terminals, services and orders which is continuously updated based on the generated solutions. The data exchange between the models is done through text files in which the output of one model is saved and then used as the input for the second model. In order to start the process, a database with the relevant transportation network for the given set of orders \( p \in \mathcal{P} \) consisting of services \( s \in \mathcal{S} \) and terminals \( i,j \in \mathcal{N} \) has to be defined. In this database, for each terminal its ID, costs, time and emissions for each transhipment are recorded. Services are defined by their ID, origin, destination, capacity, time window for departure, travel time, costs emissions and vehicle ID which is operated on a certain service. The created orders are defined by origin,
destination, demand (number of TEU), release and due time, in-transit inventory and penalty costs. Based on this input data, the optimization model is used to obtain the transportation plans based on uncongested deterministic travel times assuming the ideal state of the transportation network without any disruptions. In this way the optimal plan is obtained relatively quickly and additional constraints connected to travel time uncertainty which might limit the size of the instances that can be solved to optimality can be avoided. For each order in the optimal deterministic plans, the optimal sequence of services \( y_{sp} > 0 \) is recorded that serves as an input for the simulation model.

After obtaining the routes from the optimization model, the simulation model is used for evaluating the reliability of the created plans under uncertainty. For this, the minimal optimal costs for each deterministic plan are calculated and the travel time of each service is modeled as a random variable that can take one of three possible travel time values (i.e., uncongested, congested or disrupted) which can be chosen according to their probability of occurrence. During the simulation, multiple runs of the simulation model are executed in order to consider different possible travel time combinations for all services in the transportation network. Within each run, the optimal deterministic route for each order is simulated in order to see whether the plan is still feasible under the chosen travel time realizations. If a plan becomes infeasible (e.g., the subsequent service is missed due to the late arrival of the container to a terminal), it is assumed that containers are delivered directly from the current terminal to the destination by an extraordinary truck service. In this way not only the reliability of the plan but also the additional costs in case of infeasibility can be estimated.

At the end of the simulation phase, the reliability of the plans is evaluated based on the number of scenarios with infeasible plans and the additional costs of infeasibility as described in Sect. 3.2. These criteria are decisive for classifying a transportation plan as reliable or not reliable. Transportation plans which are
reliable \( (P_{rel}) \) leave the optimization process and are fixed for execution which means that the service capacity used by these plans has to be blocked and cannot be used for other orders. Therefore, the free capacities of the services used by reliable transportation plans have to be updated in the database before the next optimization is executed. After that transportation plans which are not reliable \( (P_{unrel}) \) are sent back to the deterministic optimization model and the whole optimization process starts from the beginning. In order to prevent the repeated choice of the unreliable plan by the optimization model, the service sequence of the unreliable plan is also used as input for the optimization model and is handled in an additional constraint (see Sect. 3.2) so that an alternative plan has to be chosen. This process is repeated until a reliable plan is found for all orders. If there is no feasible and reliable route for an order in the considered network, a direct transportation by truck is used as a default option.

4 Computational results

In order to test the capabilities of the presented solution approach, a case study using intermodal transportation network with services based on real-world schedules and routes has been developed. The network covers mainly the region of Central Europe with terminals located in Hungary, Slovakia, Czech Republic, Austria and Germany but also two maritime ports in Koper (Slovenia) and Trieste (Italy). These terminals are connected by a number of services which are usually used as feeder services to collect containers for direct trains going to the big maritime harbors in Western Europe (e.g., Rotterdam, Hamburg, Antwerp). The network consists mainly of rail services based on publicly available schedules which are extended by planned truck connections. Since the examined region is located around the Danube and other important European rivers, inland waterway services between different ports on the Rhine–Main–Danube axis were also considered as an option. The basic transportation network with railway and inland waterway connections and terminals is displayed in Fig. 4. Considered road services are not depicted in the figure due to the complexity reasons.

