Global synchromodal shipment matching problem with dynamic and stochastic travel times
a reinforcement learning approach

Guo, W.; Atasoy, B.; Negenborn, R. R.

DOI
10.1007/s10479-021-04489-z

Publication date
2022

Document Version
Final published version

Published in
Annals of Operations Research

Annals of Operations Research
https://doi.org/10.1007/s10479-021-04489-z

ORIGINAL RESEARCH

Global synchromodal shipment matching problem with dynamic and stochastic travel times: a reinforcement learning approach

W. Guo[1] · B. Atasoy[2] · R. R. Negenborn[2]

Accepted: 7 December 2021

Abstract

Global synchromodal transportation involves the movement of container shipments between inland terminals located in different continents using ships, barges, trains, trucks, or any combination among them through integrated planning at a network level. One of the challenges faced by global operators is the matching of accepted shipments with services in an integrated global synchromodal transport network with dynamic and stochastic travel times. The travel times of services are unknown and revealed dynamically during the execution of transport plans, but the stochastic information of travel times are assumed available. Matching decisions can be updated before shipments arrive at their destination terminals. The objective of the problem is to maximize the total profits that are expressed in terms of a combination of revenues, travel costs, transfer costs, storage costs, delay costs, and carbon tax over a given planning horizon. We propose a sequential decision process model to describe the problem. In order to address the curse of dimensionality, we develop a reinforcement learning approach to learn the value of matching a shipment with a service through simulations. Specifically, we adopt the Q-learning algorithm to update value function estimations and use the ϵ-greedy strategy to balance exploitation and exploration. Online decisions are created based on the estimated value functions. The performance of the reinforcement learning approach is evaluated in comparison to a myopic approach that does not consider uncertainties and a stochastic approach that sets chance constraints on feasible transshipment under a rolling horizon framework.

Keywords Global synchromodal shipment matching · Dynamic and stochastic travel times · Sequential decision process · Reinforcement learning · Q-learning

✉ W. Guo
guo.wenjing@courrier.uqam.ca

[1] CIRRELT and Department of Analytics, Operations and Information Technologies, School of Management Sciences, University of Quebec at Montreal, Montreal, Canada

[2] Department of Maritime and Transport Technology, Delft University of Technology, Delft, The Netherlands

Published online: 21 January 2022

1 Introduction

With the increasing volumes of global trade and the trend towards time-sensitive shipments, efficient global transportation becomes increasingly important in global supply chains (Yang et al. 2018). Synchromodal transportation is the provision of efficient, effective, and sustainable transport services thanks to the horizontal and vertical collaboration among players (SteadieSeifi et al. 2014). However, implementing synchromodality in global transport is still challenging from several aspects, including: the design of collaboration contracts and pricing strategies that ensure fairness and attractiveness among players at the strategic level (Lee and Song 2017); integrated service network design that determines service frequencies and time schedules at the tactical level (Meng et al. 2014); and integrated transport plan that assigns specific shipments with transport services under a dynamic and stochastic environment at the operational level (SteadieSeifi et al. 2014). This paper investigates a global synchromodal shipment matching problem with dynamic and stochastic travel times at the operational level.

With the development of digitization in the logistics industry, increasingly online platforms have appeared in freight transportation (Meng et al. 2019), such as Uber Freight and Quicargo. We consider a platform owned by a global operator that receives shipment requests from shippers and receives service offers from carriers, as shown in Fig. 1. The global operator could be a logistics service provider or an alliance formed by multiple carriers, such as Maersk and COSCO Shipping lines. A shipment is defined as a batch of containers that must be transported from its origin to its destination within a specific time window. For example, shipment r1 consists of 30 containers which require to be transported from origin terminal 1 to destination terminal 5 with a release time of Jan 1, 9:00, and a lead time of 840 h. A service is characterized by its mode, origin terminal, destination terminal, time schedule, and free capacity. For example, ship service s1 with capacity 200 TEU (twenty-foot equivalent unit) will depart from terminal 1 on Jan 2, 11:00, and arrive at terminal 5 with an estimated travel time of 680 h. The platform aims to provide optimal acceptance and matching decisions for all shipments involved in the global synchromodal transport network. A match between a shipment and a service means that the shipment will be transported by the service from the service's origin to the service's destination. The platform combines the matched services into itineraries to provide integrated transport for global shipments. For instance, shipment r2 will be transported by barge service s2 from origin terminal 2 to transshipment terminal 1 and by ship service s1 from terminal 1 to destination terminal 5. The objective of the platform is to maximize the total profits.

In this paper, the matching of shipments with services in an integrated global synchromodal transport network with the aim to maximize total profits is defined as the global synchromodal shipment matching (GSSM) problem. The GSSM problem considers multiple shipments with soft time windows, multimodal services with capacity limitations and time schedules. Shipments with different origins and destinations can be consolidated into the same service; the transshipment operations between different services are available for all shipments. From the mathematical modeling perspective, the GSSM problem belongs to the category of multi-commodity multimodal container routing problems (Sun et al. 2015). In the literature, Chang (2008) considered the routing choices for multiple commodities in a multimodal network; Sun and Lang (2015) investigated the transshipment operations between time scheduled services (i.e., trains) and time flexible services (i.e., trucks); Guo et al. (2020a) considered carbon tax charged by governmental institutions and delay costs paid to shippers in addition to travel costs, transfer costs, and storage costs in the objective function. In these studies, travel times are considered as static and deterministic information.

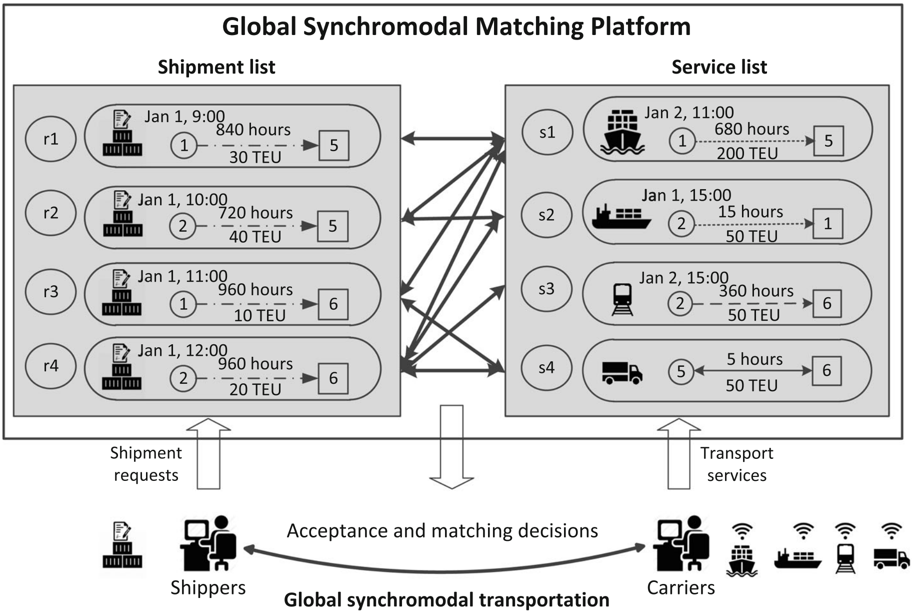


Fig. 1 A global synchromodal matching platform

In practice, at the time of transport planning, travel time uncertainties are quite common resulting from weather conditions, traffic congestion, and infrastructure capacity uncertainties (Demir et al. 2016). Due to travel time uncertainty and the utilization of multimodal services in global transportation, the services assigned to shipments might become suboptimal or even infeasible at transshipment terminals. Thanks to the development in data analytics, probability distributions of uncertainties are often available to transport systems (Gendreau et al. 2016). Stochastic approaches that incorporate stochastic information of travel times have been well investigated in vehicle routing problems (Ehmke et al. 2015; Li et al. 2010). However, only a few studies investigated synchromodal shipment routing with stochastic travel times. Specifically, Demir et al. (2016) developed a sample average approximation method to generate robust transport plans for all shipments involved in an inland synchromodal transportation network by considering possible delays and the probability of missing a service. Hrušovský et al. (2016) proposed a hybrid approach that combines an optimization model generating deterministic routes and a simulation model evaluating the feasibility of transport plans under travel time uncertainties. Generally, the decisions made by stochastic approaches are referred to as a-prior or non-adaptive decisions since decisions are not updated during the transport processes (Ritzinger et al. 2015).

In addition to stochasticity, the travel times of services are revealed dynamically during the execution of transport plans. Matching decisions can be updated before shipments arrive at their destination terminals. Methods that create adaptive decisions by incorporating dynamic as well as stochastic information in decision-making processes have attracted increasing interest in the literature (Ritzinger et al. 2015). With regards to dynamic and stochastic synchromodal shipment routing, van Riessen et al. (2016) designed a decision tree to instantaneously allocate incoming containers to inland services by analyzing the solution structure of an optimization model on historical data of transport demand. Rivera and

Mes (2017) developed an approximate dynamic programming algorithm to assign the newly arrived shipments to a barge or trucks by incorporating the probability distributions of future shipments. Guo et al. (2021a) proposed a stochastic programming-based rolling horizon approach to create online matches between newly received shipments and services in an inland synchromodal transportation network by integrating sampled shipments appearing in the near future. However, none of the above studies considered dynamic and stochastic travel times.

In the literature, most similar to our work are the work of Yee et al. (2021), Guo et al. (2020b), and Guo et al. (2021b). Yee et al. (2021) developed a Markov decision process (MDP) model to determine the optimal modal choice for a single shipment in a multimodal network based on real-time and stochastic information of travel times. As the MDP model runs for a single shipment and uses a limited number of scenarios to represent stochastic travel times, it is solved optimally by means of backtracking. In contrast, our work considers multiple shipments that can be consolidated into the same service at transshipment terminals. While some of the shipments arrive at intermediate terminals, other shipments might be in transit. Therefore, we develop a sequential decision process (SDP) model to track the states of multiple shipments and the assigned services. Since the positions of shipments at the next stage are not only decided by the decisions made at the current stage but also the decisions made in history, the SDP model developed in this paper does not have the Markov property. On the other hand, we adopt continuous probability distributions to describe the stochasticity of travel times which cause the infinite number of scenarios for each service. To be able to address the curse of dimensionality, we develop a reinforcement learning approach (RLA) to estimate the value functions.

Guo et al. (2020b) consider the same problem settings as our work. However, they developed a chance-constrained programming model to address travel time uncertainties in global synchromodal transportation at each decision epoch of a rolling horizon framework. As an extension of Guo et al. (2020b), Guo et al. (2021b) consider dynamic and stochastic travel times as well as shipment requests in global synchromodal transportation. They developed a hybrid stochastic approach to address travel time and shipment request uncertainties integrally. Under both of the approaches, shipment routes are updated only when infeasible transshipment happens. Besides, while their approaches are restricted to normal distributions of travel times, the RLA developed in this paper can be applied to any distributions.

This paper contributes to the state of the art by developing the RLA to solve the shipment routing problem in global synchromodal transportation with dynamic and stochastic travel times. The RLA learns the value of matching a shipment with a service through simulations. Online decisions are created based on the estimated value functions. To the best of our knowledge, this is the first work that applies RLA in the synchromodal shipment routing domain. The performance of the RLA is evaluated in comparison to the myopic approach (MA) proposed by Guo et al. (2020a) that does not consider travel time uncertainties and the stochastic approach (SA) proposed by Guo et al. (2020b) that sets chance constraints on feasible transshipment under a rolling horizon framework. While MA and SA require online computations when dynamic travel times are revealed, RLA determines the behavior policy that maps a perceived state to a decision before the execution of transport plans. Thanks to the developed methodology, the platform can adapt shipment matching decisions immediately based on real-time travel time information to achieve better performance in total profits over a given planning horizon.

