# How to optimize container withholding decisions for reuse in the hinterland?

Benjamin Legros, Jan Fransoo, Oualid Jouini

European Journal of Operational Research 316 (2024) 930–941

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/eor

Production, Manufacturing, Transportation and Logistics

# How to optimize container withholding decisions for reuse in the hinterland?

Benjamin Legros [a,*], Jan Fransoo [b], Oualid Jouini [c]

[a] EM Normandie Business School, Métis Lab, 92110 Clichy, France
[b] Tilburg School of Economics and Managemnt, Tilburg University, 5000 LE Tilburg, Netherlands
[c] Université Paris-Saclay, CentraleSupélec, Laboratoire Genie Industriel, 91190 Gif-sur-Yvette, France

ARTICLE INFO

*Keywords:*
(O) Transportation
Container
Double-ended queue
Markov decision process
Street turn

ABSTRACT

This study investigates how a hinterland consignee (importer) makes decisions regarding the storage of empty containers for reuse by a shipper (exporter). The system is modeled as a double-ended queue with non-zero matching times, limited truck resources, and both consignee and shipper having fixed withholding capacities. The consignee's withholding threshold is strategically set to minimize overall transport and detention costs. We derive closed-form expressions for performance measures in the case of a single storage facility at the shipper, utilizing a matrix-based approach. We extend this methodology numerically to the general case. Additionally, we present an accurate fixed-point approximation facilitating the determination of performance measures and optimal threshold. Our findings show the importance of withholding decisions in import and balanced areas for cost reduction. In export areas, a policy of full reuse proves nearly optimal. Analyzing dynamic state-dependent consignee decisions via a Markov decision process, we establish that the optimal policy involves withholding thresholds increasing with stored quantity at the shipper. While optimal, state-dependent thresholds yield limited cost savings compared to a fixed threshold, suggesting minimal impact from consignee-shipper information sharing. Additionally, we examine the influence of variability in matching and production times, observing decreased costs with reduced variability, particularly in export areas, but with minor impacts on withholding decisions.

## 1. Introduction

Containerization is a standardized system of freight transport that moves containers from door to door. This includes container ships, deep sea terminals with special handling equipment, and intermodal infrastructure in the hinterland such as inland terminals. The United Nations Conference on Trade and Development stated that in 2017 around 80% of global trade by volume or 70% by value was carried by sea and handled by ports. In line with the growth in intercontinental maritime transport, hinterland container traffic has grown substantially.

Once a container has been unloaded at its destination in the hinterland, another transport leg must be found, as moving an empty container is almost as costly as moving a full container. In an ideal situation, an inbound container should find an outbound load once it has been unloaded before being sent back to the sea terminal. The strategy of matching an empty container from an importer with a load from an exporter, so that the container is full in both directions, is called a *street turn strategy*. However, containers are often immediately sent back empty from the hinterland to the sea terminal, leading to additional transportation costs and pollution. An important reason for these empty movements is the imbalance between imports and exports. In an import-heavy area, containers are often sent back empty due to the impossibility of finding an outbound load. However, other factors incentivize the immediate return of empty containers to the sea terminal. In particular, distance and a lack of coordination between importers and exporters in the hinterland may discourage operating matches between empty containers and outbound loads. More importantly, the high detention costs imposed by shipping lines create an urgency to send back empty containers instead of storing them until a match can be found.

With the rapid increase in global container shipments in late 2020 and 2021, a global shortage of containers has been reported. As a result, ocean carriers increased the demurrage and detention fees and have limited the free detention periods. Not only does this lead to additional movements of empty containers, but also this makes it more difficult for shippers to get access to containers for their exports. Until 2020, shippers could safely assume that containers would always be available on a more or less just-in-time basis. In the newly developed situation by late 2020, shippers might need to ensure they have product ready for shipment whenever a container might be available in their immediate neighborhood after imported cargo has been unloaded. While in 2023

* Corresponding author.
*E-mail addresses:* benjamin.legros@centraliens.net (B. Legros), jan.fransoo@tilburguniversity.edu (J. Fransoo), oualid.jouini@centralesupelec.fr (O. Jouini).

https://doi.org/10.1016/j.ejor.2024.02.035
Received 16 August 2023; Accepted 29 February 2024
Available online 1 March 2024

the global containerized transport volume has substantially decreased, experts expect that period of scarce container availability may repeat itself in the future.

In this paper, we investigate the optimization of container matching within the hinterland, focusing specifically on a strategic decision by shippers to ensure the availability of export goods ahead of container availability. The consignee-shipper relationship, often overlooked in consignee-centric literature, is central to our exploration. By incorporating a detailed account of shipper inventory decisions and constraints on resources for container reuse operations, we assess the determinants influencing consignee decision-making. To this end, we model the system as a double-ended queue with a non-zero matching time and a finite number of resources, where the consignee's objective is to select a withholding threshold for containers that minimizes the sum of holding and travel costs.

*Main contributions.* The main contributions of our study can be categorized as follows:

(i) *Performance analysis.* We analyze the double-ended queue with non-zero matching time and finite number of resources. In the single storage case, we obtain closed-form performance measures through matrix computation and inversion. This approach is extended to the multi-truck and larger storage case using a numerical method that, although exact, becomes computationally intensive as the number of states increases. To address this, an iterative fixed-point approximation is developed, reintroducing some blocked containers from one time interval into the next interval, in a model where the two sides of the queue are viewed as independent. We demonstrate the convergence of this process to a modified arrival rate, enabling the approximation of performance measures with an accuracy of 5% in approximately 90% of cases. Additionally, in the single-storage case, we extend the analysis to Erlang matching time and production time distributions, employing a $z$-transform to determine closed-form performance measures as functions of the roots of a polynomial.

(ii) *Admission control.* First, we demonstrate the accurate derivation of the optimal withholding threshold using the fixed-point approximation. Next, in the single-truck case, the dynamic counterpart of the admission control problem is formulated as a Markov decision process. We prove that the optimal inventory policy at the consignee is a state-dependent threshold policy with the withholding threshold increasing with the inventory level at the shipper.

(iii) *Insights for the management of empty containers.* We identify three contexts where the withholding decision plays distinct roles. In export-heavy areas, there is only marginal gain in well selecting the withholding threshold, and a near-optimal solution is to store all arriving containers. In balanced areas, the withholding threshold balances holding and travel costs. In import-heavy areas, the withholding threshold controls the flow of containers either in immediate return or in the direction of the shipper. The consignee benefits from the inventory capacity of the shipper, but the effect is highly concave for the selection of the optimal threshold, allowing the consignee, in many cases, to ignore information about the shipper's inventory capacity. Moreover, in export-heavy areas, a larger traffic intensity or lower service capacity tends to reduce withholding thresholds, while the opposite is true in import-heavy areas. Finally, while the reduction of variability in travel time or production time at the shipper tends to reduce costs, it marginally influences the selection of the optimal withholding threshold.

*Structure of the paper.* The next section reviews the relevant literature. Section 3 formulates the model and the optimization problem. Section 4 derives the performance measure of the double-ended queue with non-zero matching time. Section 5 presents an approximation for computing the performance measures. Section 6 evaluates the withholding policy and discusses its main drivers. Section 7 investigates the dynamic version of the control problem. Section 8 extends the model formulation to include Erlang matching and production times. Finally, Section 9 concludes the paper. The appendix of this paper contains a numerical method for performance evaluation, some supporting numerical results, and the mathematical proofs.

## 2. Literature review

The related literature can be categorized into three areas: (i) performance analysis of double-ended queues, (ii) admission control of queueing systems, and (iii) management of empty containers in the hinterland.

*Performance evaluation.* The concept of the double-ended queue was first introduced by Kendall (1951). This queueing model considers independent processes for both customer arrivals and server operations, making it applicable to a diverse range of applications such as shared-mobility systems (He et al., 2021), disaster and repair management (Di Crescenzo et al., 2012), passenger and taxi queues (Shi & Lian, 2016), buyers and sellers interaction (Liu et al., 2015), and the allocation of live organs (Elalouf et al., 2018). Several studies have explored aspects of the double-ended queue, including customer joining behavior (Jiang et al., 2021), performance evaluation (Diamant & Baron, 2019), and congestion control policies (Liu & Weerasinghe, 2021). However, most of these studies assume zero matching times, a limitation that we address in this paper.

With a zero matching time, the performance analysis is already challenging due to the bidirectional state space. Early studies established and analyzed the Chapman–Kolmogorov forward differential-difference equations for this queue (Sasieni, 1961). When at least one of the two queues has an infinite capacity, the matrix-geometric method proposed by Neuts (1981) enables the derivation of performance measures, as seen in Liu et al. (2020). For transient analyses, the Laplace transform method is also effective in performance evaluation (Conolly et al., 2002). However, in our case where both sides of the queue have finite capacities, these methods are not applicable. With two finite queues, the supplementary variable method was employed in Kashyap (1966). Nevertheless, with a non-zero matching time, the supplementary variable method leads to a state space with excessively high dimensionality, rendering performance evaluation impractical.