The process of finding the optimal and reliable transportation plans was tested on different instances with varying number of services and orders. The basic network consists of 20 terminals, 50 services and one order, but the number of services can be increased up to 500 and the maximal number of orders considered at the same time was 20. The examined instances are based on instances presented in Demir et al. (2016) in order to be able to compare the results between SAA and the presented hybrid simulation–optimization approach. The schedules and planned travel times of trains are based on Kombiverkehr (2014) and Metrans (2014). For travel times and distances on road PTV Map and Guide (PTV 2014) was used and data regarding inland waterway services is based on via donau (2007). The transportation costs for each service were estimated using information from PTV (2014), PLANCO (2007) and via donau (2007). For the in-transit inventory costs 1 EUR per hour and order was assumed in order to be able to differentiate between different routes with different transportation times. The simulation model was run
100 times for each instance and this was repeated 10 times in order to get statistically significant results. Transportation plans were considered as not reliable if the average number of infeasible plans was higher than 5% and the difference in average total costs in comparison to deterministic solution was also more than 5%.

Since transportation planning is an operative activity which has to be conducted multiple times per day, the time needed for obtaining the results is an important factor when evaluating the performance of the model. Therefore, at the beginning an analysis of computational times was performed for all instances which is displayed in Table 3. The analysis was conducted on an Intel(R) Core(TM) i5-5300U CPU with 2.3 GHz and 8 GB of memory. The deterministic optimization model was solved using CPLEX 12.6 (IBM ILOG 2016) and the simulation was run in Anylogic University 7.2.0 (AnyLogic 2016). For the simulation model, three different sample sizes were analyzed: 10 runs which were also used by Demir et al. (2016), 100 runs which offer more stable results in comparison to 10 runs, and 1000 runs which lead to similar results as 100 runs. In order to compare these computational times with the SAA method, also the reported average computational times for stochastic model in Demir et al. (2016) are included in the table.

Fig. 4 Rail and inland waterway services within the intermodal transportation network
As it can be seen in Table 3, we could solve all instances up to 250 services and 20 orders with the deterministic optimization model, instances with 500 services could not be solved by the optimization model due to memory problems (indicated by X in the table). However, the computational times for 500 orders could be determined for simulation model. In general the results show that the computational times are increasing with the complexity of the instances, however, the increase differs significantly between the presented models. Whereas the deterministic optimization model finds an optimal solution in less than 5 seconds for most of the instances, the stochastic SAA method also starts with low computational times for small instances but these times tend to grow exponentially as it seems to be very sensitive to the increasing number of orders. In contrast to that, the computational times for simulation model are higher for smaller instances, but they are relatively stable with the growing complexity. Besides that, increasing the number of scenarios is not a problem for the simulation model whereas it is a limiting factor for the SAA approach. Therefore the simulation–optimization approach is more

Table 3  Computational times for network with 20 terminals

<table>
<thead>
<tr>
<th>Service</th>
<th>Order</th>
<th>Deterministic optimization CPU time in seconds</th>
<th>SAA CPU time for 10 scenarios in seconds</th>
<th>Simulation CPU time in seconds for 10 runs</th>
<th>100 runs</th>
<th>1000 runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>1</td>
<td>0.06</td>
<td>0.64</td>
<td></td>
<td>7.18</td>
<td>28.70 221.32</td>
</tr>
<tr>
<td>50</td>
<td>2</td>
<td>0.08</td>
<td>0.60</td>
<td></td>
<td>7.20</td>
<td>29.18 201.75</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>0.24</td>
<td>0.93</td>
<td></td>
<td>7.55</td>
<td>28.48 202.02</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>0.30</td>
<td>1.48</td>
<td></td>
<td>9.40</td>
<td>34.50 240.00</td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>0.57</td>
<td>1.99</td>
<td></td>
<td>9.40</td>
<td>36.50 236.70</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>0.18</td>
<td>0.89</td>
<td></td>
<td>8.90</td>
<td>34.00 250.10</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>0.25</td>
<td>1.27</td>
<td></td>
<td>8.80</td>
<td>35.80 247.40</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>0.39</td>
<td>2.48</td>
<td></td>
<td>9.00</td>
<td>35.40 248.70</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>0.68</td>
<td>38.97</td>
<td></td>
<td>9.60</td>
<td>38.50 258.70</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>1.40</td>
<td>112.27</td>
<td></td>
<td>9.90</td>
<td>36.80 264.60</td>
</tr>
<tr>
<td>250</td>
<td>1</td>
<td>0.39</td>
<td>1.48</td>
<td></td>
<td>8.90</td>
<td>34.40 261.20</td>
</tr>
<tr>
<td>250</td>
<td>2</td>
<td>0.71</td>
<td>5.00</td>
<td></td>
<td>9.80</td>
<td>37.00 259.70</td>
</tr>
<tr>
<td>250</td>
<td>5</td>
<td>1.62</td>
<td>24.86</td>
<td></td>
<td>9.20</td>
<td>37.20 262.20</td>
</tr>
<tr>
<td>250</td>
<td>10</td>
<td>4.29</td>
<td>X</td>
<td></td>
<td>10.50</td>
<td>37.90 255.90</td>
</tr>
<tr>
<td>250</td>
<td>20</td>
<td>26.38</td>
<td>X</td>
<td></td>
<td>10.20</td>
<td>43.90 281.00</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>X</td>
<td>10.20</td>
<td></td>
<td>11.90</td>
<td>54.40 423.30</td>
</tr>
<tr>
<td>500</td>
<td>2</td>
<td>X</td>
<td>X</td>
<td></td>
<td>11.90</td>
<td>53.90 432.40</td>
</tr>
<tr>
<td>500</td>
<td>5</td>
<td>X</td>
<td>X</td>
<td></td>
<td>13.20</td>
<td>54.10 430.90</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
<td>X</td>
<td>X</td>
<td></td>
<td>13.10</td>
<td>54.80 427.90</td>
</tr>
<tr>
<td>500</td>
<td>20</td>
<td>X</td>
<td>X</td>
<td></td>
<td>13.50</td>
<td>55.90 454.40</td>
</tr>
</tbody>
</table>