The remainder of this paper is structured as follows. In Sect. 2, we provide a detailed problem description, followed by a sequential decision process model in Sect. 3. In Sect. 4, we develop the reinforcement learning approach. In Sect. 5, we present the experimental results. Finally, in Sect. 6, we provide concluding remarks and directions for future research.

2 Problem description

Let N be the set of terminals. Each terminal $i \in N$ is characterized by its loading/unloading cost lc_i^m, loading/unloading time lt_i^m with mode $m \in M = \{\text{ship}, \text{ barge}, \text{ train}, \text{ truck}\}$, and storage cost per container per hour c_i^{storage}. We assume terminal operators provide unlimited loading/unloading and storage capacity to the global operator.

Let R be the set of shipments. Each shipment $r \in R$ is characterized by its origin terminal o_r, destination terminal d_r, container volume u_r, release time $\mathbb{T}_r^{\text{release}}$ (i.e., the time when the shipment is available for transport process), lead time $\mathbb{T}_r^{\text{lead}}$, freight rate p_r, and delay cost c_r^{delay}. The due time of shipment r is $\mathbb{T}_r^{\text{due}} = \mathbb{T}_r^{\text{release}} + \mathbb{T}_r^{\text{lead}}$.

Let S be the set of services. Each service $s \in S$ is characterized by its mode $m_s \in M$, origin terminal o_s, destination terminal d_s, time-dependent free capacity U_s^t at decision epoch t, estimated travel time t_s, travel cost c_s, and generation of carbon emissions e_s. We consider ship, barge and train services as time scheduled services with scheduled departure time D_s and scheduled arrival time A_s for $s \in S^{\text{ship}} \cup S^{\text{barge}} \cup S^{\text{train}}$. Each truck service consists of a fleet of trucks that have flexible departure times.

Due to travel time uncertainty at the time of planning, the arrival times of services are also uncertain. The probability distributions of travel and arrival times of services are assumed available. Travel time uncertainty in global synchromodal transportation may lead to infeasible transshipment during the transport process in addition to the commonly studied outcome of late or early delivery at destinations (Li et al. 2010; Rodrigues et al. 2019). An illustrative example is shown in Fig. 2. A shipment is planned to be transported by a train service from its origin terminal to port A, by a ship service from port A to port B, and by two barge services from port B to its destination terminal according to fixed time schedules. The outcomes of travel time uncertainty in global synchromodal transportation include late delivery at destination terminal under realization 1, which causes delayed costs; early delivery at destination terminal under realization 2, which causes storage costs; and infeasible transshipment at port B under realization 3, which requires re-planning from port B to destination terminal.

The actual travel and arrival times of services are assumed known immediately when services arrive at their destination terminals. Once shipments arrive at a new terminal, the platform needs to decide on the next service that moves a shipment leaving its current terminal. Shipments might be moved following their transportation plans, or they might be moved by a new service with updated plans. An illustrative example of dynamic shipment routing in global synchromodal transportation is shown in Fig. 3. At time 100, a shipment arrives at inland terminal A with a truck service. Instead of following the transport plan that moves the shipment from inland terminal A to port B, the platform selects a train service that moves the shipment to inland terminal B. At time 120, the shipment arrives at inland terminal B. The same decision process continues until all shipments arrive at their destination terminals.

The objective of the global synchromodal matching platform is to maximize the total profits by optimizing acceptance and matching decisions over a given planning horizon T. The total profits are formed by a combination of revenues received from shippers, travel costs paid to carriers, transfer costs and storage costs paid to terminal operators, delay costs paid to shippers, and carbon tax charged by institutional authorities.

The notation used in this paper is shown in Table 1.

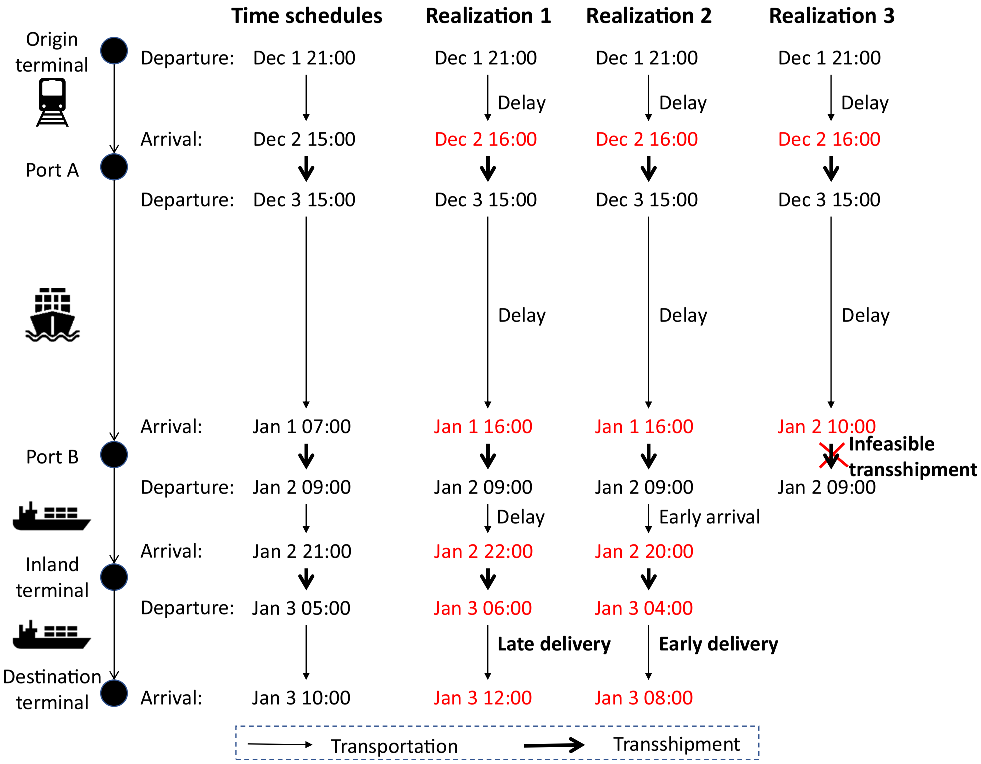


Fig. 2 Possible outcomes of travel time uncertainty in global synchromodal transportation

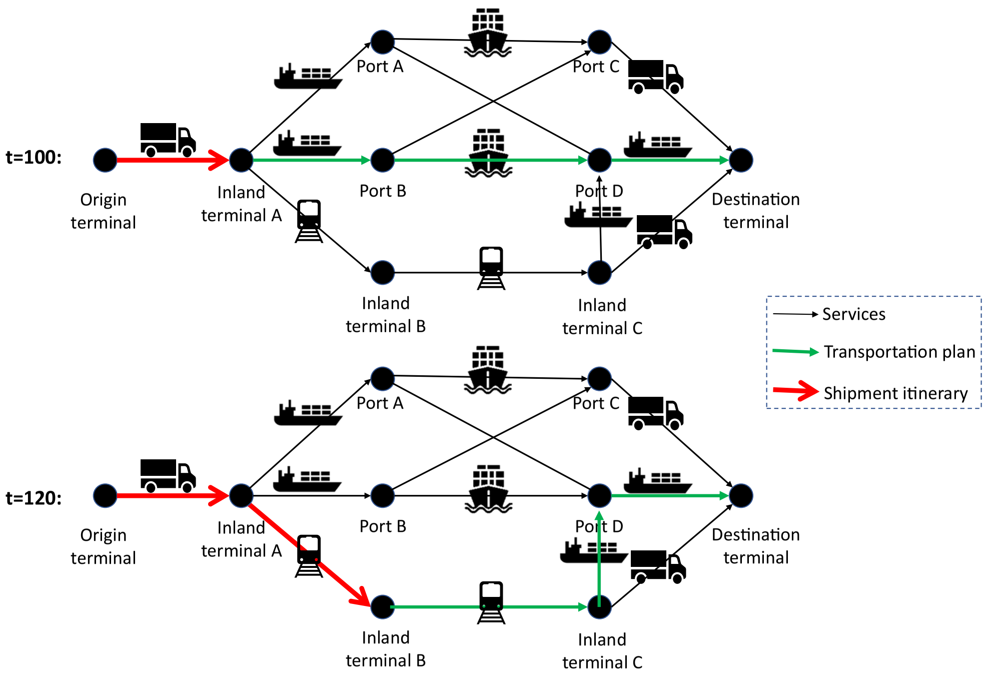


Fig. 3 Dynamic shipment routing in global synchromodal transportation

Table 1 Notation

Sets	
N	Terminals
R	Shipments
R^t	Accepted shipments arrival at new terminals during time interval $(t-1, t]$
M	Modes, $M = \{\text{ship}, \text{ barge}, \text{ train}, \text{ truck}\}$
S	Services, $S = S^{\text{ship}} \cup S^{\text{barge}} \cup S^{\text{train}} \cup S^{\text{truck}}$
S_i^+	Services departing from terminal i, $S_i^+ = S_i^{+\text{ship}} \cup S_i^{+\text{barge}} \cup S_i^{+\text{train}} \cup S_i^{+\text{truck}}$
S_i^-	Services arriving at terminal i, $S_i^- = S_i^{-\text{ship}} \cup S_i^{-\text{barge}} \cup S_i^{-\text{train}} \cup S_i^{-\text{truck}}$
S^{-t}	Services arriving at their destination terminals during time interval $(t-1, t]$
$\mathbf{W}^t$	Exogenous information received during time interval $(t-1, t]$
$\mathbf{F}^t$	State of the platform at stage t
Deterministic parameters	
T	Length of the planning horizon
o_r	Origin terminal of shipment $r \in R$, $o_r \in N$
d_r	Destination terminal of shipment $r \in R$, $d_r \in N$
u_r	Container volume of shipment $r \in R$
$\mathbb{T}_r^{\text{release}}$	Release time of shipment $r \in R$
$\mathbb{T}_r^{\text{due}}$	Due time of shipment $r \in R$
$\mathbb{T}_r^{\text{lead}}$	Lead time of shipment $r \in R$, $\mathbb{T}_r^{\text{lead}} = \mathbb{T}_r^{\text{due}} - \mathbb{T}_r^{\text{release}}$
p_r	Freight rate of shipment $r \in R$
c_r^{delay}	Delay cost of shipment $r \in R$ per container per hour overdue
I_r^t	Itinerary of request $r \in R$ consists of matched services
ϕ_r^t	Position of shipment $r \in R$ at decision epoch $t \in \{0, 1, ..., T\}$
$A_{r\phi_r^t}$	Arrival time of request $r \in R^t$ at terminal ϕ_r^t.
m_s	Mode of service $s \in S$, $m_s \in M$
o_s	Origin terminal of service $s \in S$, $o_s \in N$
d_s	Destination terminal of service $s \in S$, $d_s \in N$
U_s^t	Free capacity of service $s \in S$ at decision epoch $t \in \{0, 1, ...T\}$
c_s	Travel cost of service $s \in S$ per container
e_s	Carbon emissions of service $s \in S$ per container
D_s	Scheduled departure time of service $s \in S \backslash S^{\text{truck}}$
A_s	Scheduled arrival time of service $s \in S \backslash S^{\text{truck}}$
t_s	Estimated travel time of service $s \in S$
$\bar{A}_s$	Actual arrival time of service $s \in S \backslash S^{\text{truck}}$
$\bar{A}_{rs}$	Actual arrival time of service $s \in S^{\text{truck}}$ with shipment $r \in R$
$\bar{t}_s$	Actual travel time of service $s \in S$
lc_i^m	Loading/unloading cost per container at terminal $i \in N$ with mode $m \in M$
lt_i^m	Loading/unloading time at terminal $i \in N$ with mode $m \in M$
c_i^{storage}	Storage cost at terminal i per container per hour
c^{emission}	Activity-based carbon tax charged by institutional authorities
B	A large number used for binary constraints