Analyzing a double-ended queue with non-zero matching time poses inherent challenges. Notably, the work of Kim et al. (2010) is cited for its development of a simulation model aimed at evaluating the performance measures. Additionally, Shi et al. (2015) contributed to this discourse by employing a numerical method based on the matrix geometric approach introduced by Neuts (1981). It is important to highlight that the latter method is specifically tailored to scenarios where matches can only be operated singularly and one of the queues has an infinite capacity. When confronted with a finite queue, direct application of this method is precluded, as the proportional relationship between two adjacent line vectors of probabilities is not maintained. Wang et al. (2023) analyzed a specific double-ended queue with a two-mass point distribution for the matching time. Finally, Nguyen and Phung-Duc (2022) investigated customers' strategic joining decisions in a double-ended queue with a non-zero matching time for a passenger-taxi system. They considered the equilibrium joining behavior of customers, requiring an evaluation of the expected wait in a given state at arrival without the need for the expected wait to be averaged across all arriving states.

*Admission control.* The admission control problem for a social planner involves determining whether to allow an arriving customer to join a queue, aiming to achieve an optimal trade-off between congestion and rejection. This problem has garnered significant attention in the literature. For an overview of admission control studies, we refer to the book by Boucherie and Van Dijk (2017), and the references therein.

One common approach is to employ a performance evaluation method for a given family of joining policies, often threshold or randomizing policies. This method is used to optimize joining parameters such as the joining probability or the joining threshold (Stidham, 1985). It relies on a performance evaluation method with properties suitable for optimizing joining parameters, necessitating consideration of a sufficiently simple queueing model.

Optimal admission policies can also be obtained without performance evaluation. Dynamic programming is a standard tool for computing and proving optimal policies (Boucherie & Van Dijk, 2017). In this paper, we employ the uniformization technique explained in Puterman (1994) to transform the continuous-time Markov chain into a discrete-time one. This transformation enables us to compute the optimal dynamic policy and corresponding performance measures. In the single-truck case, we additionally prove the threshold form of the optimal policy by establishing the propagation of certain properties for the value function, such as convexity and submodularity, using the approach of Koole (2007). To the best of our knowledge, the solution to the admission control problem in the single-truck case constitutes a novel contribution. While the admission control problem for this queue has been considered by Liu and Weerasinghe (2021), Lee et al. (2021), and Su and Li (2023), these studies focused on scenarios with a zero matching time.

*Container management in the hinterland.* Container management has been a prominent area of interest within the transport and maritime economics communities. Early research in the operations management and transportation science fields on containerization is reviewed by Dejax and Crainic (1987), while a recent survey can be found in Lee and Song (2017). Numerous studies have concentrated on sea terminal container management, addressing scheduling, estimating, and modeling aspects (Bakshi et al., 2011; Roy et al., 2020; Vis & Roodbergen, 2009). The transportation of empty containers has also garnered significant attention. However, few studies, including this paper, have explored the inventory theory perspective of managing empty containers in the hinterland.

Li et al. (2004) investigated strategies for importing and exporting empty containers to address shortages and reduce redundancy in the port. Using a Markov decision process approach, similar to this study, they find that a two-threshold policy is optimal for controlling import and export. Song (2007), Song and Zhang (2010) focused on optimizing the repositioning of empty containers within a single port, considering a two-state Markov chain demand model and aiming to minimize costs associated with holding, leasing, and repositioning. One difference with our model is that decisions are taken periodically instead of being taken at any point in time. Using a two-threshold policy, similar to the one of Li et al. (2004), Zhang et al. (2014) proposed an approximated solution approach for repositioning empty containers between multiple ports over multiple periods, taking into account stochastic demand, lost sales, and various operating costs. Later, Xie et al. (2017) investigated the repositioning problem not only through an inventory management angle but also through contracting. They characterized the optimal delivery policy between a dry port and seaport in a centralized model and proposed a bilateral buy-back contract to coordinate the decentralized system. Using a two-stage game model, Yu et al. (2018) analyzed optimal delivery policies and detention time decisions in an export-heavy area consisting of sea and inland container terminals, a container operator, and an ocean carrier, revealing coordination challenges in the decentralized setup.

While the aforementioned studies examined the inventory management of containers in the hinterland, they analyzed it only through the relationship between the shipping lines and the consignee. While this relationship is essential in the management of empty containers, it ignores the possibilities of street turn strategies that involve an exporter (a shipper) for the reuse of containers. In this study, we turn our attention to the consignee-shipper interaction for the decision to keep a container at the consignee's location. Legros et al. (2019) investigated this problem but in a context of an import-heavy area, where the shipper's inventory decision could be ignored. We instead propose a more general setting which accounts for any import-export ratio and allows understanding the impact of the shipper's decision on the consignee's inventory policy.

## 3. Formulation of the problem

We analyze the management of empty containers by a consignee in the hinterland. A consignee imports products via containers from a sea terminal. Later, empty containers are sent back to the sea terminal to be reloaded for a new import operation. Due to increased fuel prices and a shortage of trucks and drivers, road transportation costs have recently gone up and now represent a major part of import costs. To reduce these costs, consignees send their containers to shippers in need of sending their products to the sea terminal. The policy of reusing containers for the return trip to the sea terminal is referred to as a street turn strategy, as opposed to an immediate return policy, in which all containers are sent back empty to the sea terminal. For simplicity of modeling, we assume that we have a single shipper. We could instead consider a group of shippers viewed as a single entity, as in areas where shippers are located close to each other and face similar costs.

We assume that the containers' arrival process at the consignee is Poisson with constant parameter $\lambda_c$. This parameter is called the arrival rate of containers and represents the expected number of containers arriving per time unit. The Poisson assumption is justified for the arrival processes at sea terminals. Some statistical analyses have revealed that vessel arrivals fit well with a Poisson distribution (Kozan, 1997). In addition, truck arrivals at the sea terminal can be modeled by Poisson distributions (Roy et al., 2022). We make the further assumption that the arrival rate is constant over time. This may not be realistic because in some areas there is a very pronounced variation by time of day and day of the week, due to the activity at the sea terminal. If time dependency varies slowly relative to the system dynamics, then such systems have been typically analyzed using a point-wise stationary approximation, where the performance at a given time is approximated by the steady-state performance of the stationary system with a constant arrival rate (Green & Kolesar, 1991).

The shipper needs to send its loads to the sea terminal. It either asks the shipping line or the consignee to send an empty container. The need for empty containers at the shipper follows a Poisson process with rate $\lambda_s$. The demand rate $\lambda_s$ is the quantity of products produced by the shipper per time unit, measured in equivalent container volume. This parameter is also referred to as the production rate. We assume that it is cheaper for the shipper to use a container from the consignee than to request a container from the shipping line and pay for the travel time between the sea terminal and its location. The time to send a container from the consignee's location to the shipper is non-zero because it includes the time for the consignee and shipper to make an agreement, the time to find an available truck, and the transportation time between the consignee and shipper. To account for the variability of these durations, we assume that the total time to send a container from the consignee's location to the shipper, known as the matching time, is exponentially distributed with rate $\mu$. The matching rate $\mu$ is the inverse of the expected matching time. We further assume that there are $m$ trucks devoted to the matches, which creates a bound on the number of simultaneous matches that can be carried out. Furthermore, the shipper stores part of its stockpile for future matches. Specifically, the shipper has an inventory capacity of $q$ container volumes. If $q = 0$, the shipper declines to reuse containers from the consignee. If $q > 0$, then a quantity of at most $q$ equivalent containers is stored at the shipper's location, either through matching processes or waiting for a match to occur.

The consignee's objective is to determine the optimal withholding policy that minimizes the operational cost per time unit associated with

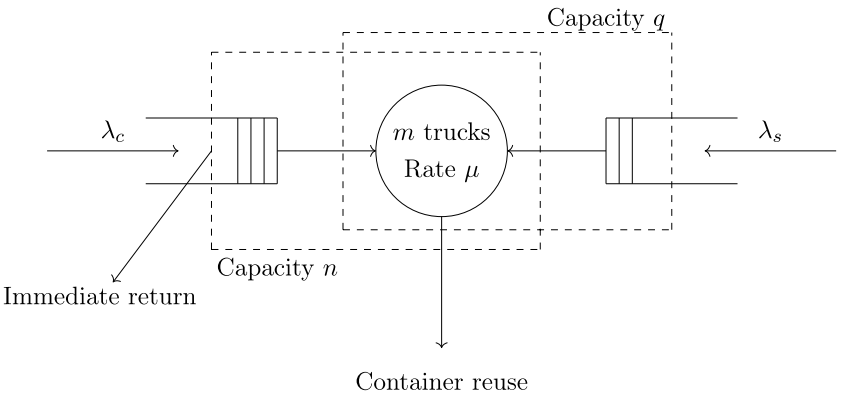



**Fig. 1.** Queueing model for the street-turn policy.

a street turn strategy. We use $\lambda_r$ to represent the expected rate of containers immediately returned to the sea terminal. This represents the expected number of containers that return empty to the sea terminal per time unit. Consequently, the average total extra travel cost of the street turn strategy is defined as $\lambda_r t_r d$, where $d$ is the additional distance in kilometers traveled if a container is immediately returned. This is the difference between the distance from the consignee's location to the sea terminal and the distance between the consignee's location and the shipper's location. The parameter $t_r$ represents the travel cost per kilometer. According to estimates by Pakulniewicz (2021), the average cost to transport a container on the road was €1.77 per kilometer in Europe in the second quarter of 2021. A similar range of values holds in the United States, with a noticeable upward trend. For our numerical estimations, we adopt a reference extra distance of $d = 120$ kilometers, resulting in an extra travel cost of $dt_r = €212.4$ per container.