X: not solved to optimality due to the time limit (60 min) or memory problem
favorable for more complex instances which correspond to the realistic problem sizes usually solved by transportation planners.

Besides the computational times, the quality of the solutions has also been evaluated. Transportation plans on the network with 20 terminals for each of the instances between 50 and 250 services and 1 to 20 orders were optimized according to the three different objectives: transportation costs (1, 0, 0), time (0, 1, 0) and $CO_2e$ emissions (0, 0, 1). In addition to that, also the combination of all three objectives with equal weights (1, 1, 1) has been tested, but the results showed that the plans for this alternative are equal to the optimization according to transportation costs due to the fact that the transportation costs are much higher in comparison to time and emission costs. Table 4 presents the results for transportation costs optimization and emissions optimization which can be easily compared to the results of the SAA approach since the objective function of the optimization model was not changed for these two objectives. This is also confirmed by the fact that the results of optimization match the results of the deterministic model reported in Demir et al. (2016) and are shown in the first column for each objective.

Whereas the results for optimization model are the same, the results for SAA and simulation model show differences. This is due to the fact that the simulation model evaluates the reliability of the plans based on two criteria and therefore also plans might be accepted that become infeasible in more than 5% of the simulation runs, but the total cost change is very small. This might be the case when, e.g., there is a service which occasionally arrives slightly late to the destination of the order. Nevertheless, this small delay might be acceptable if the next alternative plan is much more expensive. Whereas this plan can be accepted by the simulation–optimization model, the SAA approach would ignore this alternative and choose a more expensive plan since the decision there is based solely on the number of infeasible scenarios.

The differences in results for transportation costs are mainly visible for the network of 20 terminals and 50 services which offers only limited number of alternatives (mostly one or two) for each order. In this case if the alternatives are evaluated as unreliable, the simulation model offers a direct truck connection (based on the truck service matrix for extraordinary services) as the best alternative which might result in higher but also in lower costs compared to the original plan depending on the origin and destination of the order. Therefore the total costs of the plans after simulation are increasing for the instance with 5 and 10 orders, but there is a slight decrease of the costs for 20 orders since for some of them the direct truck (which is not available in the initial network with 50 services) is a cheaper and more reliable option. A similar situation can be observed in case of emission minimization, where the emissions after simulation are increasing due to the use of direct truck transportation on routes where combination of train or vessels is evaluated as unreliable. Although this might be not completely supporting the idea of combining different modes in intermodal transportaqtion, avoiding transportation plans which are not reliable and using direct trucks in cases without other alternatives helps to reduce the disruptions in supply chains. In this way the decision maker can compare different options and choose the intermodal transportation only
if it is reliable. The information about the reliability of the plan might then increase
the motivation of transportation planners to choose the more complex but maybe
also cheaper intermodal alternative in comparison to direct transportation by truck.