Table 1 continued

Sets	
Random variables	
$\tilde{t}_s$	Travel time of service $s \in S$
$\tilde{A}_s$	Arrival time of service $s \in S \backslash S^{\text{truck}}$ at destination terminal d_s
Variables	
y_r	Binary variable; 1 if shipment $r \in R$ is accepted
x_{rs}^t	Binary variable; 1 if shipment $r \in R^t$ is matched with service $s \in S$, 0 otherwise
D_{rs}	Departure time of truck service $s \in S^{\text{truck}}$ with shipment $r \in R$
$\tilde{\mathbb{T}}_r^{\text{delay}}$	Delay of shipment $r \in R$ at destination terminal $d_r \in N$

3 Sequential decision process model

In this section, we formulate a sequential decision process (SDP) model to describe the interaction between the global synchromodal matching platform and the transport network. The flowchart of the SDP is shown in Fig. 4. There are seven fundamental elements in the SDP model: decision epochs, exogenous information, states, decision variables, transition functions, costs, and objective functions (Powell 2019). A brief summary of these elements is as follows:

- *Decision epochs* We define t as the points in time at which decisions are made, referred to as the decision epoch, $t \in \{0, 1, ..., T\}$. Therefore, the planning horizon is divided into T consecutive time intervals.

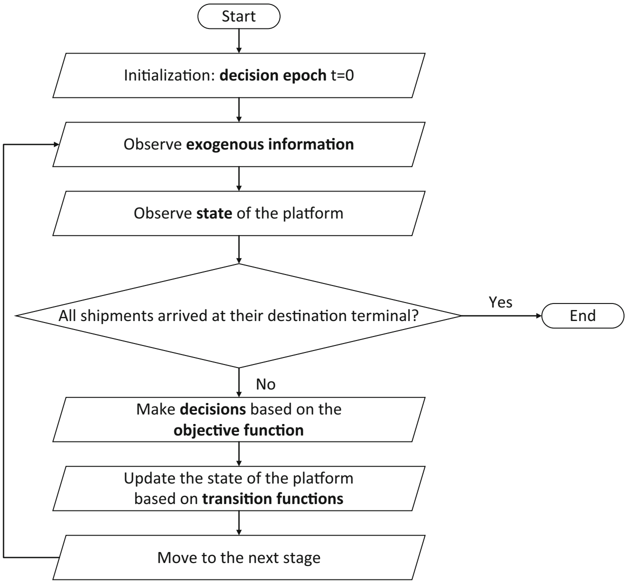

Start
Initialization: decision epoch t=0
Observe exogenous information
Observe state of the platform
All shipments arrived at their destination terminal?
Yes
End
No
Make decisions based on the objective function
Update the state of the platform based on transition functions
Move to the next stage

Fig. 4 Flowchart of the sequential decision process model

- *Exogenous information* The exogenous information $\mathbf{W}^t$ consists of all the new information that first becomes known at decision epoch t. We define $\bar{A}_s$ as the actual arrival time of service $s \in S \backslash S^{\text{truck}}$, and define $\bar{A}_{rs}$ as the actual arrival time of service $s \in S^{\text{truck}}$ with shipment $r \in R$. We represent $\mathbf{W}^t = \left[\bar{A}_s\right]_{s \in S^{-t}} \cup \left[\bar{A}_{rs}\right]_{s \in S^{-t}}$, where $S^{-t} = \{s \in S | t - 1 < \bar{A}_s / \bar{A}_{rs} \leq t\}$ is the set of services arriving their destination terminals during time interval $(t-1, t]$.
- *States* The state $\mathbf{F}^t$ of the global synchromodal matching platform contains all the information that is necessary and sufficient to model the platform at decision epoch t. We distinguish between the initial state $\mathbf{F}^0$ and the dynamic state $\mathbf{F}^t$ for $t > 0$. The initial state contains all the deterministic sets and parameters $\{N, R, S, T\}$, and probability distributions of random variables $\left[\tilde{t}_s\right]_{\forall s \in S}$ and $\left[\tilde{A}_s\right]_{\forall s \in S \backslash S^{\text{truck}}}$. The dynamic state $\mathbf{F}^t$ contains the information that is evolving over time, including the free capacity U_s^t of service $s \in S$, the itinerary I_r^t of shipment $r \in R$ at decision epoch t, the position ϕ_r^t of shipment $r \in R$ at decision epoch t, the set of accepted shipments R^t arrived at new terminals during time interval $(t-1, t]$, and the arrival time $A_{r\phi_r^t}$ of shipment $r \in R^t$ at terminal ϕ_r^t. We define $\phi_r^0 = o_r$, $A_{ro_r} = \mathbb{T}_r^{\text{release}}, \forall r \in R$. The system is terminated when all the shipments arrive at their destination terminals.
- *Decision variables* Let y_r be the binary variable which is 1 if shipment $r \in R$ is accepted, 0 otherwise. Let x_{rs}^t represent the match between shipment $r \in R$ and service $s \in S$. At decision epoch 0, the platform needs to decide on acceptance decision y_r and matching decision $\left[x_{rs}^0\right]_{s \in S_{o_r}^+}$ for shipment $r \in R$. At decision epoch $t \in \{1, ..., T\}$, the platform needs to decide on the matching decision x^t for accepted shipments R^t. Let D_{rs} be the variable that indicates the departure time of service $s \in S^{\text{truck}}$ with shipment $r \in R^t$. The decisions are restricted by the time-spatial compatibility between shipments and services, and free capacities of services at decision epoch t. The decision vectors x^t consist of all the decisions at decision epoch t as seen in (1), subject to constraints (2–5), which define the feasible decision space.

$$x^t = \left[x_{rs}^t\right]_{\forall r \in R^t, s \in S^+_{\phi_r^t}} \tag{1}$$

subject to

$$\sum_{s \in S^+_{\phi_r^t}} x_{rs}^t = 1, \quad \forall r \in R^t, \tag{2}$$

$$\sum_{s \in S^+_{\phi_r^t}} x_{rs}^t = 0, \quad \forall r \in R^t, d_s \in \{\phi_r^0, ..., \phi_r^t\}, \tag{3}$$

$$\sum_{r \in R^t} x_{rs}^t u_r \leq U_s^t, \quad \forall s \in S, \tag{4}$$

$$A_{r\phi_r^t} + lt^{m_s}_{\phi_r^t} \leq D_s + B(1 - x_{rs}^t), \quad \forall r \in R^t, s \in S^+_{\phi_r^t} \backslash S^{+\text{truck}}_{\phi_r^t}, \tag{5}$$

$$A_{r\phi_r^t} + lt^{m_s}_{\phi_r^t} \leq D_{rs} + B(1 - x_{rs}^t), \quad \forall r \in R^t, s \in S^{+\text{truck}}_{\phi_r^t}. \tag{6}$$

Constraints (2) ensure that a service will be selected to transport shipment $r \in R^t$ departing from its current terminal ϕ_r^t. Constraints (3) are designed to eliminate subtours. Constraints (4) ensure that the total container volumes of shipments matched with service s do not exceed its free capacity at decision epoch t. Constraints (5-6) ensure that the arrival time of shipment r at terminal ϕ_r^t plus loading time must be earlier than the scheduled departure time of service $s \in S_{\phi_r^t}^{+}$ if shipment r will be transported by service s leaving terminal ϕ_r^t. Here, B is a large number used for binary constraints.

- **Transition function**. Following state $\mathbf{F}^t$, decisions $\left[y, x^0, ..., x^t\right]$, and exogenous information $\mathbf{W}^{t+1}$, the system transitions to a new state. We denote the transition function by $\mathbf{F}^{t+1} = f(\mathbf{F}^t, \left[y, x^0, ..., x^t\right], \mathbf{W}^{t+1})$. The free capacity of service $s \in S$ at stage $t+1$ is decided by the free capacity of service s at decision epoch t and the matching decisions made for shipments R^t, as shown in (7). The itinerary of shipment r at stage $t+1$ is decided by the current itinerary and the matching decisions, as shown in (8). The position of shipment r at stage $t+1$ is decided by the arrival time of the matched service $s \in S^{-(t+1)}$, as shown in (9–10). Set R^{t+1} consists of the accepted shipments that arrive at new terminals at stage $t+1$, as shown in (11). Equations (12–13) represent that the arrival time of shipment $r \in R^{t+1}$ at terminal ϕ_r^{t+1} equals the actual arrival time of service $s \in S^{-(t+1)}$ plus unloading time if shipment r is transported by service s arriving terminal ϕ_r^{t+1}.

$$U_s^{t+1} = U_s^t - \sum_{r \in R^t} u_r x_{rs}^t, \quad \forall s \in S, \tag{7}$$

$$I_r^{t+1} = I_r^t \cup s, \quad \forall r \in R^t, s \in S_{\phi_r^t}^{+}, x_{rs}^t = 1, \tag{8}$$

$$\phi_r^{t+1} = d_s, \quad \forall r \in R, s \in S^{-(t+1)}, \sum_{t'=0}^{t} x_{rs}^{t'} = 1, \tag{9}$$

$$\phi_r^{t+1} = \phi_r^t, \quad \forall r \in R, \sum_{t'=0}^{t} \sum_{s \in S^{-(t+1)}} x_{rs}^{t'} = 0, \tag{10}$$

$$R^{t+1} = \{r \in R | y_r = 1, \phi_r^{t+1} \neq \phi_r^t, \phi_r^{t+1} \neq d_r\}, \tag{11}$$

$$A_{r\phi_r^{t+1}} = \bar{A}_s + lt_{\phi_r^{t+1}}^{m_s}, \quad \forall r \in R^{t+1}, s \in S^{-(t+1)} \backslash S^{\text{truck}}, \sum_{t'=0}^{t} x_{rs}^{t'} = 1, \tag{12}$$

$$A_{r\phi_r^{t+1}} = \bar{A}_{rs} + lt_{\phi_r^{t+1}}^{m_s}, \quad \forall r \in R^{t+1}, s \in S^{-(t+1)} \cap S^{\text{truck}}, \sum_{t'=0}^{t} x_{rs}^{t'} = 1. \tag{13}$$

- *Costs* Based on the state $\mathbf{F}^t$, and the decision x^t, the costs at decision epoch t can be defined as a function of $\mathbf{F}^t$ and x^t, as shown in (14).

$$\begin{aligned} \tilde{\mathbf{C}}^t(\mathbf{F}^t, x^t) = & \sum_{r \in R^t} \sum_{s \in S_{\phi_r^t}^{+}} x_{rs}^t u_r \left(c_s + lc_{o_s}^{m_s} + lc_{d_s}^{m_s} + \left(D_s - lt_{o_s}^{m_s} - A_{ro_s}\right) c_{o_s}^{\text{storage}}\right) \\ & + \sum_{r \in R^t} \sum_{s \in S_{\phi_r^t}^{+}} x_{rs}^t u_r c^{\text{emission}} e_s + \sum_{r \in R^t} \sum_{s \in S_{\phi_r^t}^{+} \cap S_{d_r}^{-}} c_r^{\text{delay}} \tilde{\mathbb{T}}_r^{\text{delay}} u_r, \end{aligned} \tag{14}$$

where

$$\tilde{\mathbb{T}}_r^{\text{delay}} \geq \tilde{A}_s + lt_{d_s}^{m_s} - \mathbb{T}_r^{\text{due}} + B(x_{rs}^t - 1), \forall r \in R^t, s \in S_{\phi_r^t}^{+}, d_s = d_r. \tag{15}$$

We use $\tilde{\mathbf{C}}^t(\mathbf{F}^t, x^t)$ to denote the costs which consist of travel costs, loading and unloading costs, storage costs, carbon tax, and delay costs. Let $\tilde{\mathbb{T}}_r^{\text{delay}}$ represent the delay in delivery of shipment $r \in R^t$ at its destination terminal d_r, which is decided by the matching decisions x^t and the arrival time $\tilde{A}_s$ of matched service $s \in S_{\phi_r^t}^{+}, d_s = d_r$.