In addition to the extra travel cost, a street turn policy incurs expenses related to the storage and detention of empty containers. BigRentz (2022) found that self-storage costs range from $130 to $175 per month per container, leading to an average estimate of $5 per container per day for storage. The equivalent detention cost per container is estimated to be €227.29 per day in New York (xChange, 2023), and slightly lower in Europe. This observation suggests that storage costs can be considered negligible compared to detention costs, resulting in a holding cost of $hE(N)$, where $E(N)$ is the expected number of containers in the system (at the location or operating a match). For our numerical illustration, we consider a detention cost $h$ of €200 per day, or equivalently €8.33 per hour, assuming the unit of time to be an hour.

The consignee has control over the maximal number of containers in the system, through the selection of a withholding threshold $n$. The withholding threshold is optimized such that the expected cost $E(C)$ defined as

$$E(C) := \lambda_r t_r d + hE(N), \tag{1}$$

is minimized. The withholding threshold can be selected as low as desired. However, it should be noted that if $n < m$, then $m - n$ trucks are never used and can be removed, which turns the situation to the case where the threshold level is higher than the number of trucks. Fig. 1 depicts the queueing model under consideration. In this figure, the queue on the left side represents the inventory at the consignee's location with containers arriving over time. The queue on the right side represents the inventory of production at the shipper's. In the center, the resources are represented by the trucks that can transport the empty containers from the consignee's location to the shipper's location.

In Section 7, we investigate the dynamic version of this control problem, where the withholding threshold can be selected as a function of the system state at the shipper. Next, in Section 8, we extend the model to Erlang distributions for the matching time and production time to investigate the role of the variability of these durations in decision making. We end this section with a list of notations used throughout the paper (Table 1).

## 4. Performance evaluation

In this section, we evaluate the performance measures of the double-ended queue. First, using a matrix approach, we derive closed-form expressions for stationary probabilities and performance measures when $q = 1$. This approach can be extended numerically to scenarios with multiple trucks and more than one storing space at the shipper. Next, we present some lower bounds for the performance measures in the general case.

A state of the system is defined by the pair $(x, y)$, where $x$ represents the number of containers in the system (either in the inventory or on the road for a match to be operated) with $x = 0, 1, \dots, n$, and $y$ is the equivalent container volume of goods at the shipper for $y = 0, 1, \dots, q$. We introduce the ratios $c := \frac{\lambda_c}{\mu}$ and $s := \frac{\lambda_s}{\mu}$ to simplify the involved expressions. The stationary probability of being in state $(x, y)$ is denoted by $p_{x,y}$. The transition rate from state $(x, y)$ to state $(x', y')$, denoted as $r_{(x,y),(x',y')}$ for $x = 0, 1, \dots, n$ and $y = 0, 1, \dots, q$, is defined as follows:

$$r_{(x,y),(x',y')} = \begin{cases} \lambda_c \text{ if } x = 0, 1, \dots, n-1 \text{ and } y = 0, 1, \dots, q, \\ \qquad \text{with } (x', y') = (x+1, y), \\ \lambda_s \text{ if } x = 0, 1, \dots, n \text{ and } y = 0, 1, \dots, q-1, \\ \qquad \text{with } (x', y') = (x, y+1), \\ \min(x, y, m)\mu \text{ if } x = 0, 1, \dots, n \text{ and } y = 0, 1, \dots, q, \text{ with} \\ \quad (x', y') = (x-1, y-1), \text{ and} \\ 0, \text{ otherwise.} \end{cases}$$

In Fig. 2, we present the Markov chain corresponding to the case $q = 1$. The system of equations derived from this Markov chain description can be analytically solved, enabling the expression of stationary probabilities as given in Theorem 1.

**Theorem 1** (*Stationary Probabilities*). *The stationary probabilities are given by*

$$p_{x,0} = \frac{\left(\frac{c(s+1)}{s} - 1\right)\left[r_2^x(r_2 - c)(c(c+s) - r_1 s) - r_1^x(r_1 - c)(c(c+s) - r_2 s)\right]}{r_2^{n+2}(s+1)(1-r_1)(c-r1) - r_1^{n+2}(s+1)(r_2-1)(r_2-c) - cs(r_2 - r_1)}, \quad \text{for } x = 0, 1, \dots, n-1, \tag{2}$$

$$p_{n,0} = \frac{c}{s}\frac{\left(\frac{c(s+1)}{s} - 1\right)\left[r_2^{n-1}(r_2 - c)(c(c+s) - r_1 s) - r_1^{n-1}(r_1 - c)(c(c+s) - r_2 s)\right]}{r_2^{n+2}(s+1)(1-r_1)(c-r1) - r_1^{n+2}(s+1)(r_2-1)(r_2-c) - cs(r_2 - r_1)} \quad \textit{and} \tag{3}$$

$$p_{x,1} = \frac{c\left(\frac{c(s+1)}{s} - 1\right)\left[r_2^x(c(c+s) - r_1 s) - r_1^x(c(c+s) - r_2 s)\right]}{r_2^{n+2}(s+1)(1-r_1)(c-r1) - r_1^{n+2}(s+1)(r_2-1)(r_2-c) - cs(r_2 - r_1)}, \quad \text{for } x = 0, 1, 2, \dots, n, \tag{4}$$

**Table 1**
Table of notations.

| | |
|---|---|
| System state | |
| $x$ | Number of containers at the consignee's location or carrying out a match |
| $y$ | Amount of goods at the shipper in container volumes |
| System parameters | |
| $\lambda_c$ | Containers' arrival rate at the consignee's location |
| $\lambda_s$ | Containers' demand rate at the shipper's location |
| $\mu$ | Matching rate |
| $m$ | Number of trucks to carry out matches |
| $c, s$ | Ratios $\lambda_c/\mu$ and $\lambda_s/\mu$ |
| $q$ | Maximum inventory for the shipper to hold (expressed in container equivalents) |
| $n$ | Withholding threshold for empty containers |
| $n_y$ | State-dependent admission withholding thresholds for empty containers |
| Cost parameters and distances | |
| $t_r d$ | Extra transportation cost per container (€212.4 per container) |
| $h$ | Holding cost per time unit and per container (€8.33 per hour) |
| Performance measures | |
| $p_{x,y}$ | Stationary probability to be in state $(x, y)$ |
| $\lambda_r$ | Rate of empty containers returned immediately to the sea terminal per time unit |
| $E(N)$ | Expected number of containers in the system |
| $E(C)$ | Expected operational cost per time unit |

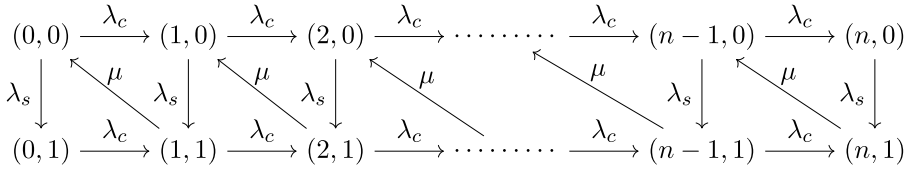



**Fig. 2.** Markov chain with $q = 1$.

*where* $\Delta = (c+s+1)^2 - 4s$, $r_1 = \frac{c\left(c+s+1-\sqrt{(c+s+1)^2-4s}\right)}{2s}$, *and* $r_2 = \frac{c\left(c+s+1+\sqrt{(c+s+1)^2-4s}\right)}{2s}$.

To prove this result, we begin by expressing the vector $(p_{x,0}, p_{x,1})$ as a function of $(p_{n,0}, p_{n,1})$ for $x = 1, 2, \ldots, n$. This enables us to determine a matrix that establishes a proportionality between $(p_{x,0}, p_{x,1})$ and $(p_{x+1,0}, p_{x+1,1})$. Using the relations at $x = 0$ and inverting the considered matrix, we subsequently express $(p_{x,0}, p_{x,1})$ as a function of $(p_{x-1,0}, p_{x-1,1})$. Next, we calculate the equivalent diagonal matrix that relates the two vectors and observe that its eigenvalues are solutions to the quadratic equation in $z$: $sz^2 - c(c+s+1)z + c^2 = 0$. The solutions to this equation are denoted as $r_1$ and $r_2$. Consequently, each probability $p_{x,0}$ and $p_{x,1}$ can be expressed as functions of $r_1^x$ and $r_2^x$. The coefficients of $r_1^x$ and $r_2^x$ are determined by the initial relation at $x = 0$ and the normalizing condition.