Table 4 also shows that the increasing number of services has a positive impact
on costs as well as emissions since the number of transportation alternatives is
increasing. Therefore not only cheaper but also more reliable alternatives might
become available which means that the differences between the optimization result
and the costs of the plans after simulation are decreasing. If a network of 250
services is used, the results show that only very small additional costs (less than 1%)
have to be paid due to unreliability of the transportation plans.

In addition to transportation costs and emissions, also the time objective was
considered where the in-transit inventory costs for transportation time have been
added as described in Sect. 3. Due to this change, a direct comparison to the results
of SAA analysis is not possible. Therefore, instead of comparing SAA and
simulation results, only results for simulation are presented using two different
travel time distributions as mentioned in Sect. 2: a discrete three-point distribution
used in the whole computational study and a continuous shifted exponential
distribution. In order to obtain the shifted exponential distribution, three intervals
(uncongested, congested, disrupted) were created with borders located in the middle
between the discrete travel times for each state and probabilities corresponding to
the discrete ones. After that the exponential distribution was fitted to these intervals
so that it is shifted to the right starting at the point of the uncongested travel time
that is the shortest possible travel time a service can have. The two different
distributions are illustrated in Fig. 5.

The results for time optimization are presented in Table 5, where only instances
with up to 100 services could be solved by the simulation–optimization approach.
For instances with 250 services no optimal solution could be found within one hour
of running the model. The results show that there is a strong preference for using
truck services since their departures are flexible and travel times are usually very
fast. Therefore sometimes combination of three trucks can be preferred before a
direct train connection which might leave too late or take too long. Similarly, a
vessel connection from the transportation costs optimal plan can be replaced by
various truck or train services.

When comparing the discrete and continuous travel time distribution, three trends
could be observed: firstly, the average time costs after simulation are in general
slightly higher for the continuous distribution due to the fact that the uncongested
time is only the minimal travel time and most travel times are higher. Secondly, in
cases where the number of infeasible scenarios is higher than 5% but the cost
increase is very low, the exponential distribution tends to decrease the number of
infeasible scenarios. The explanation for this is probably that here only the discrete
disrupted travel time causes infeasibility which has now a lower probability of
occurrence since the probability is distributed over a bigger interval. Thirdly, if the
travel time border causing infeasibility is located between uncongested and
congested time, the number of infeasible runs is increasing in case of continuous
distribution. In that case more re-planning is necessary which results in more direct
trucks especially in smaller networks. This is also visible in Table 5 where the
<table>
<thead>
<tr>
<th>Services, orders</th>
<th>Weight = (1, 0, 0) (transportation costs)</th>
<th>Weight = (0, 0, 1) (CO$_2$e emissions)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Optimization objective value (€)</td>
<td>SAA objective value (€)</td>
</tr>
<tr>
<td>50, 1</td>
<td>1740</td>
<td>1740</td>
</tr>
<tr>
<td>50, 2</td>
<td>2485</td>
<td>2485</td>
</tr>
<tr>
<td>50, 5</td>
<td>7775.5</td>
<td>7775.5</td>
</tr>
<tr>
<td>50, 10</td>
<td>14,746.5</td>
<td>14,866.5</td>
</tr>
<tr>
<td>50, 20</td>
<td>31,072</td>
<td>31,192</td>
</tr>
<tr>
<td>100, 1</td>
<td>1740</td>
<td>1740</td>
</tr>
<tr>
<td>100, 2</td>
<td>2485</td>
<td>2485</td>
</tr>
<tr>
<td>100, 5</td>
<td>7205.5</td>
<td>7205.5</td>
</tr>
<tr>
<td>100, 10</td>
<td>13,651.5</td>
<td>13,768.5</td>
</tr>
<tr>
<td>100, 20</td>
<td>27,463</td>
<td>27,580</td>
</tr>
<tr>
<td>250, 1</td>
<td>1740</td>
<td>1740</td>
</tr>
<tr>
<td>250, 2</td>
<td>2485</td>
<td>2485</td>
</tr>
<tr>
<td>250, 5</td>
<td>7205.5</td>
<td>7205.5</td>
</tr>
<tr>
<td>250, 10</td>
<td>13,337.5</td>
<td>X</td>
</tr>
<tr>
<td>250, 20</td>
<td>26,991</td>
<td>X</td>
</tr>
</tbody>
</table>

X: not solved to optimality due to the time limit (60 min) or memory problem
unreliable plans are included in instances with 10 and especially with 20 orders, so that the costs in case of exponential distribution are lower due to increased use of cheaper direct trucks in comparison to discrete distribution for the instance with 100 services and 20 orders.