- **Objective functions**. The objective of the SDP model is to maximize the expected total profits over the planning horizon given as follows:

$$\max_{y,x^t} \sum_{r \in R} p_r u_r y_r - \mathbb{E}_{\mathbf{W}^1,\ldots,\mathbf{W}^T|\mathbf{F}^0} \{ \sum_{t=0}^{T} \tilde{\mathbf{C}}^t (\mathbf{F}^t, x^t) | \mathbf{F}^0 \} \tag{16}$$

We refer to the objective function in (16) as the cumulative formulation. Using Bellman's principle of optimality, the optimal profits can be computed through a set of recursive equations, as seen in (17–18).

$$V(\mathbf{F}^0) = \max_{y,x^0} \sum_{r \in R} p_r u_r y_r - \mathbb{E}\{\tilde{\mathbf{C}}^0(\mathbf{F}^0, x^0)\} - \mathbb{E}_{\mathbf{W}^1}\{V(\mathbf{F}^1|\mathbf{F}^0, [y, x^0], \mathbf{W}^1)\}, \tag{17}$$

$$V(\mathbf{F}^t) = \min_{x^t} \mathbb{E}\{\tilde{\mathbf{C}}^t(\mathbf{F}^t, x^t)\} + \mathbb{E}_{\mathbf{W}^{t+1}}\{V(\mathbf{F}^{t+1}|\mathbf{F}^t, [y, x^0, ..., x^t], \mathbf{W}^{t+1})\}, \ \forall t > 0. \tag{18}$$

Here, $V(\mathbf{F}^t)$ represents the value function of being in state $\mathbf{F}^t$ at decision epoch t in the SDP model, which evaluates how good it is for the platform to be in a given state.

The complexity of the SDP model lies in several aspects. First, at each decision epoch, the costs caused by state $\mathbf{F}^t$ and decision x^t are uncertain, which relies on the arrival time of matched services. Therefore, the stage that costs $\tilde{\mathbf{C}}^t(\mathbf{F}^t, x^t)$ will be fully observed is also uncertain. Second, the decisions made at stage t not only have influence on the costs generated at the current stage $\tilde{\mathbf{C}}^t(\mathbf{F}^t, x^t)$ but also affect the future costs $V(\mathbf{F}^{t+1})$. Third, the state $\mathbf{F}^{t+1}$ of the platform at stage $t+1$ depends not only on the decisions made at stage t but also on the decisions made at previous stages $\{0, 1, ..., t-1\}$.

4 Reinforcement learning approach

Although the probability distributions of the travel and arrival times of services are assumed available, it is very difficult to obtain optimal solutions by solving the Bellman equations (18) directly via dynamic programming algorithms, known as "the curse of dimensionality" (Mes and Rivera 2017). Methods based on approximation strategies to solve SDP models have attracted increasing interest in the literature (Powell 2019). These methods can be divided into two groups: methods based on online decisions which focus on the computation when a dynamic event occurs with respect to the current system state and the available stochastic information, such as stochastic programming-based rolling horizon approaches (Guo et al. 2021a); and methods based on preprocessed decisions which estimate the value functions and determine the behavior policies before the execution of transport plans, such as reinforcement learning approaches (Sutton and Barto 2018). A policy is defined as a mapping from perceived states of the environment to decisions to be taken when in those states (Sutton and Barto 2018). In this paper, we develop a reinforcement learning approach (RLA) to solve the GSSM problem with dynamic and stochastic travel times.

The key idea of the RLA is to learn the value functions through simulations and to determine the policy that maps a state to a decision. However, estimating $V(\mathbf{F}^t)$ in the SDP model requires storing the information on travel time, arrival time, free capacity of all services, and storing the information on the position and itinerary of all shipments in the value functions. Besides, at each decision epoch, a mixed integer linear programming model needs to be solved to obtain x^t, which further increases the computational burden. To reduce memory and computation time requirements, we estimate the value function $Q(r,s)$ for matching shipment $r \in R$ with service $s \in S$. The relationship between $V(\mathbf{F}^t)$ and $Q(r,s)$ can be represented as:

$$V(\mathbf{F}^t) = \min_{x^t} \sum_{r \in R^t} \sum_{s \in S^{+}_{\phi_r^t}} Q(r,s). \tag{19}$$

The first term in the value function $Q(r,s)$ is the cost $\tilde{\theta}_{rs}$ of moving shipment r from terminal o_s to terminal d_s via service s, and the second term is the value function from terminal d_s to other terminals, as shown in equations (20). If the destination terminal of service s is the destination terminal of shipment r, then the value function $Q(r,s)$ only includes the cost $\tilde{\theta}_{rs}$, as shown in equations (22).

$$Q(r,s) = \tilde{\theta}_{rs} + \min_{q \in S^{+}_{d_s}} Q(r,q) \quad \forall r \in R, s \in S, d_s \neq d_r, \tag{20}$$

where

$$\tilde{\theta}_{rs} = \left(c_s + lc_{o_s}^{m_s} + lc_{d_s}^{m_s} + \left(D_s - lt_{o_s}^{m_s} - A_{ro_s}\right) c_{o_s}^{\text{storage}} + c^{\text{emission}} e_s\right) u_r, \tag{21}$$

$$Q(r,s) = \tilde{\theta}_{rs} \quad \forall r \in R, s \in S, d_s = d_r, \tag{22}$$

where

$$\begin{aligned} \tilde{\theta}_{rs} = &\left(c_s + lc_{o_s}^{m_s} + lc_{d_s}^{m_s} + \left(D_s - lt_{o_s}^{m_s} - A_{ro_s}\right) c_{o_s}^{\text{storage}} + c^{\text{emission}} e_s\right) u_r \\ &+ c_r^{\text{delay}} \tilde{\mathbb{T}}_r^{\text{delay}} u_r. \end{aligned} \tag{23}$$

Under a given strategy, the value function estimation $Q(r,s)$ is updated once shipment r arrives at terminal d_s by using service s in a simulation. The typically used updating strategies in reinforcement learning include Monte Carlo learning (MC), on-policy temporal difference learning (i.e., SARSA), and off-policy temporal difference learning (i.e., Q-learning) (Sutton and Barto 2018). While MC learns from complete episodes, SARSA and Q-learning learn from incomplete episodes by bootstrapping, namely, the value function is updated based on the estimates of the values of successor states (Abdulhai and Kattan 2003). Compared with SARSA in which the target policy and the behavior policy are the same, Q-learning learns from the optimal policy while following a given behavior policy (Mao and Shen 2018). In this paper, we adopt the Q-learning algorithm to update value function estimation $Q(r,s)$ for shipment $r \in R$ and service $s \in S$, as follows:

$$Q(r,s) \leftarrow Q(r,s) + \alpha \left[\bar{\theta}_{rs} + \max_{q \in \Xi_{r\phi_r}} Q(r,q) - Q(r,s)\right] \quad \text{if} \quad d_s \neq d_r \tag{24}$$

$$Q(r,s) \leftarrow Q(r,s) + \alpha \left[\bar{\theta}_{rs} - Q(r,s)\right] \quad \text{if} \quad d_s = d_r \tag{25}$$

Here, α represents the step-size which controls the learning rate from simulations, $0 \leq \alpha \leq 1$; $\bar{\theta}_{rs}$ denotes the observed cost of traveling shipment r from o_s to d_s by service s; $\Xi_{r\phi_r}$ represents the set of feasible services for shipment r at terminal ϕ_r that satisfy time, spatial, and capacity constraints (2–6); $\delta = \left[\bar{\theta}_{rs} + \max_{q \in \Xi_{r\phi_r}} Q(r,q) - Q(r,s)\right]$ is the temporal difference between random observations and the current value function estimations.

At each decision epoch of a simulation, the next service that moves a shipment from the current terminal to the next terminal is selected based on a given behavior policy. The important aspect of the RLA is the trade-off between exploitation and exploration (Sutton and Barto 2018). The RLA has to exploit the services that minimize the total costs based on the current value function estimations. However, due to travel time uncertainties, the RLA has to explore new services that might be a better choice than the current best service. One of the behavior policies that balance exploitation and exploration is the ϵ-greedy policy. Under the ϵ-greedy policy, the RLA selects the best service based on the current value function estimations with probability $1-\epsilon$, and selects randomly with probability ϵ (Çimen and Soysal 2017).

The RLA that estimates value functions for the GSSM problem with dynamic and stochastic travel times is briefly presented in Algorithm 1. The algorithm mainly consists of five steps at each simulation: sampling random variables; observing exogenous information; updating value function estimations; selecting services; updating states.

- *Sampling random variables* At the beginning of each simulation, the actual travel and arrival times of all the services are sampled from given probability distributions.
- *Observing exogenous information* At each decision epoch $t \in \{1, ..., T\}$ of a simulation, the actual travel times $\left[\bar{t}_s\right]_{s\in S^{-t}}$ and actual arrival times $\left[\bar{A}_s\right]_{s\in S^{-t}}$ are observed. The position of shipment r is updated if service $s \in S^{-t}$ was selected, i.e., $x_{rs} = 1$; the arrival time of shipment r at the new position is updated accordingly; the actual cost of matching shipment r with service s is calculated based on Eqs. (21,23).
- *Updating value function estimations* Based on the observations, the value function estimations are updated based on Eqs. (24–25) for shipments that arrive at a new terminal at each decision epoch.
- *Selecting services* For shipments arriving at a new terminal but not its destination terminal (i.e., $\phi_r \neq d_r$), the next service that moves the shipment leaving the current terminal needs to be selected. For shipment $r \in R^t$ and service $s \in S^+_{\phi_r}$, service s is a feasible choice if: the arrival time of shipment r at terminal ϕ_r plus loading time is earlier than the scheduled departure time of service s, the container volume of shipment r does not exceed the free capacity of service s; and the destination terminal of service s hasn't been visited before, $d_s \neq o_q$ for $q \in I_r$. The next service q is determined based on a ϵ-greedy policy: $q \leftarrow \arg\min_{q\in\ \Xi_{r\phi_r}} Q(r,q)$ with probability $1-\epsilon+\frac{\epsilon}{|\Xi_{r\phi_r}|}$; $q \leftarrow$ other choice from $\Xi_{r\phi_r}$ with probability $\frac{\epsilon}{|\Xi_{r\phi_r}|}$.
- **Updating states.** The itinerary of shipment $r \in R^t$ and the free capacity of service $q \in S^+_{\phi_r}$ are updated if service q is selected to move shipment r leaving terminal ϕ_r.