In Proposition 1, we deduce the expected number of empty containers and the rate of immediately returned containers.

**Proposition 1** (*Performance Measures*). *The expected number of containers* $E(N)$ *and the rate of immediately returned containers* $\lambda_r$ *are given by*

$$E(N) = \frac{n\alpha\frac{s}{c}f(n) + \frac{r_2(1-r_2^n)((c+s)r_2-c)}{(r_2-1)^2} - \frac{r_1(1-r_1^n)((c+s)r_1-c)}{(r_1-1)^2}}{(s+1)f(n)+sB}, \text{ and}$$
$$\lambda_r = \lambda_c \frac{\frac{s}{c}(\alpha-1)f(n)}{(s+1)f(n)+sB}, \tag{5}$$

*where* $f(n) := \frac{r_2^{n+1}}{r_2-1}(r_2 - c/s) + \frac{r_1^{n+1}}{1-r_1}(r_1 - c/s)$, $B := \frac{c\sqrt{\Delta}}{s-c(s+1)}$, *and* $\alpha := \frac{c(s+1)}{s}$.

We deduce these expressions from the stationary probabilities of Theorem 1. We find $\lambda_r$ with $\lambda_r = \lambda_c(p_{n,0} + p_{n,1})$. Next, the expected number of containers is given by $E(N) = \sum_{x=1}^{n} x(p_{x,0} + p_{x,1})$. Note that we simplify the involved expressions using $r_1 r_2 = \frac{c^2}{s}$ and $r_1 + r_2 = \frac{c(c+s+1)}{s}$.

*Lower bounds in the general case.* The method to obtain the performance measures can be extended to the general case. However, this does not lead to closed-form solutions. Instead, we end this section by relating the double-ended queue with two canonical queues that can be used as lower bounds for the system performance. When one of the system parameters, $\lambda_s$ or $\mu$, is high as compared to the others, we can obtain lower bounds for the performance measures as follows:

- *Export-heavy area.* When $\lambda_s \gg \lambda_c, \mu$, the system can be viewed as an $M/M/m/n$ queue with arrival rate $\lambda_c$ and service rate $\mu$ as it asymptotically never happens that an available empty container has to wait before a production is being made. In this case, the performance measures can be approximated by the formulas provided in (6) with $c_k = \frac{\lambda_c}{\mu}$.
- *Instantaneous match.* When $\mu \gg \lambda_c, \lambda_s$, we approximate the model by a double-ended queue with zero matching time. We thus obtain $\lambda_r = \lambda_c \frac{1-\frac{s}{c}}{1-\left(\frac{s}{c}\right)^{n+q+1}}$, and $E(N) = \frac{1-\frac{s}{c}}{1-\left(\frac{s}{c}\right)^{n+q+1}} \sum_{x=1}^{n} x \left(\frac{c}{s}\right)^{x-n}$.

## 5. Fixed-point approximation

When $n$ and $q$ become large, deriving the stationary probabilities becomes computationally intensive. To avoid this difficulty, we propose a fixed-point approximation for the computation of performance measures.

Consider a given interval $k$ of observation. We approximate the double-ended queue by two independent queues, each with $m$ exponential servers operating at a service rate of 1. We designate Queue 1 as the queue of empty containers at the consignee and Queue 2 as the queue of goods at the shipper's location in equivalent container volume. At interval $k$, the arrival of containers at Queue 1 follows a Poisson distribution with rate $c_k$, and the arrival of production in equivalent containers at Queue 2 follows a Poisson distribution with rate $s_k$. Consequently, we estimate performance measure by leveraging the performance metrics of Queue 1, which is characterized as an $M/M/m/n$ queue, while

Queue 2 is an $M/M/m/q$ queue. These metrics are detailed in Gross and Harris (1985), pages 75–76, and are expressed as follows:

$$E(N)^k = \frac{\frac{c_k^m\left(\frac{c_k}{m}\right)}{m!\left(1-\frac{c_k}{m}\right)^2}\left[1-\left(\frac{c_k}{m}\right)^{n-m+1}-\left(1-\frac{c_k}{m}\right)(n-m+1)\left(\frac{c_k}{m}\right)^{n-m}\right]}{\sum_{x=0}^{m-1}\frac{c_k^x}{x!}+\frac{c_k^m}{m!}\frac{1-(c_k/m)^{n-m+1}}{1-c_k/m}} + c_k\left(1-\frac{\lambda_r^k}{\lambda_c}\right), \text{ and} \tag{6}$$

$$\lambda_r^k = \lambda_c \frac{\frac{c_k^n}{m!m^{n-m}}}{\sum_{x=0}^{m-1}\frac{c_k^x}{x!}+\frac{c_k^m}{m!}\frac{1-(c_k/m)^{n-m+1}}{1-c_k/m}}, \text{ assuming that } n \geq m.$$

We initialize the computation of the sequences $(c_k)_{k\geq 0}$ and $(s_k)_{k\geq 0}$ with $c_0 = c$ and $s_0 = s$. This initialization assumes that a container in Queue 1 never has to wait for the production from the shipper, and vice versa, the shipper never has to wait for the arrival of a container at the consignee. This initialization then leads to an underestimation of the expected number of containers in the system.

At a given interval, even if a server should not operate a match, we assume that a match is processed. A correction is made at the next interval by reintroducing the containers that should not have been processed into the flow of arriving containers. The probability that a container from Queue 1 cannot operate a match is assumed to be the probability that a server at Queue 2 is idle. This idling probability, denoted as $I(z, y)$ with arrival rate $z$ and system capacity $y \geq m$, is given by

$$I(z,y) = \frac{\sum_{x=0}^{m-1}\frac{m-x}{m}\frac{z^x}{x!}}{\sum_{x=0}^{m-1}\frac{z^x}{x!}+\frac{z^m}{m!}\frac{1-(z/m)^{y-m+1}}{1-z/m}}.$$

Consequently, the arrival rate at iteration $k+1$ at Queue 1 is the sum of the new arrivals with arrival rate $c$ and those which were blocked at the last interval, $c_k I(s_k, q)$. Similarly, the arrival rate at Queue 2 is the sum of $s$ and $s_k I(c_k, n)$. Thus, we estimate the system performance by letting $k$ tend to infinity in the following iterative definition:

$$c_{k+1} = c + c_k I(s_k, q), \text{ and } s_{k+1} = s + s_k I(c_k, n), \text{ with } c_0 = c \text{ and } s_0 = s. \tag{7}$$

In Theorem 2, we prove that both $c_k$ and $s_k$ converge to a finite limit as $k$ grows to infinity.

**Theorem 2** (*Convergence Result*). *The sequences $(c_k)_{k\geq 0}$ and $(s_k)_{k\geq 0}$ converge to finite limits $c^*$ and $s^*$, respectively, which are solutions of the following system:*

$$c^* = c + c^* I(s^*, q), \text{ and } s^* = s + s^* I(c^*, n). \tag{8}$$

One difficulty to prove the convergence result is that $(c_k)_{k\geq 0}$ and $(s_k)_{k\geq 0}$ are not monotonous in $k$. Instead, we first prove that these two sequences are bounded. Next, we consider the function from $\mathbb{R}^2$ to $\mathbb{R}^2$ defined as $(u, v) :\to (c + uI(v, q), s + vI(u, n))$, and prove that this function is a contraction mapping. To establish this property, we use the expression of the upper bounds for $c_k$ and $s_k$ and a novel property of the idling probability, as given in Lemma 1.

**Lemma 1** (*Property of the Idling Probability*). *For $z > 0$ and $y \geq m$, we have*

$$1 - I(z,y) + z\frac{\partial I(z,y)}{\partial z} > 0. \tag{9}$$

The proof of Lemma 1 relies on the concavity result of the throughput out of an $M/M/m/n$ queue from Meester and Shanthikumar (1990).

**Table 2**
Computation of the optimal threshold level $n$ ($q = 20$, $m = 10$, $\mu = 1$).

| $s \backslash c$ | 5 | 8 | 9 | 10 | 11 | 12 | 15 |
|---|---|---|---|---|---|---|---|
| 5 | 15\15 | 10\10 | 10\10 | 10\10 | 10\10 | 10\10 | 10\10 |
| 8 | **60\63** | 21\21 | 15\15 | 13\13 | 12\12 | 12\12 | 11\11 |
| 9 | **124\128** | 48\48 | 22\22 | 17\17 | 15\15 | 14\14 | 13\13 |
| 10 | **125\131** | 57\57 | 30\30 | 21\21 | 18\18 | 16\16 | 14\14 |
| 11 | **128\131** | **58\59** | 36\36 | 24\24 | 20\20 | 18\18 | 15\15 |
| 12 | **131\132** | **59\60** | 38\38 | 26\26 | 21\21 | 18\18 | 15\15 |
| 15 | 132\132 | 60\60 | 40\40 | 28\28 | 22\22 | 19\19 | 16\16 |

*Accuracy.* The value of the proposed approximation lies in its capacity to accurately evaluate performance measures across diverse parameter sets. In the appendix, we present a comparison between the fixed-point approximation and exact performance measures. The approximation does not function as an upper or lower bound. Its limitation is its neglect of correlation between the two queues. However, around two-thirds of our numerical results exhibit a relative difference of less than 1%, and approximately 90% display a relative difference of less than 5%, indicating the approximation's relative accuracy. Furthermore, the accuracy tends to improve, with some counterexamples, with a high arrival rate at the importer compared to the production rate at the shipper (i.e., in import areas), increasing inventory capacities at the shipper and consignee (parameters $n$ and $q$), and augmenting the transport capacity (i.e., the number of trucks, $m$). It should be noted that as the number of trucks increases, the approximation tends to perform better in cases where $c < s$, while situations with $c = s$ remain less accurately approximated. Moreover, when $s$ becomes very large as in export-heavy area, then the approximation also becomes accurate, even if the transport capacity is low.