In order to better illustrate some of the transportation plans and possible changes due to unreliability as described in the previous paragraphs, selected orders are presented in Table 6 with their respective plans for each individual objective resulting from optimization model and also from the changes after their evaluation by the simulation model. These orders were tested on a network with 20 terminals and 100 services.

![Travel time distribution](image)

**Table 5** Results for time optimization for different number of services and different distributions

<table>
<thead>
<tr>
<th>Services</th>
<th>Orders</th>
<th>Optimization objective value (€)</th>
<th>Discrete three-point distribution</th>
<th>Continuous shifted exponential distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Simulation objective value (€)</td>
<td>Difference optimization versus simulation (%)</td>
<td>Simulation objective value (€)</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>65.24</td>
<td>67.43</td>
<td>3.36</td>
</tr>
<tr>
<td>50</td>
<td>2</td>
<td>149.69</td>
<td>152.35</td>
<td>1.77</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>541.82</td>
<td>390.45</td>
<td>−27.94</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>1099.54</td>
<td>960.46</td>
<td>−12.65</td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>2300.92</td>
<td>2001.10</td>
<td>−13.03</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>65.24</td>
<td>67.43</td>
<td>3.36</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>149.69</td>
<td>152.35</td>
<td>1.77</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>366.57</td>
<td>372.96</td>
<td>1.74</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>533.63</td>
<td>547.63</td>
<td>2.62</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>1229.94</td>
<td>1368.77</td>
<td>11.28</td>
</tr>
</tbody>
</table>

![Discrete versus continuous travel time distribution](image)

**Fig. 5** Discrete versus continuous travel time distribution. a Discrete three-point distribution, b continuous shifted exponential distribution
As the routes for individual orders show, the plans and their reliability can vary significantly. Whereas in case of order 1 a direct truck service 94 is the best option for all objectives to transport goods on the short distance between Koper and Trieste, a direct train service 40 is not the preferred option for time optimization in case of order 2. The main reason for this is the late departure of the cost and emissions minimizing train service, so that a combination of three trucks (84, 100 and 81) with flexible departures brings the goods faster to the customer but results in more than 115% increase in total costs. However, as this route does not show any infeasible runs in the simulation model, it is evaluated as reliable and therefore can be used since time is the only important factor here.

This is not true for order 3, where the combination of train services 34 and 37 between Duisburg and Plzen is not reliable because of the delays of the first service and short buffer time at the intermediate terminal. In addition, the due time of the order might be also violated if service 37 is delayed. These infeasibilities lead to transportation costs which are on average more than 5% higher in comparison to the costs resulting from optimization. Therefore an alternative plan has to be found, which in this case might be the combination of train services 71 and 37 that is also used for emission minimization. The only difference here is that service 71 departs earlier than the original service 34 and therefore there is enough time for transshipment at the intermediate terminal. The unreliable simulation runs result only from an occasional delay of service 37 at the destination which might be