Using the estimated value functions, the global synchromodal matching platform selects the greedy services that minimize the total costs in the online decision-making processes, as shown in Algorithm 2. Different from the simulation process, at decision epoch 0, the platform needs to decide on the acceptance decisions for all shipments. Shipment r is rejected if revenue $p_r u_r$ is lower than the estimated minimum total costs $\min_{q\in\ \Xi_{ro_r}} Q(r,q)$. At each decision epoch, the next service q that moves shipment r leaving its current terminal ϕ_r is determined based on a greedy policy: $q \leftarrow \arg\min_{q\in\Xi_{r\phi_r}} Q(r,q)$.

5 Numerical experiments

In this section, we evaluate the performance of the reinforcement learning approach (RLA) in comparison to the myopic approach (MA) proposed by Guo et al. (2020a) and the stochastic approach (SA) proposed by Guo et al. (2020b). While MA uses average travel times for transport planning, SA sets chance constraints for feasible transshipment. Both MA and SA use a heuristic algorithm to generate timely solutions at each decision epoch of a rolling horizon framework. Matching decisions under MA and SA are updated only when shipments

Algorithm 1 The RLA for estimating value functions.

```
Input. Terminals $N$; shipments $R$; services $S$; planning horizon $T$; simulation length $L$; probability of random choices $\epsilon$; step-size $\alpha$;
Output. Value function estimations $Q(r, s)$ for all shipments $r \in R$, services $s \in S$.
Initialization. Let $Q(r, s) = 0$ for $r \in R$, $s \in S$.
for Simulation counter $l = 1$ to Simulation Length $L$ do
  Reset simulation parameters. Set decision $x_{rs} = 0$; set position of shipments $\phi_r = o_r$; set arrival time of shipment $r$ at origin terminal $o_r$ as $A_{ro_r} = \mathbb{T}_r^{\text{release}}$; reset free capacity $U_s$ of service $s$; set decision space $\Xi_{ri} = \emptyset$ for shipment $r$ at terminal $i \in N$; set itinerary $I_r = \emptyset$ of shipment $r \in R$.
  Sampling random variables. Sample arrival and travel times of services based on given probability distributions.
  for Shipment $r \in R$ do
    for Service $s \in S^+_{o_r}$ do
      if Service $s$ satisfies time, spatial, and capacity constraints (2-6) then
         Update decision space $\Xi_{ro_r} \leftarrow \Xi_{ro_r} \cup s$.
    Selecting service. Select the next service $q$ to travel for shipment $r$ using a $\epsilon$-greedy policy. Set decision $x_{rq} = 1$.
    Updating states. Update itinerary$I_r \leftarrow I_r \cup q$; free capacity $U_q \leftarrow U_q - u_r$.
  for Decision epoch $t = 1$ to Planning horizon $T$ do
    Observing exogenous information. Observe actual arrival time $\bar{A}_s$ and actual travel time $\bar{t}_s$ for $s \in S^{-t}$.
    for Shipment $r \in R$ do
      for Service $s \in S^{-t}$ do
        if Decision $x_{rs} = 1$ then
          Update shipment position $\phi_r \leftarrow d_s$; update arrival time $A_{rd_s} \leftarrow \bar{A}_s + lt^{ms}_{d_s}$.
          Calculate actual cost $\bar{\theta}_{rs}$ based on equations (21,23).
          if $\phi_r = d_r$ then
            Updating value function estimations. Temporal difference $\delta = \bar{\theta}_{rs} - Q(r, s)$; value function $Q(r, s) \leftarrow \alpha\delta + Q(r, s)$.
          else
            Update set of shipments need further decisions: $R^t \leftarrow R^t \cup r$
    if Shipments are not all at their destination terminals. then
      for shipments $r \in R^t$ do
        for service $s \in S^+_{\phi_r}$ do
          if Service $s$ satisfies time, spatial, and capacity constraints (2-6) then
            Update decision space $\Xi_{r\phi_r} \leftarrow \Xi_{r\phi_r} \cup s$.
        Updating value function estimations. Temporal difference $\delta = \bar{\theta}_{rs} + \max_{q \in \Xi_{r\phi_r}} Q(r, q) - Q(r, s)$; value function $Q(r, s) \leftarrow \alpha\delta + Q(r, s)$.
        Selecting service. Select the next service $q$ to travel for shipment $r$ using a $\epsilon$-greedy policy. Set decision $x_{rq} = 1$.
        Updating states. Update itinerary$I_r \leftarrow I_r \cup q$; free capacity $U_q \leftarrow U_q - u_r$.
    else
      Go to the next simulation.
Return the value functions $Q(r, s)$ for shipment $r \in R$, service $s \in S$.
```

face infeasible transshipment during the transport processes. The approaches are implemented in MATLAB and all experiments are performed on a computer with 2.50GHz Intel Core i5-7200U CPU and 8 GB RAM. The optimization problems in MA and SA are solved with CPLEX 12.6.3.

5.1 Experimental setup

We consider a global transport network that consists of eight terminals in Europe and four terminals in Asia that are connected by Suez Canal Route (SCR), Northern Sea Route (NSR),

Algorithm 2 Online decision making using the estimated value functions.

```
**Input**. Terminals $N$; shipments $R$; services $S$; planning horizon $T$; value function estimations $Q(r,s)$.
**Output**. Acceptance decision $y_r$ and matching decision $x_{rs}$ for $r \in R, s \in S$.
**Initialization**. Set $x_{rs}=0$, $\phi_r = o_r$, $A_{ro_r} = \mathbb{T}_r^{\text{release}}$, $\Xi_{ri} = \emptyset$, $I_r = \emptyset$.
**for** Shipment $r \in R$ **do**
  **for** Service $s \in S_{o_r}^{+}$ **do**
    **if** Service $s$ satisfies time, spatial, and capacity constraints (2-6) **then**
      Update decision space $\Xi_{ro_r} \leftarrow \Xi_{ro_r} \cup s$.
  **if** $p_r u_r < \min_{q \in \Xi_{ro_r}} Q(r,q)$ **then**
    Reject shipment $r$, $y_r = 0$.
  **else**
    **Selecting service**. Accept shipment $r$, $y_r = 1$; select the next service $q$ to travel using greedy policy: $q \leftarrow \arg\min_{q \in \Xi_{ro_r}} Q(r,q)$. Set decision $x_{rq} = 1$.
    **Updating states**. Update itinerary$I_r \leftarrow I_r \cup q$; free capacity $U_q \leftarrow U_q - u_r$.
**for** Decision epoch $t = 1$ to Planning horizon $T$ **do**
  **Observing exogenous information**. Observe actual arrival time $\bar{A}_s$ and actual travel time $\bar{t}_s$ for $s \in S^{-t}$.
  **for** Shipment $r \in R$ and $y_r = 1$ **do**
    **for** Service $s \in S^{-t}$ **do**
      **if** Decision $x_{rs} = 1$ **then**
        Update shipment position $\phi_r \leftarrow d_s$; update arrival time $A_{rd_s} \leftarrow \bar{A}_s + lt_{d_s}^{m_s}$.
        Calculate actual cost $\bar{\theta}_{rs}$ based on equations (21,23).
        **if** $\phi_r \neq d_r$ **then**
          Update set of shipments need further decisions: $R^t \leftarrow R^t \cup r$
  **if** Shipments are not all at their destination terminals. **then**
    **for** shipments $r \in R^t$ **do**
      **for** service $s \in S_{\phi_r}^{+}$ **do**
        **if** Service $s$ satisfies time, spatial, and capacity constraints (2-6) **then**
          Update decision space $\Xi_{r\phi_r} \leftarrow \Xi_{r\phi_r} \cup s$.
      **Selecting service**. Select the next service $q$ to travel for shipment $r$ using greedy policy: $q \leftarrow \arg\min_{q \in \Xi_{r\phi_r}} Q(r,q)$. Set decision $x_{rq} = 1$.
      **Updating states**. Update itinerary$I_r \leftarrow I_r \cup q$; free capacity $U_q \leftarrow U_q - u_r$.
  **else**
    Break.
Calculate the total profits:
    $$TP = \sum_{r \in R} p_r u_r y_r - \sum_{r \in R} \sum_{s \in S} \bar{\theta}_{rs} x_{rs}$$
```

and Eurasia Land Bridge (ELB), as shown in Fig. 5. Compared with SCR, NSR has a shorter travel time but a higher travel cost caused by ice-breaking fees (Lin and Chang 2018). With the implementation of IMO 2020 regulations, shipping liner companies are required to use low-sulfur fuels on the sea, which in turn increases travel costs in SCR and NSR (Lian et al. 2020). As an alternative, ELB becomes more and more competitive thanks to its shortest travel time.

Unless otherwise stated, the benchmark values of coefficients are set as follows: planning horizon (unit: hours) $T = 1400$; loading cost (unit: €/TEU) $lc_i^{\text{ship}} = 18$, $lc_i^{\text{barge}} = 18$, $lc_i^{\text{train}} = 12$, $lc_i^{\text{truck}} = 12$ for $i \in N$; loading time (unit: hours) $lt_i^{\text{ship}} = 12$, $lt_i^{\text{barge}} = 4$, $lt_i^{\text{train}} = 2$, $lt_i^{\text{truck}} = 1$ for $i \in N$; storage cost (unit: €/TEU-h) $c_i^{\text{storage}} = 1$ for $i \in N$; carbon tax (unit: €/kg) $c^{\text{emission}} = 0.07$. The travel times of all the services follow normal distributions: $\tilde{t}_s \sim N(\mu_s, \sigma_s^2)$ for $s \in S$. The mean of travel times $\mu_s = t_s$ for $s \in S$, standard deviation of travel times $\sigma_s = 0.1 t_s$ for $s \in S \backslash S^{\text{truck}}$, $\sigma_s = 0.5 t_s$ for $s \in S^{\text{truck}}$. Besides, we let $0.9 t_s$ be the fixed lower bound for travel times of service $s \in S$. Regarding SA, we set

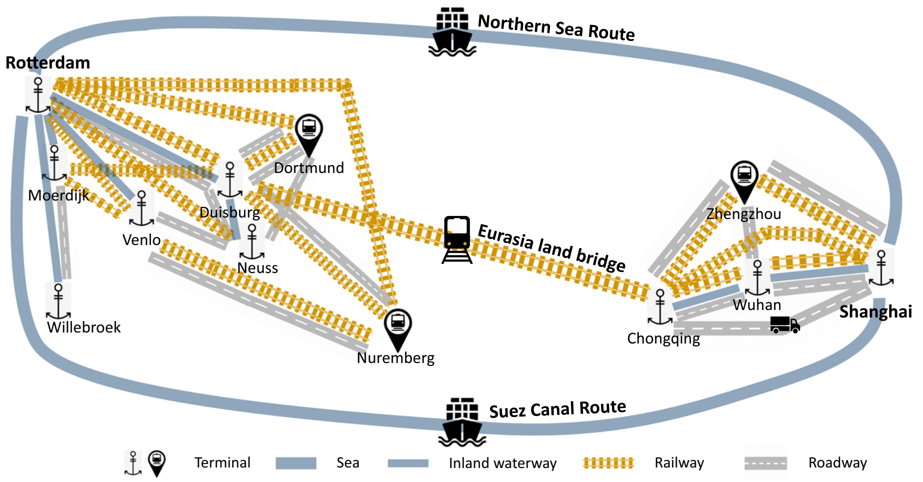


Fig. 5 The topology of a global synchromodal network

the confidence level to 0.7. In general, SA performs the best under this setting (Guo et al. 2020b). Regarding RLA, we set the simulation length $L = 100000$, probability of random choices $\epsilon = 0.3$, step-size $\alpha = 0.01$.