*Approximation of the optimal threshold level.* In Table 2, we calculate the threshold level for various values of $c$ and $s$. Additional numerical illustrations are provided in the appendix. The first value before the backslash represents the exact threshold, while the second value after the backslash indicates the approximate threshold obtained using the fixed-point approximation.

This shows that the fixed-point approximation method gives the exact threshold in most cases. However, in some instances, particularly when $c$ is low (indicating an export-heavy area) and when $s$ approaches the system capacity $m$, the approximation tends to overestimate the optimal threshold. It is important to note that in these cases, the sensitivity of the performance measures to the threshold $n$ is very low. Consequently, any inaccuracies in determining the optimal threshold do not result in significant consequences for the evaluation of the performance measures.

## 6. Optimal withholding decisions

We now evaluate the expected cost of the optimal withholding policy in order to determine the contexts where reusing containers is most advantageous as compared to an immediate return policy. We focus on (i) the import-export ratio, (ii) the intensity of demand relative to the service capacity, and (iii) the inventory capacity at the shipper. First, we identify cases in which the immediate return policy is optimal. Adjusting the result of Theorem 1, we establish in Corollary 1 a condition ensuring the optimality of the immediate return policy.

**Corollary 1** (*Optimality of the Immediate Return Policy*). *The full rejection policy is optimal if and only if*

$$\lambda_c t_r d(\phi(q) + s\overline{B}) < h(c\phi(q) + \overline{B}(c - s)), \tag{10}$$

*where* $\phi(q) = \frac{t_2^{q+1}(t_2 - s/c)}{t_2 - 1} + \frac{t_1^{q+1}(t_1 - s/c)}{1 - t_1}$, $\overline{B} = \frac{s\sqrt{(c+s+1)^2-4c}}{c-s(c+1)}$, $t_1 = \frac{s(s+c+1-\sqrt{(c+s+1)^2-4c})}{2c}$, *and* $t_2 = \frac{s(s+c+1+\sqrt{(c+s+1)^2-4c})}{2c}$.

**Table 3**
Impact of the import-export balance and traffic intensity ($m = 5,\ q = 5,\ \mu = 1$).

| $s$ | $c=1$ | $c=2$ | $c=5$ | $c=8$ | $c=10$ | $c=1$ | $c=2$ | $c=5$ | $c=8$ | $c=10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | Optimal threshold ($n$) | | | | | Optimal cost $E(C)$ | | | | |
| 1 | 5 | 3 | 2 | 2 | 2 | 40.949 | 236.002 | 870.836 | 1507.841 | 1932.634 |
| 2 | 24 | 8 | 4 | 4 | 4 | 9.065 | 81.394 | 690.631 | 1326.650 | 1751.388 |
| 5 | 65 | 41 | 8 | 6 | 6 | 8.367 | 18.117 | 370.296 | 990.930 | 1413.231 |
| 8 | 81 | 56 | 11 | 7 | 6 | 8.351 | 17.375 | 267.433 | 874.686 | 1294.372 |
| 10 | 86 | 61 | 12 | 8 | 7 | 8.348 | 17.251 | 236.770 | 838.582 | 1256.199 |
| | | | Saving= $\frac{E(C)}{\lambda_c t_r d}$ | | | Matching proportion | | | | |
| 1 | 19.279% | 55.556% | 82.000% | 88.738% | 90.990% | 88.754% | 97.718% | 96.994% | 97.425% | 97.537% |
| 2 | 4.268% | 19.161% | 65.031% | 78.075% | 82.457% | 50.000% | 88.989% | 94.213% | 94.966% | 95.116% |
| 5 | 3.939% | 4.265% | 34.868% | 58.317% | 66.536% | 20.000% | 40.000% | 70.035% | 70.852% | 71.238% |
| 8 | 3.932% | 4.090% | 25.182% | 51.476% | 60.940% | 12.500% | 25.000% | 50.696% | 51.508% | 51.462% |
| 10 | 3.930% | 4.061% | 22.295% | 49.352% | 59.143% | 10.000% | 20.000% | 42.157% | 43.254% | 43.332% |
| | Holding cost | | | | | Proportion of holding cost= $\frac{\text{Holding cost}}{E(C)}$ | | | | |
| 1 | 17.062 | 18.754 | 14.851 | 15.572 | 15.802 | 41.667% | 7.947% | 1.705% | 1.033% | 0.818% |
| 2 | 9.065 | 34.621 | 28.846 | 30.865 | 31.443 | 100.000% | 42.535% | 4.177% | 2.327% | 1.795% |
| 5 | 8.367 | 18.117 | 52.063 | 44.183 | 45.776 | 100.000% | 100.000% | 14.060% | 4.459% | 3.239% |
| 8 | 8.351 | 17.375 | 66.861 | 50.708 | 44.818 | 100.000% | 100.000% | 25.001% | 5.797% | 3.463% |
| 10 | 8.348 | 17.251 | 70.188 | 58.088 | 52.580 | 100.000% | 100.000% | 29.644% | 6.927% | 4.186% |

To prove this condition, we modify the formulas of Theorem 1 by interchanging the roles of $s$ and $c$ and the roles of $n$ and $q$. Corollary 1 suggests that the immediate return policy tends to be optimal when the holding cost $h$ is high compared to the rejection cost $t_r d$, and in import-heavy areas where $\lambda_c$ is high.

*Import-export ratio.* In Table 3, we present various scenarios involving export–import ratios and parameter intensities by selecting $c$ and $s$ from the set $\{1, 2, 5, 8, 10\}$ within a context where there are $m = 5$ trucks and it takes one hour to transport a container from the consignee to the shipper. For each $(c, s)$ combination, we determine the optimal threshold level, the corresponding expected cost $E(C)$, the savings measured in terms of the ratio between the optimal cost and the cost of an immediate return policy, the holding cost, the proportion of the holding cost in the expected cost $E(C)$, and the matching proportion computed as $\frac{\lambda_c - \lambda_r}{\lambda_s}$. This latter metric evaluates the proportion of production at the shipper that can be sent to the sea terminal through the reuse of containers.

We observe that the optimal threshold increases with $s$ and decreases with $c$. This observation is justified by Proposition 2, which establishes that the same outcome holds for the fixed-point approximation.

**Proposition 2** (*Effect of $c$ and $s$*). *In the fixed-point approximation, the optimal threshold is increasing in $s$ and decreasing in $c$.*

We prove this proposition by differentiating Eq. (8), in combination with the monotonicity properties of the $M/M/m/n$ queue.

This result is intuitive, implying that when there is more demand from the shipper compared to container arrivals (as occurs when $s$ is large and $c$ is low, characterizing an export-heavy area), the consignee should store more containers. These contexts also yield the greatest savings through the implementation of a street-turn policy, as indicated by the low values of the ratio $\frac{E(C)}{\lambda_c t_r d}$. In such cases, the cost is primarily constituted of holding costs, signifying that almost no containers are immediately returned. Consequently, the selection of the withholding threshold plays a marginal role in cost, as long as it is sufficiently high to ensure close-to-full reuse.

Even in contexts that do not inherently encourage reuse, such as import-heavy areas, substantial savings can be observed. In the worst-case scenario, around 10% of the cost can be saved compared to a full rejection policy. In these cases, the holding proportion is marginal, so the role of the withholding threshold is to organize the routing of containers either directly back to the sea terminal or to the shipper. In contrast in balanced-area, the withholding threshold acts to balance holding and transportation costs.