<table>
<thead>
<tr>
<th>Order</th>
<th>Weight $(w_1, w_2, w_3)$</th>
<th>Origin, destination</th>
<th>Optimal services per order after optimization</th>
<th>Total optimization costs (€)</th>
<th>Average number of infeasible simulation runs</th>
<th>Average total costs after simulation (€)</th>
<th>Diff. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(1, 0, 0)</td>
<td>Koper(SI)</td>
<td>94</td>
<td>515.94</td>
<td>0</td>
<td>516.08</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0, 1, 0)</td>
<td>Trieste(IT)</td>
<td>94</td>
<td>515.94</td>
<td>0</td>
<td>516.07</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0, 0, 1)</td>
<td></td>
<td>94</td>
<td>515.94</td>
<td>0</td>
<td>516.07</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>(1, 0, 0)</td>
<td>Villach(AT)</td>
<td>40</td>
<td>1517.14</td>
<td>0</td>
<td>1521.4</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>(0, 1, 0)</td>
<td>Wels(AT)</td>
<td>84, 100, 81</td>
<td>3301.39</td>
<td>0</td>
<td>3302.47</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0, 0, 1)</td>
<td></td>
<td>40</td>
<td>1517.14</td>
<td>0</td>
<td>1521.92</td>
<td>0.31</td>
</tr>
<tr>
<td>3</td>
<td>(1, 0, 0)</td>
<td>Duisburg(DE)</td>
<td>34, 37</td>
<td>953.07</td>
<td>11</td>
<td>1002.93</td>
<td>5.23</td>
</tr>
<tr>
<td></td>
<td>(0, 1, 0)</td>
<td>Plzen(CZ)</td>
<td>95, 96, 98, 75</td>
<td>2823.83</td>
<td>0</td>
<td>2826.69</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>(0, 0, 1)</td>
<td></td>
<td>71, 37</td>
<td>960.79</td>
<td>6</td>
<td>961.58</td>
<td>0.08</td>
</tr>
<tr>
<td>4</td>
<td>(1, 0, 0)</td>
<td>Budapest(HU)</td>
<td>47, 1, 60</td>
<td>2332.32</td>
<td>0</td>
<td>2333.33</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(0, 1, 0)</td>
<td>Villach(AT)</td>
<td>29, 92</td>
<td>2267.94</td>
<td>0</td>
<td>2269.44</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>(0, 0, 1)</td>
<td></td>
<td>32, 17</td>
<td>2460.72</td>
<td>0</td>
<td>2463.56</td>
<td>0.12</td>
</tr>
<tr>
<td>5</td>
<td>(1, 0, 0)</td>
<td>Villach(AT)</td>
<td>23, 24, 38</td>
<td>675.09</td>
<td>24</td>
<td>662.77</td>
<td>-1.82</td>
</tr>
<tr>
<td></td>
<td>(0, 1, 0)</td>
<td>D. Streda(SK)</td>
<td>23, 24, 38</td>
<td>675.09</td>
<td>26</td>
<td>658.7</td>
<td>-2.42</td>
</tr>
<tr>
<td></td>
<td>(0, 0, 1)</td>
<td></td>
<td>23, 24, 38</td>
<td>675.09</td>
<td>25</td>
<td>660.56</td>
<td>-2.15</td>
</tr>
</tbody>
</table>
acceptable since the only more reliable alternative in this case would be the direct truck with higher costs and emissions.

Another unreliable route is computed by the optimization model for order 5 where the same route is proposed for all three objectives, combining three train services 23, 24 and 38 on a route between Austria and Slovakia. Due to the delays of the trains, the plan becomes infeasible in 25% of the cases on average and the goods have to be delivered to the destination by an extraordinary truck. However, the total costs suggest that considering a direct truck (originally not included in the network) might be even a more favourable option since the total costs are decreasing.

The example of order 4 shows how different objectives influence the optimal routes when multiple reliable alternatives exist. In case of cost optimization, goods are transported by truck 47 to the port in Budapest from where they are brought by vessel 1 to Vienna and then delivered to Villach by train 60. If time is important, it might be faster to go by train 29 to Munich and then use a fast truck 92 to Villach. In case of emission minimization, the cost-optimal route is replaced by two electric trains 32 and 17 with transshipment in Wels due to the higher emissions of inland vessel in the upstream direction between Budapest and Vienna.

5 Conclusions

Intermodal transportation can be considered as an alternative to the traditionally used road transportation which is exposed to delays and disruptions due to increasing traffic volumes on the transportation network. Although intermodal solutions offer efficient transportation chains combining cheap and sustainable modes (e.g., train, inland waterway) with last-mile truck services, the complexity of the network requires higher coordination effort and reliable transportation plans that can ensure timely delivery of the goods to the terminals. In order to solve this problem we proposed a hybrid simulation and optimization approach to investigate the intermodal transportation planning problem in a stochastic and dynamic environment. In order to provide greener transportation solutions for the investigated problem, we aimed to optimize the transportation plan by increasing the reliability in terms of travel time uncertainty. We have validated the proposed methodology on a real-life case study. Our results indicate that the proposed methodology is capable of generating reliable solutions that compare favorably with solutions generated by a deterministic formulation. To evaluate the effectiveness of the proposed approach, we used different sets of instances based on real geographic data. The results show that the proposed algorithm is highly effective in finding good-quality and reliable solutions on instances with up to 20 locations and 250 transportation services.