We use I-n_1-n_2-n_3 to represent an instance with n_1 terminals, n_2 services, and n_3 shipments under the global transport network.

5.2 Case study

We use a small instance I-5-18-6 with 5 terminals, 18 services, and 6 shipments to do the case study. The service data is presented in Table 2. The shipment data is shown in Table 3. Compared with reefer shipments (1, 3, 5), dry shipments (2, 4, 6) have longer lead times, lower freight rates, and lower delay costs.

5.2.1 Sensitivity analysis of problem parameters

The sensitivity analysis of problem parameters is investigated under a static and deterministic environment, i.e., the realization of travel times equals the expected values.

To test the impact of the carbon tax coefficient c^{emission} on costs, delays, emissions, and shipment itineraries, we set the objective function to minimize total costs without rejections. Table 4 shows that increasing the carbon tax coefficient, the total costs will be increased but emissions will be reduced. It is interesting to note that, the delay in deliveries grows as emission decreases. The reason is that, with a large value of carbon tax coefficient, shipments will be assigned to 'greener' services with lower emissions but mostly longer travel times. For example, reefer shipments 1, 3 will be switched from Eurasia Land Bridge (service 17) to Northern Sea Route (service 16). Besides, we notice that with the increase of carbon tax coefficient, shipments 1 and 6 will be switched from barge transportation (services 1, 2, 3, 4) to train transportation (services 5, 6) which generates lower emissions and travel costs but higher storage costs at transshipment terminals.

Table 2 Service data

Service. ID	Mode	Origin	Destination	Total capacity (TEU)	Scheduled departure time	Scheduled arrival time	Travel time (h)	Travel cost (€)	Carbon emissions (kg)
1	Barge	Chongqing	Wuhan	160	144	235	91	192	313
2	Barge	Wuhan	Shanghai	160	243	328	85	178	291
3	Barge	Shanghai	Wuhan	160	144	229	85	178	291
4	Barge	Wuhan	Chongqing	160	237	328	91	192	313
5	Train	Chongqing	Shanghai	90	144	181	37	269	526
6	Train	Shanghai	Chongqing	90	144	181	37	269	526
7	truck	Shanghai	Chongqing	200			22	1823	1489
8	Truck	Chongqing	Shanghai	200			22	1823	1489
9	Barge	Rotterdam	Duisburg	160	1010	1027	17	35	57
10	Barge	Duisburg	Rotterdam	160	750	767	17	35	57
11	Train	Rotterdam	Duisburg	90	910	917	7	48	92
12	Train	Duisburg	Rotterdam	90	750	757	7	48	92
13	Truck	Rotterdam	Duisburg	200			3	334	219
14	Truck	Duisburg	Rotterdam	200			3	334	219
15	Ship (SCR)	Shanghai	Rotterdam	200	350	988	638	1441	2161
16	Ship (NSR)	Shanghai	Rotterdam	200	350	900	550	2240	1631
17	Train (ELB)	Chongqing	Duisburg	90	350	723	373	2007	3517
18	Ship (SCR)	Shanghai	Rotterdam	200	518	1156	638	1441	2161

Table 3 Shipment data

Shipments	Container type	Origin	Destination	Container volume (TEU)	Release time	Lead time (h)	Freight rate (€/TEU)	Delay cost (€/TEU-h)
1	Reefer	Shanghai	Rotterdam	5	100	720	4000	20
2	Dry	Shanghai	Rotterdam	5	100	840	3500	17.5
3	Reefer	Wuhan	Rotterdam	5	100	600	4500	22.5
4	Dry	Wuhan	Rotterdam	5	100	960	3000	15
5	Reefer	Chongqing	Duisburg	5	100	480	5000	25
6	Dry	Chongqing	Duisburg	5	100	1080	2500	12.5

Table 4 Sensitivity analysis of carbon tax coefficient

c^{emission}	Total	Travel	Transfer	Storage	Delay	Carbon	Delay	Emission	Shipment itineraries					
(€/ton)	Costs (€)	Costs	Costs	Costs	Costs	Tax	(TEU-h)	(kg)	1	2	3	4	5	6
0	92804	63282	2100	5983	21439	0	873	210711	[3,4,17,10]	16	[4,17,14]	[2,15]	17	[1,2,15,9]
70	107554	63282	2100	5983	21439	14750	873	210711	[3,4,17,10]	16	[4,17,14]	[2,15]	17	[1,2,15,9]
210	136995	62777	2040	6739	21439	44001	873	209527	[6,17,10]	16	[4,17,14]	[2,15]	17	[1,2,15,9]
350	161782	62427	1800	6540	30639	60377	1333	172504	16	16	[4,17,14]	[2,15]	17	[1,2,15,9]
700	219580	61353	1680	7277	51125	98144	2243	140206	16	16	[2,16]	[2,15]	17	[5,15,9]

Table 5 Sensitivity analysis of delay cost coefficient

Coefficient	Total	Travel	Transfer	Storage	Delay	Carbon	Delay	Emission	Shipment itineraries					
	costs (€)	costs	costs	costs	costs	tax	(TEU-h)	(kg)	1	2	3	4	5	6
0.0000%	67046	49072	2160	5403	0	10411	4945	148725	15	15	[2,15]	[2,15]	[1,2,15,9]	[1,2,15,9]
0.0625%	78501	49874	1740	6382	9366	11140	3416	159139	15	15	[2,15]	[2,15]	17	[1,2,15,9]
0.1250%	85963	53010	1800	6517	11953	12683	2211	181187	15	15	[4,17,12]	[2,15]	17	[1,2,15,9]
0.1875%	91920	57860	2100	5960	11198	14801	1313	211448	[3,4,17,10]	15	[4,17,12]	[2,15]	17	[1,2,15,9]
0.2500%	95488	59287	2100	5843	13323	14935	1170	213359	[3,4,17,10]	15	[4,17,14]	[2,15]	17	[1,2,15,9]
0.3750%	102149	59287	2100	5843	19984	14935	1170	213359	[3,4,17,10]	15	[4,17,14]	[2,15]	17	[1,2,15,9]
0.5000%	107554	63282	2100	5983	21439	14750	873	210711	[3,4,17,10]	16	[4,17,14]	[2,15]	17	[1,2,15,9]
1.0000%	128993	63282	2100	5983	42878	14750	873	210711	[3,4,17,10]	16	[4,17,14]	[2,15]	17	[1,2,15,9]
1.5000%	150431	63282	2100	5983	64316	14750	873	210711	[3,4,17,10]	16	[4,17,14]	[2,15]	17	[1,2,15,9]

$c_r^{\text{delay}} = \text{Coefficient} * p_r$

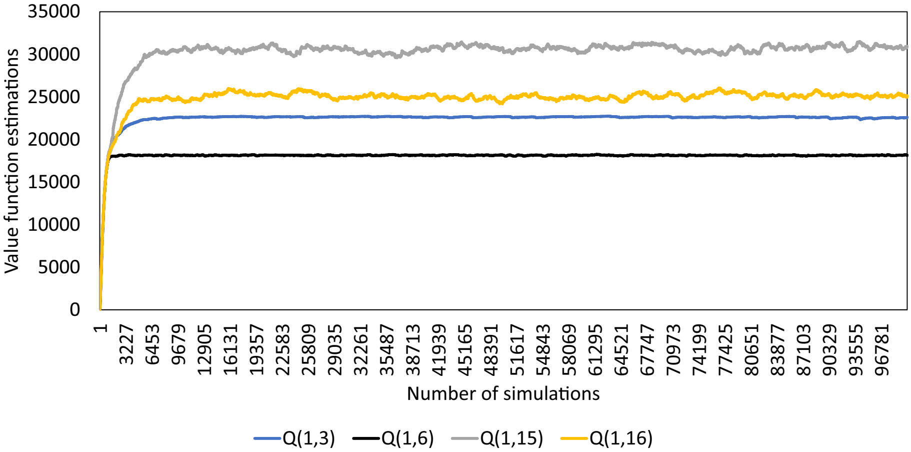


Fig. 6 Sensitivity analysis of simulation length L

To investigate the impact of delay costs on shipment itineraries, we varied its value from 0.0000% to 1.5000%*freight rate. Table 5 shows that the increase in delay cost coefficient can reduce delay in deliveries and impact service choices. Shipments will be switched from slower to faster services, which are more expensive. For example, reefer shipments 1, 3, 5 will be switched from maritime transportation (service 15) to rail transportation (service 17); dry shipment 2 will be switched from Suez Canal Route (service 15) to Northern Sea Route (service 16). It is interesting to observe again that there is a clear trade-off between delays and emissions.

5.2.2 Sensitivity analysis of algorithm parameters

To show the impact of simulation length on the value function estimations under RLA, we varied its value from 1 to 100,000. Figure 6 shows that the larger the number of simulations, the more accurate the estimation of value functions, however the higher the computation time requirements. When $L = 100,000$, the CPU required for the case study is 461 s.

To investigate the sensitivity of step-size α which controls the learning rate under RLA, we varied its value from 0.001 to 0.1. Figure 7 shows that the larger the step-size, the faster the convergence of value function estimations $Q(1, 3)$. The reason is that when the step-size is large, the platform learns fast from simulations. However, when the step-size is too large (i.e., $\alpha = 0.1$), value function estimation $Q(1, 3)$ fluctuates.

To test the sensitivity of random probability ϵ which controls the exploration rate under RLA, we varied its value from 0.1 to 0.9. Figure 8 shows that the smaller the value of ϵ, the smaller the degree of exploration, however the faster the convergence. Further decreasing ϵ from 0.3 to 0.1, the changes become quite small.

5.2.3 Comparison between MA, SA, and RLA under the case study

In this section, we compare the performance of MA, SA, and RLA under the given case study with dynamic and stochastic travel times. Table 6 presents the estimated value functions by RLA before the execution of transport plans.