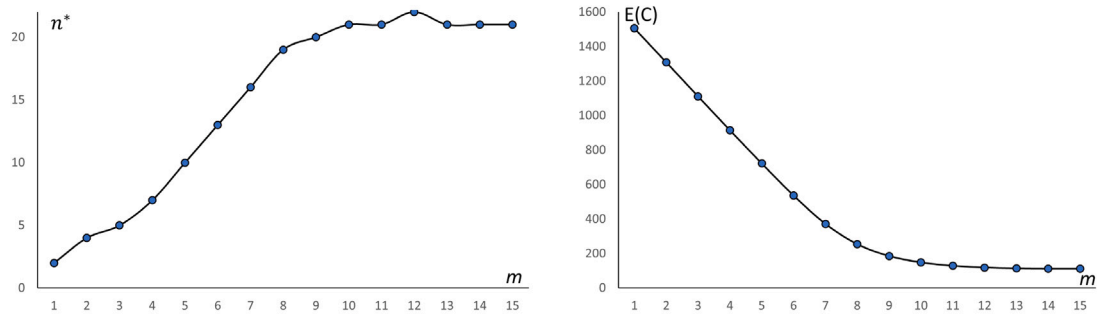
*Traffic intensity and service capacity.* In Table 3, we observe that the optimal threshold is non-monotonic in traffic intensity, exemplified when $c = s$. This non-monotonicity can be attributed to the interplay of two competing phenomena. As the flow of container arrivals increases, holding costs tend to rise, motivating a reduction in the withholding threshold. Simultaneously, if the demand from the shipper also increases, the withholding threshold should increase. However, when combined with the import-export ratio, an increase in traffic intensity results in an increased withholding threshold in import-heavy areas, as the impact of an increase in $s$ becomes more significant than an increase in $c$. Conversely, the opposite holds in export-heavy areas.

An increase in service capacity has a similar effect to a reduction in traffic intensity, as illustrated in Fig. 3 by the non-monotone behavior of the optimal threshold. Furthermore, consistent with the impact of traffic intensity, the optimal cost is decreasing and convex in the number of trucks, as illustrated in Fig. 3(b).

*Inventory capacity of the shipper.* In practice, the consignee is not necessarily informed by the shipper on the value of the inventory capacity $q$. Thus, it is interesting to know whether the knowledge of $q$ is essential in decision making. In Table 4, we determine the optimal threshold level for different values of $q$. We observe that the consignee benefits from the shipper's inventory capacity. This result is proven with the fixed-point approximation in Proposition 3. As the shipper and consignee are in a symmetric situation, the shipper also benefits from the consignee's inventory, which makes the street turn strategy a win-win situation for both participants, incentivizing the shipper to store part of the production for street turn.

**Proposition 3** (*Effect of $q$*). *In the fixed-point approximation, the optimal threshold is increasing in $q$.*

As $q$ increases, the expected cost decreases and a larger number of containers is stored. However, the impact of an increased inventory level at the shipper has a limited effect on the optimal cost. This is particularly true when the imbalance between imports and exports is high. In an import-heavy area, the inventory at the exporter does not have time to reach its maximum level, which limits the effect of an increase of $q$. In export-heavy areas, most containers from the consignee are already used for matches when $q$ is small, so there is scant opportunity for improved cost savings when $q$ increases. The role of $q$ is a bit more apparent in balanced areas, where the size of shipper inventory plays a role in the effective demand as viewed by the consignee. The low sensitivity of the optimal decisions and optimal costs to $q$ reveals that knowing the inventory policy of the shipper (i.e., the value of $q$) has in many cases a minor impact.




(a) Optimal withholding threshold (b) Expected cost

**Fig. 3.** Impact of the number of trucks ($c = s = 8$, $q = 20$).

**Table 4**
Effect of the inventory capacity at the shipper ($m = q = 1$, $\mu = 1$).

| $q$ | $c = 2, s = 0.5$ | | | | $c = 1, s = 1$ | | | | $c = 0.5, s = 2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | $\lambda_r/\lambda_c$ | $E(N)$ | $E(C)$ | $n$ | $\lambda_r/\lambda_c$ | $E(N)$ | $E(C)$ | $n$ | $\lambda_r/\lambda_c$ | $E(N)$ | $E(C)$ |
| 1 | 1 | 0.839 | 0.839 | 363.273 | 2 | 0.533 | 1.400 | 124.947 | 6 | 0.030 | 1.632 | 16.822 |
| 2 | 1 | 0.799 | 0.799 | 345.884 | 3 | 0.385 | 1.905 | 97.700 | 12 | 0.000 | 1.113 | 9.301 |
| 3 | 1 | 0.779 | 0.779 | 337.215 | 4 | 0.300 | 2.447 | 84.090 | 13 | 0.000 | 1.024 | 8.540 |
| 4 | 1 | 0.768 | 0.768 | 332.544 | 4 | 0.266 | 2.306 | 75.740 | 13 | 0.000 | 1.005 | 8.382 |
| 5 | 1 | 0.761 | 0.761 | 329.786 | 4 | 0.246 | 2.218 | 70.719 | 13 | 0.000 | 1.001 | 8.345 |
| 10 | 1 | 0.751 | 0.751 | 325.473 | 5 | 0.182 | 2.603 | 60.308 | 13 | 0.000 | 0.999 | 8.333 |

## 7. Optimal dynamic withholding policy in the single-truck case

Shifting to the dynamic admission control problem, we explore optimal policies for managing empty containers when the shipper's stored production quantity is communicated to the consignee. Using a Markov decision process approach, we find that the optimal dynamic withholding policy is a state-dependent threshold policy in the single-truck case. In Definition 1, we define a state-dependent threshold policy.

**Definition 1** (*Definition of a State-Dependent Threshold Policy*)**.** A state-dependent threshold policy is defined by the thresholds $n_0, n_1, \dots, n_q$. These thresholds separate the states in which a container is kept in the system from those where it is returned to the sea terminal. Specifically, for $y = 0, 1, \dots, q$, when $x$ containers are already present in the system with $x \geq 0$, the handling of a new container at arrival is determined as follows:

- If $x < n_y$, then the container is either stored at the consignee or sent for matching.
- If $x \geq n_y$, then the container is directly returned to the sea terminal.

We deduce from this definition that when $n_0 = n_1 = \cdots = n_q$, the state-dependent threshold policy becomes the static threshold policy analyzed in the previous section.

First, we provide the optimality equations for the long-run relative value function $V(x, y)$, and average optimal cost $E(C)^*$ in the case $m = 1$. This is possible because the maximal event rate $\lambda_c + \lambda_s + \mu$ is bounded. Thus, for $x \geq 0$ and $y = 0, 1, \dots, q$, we have

$$\begin{aligned} E(C)^* = {} & \frac{hx}{\lambda_c + \lambda_s + \mu} + \frac{\lambda_c}{\lambda_c + \lambda_s + \mu} \min(V(x+1, y) - V(x, y), t_r d) \\ & + \frac{\lambda_s}{\lambda_c + \lambda_s + \mu} \mathbb{1}_{y<q}(V(x, y+1) - V(x, y)) \\ & + \frac{\min(x, y, 1)\mu}{\lambda_c + \lambda_s + \mu}(V(x-1, y-1) - V(x, y)). \end{aligned} \tag{11}$$

The minimizing operator in (11) represents the control action to either store or return a container.

In Theorem 3, we prove the threshold form of the optimal policy for $m = 1$.

**Table 5**
Properties of the function $f(x, y)$.

| Property | Inequality |
|---|---|
| Increasing in $x$ | $f(x+1, y) \geq f(x, y)$ |
| Decreasing in $y$ | $f(x, y+1) \leq f(x, y)$ |
| Increasing in $(x, y)$ | $f(x+1, y+1) \geq f(x, y)$ |
| Convex in $x$ | $f(x+2, y) + f(x, y) \geq 2f(x+1, y)$ |
| Convex in $y$ | $f(x, y+2) + f(x, y) \geq 2f(x, y+1)$ |
| Submodular in $(x, y)$ | $f(x, y+1) + f(x+1, y) \geq f(x+1, y+1) + f(x, y)$ |
| Relation A | $f(x+2, y+1) + f(x, y) - f(x+1, y) - f(x+1, y+1) \geq 0$ |
| Relation B | $f(x+1, y+2) + f(x, y) - f(x, y+1) - f(x+1, y+1) \geq 0$ |

**Theorem 3** (*Optimal Withholding Policy*)**.** *In the single-truck case (i.e., $m = 1$), the optimal dynamic withholding policy for the consignee is a state-dependent threshold that follows Definition 1.*

To prove Theorem 3, we prove that if it is optimal to return a container in state $(x, y)$, then the same action is optimal in state $(x+1, y)$. A necessary condition for this is that if $V(x+1, y) - V(x, y) - t_r d \geq 0$, then $V(x+2, y) - V(x+1, y) - t_r d \geq 0$, or equivalently $V(x+2, y) - V(x+1, y) - t_r d \geq V(x+1, y) - V(x, y) - t_r d$, which can be rewritten as $V(x+2, y) + V(x, y) - 2V(x+1, y) \geq 0$.

Therefore, by showing that $V(x, y)$ is convex in $x$, we prove that the optimal policy converges to the unique average optimal policy, as defined in Definition 1. We prove this result by considering the equivalent discrete time Markov using uniformization. This allows us to define the value function in an iterative way and prove the threshold form of the optimal policy by iteration. However, the convexity property in $x$ of $V(x, y)$ cannot be proven in isolation but has to be proven with a set of other properties. This set of properties $\mathcal{C}$ is defined for a given function $f$ in Table 5.