Based on the used methodology, the following conclusions can be presented:

- The integration among transportation modes, such as road, rail and waterway transportation, potentially increases harmonization efforts for multiple transportation users. This could be achieved by considering not only monetary costs
but also customer satisfaction (i.e., timely delivery), and environmental concerns of the transportation.

- Despite the advantages of shifting of freight from road to rail or waterway transportation, there are still a number of challenges which need to be overcome. It is especially important to improve the coordination and reliability of the transportation by providing reliable and consistent data (e.g., lack of real-time data) and developing models which are capable of creating reliable transportation plans. In this way the number of disruptions within supply chains can be reduced which results in smoother processes and cost savings.

- Although the presented approach attempts to eliminate the need for re-planning in case of disruptions during the transportation execution, it is not capable of capturing all possible disruptions. Especially disruptions which occur very rarely but have high impact on travel times cannot be considered directly in the proposed model since the resulting plans would include unnecessary long buffer times. Therefore an effective re-planning algorithm has to be developed which can be executed quickly in cases where a disruption not covered by the travel time distribution happens.

- The presented results show that the combination of optimization and simulation is capable to find the solution of complex instances and gives a better overview about the reliability of the plans and possible reasons for disruptions. The combination of two criteria for reliability evaluation enables to identify plans that might become infeasible but where the additional costs for unreliability are very low and much cheaper in comparison to the next best alternative. In contrast to that, the model also shows routes where the intermodal alternative is highly unreliable and therefore a use of direct truck transportation might be more favorable for the customer.

Acknowledgements We thank the Editor-in-Chief Professor Hans-Otto Guenther and the two anonymous reviewers for their constructive comments, which helped us in improving the paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References


Hybrid simulation and optimization approach for green…

AnyLogic (2016) Copyright © AnyLogic North America, LLC
IBM ILOG (2016) Copyright © International Business Machines Corporation 1987


Hybrid simulation and optimization approach for green…


Martin Hrušovský is a PhD student and also research and teaching associate working at the Institute for Production Management (Vienna University of Economics and Business) since 2011. In his research he is currently dealing with mathematical optimization models for intermodal transport planning including uncertainty using stochastic methods. Besides that his research interests include risk management and sustainability in supply chains. Martin is and has been teaching bachelor courses with the focus on basics in Operations Management, such as capacity management, inventory management, production planning and demand forecasting.

Emrah Demir is a Senior Lecturer (Associate Professor) in the Logistics and Operations Management Section of the Cardiff Business School. He holds BSc and MSc degrees in Industrial Engineering from Baskent University (Turkey), and a PhD in Management Science from University of Southampton. Dr Demir’s main research interest is positioned within the field of green logistics with respect to negative externalities of freight transportation. He is author and co-author to numerous research papers, book chapters and technical reports. He also serves as a reviewer in several international journals of operational research, transportation and logistics.

Werner Jammernegg studied Business Administration at University of Graz and Technical Mathematics (Information Processing) at Graz University of Technology, Austria. Besides two master degrees, he holds a doctoral degree in Business Administration and a habilitation degree in Operations Research and Statistics. He has been Professor of Operations Management at WU (Vienna University of Economics and Business), Austria, since 1990. From 2010 to 2015 he was director of the Master’s of Science program in Supply Chain Management at WU. Werner research interests include Production and Operations Management, Supply Chain Management and Stochastic Modelling. Examples of current research topics are stochastic inventory models, economic and environmental sustainability in supply chains, resource efficiency in supply chains and dependency of supply chain risks.

Tom Van Woensel is a full professor of freight transport and logistics in the OPAC group of the department of Industrial Engineering and Innovation Sciences at the Eindhoven University of Technology in the Netherlands. Professor Tom Van Woensel holds a M.Sc. in Applied Economics and a Ph.D. in Operations Management from the University of Antwerp, Belgium. He is a board member of the European Supply Chain Forum and an account manager for Bausch and Lomb, Kuehne+Nagel and Access Business Group with a specialization in distribution logistics, retail operations, and manufacturing systems. He also heads the Smart Logistics Lab, a joint effort of around 16 people doing research in transport and logistics. His research is mainly focused on Freight Transport and Logistics.