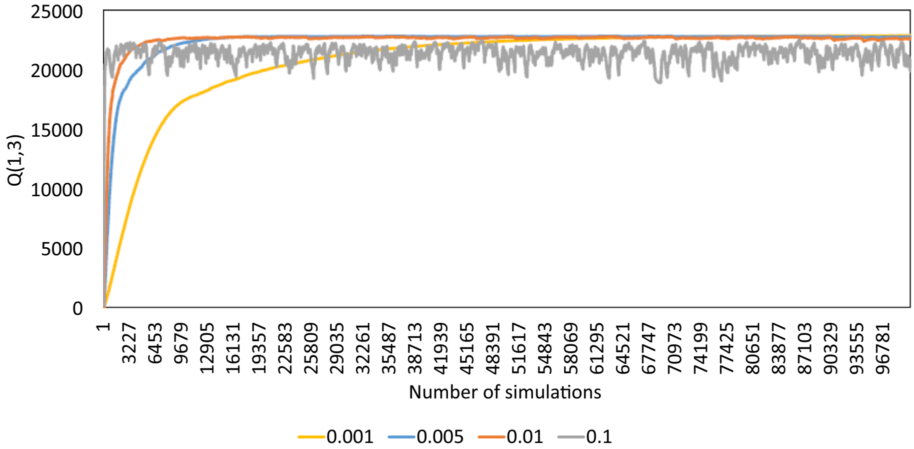


Fig. 7 Sensitivity analysis of step-size α

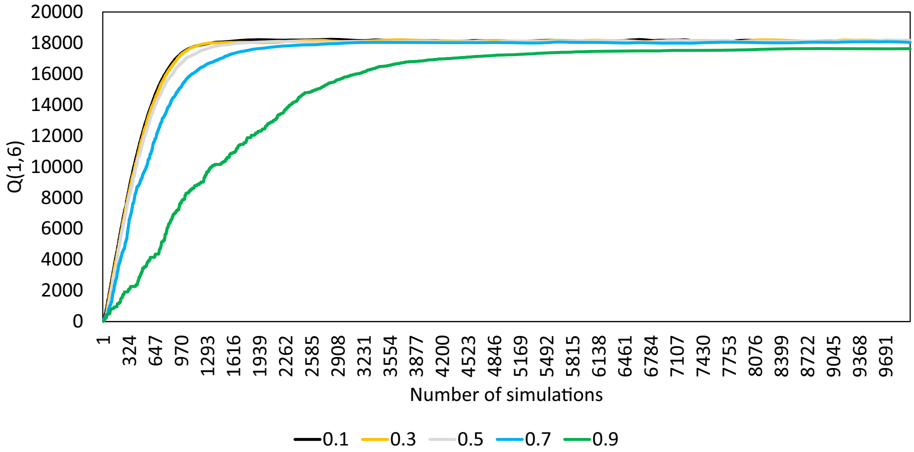
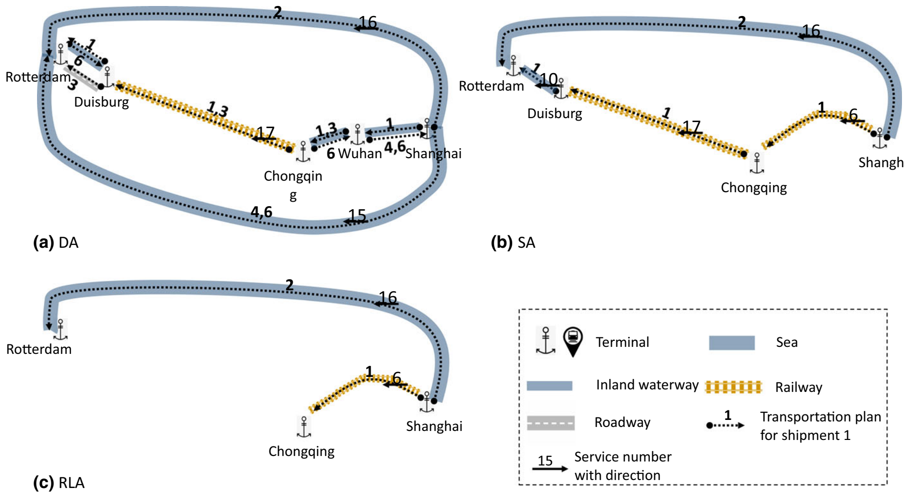


Fig. 8 Sensitivity analysis of random probability ϵ

At decision epoch $t = 0$, under RLA, the platform rejects shipments 3, 4, 5, 6 since their revenues are lower than the estimated minimum total costs. For example, for shipment 3, the revenue is 5*4500 €, the feasible services at its origin terminal (i.e., Wuhan) include services 2 and 4, the estimated minimum total cost is $Q(3, 4) = 30703 > 5 * 4500$, shipment 3 is therefore rejected. The platform accepts profitable shipments 1 and 2, and selects the services that move shipments 1 and 2 departing their origin terminals based on the estimated value functions. Under MA and SA, an optimization model needs to be solved, and the platform generates the transport plans for shipments moving from their origin to destination terminals. Figure 9 shows the initial transport plans generated by MA, SA, and RLA at decision epoch $t = 0$. Without the consideration of travel time uncertainties, MA almost accepts all the shipments. In comparison, SA and RLA consider stochastic travel times by using chance constraints and simulations, respectively.

Table 6 Estimated value functions by RLA

Q(r,s)	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15	s16	s17	s18
r1	21550	0	22635	21086	0	18119	26714	0	0	795	0	896	0	2376	30546	25327	15902	48441
r2	18837	0	19751	18497	0	15681	23578	0	0	1359	0	1427	0	2737	16447	14492	13871	31159
r3	0	35387	0	30703	0	0	33296	75931	0	8531	0	7210	0	9108	45025	38761	21669	65253
r4	0	15450	0	17885	0	0	24754	28191	0	1957	0	2014	0	3343	8988	12680	13844	18578
r5	76252	74442	26439	0	62622	0	0	37249	57088	0	42920	0	66440	0	68991	60366	34411	89418
r6	14385	12890	13730	0	12812	0	0	20699	1385	0	1811	0	2895	0	11014	14116	14845	13306



(a) DA

(b) SA

(c) RLA

Fig. 9 Initial transport plans by MA, SA, and RLA

The realization of travel times designed in this case study is shown in Table 7. Under this realization, services 1, 2, 5, 6, 15, and 17 are delayed.

The realization of shipment itineraries under MA, SA, and RLA is presented in Table 8. Under MA and SA, the transport plan for a shipment is updated once infeasible transshipment happens. Under MA, at decision epoch $t = 350$, shipments 4 and 6 meet infeasible transshipment at Shanghai terminal between services 2 and 15 due to the delay of service 2. Transport plans for shipments 4 and 6 are updated by solving an optimization model, service 18 is then selected to move shipments 4 and 6 to their destination terminals. Under SA, at decision epoch $t = 750$, shipment 1 faces infeasible transshipment at Duisburg terminal between service 17 and service 10. Service 14 is selected to replace service 10 to move the shipment to its destination terminal. Under RLA, the next service is selected once shipments arrive at new terminals by using the estimated value functions of matching the shipments with feasible services. Specifically, at decision epoch $t = 187$, shipment 1 arrives Chongqing terminal, service 17 is selected to move the shipment to Duisburg terminal; at decision epoch $t = 747$, shipment 1 arrives at Duisburg terminal, service 12 is selected to move the shipment to its destination terminal. Compared with MA and SA, RLA generates the highest total profits under the designed case.

5.3 Impact of travel time distributions

In this section, we aim to investigate the performance of RLA under scenarios with different types of travel time distributions. We consider 3 instances with different numbers of shipments. Each instance is tested under 20 realizations of travel times sampled from three types of distributions: normal distribution, gamma distribution, and lognormal distribution. These distributions are selected from the most commonly used travel time distributions in the literature (Chen and Fan 2020). For each service, we set the same means and variances of travel times under different distributions to ensure the fairness for comparisons. To avoid the generation of too small values, we set the same lower bounds for the realization of travel times under all types of distributions. We use MA as the benchmark and use the gaps in

Table 7 The realization of travel times

Service. ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Actual departure time	144	258	144	230	144	144			1010	750	910	750			350	350	350	518
Actual travel time (h)	106	84	78	97	40	40	22	22	17	16	7	6	3	3	656	500	394	611
Actual arrival time	250	342	222	327	184	184			1027	766	917	756			1006	850	744	1129
Service delay (h)	15	14			3	3									19		21	

Table 8 Realized shipment itineraries under MA, SA, and RLA

Approaches	Total profits	Infeasible	Rejection	Shipment itineraries					
		transshipments		1	2	3	4	5	6
MA	−3113	2	1	3,4,17,14	16	4,17,14	2,18		1,2,18,13
SA	970	0	4	6,17,14	16				
RLA	6305	0	4	6,17,12	16				

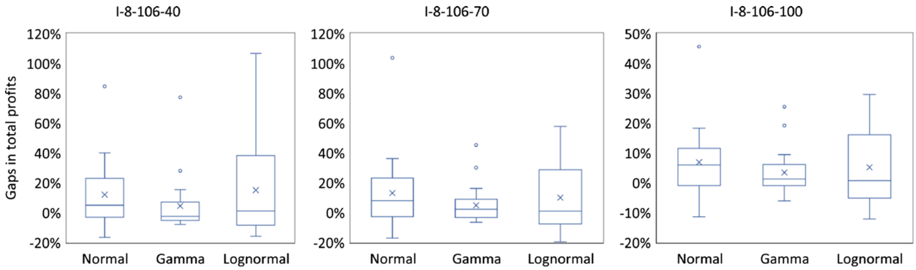


Fig. 10 Performance of RLA under scenarios with different types of probability distributions

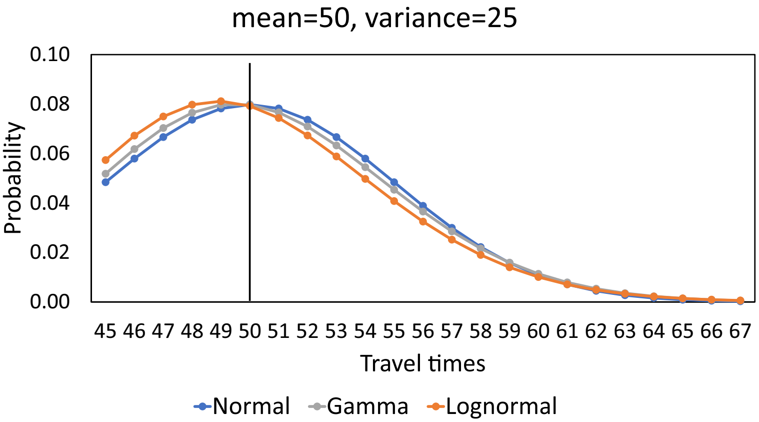


Fig. 11 Probability distributions of normal, gamma, and lognormal

total profits between MA and RLA as the performance indicator. Figure 10 shows that on average, RLA outperforms MA under all types of distributions. It is interesting to note that RLA has the best performance on scenarios with Normal distributions. The reason might be that under normal distributions, services might have more realizations with delays (as shown in Fig. 11), and RLA performs better than MA when delay happens since RLA learns from realizations to avoid infeasible transshipments caused by delays.