**Remarks:**

- The induction approach to prove Theorem 3 cannot be extended to the case $m > 1$, although we observe numerically that Theorem 3 also holds for $m > 1$. The first-order monotonicity properties can be proven for $m > 1$. However, after a $\mu$-transition, Relations A and B induce a negative term. The problem is mainly due

to the combination of the convex properties together with the submodular property.

- The optimal dynamic policy differs from the static one. However, the difference between these two policies in terms of cost is minor. Some supporting numerical data are provided in the appendix. This underscores that a dynamic threshold policy might be unnecessarily complex to implement, demanding real-time information about the shipper inventory. Instead, a static threshold policy proves to be nearly optimal and simpler to implement.
- The decision variable is the number of containers in the system, not the number of containers at the location. The result could not be proven if the latter was used to define the state space. This distinction is important compared to other admission control problems where one can equivalently consider the quantity in the queue or the quantity in the system.

## 8. Model extensions

In this section, we explore two extensions of the initial model where either the matching time (Section 8.1) or the production time (Section 8.2) follows an Erlang distribution rather than an exponential one. The Erlang distribution is known for its lower variability compared to the exponential distribution and can tend to a deterministic distribution. This characteristic may better reflect the reality of transportation time, where although randomness exists, the variability might not be as pronounced as that in an exponential distribution. For production time, the inventory at the shipper often accumulates gradually in quantities smaller than an equivalent container volume. As a result, one may need to wait through several exponential phases until an equivalent container volume can be stored.

The aim of this section is to evaluate the influence of the variability in matching or production time on the optimal threshold level and corresponding expected cost. To this end, we analyze the simplest case where $q = 1$. We employ a $z$-transform to derive performance measures. This technique enables the expression of performance measures in closed-form as functions of the roots of a polynomial, facilitating a more comprehensive understanding of the impact of distribution variability on the optimal threshold level.

### 8.1. Erlang matching time

We consider an Erlang matching time with $r$ phases and rate $r\mu$ per phase. In this way, by varying $r$, we change the variability of the distribution without modifying its mean. We redefine a state of the system by the couple $(x, y)$ where $x$ is the number of containers and $y$ is the number of matching phases that remains to be done before the goods from the shipper can be sent to the shipping line. We have $x = 0, 1, \dots, n$ and $y = 0, 1, \dots, r$. State $y = 0$ corresponds to the case where there are no goods available at the shipper for operating a match, while with states $y = 1, 2, \dots, r$, there are still $y$ phases that need to be completed. It should be noted that if $x = 0$, then either $y = 0$ (i.e., the system is empty), or $y = r$ (i.e., there is no container available but a quantity of goods is waiting for a match to be operated). The cases where $y = 1, 2, \dots, r-1$ are only possible if at least one container is available (i.e., $x \geq 1$).

The transition rate from state $(x, y)$ to state $(x', y')$, denoted by $t_{(x,y),(x',y')}$ for $x = 0, 1, \dots, n$ and $y = 0, 1, \dots, r$, is defined by

$$t_{(x,y),(x',y')} = \begin{cases} \lambda_c & \text{if } x = 0, 1, \dots, n-1 \text{ and } y = 0, 1, \dots, r, \\ & \quad \text{with } (x', y') = (x+1, y), \\ \lambda_s & \text{if } x = 0, 1, \dots, n \text{ and } y = 0, \text{ with } (x', y') = (x, r), \\ r\mu & \text{if } x = 1, 2, \dots, n \text{ and } y = 2, \dots, r, \\ & \quad \text{with } (x', y') = (x, y-1), \\ r\mu & \text{if } x = 1, 2, \dots, n \text{ and } y = 1, \\ & \quad \text{with } (x', y') = (x-1, 0), \\ 0 & \text{otherwise.} \end{cases}$$

**Table 6**
Performance measures with Erlang matching time.

| Cases | $E(N)$ |
|---|---|
| $r = 1$ | See Proposition 1 |
| $n = 1$ | $\dfrac{s + c + \frac{c}{s}}{s + c + 1 + s/c + c/s}$ |
| $n = \infty$, $\dfrac{c(s+1)}{s} < 1$ | $\dfrac{c}{s}\dfrac{-s^2(c(r-1)-2r)+2cr}{2r(s-c(s+1))}$ |
| $n = \infty$, $r = \infty$, $\dfrac{c(s+1)}{s} < 1$ | $\dfrac{c}{s}\dfrac{2c+2s^2-cs^2}{2(s-c(s+1))}$ |

The stationary probabilities are denoted by $p_{x,y}$. They can be found using the same approach as in the multi-server case. However, in this context, we propose employing a $z$-transform, allowing us to derive performance measures in closed-form as functions of the roots of a polynomial. We define the polynomials $P_0(z)$, $P_1(z)$, $P_2(z)$, …, $P_r(z)$ as $P_k(z) := \sum_{x=0}^{n} p_{x,k} z^x$, for $k = 0, 1, \dots, r$. From the balance equations that relate the stationary probabilities, we obtain

$$\begin{aligned} &(c(1-z)+s)P_0(z) - c(1-z)p_{n,0}z^n = rz^{-1}P_1(z), \\ &(c(1-z)+r)P_k(z) - c(1-z)p_{n,k}z^n = rP_{k+1}(z), \text{ for } k = 1, 2, \dots, r-2, \\ &(c(1-z)+r)P_{r-1}(z) - c(1-z)p_{n,r-1}z^n = rP_r(z) - rp_{0,r}, \text{ and} \\ &(c(1-z)+r)P_r(z) - c(1-z)p_{n,r}z^n - rp_{0,r} = sP_0(z). \end{aligned} \tag{12}$$

This set of equations allows us to express the polynomials $P_0(z)$, $P_1(z)$, …, $P_r(z)$ as functions of the probabilities $p_{0,r}$, $p_{n,0}$, $p_{n,1}$, …, $p_{n,r}$. Subsequently, the stationary probabilities $p_{0,r}$, $p_{n,0}$, $p_{n,1}$, …, $p_{n,r}$ can be deduced by solving a linear system of equations, as given in Proposition 4.

**Proposition 4** (*Stationary Probabilities*). *The polynomial in $z$ defined as*

$$Q_r(z) = cz(c(1-z)+r)^r + sz\sum_{k=1}^{r}\binom{r}{k}c^k(1-z)^{k-1}r^{r-k} - sr^r$$

*has $r+1$ distinct roots, $z_1, z_2, \dots, z_{r+1}$ from which we deduce the probabilities $p_{0,r}$, $p_{n,0}$, $p_{n,1}$, …, $p_{n,r}$ as the solution of the following system of linear equations:*

$$\begin{aligned} &c\sum_{k=0}^{r}p_{n,k} - \frac{s}{s+1}p_{0,r} = c - \frac{s}{s+1}, \text{ and} \\ &sp_{0,r}z_i^n + sz_i^{n-1}\sum_{k=1}^{r}\left(\frac{r}{c(1-z_i)+r}\right)^k p_{n,k} - (c(1-z_i)+s)p_{0,r} = 0, \\ &\quad \text{for } i = 1, 2, \dots, r+1. \end{aligned} \tag{13}$$

Next, we deduce the performance measures in Corollary 2.

**Corollary 2** (*Performance Measures*). *The performance measures are given by*

$$\begin{aligned} \lambda_r &= \lambda_c\left(1 - \frac{s}{c(s+1)} + \frac{s}{c(s+1)}p_{0,r}\right), \text{ and} \\ E(N) &= \frac{1-p_{0,r}}{s+1}\frac{s^2(c(r-1)-2r)-2cr}{2r(c(s+1)-s)} + \frac{c(n(s+1)+1)}{c(s+1)-s}p_{n,0} \\ &\quad + \sum_{k=1}^{r}\frac{c(nr(s+1)-s(r-k))}{r(c(s+1)-s)}p_{n,k}. \end{aligned}$$

From Proposition 4 and Corollary 2, we can derive the performance measures in closed-form in some cases in Table 6.