5.4 Comparison between MA, SA, and RLA under large instances

To compare the performance of MA, SA, and RLA not only in terms of solution quality but also computation time, we designed 15 large instances with up to 12 terminals, 167 services, and 200 shipments under the global synchromodal network. For each instance, we report the average results over 20 realizations of travel times sampled from normal distributions. We denote 'CPU' as the total computation time in seconds for each instance. Let 'γ_1' be the gap in total profits between MA and RLA, i.e., $\gamma_1 = \frac{\text{Total profits(RLA)}-\text{Total profits(MA)}}{\text{Total profits(MA)}}$. Let '$\gamma_2$' be the gap between SA and RLA, i.e., $\gamma_2 = \frac{\text{Total profits(RLA)}-\text{Total profits(SA)}}{\text{Total profits(SA)}}$. Table 9 shows that RLA has better performance than MA and SA in total profits and computation time under all instances. On average, RLA has 11.37% improvement in total profits in comparison to MA and has 3.16% improvement in comparison to SA. Besides, we note that the computation

Table 9 Comparison between MA, SA, and RLA under large instances

Instance	MA		SA		RLA		γ_1	γ_2
	Total profits	CPU(s)	Total profits	CPU(s)	Total profits	CPU(s)		
I-8-106-10	50241	5.33	50725	5.47	52813	0.05	5.12%	4.12%
I-8-106-20	99160	14.23	101260	15.17	104998	0.08	5.89%	3.69%
I-8-106-30	142077	18.03	151858	16.49	154595	0.14	8.81%	1.80%
I-8-106-40	194694	27.22	210426	23.08	216969	0.15	11.44%	3.11%
I-8-106-50	252734	32.00	264184	31.47	273613	0.15	8.26%	3.57%
I-8-106-60	296584	28.00	319860	36.06	329951	0.17	11.25%	3.15%
I-8-106-70	327918	32.99	365483	29.40	369257	0.13	12.61%	1.03%
I-8-106-80	404406	49.38	443268	33.64	445148	0.15	10.07%	0.42%
I-8-106-90	487451	59.72	526059	38.40	529164	0.36	8.56%	0.59%
I-8-106-100	561349	60.25	600333	56.32	603178	0.17	7.45%	0.47%
I-12-167-120	487586	99.83	499609	65.64	529716	0.30	8.64%	6.03%
I-12-167-140	586013	108.74	598955	74.08	630446	0.76	7.58%	5.26%
I-12-167-160	628221	121.23	651486	85.18	687255	0.57	9.40%	5.49%
I-12-167-180	571015	131.79	711212	88.89	764926	0.55	33.96%	7.55%
I-12-167-200	642648	144.78	772528	98.32	780615	0.36	21.47%	1.05%
Average							**11.37%**	**3.16%**

time of MA and SA increases dramatically with the increasing size of instances. In contrast, all the instances can be solved by the RLA within one second.

6 Conclusions and future research

In this paper, we investigated a global synchromodal shipment matching problem with dynamic and stochastic travel times. We formulated a sequential decision process (SDP) model to describe the problem. Due to the curse of dimensionality, the SDP model is very hard to be solved directly by classical dynamic programming algorithms. To address this, we adopted one of the most basic and popular reinforcement learning approaches (RLA), i.e., the Q-learning algorithm, to estimate the value functions of matching shipments with services. During the transport process, the next service that moves a shipment departing from its current position is selected based on the estimated value functions. We conducted experiments to validate the performance of RLA in comparison to a myopic approach (MA) proposed by Guo et al. (2020a) that does not consider travel time uncertainty and a stochastic approach (SA) proposed by Guo et al. (2020b) that sets chance constraints on feasible transshipment under a rolling horizon framework. The experimental results indicate that RLA performs better than MA and SA in total profits and computation time in all instances. With the developed methodology, the global synchromodal matching platform can adapt shipment routes immediately when real-time travel times are observed to maximize the total profits over a given planning horizon.

This research can be extended in several promising directions. First, in this paper, we only considered contractual shipment requests that are received before the planning horizon. Future research can take into account dynamic and stochastic shipment requests that are received from spot markets. Second, in this paper, we considered a centralized platform that has full information and provides integrated decisions for global shipments. However, in practice, a large number of entities are involved in global container transport and they may not all be willing to give authority to a centralized platform. Instead, they would like to share limited information and make local decisions by themselves. Coordination mechanisms among them and incentives to stimulate cooperation are part of future research. Third, in this paper, we assumed the routes of services are fixed. Future research might consider flexible routes of shipments and services integrally in synchromodal transportation.

Acknowledgements This research is financially supported by the National Key R&D program of China under Grant 2019YFB1600400, by the project “Complexity Methods for Pre dictive Synchromodality” (Project 439.16.120) of the Netherlands Organisation for Scientic Research (NWO), as well as by the Natural Sciences and Engineering Council of Canada (NSERC) through its Cooperative Research and Development Grants Program.

Declarations

Conflicts of interest The authors declare that they have no conict of interest.

References

Abdulhai, B., & Kattan, L. (2003). Reinforcement learning: Introduction to theory and potential for transport applications. *Canadian Journal of Civil Engineering, 30*(6), 981–991. https://doi.org/10.1139/l03-014.

Chang, T. S. (2008). Best routes selection in international intermodal networks. *Computers &amp; Operations Research, 35*(9), 2877–2891. https://doi.org/10.1016/j.cor.2006.12.025.

Chen, Z., & Fan, W. D. (2020). Analyzing travel time distribution based on different travel time reliability patterns using probe vehicle data. *International Journal of Transportation Science and Technology, 9*(1), 64–75. https://doi.org/10.1016/j.ijtst.2019.10.001.

Çimen, M., & Soysal, M. (2017). Time-dependent green vehicle routing problem with stochastic vehicle speeds: An approximate dynamic programming algorithm. *Transportation Research Part D: Transport and Environment, 54,* 82–98. https://doi.org/10.1016/j.trd.2017.04.016.

Demir, E., Burgholzer, W., Hrušovský, M., Arıkan, E., Jammernegg, W., & van Woensel, T. (2016). A green intermodal service network design problem with travel time uncertainty. *Transportation Research Part B: Methodological, 93,* 789–807. https://doi.org/10.1016/j.trb.2015.09.007.

Ehmke, J. F., Campbell, A. M., & Urban, T. L. (2015). Ensuring service levels in routing problems with time windows and stochastic travel times. *European Journal of Operational Research, 240*(2), 539–550. https://doi.org/10.1016/j.ejor.2014.06.045.

Gendreau, M., Jabali, O., & Rei, W. (2016). 50th anniversary invited article-future research directions in stochastic vehicle routing. *Transportation Science, 50*(4), 1163–1173. https://doi.org/10.1287/trsc.2016.0709.

Guo, W., Atasoy, B., van Blokland, W.B., & Negenborn, R.R. (2020b). A global intermodal shipment matching problem under travel time uncertainty. In: Lecture notes in computer science, Springer International Publishing (pp. 553–568). https://doi.org/10.1007/978-3-030-59747-4_36

Guo, W., Atasoy, B., van Blokland, W. B., & Negenborn, R. R. (2020). A dynamic shipment matching problem in hinterland synchromodal transportation. *Decision Support Systems*. https://doi.org/10.1016/j.dss.2020.113289.

Guo, W., Atasoy, B., van Blokland, W. B., & Negenborn, R. R. (2021). Anticipatory approach for dynamic and stochastic shipment matching in hinterland synchromodal transportation. *Flexible Services and Manufacturing Journal*. https://doi.org/10.1007/s10696-021-09428-5.

Guo, W., Atasoy, B., van Blokland, W. B., & Negenborn, R. R. (2021). Global synchromodal transport with dynamic and stochastic shipment matching. *Transportation Research Part E: Logistics and Transportation Review, 152,* 102404. https://doi.org/10.1016/j.tre.2021.102404.

Hrušovský, M., Demir, E., Jammernegg, W., & van Woensel, T. (2016). Hybrid simulation and optimization approach for green intermodal transportation problem with travel time uncertainty. *Flexible Services and Manufacturing Journal, 30*(3), 486–516. https://doi.org/10.1007/s10696-016-9267-1.

Lee, C. Y., & Song, D. P. (2017). Ocean container transport in global supply chains: Overview and research opportunities. *Transportation Research Part B: Methodological, 95,* 442–474. https://doi.org/10.1016/j.trb.2016.05.001.

Lian, F., He, Y., & Yang, Z. (2020). Competitiveness of the China-Europe railway express and liner shipping under the enforced sulfur emission control convention. *Transportation Research Part E: Logistics and Transportation Review, 135,* 101861. https://doi.org/10.1016/j.tre.2020.101861.

Lin, D. Y., & Chang, Y. T. (2018). Ship routing and freight assignment problem for liner shipping: Application to the northern sea route planning problem. *Transportation Research Part E: Logistics and Transportation Review, 110,* 47–70. https://doi.org/10.1016/j.tre.2017.12.003.

Li, X., Tian, P., & Leung, S. C. (2010). Vehicle routing problems with time windows and stochastic travel and service times: Models and algorithm. *International Journal of Production Economics, 125*(1), 137–145. https://doi.org/10.1016/j.ijpe.2010.01.013.

Mao, C., & Shen, Z. (2018). A reinforcement learning framework for the adaptive routing problem in stochastic time-dependent network. *Transportation Research Part C: Emerging Technologies, 93,* 179–197. https://doi.org/10.1016/j.trc.2018.06.001.

Meng, Q., Wang, S., Andersson, H., & Thun, K. (2014). Containership routing and scheduling in liner shipping: Overview and future research directions. *Transportation Science, 48*(2), 265–280. https://doi.org/10.1287/trsc.2013.0461.

Meng, Q., Zhao, H., & Wang, Y. (2019). Revenue management for container liner shipping services: Critical review and future research directions. *Transportation Research Part E: Logistics and Transportation Review, 128,* 280–292. https://doi.org/10.1016/j.tre.2019.06.010.

Mes, M.R.K., & Rivera, A.P. (2017). Approximate dynamic programming by practical examples. In: International series in operations research & management science, Springer International Publishing (pp. 63–101). https://doi.org/10.1007/978-3-319-47766-4_3

Powell, W. B. (2019). A unified framework for stochastic optimization. *European Journal of Operational Research, 275*(3), 795–821. https://doi.org/10.1016/j.ejor.2018.07.014.

Ritzinger, U., Puchinger, J., & Hartl, R. F. (2015). A survey on dynamic and stochastic vehicle routing problems. *International Journal of Production Research, 54*(1), 215–231. https://doi.org/10.1080/00207543.2015.1043403.

Rivera, A. E. P., & Mes, M. R. (2017). Anticipatory freight selection in intermodal long-haul round-trips. *Transportation Research Part E: Logistics and Transportation Review, 105,* 176–194. https://doi.org/10.1016/j.tre.2016.09.002.

Rodrigues, F., Agra, A., Christiansen, M., Hvattum, L. M., & Requejo, C. (2019). Comparing techniques for modelling uncertainty in a maritime inventory routing problem. *European Journal of Operational Research, 277*(3), 831–845. https://doi.org/10.1016/j.ejor.2019.03.015.

SteadieSeifi, M., Dellaert, N., Nuijten, W., van Woensel, T., & Raoufi, R. (2014). Multimodal freight transportation planning: A literature review. *European Journal of Operational Research, 233*(1), 1–15. https://doi.org/10.1016/j.ejor.2013.06.055.

Sun, Y., & Lang, M. (2015). Modeling the multicommodity multimodal routing problem with schedule-based services and carbon dioxide emission costs. *Mathematical Problems in Engineering, 2015,* 1–21. https://doi.org/10.1155/2015/406218.

Sun, Y., Lang, M., & Wang, D. (2015). Optimization models and solution algorithms for freight routing planning problem in the multi-modal transportation networks: A review of the state-of-the-art. *The Open Civil Engineering Journal, 9*(1), 714–723. https://doi.org/10.2174/1874149501509010714.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. Cambridge: MIT Press.

van Riessen, B., Negenborn, R. R., & Dekker, R. (2016). Real-time container transport planning with decision trees based on offline obtained optimal solutions. *Decision Support Systems, 89,* 1–16. https://doi.org/10.1016/j.dss.2016.06.004.

Yang, D., Pan, K., & Wang, S. (2018). On service network improvement for shipping lines under the one belt one road initiative of china. *Transportation Research Part E: Logistics and Transportation Review, 117,* 82–95. https://doi.org/10.1016/j.tre.2017.07.003.

Yee, H., Gijsbrechts, J., & Boute, R. (2021). Synchromodal transportation planning using travel time information. *Computers in Industry, 125,* 103367. https://doi.org/10.1016/j.compind.2020.103367.