We observe that $r$ is not part of the expression of $E(N)$ when $n = 1$, revealing that the variability of the matching time does not play a role in this case. When $n$ tends to infinity, we find that $\frac{\partial E(N)}{\partial r} = -\frac{sc^2}{2r^2(s-c(s+1))} < 0$, revealing that the expected number of containers in

**Table 7**
Impact of the matching time variability ($\mu = 1$).

| $n$ | $r=1$ | | | $r=2$ | | | | $r=5$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $E(N)$ | $\lambda_r/\lambda_c$ | $E(C)$ | $E(N)$ | $\lambda_r/\lambda_c$ | $E(C)$ | RD | $E(N)$ | $\lambda_r/\lambda_c$ | $E(C)$ | RD |
| | | | | | | $c=2, s=0.5$ | | | | | |
| 1 | 0.839 | 0.839 | 363.273 | 0.839 | 0.839 | 363.273 | 0.000% | 0.839 | 0.839 | 363.273 | 0.000% |
| 2 | 1.822 | 0.834 | 369.349 | 1.825 | 0.834 | 369.331 | −0.005% | 1.827 | 0.834 | 369.317 | −0.009% |
| 3 | 2.820 | 0.833 | 377.513 | 2.824 | 0.833 | 377.537 | 0.006% | 2.826 | 0.833 | 377.556 | 0.011% |
| 4 | 3.820 | 0.833 | 385.834 | 3.824 | 0.833 | 385.864 | 0.008% | 3.826 | 0.833 | 385.886 | 0.013% |
| 5 | 4.820 | 0.833 | 394.166 | 4.824 | 0.833 | 394.196 | 0.008% | 4.826 | 0.833 | 394.219 | 0.013% |
| 10 | 9.820 | 0.833 | 435.832 | 9.824 | 0.833 | 435.863 | 0.007% | 9.826 | 0.833 | 435.885 | 0.012% |
| 15 | 14.820 | 0.833 | 477.499 | 14.824 | 0.833 | 477.530 | 0.006% | 14.826 | 0.833 | 477.552 | 0.011% |
| | | | | | | $c=1, s=1$ | | | | | |
| 1 | 0.600 | 0.600 | 132.440 | 0.600 | 0.600 | 132.440 | 0.000% | 0.600 | 0.600 | 132.440 | 0.000% |
| 2 | 1.400 | 0.533 | 124.947 | 1.412 | 0.529 | 124.212 | −0.588% | 1.421 | 0.526 | 123.656 | −1.033% |
| 3 | 2.293 | 0.512 | 127.896 | 2.317 | 0.510 | 127.553 | −0.268% | 2.336 | 0.508 | 127.334 | −0.440% |
| 4 | 3.239 | 0.505 | 134.162 | 3.273 | 0.503 | 134.166 | 0.003% | 3.299 | 0.502 | 134.202 | 0.029% |
| 5 | 4.213 | 0.502 | 141.675 | 4.254 | 0.501 | 141.887 | 0.150% | 4.284 | 0.501 | 142.059 | 0.271% |
| 10 | 9.191 | 0.500 | 182.797 | 9.241 | 0.500 | 183.212 | 0.227% | 9.273 | 0.500 | 183.495 | 0.382% |
| 15 | 14.191 | 0.500 | 224.458 | 14.241 | 0.500 | 224.877 | 0.187% | 14.243 | 0.500 | 225.161 | 0.313% |
| | | | | | | $c=0.5, s=2$ | | | | | |
| 1 | 0.355 | 0.355 | 40.641 | 0.355 | 0.355 | 40.641 | 0.000% | 0.355 | 0.355 | 40.641 | 0.000% |
| 2 | 0.699 | 0.185 | 25.516 | 0.698 | 0.171 | 23.982 | −6.011% | 0.698 | 0.161 | 22.861 | −10.407% |
| 3 | 0.998 | 0.110 | 20.010 | 0.989 | 0.095 | 18.352 | −8.283% | 0.982 | 0.085 | 17.216 | −13.964% |
| 4 | 1.251 | 0.070 | 17.806 | 1.226 | 0.057 | 16.250 | −8.740% | 1.206 | 0.049 | 15.223 | −14.507% |
| 5 | 1.461 | 0.045 | 16.997 | 1.415 | 0.035 | 15.528 | −8.641% | 1.379 | 0.029 | 14.574 | −14.254% |
| 10 | 2.036 | 0.007 | 17.667 | 1.875 | 0.004 | 16.044 | −9.188% | 1.767 | 0.003 | 15.009 | −15.043% |
| 15 | 2.200 | 0.001 | 18.445 | 1.970 | 0.000 | 16.468 | −10.716% | 1.838 | 0.000 | 15.344 | −16.809% |

the system (and the holding cost) is increasing in the matching time variability.

In Table 7, we calculate the expected number of containers, rejection probability, and expected cost for different values of $n$, $r$, $c$, and $s$. Moreover, we derive the relative difference in cost between a situation with $r = 2$ or $r = 5$ and the one with $r = 1$ as RD= $\frac{E(C)^{r=2,5}-E(C)^{r=1}}{E(C)^{r=1}}$.

We observe that the variability in matching time only significantly impacts the performance measures in export-heavy areas. In a queue with Erlang service, the expected time spent in the system, measured at arrival, depends on the number of remaining phases of service for the container currently operating a match and the number of containers waiting. However, the expected matching time of waiting containers is not a function of the number of phases of the matching time. Thus, the number of phases only affects the wait generated by the container currently in matching operation. Consequently, the relative importance of the number of phases tends to diminish in highly congested queues, such as in import-heavy areas.

In general, an increase in the number of phases $r$ leads to a reduction in rejection probability and congestion. However, counterexamples can be found, especially in import-heavy and balanced areas. In these cases, having lower variability in matching time allows more containers to join the inventory, leading to increased values of $E(N)$ as $r$ increases. However, the rejection probability remains almost constant. This explains why the expected cost can increase with $r$. In terms of decision making, the number of phases $r$ almost never has an impact on the optimal threshold level. In the examples in the table, the minimizer of the cost is the same for $r = 1, 2$, and 5. Some rare counterexamples can still be found, but even then, a difference of at most 1 is observed between a situation with $r = 1$ and one with $r = \infty$.

### 8.2. Erlang production time

We now employ the same approach to investigate the case of an Erlang production time. We assume that the production of goods at the shipper follows an Erlang distribution with $r$ phases, each with a rate of $r\lambda_s$. The transition rate from state $(x, y)$ to state $(x', y')$, denoted by $t_{(x,y),(x',y')}$ for $x = 0, 1, \ldots, n$ and $y = 0, 1, \ldots, r$, is defined as:

$$t_{(x,y),(x',y')} = \begin{cases} \lambda_c & \text{if } x = 0,1,\ldots,n-1 \text{ and } y = 0,1,\ldots,r, \\ & \text{with } (x',y') = (x+1,y), \\ r\lambda_s & \text{if } x = 0,1,\ldots,n \text{ and } y = 0,1,\ldots,r-1 \\ & \text{with } (x',y') = (x,y+1), \\ \mu & \text{if } x = 1,2,\ldots,n \text{ and } y = r, \text{ with } (x',y') = (x-1,0), \\ 0 & \text{otherwise.} \end{cases}$$

As in the previous section, we introduce the polynomials $P_i(z) = \sum_{x=0}^{n} p_{x,i} z^x$, where $p_{x,i}$ is the stationary probability of being in state $(x, i)$. From the balance equations, we find that:

$$\begin{aligned} &(c(1-z)+rs)P_0(z) - c(1-z)z^n p_{n,0} = z^{-1}(P_r(z) - p_{0,r}), \\ &(c(1-z)+rs)P_k(z) - c(1-z)z^n p_{n,k} = srP_{k-1}(z) \quad \text{for } k = 1,2,\ldots,r-1, \\ &(c(1-z)+1)P_r(z) - p_{0,r} - c(1-z)z^n p_{n,r} = srP_{r-1}(z). \end{aligned} \tag{14}$$

In Proposition 5, we provide the stationary probabilities $p_{0,r}$, $p_{n,0}$, $p_{n,1}$, …, $p_{n,r}$ expressed as functions of the roots of a polynomial. In Corollary 3, we present the performance measures.

**Proposition 5** (*Stationary Probabilities*). *The polynomial in $z$ defined as*

$$R_r(z) = z(c(1-z)+1)\left(\frac{c}{sr}(1-z)+1\right)^r$$

*has r+1 distinct roots, which are different from 1, denoted as $z_1, z_2, \ldots, z_{r+1}$. From these roots, we deduce the probabilities $p_{0,r}$, $p_{n,0}$, $p_{n,1}$, …, $p_{n,r}$ as the solution of the following system of equations:*

$$\begin{aligned} &-\frac{s}{c(s+1)-s}p_{0,r} + \frac{c(s+1)}{c(s+1)-s}\sum_{i=0}^{r} p_{n,i} = 1, \quad \text{and} \\ &\left[1 - z_i\left(\frac{c}{sr}(1-z_i)+1\right)^r\right]p_{0,r} - c(1-z_i)z_i^{n+1}\sum_{i=0}^{r} p_{n,i}\left(\frac{c}{sr}(1-z_i)+1\right)^i = 0, \\ &\text{for } i = 1,2,\ldots,r+1. \end{aligned} \tag{15}$$

**Corollary 3** (*Performance Measures*). *The performance measures are given by*

$$\lambda_r = \lambda_c\left(1 - \frac{s}{c(s+1)} + \frac{s}{c(s+1)}p_{0,r}\right), \quad \text{and}$$

$$E(N) = \frac{ns}{r} - \frac{s}{c(s+1)-s} + \frac{c(r-1+2rs)}{2r(s+1)(c(s+1)-s)} + \frac{\frac{ns^2}{r} + c - s + \frac{c(r-1+2rs)}{2r(s+1)}}{c(s+1)-s} p_{0,r} + \sum_{i=0}^{r} p_{n,i} \frac{ci}{r^2}.$$

As for the case of Erlang matching time, performance measures can be derived in certain scenarios. For example, when $n$ tends to infinity, we find that $E(N) = \frac{s}{s-c(s+1)} - \frac{c(r-1+2rs)}{2r(s+1)(s-c(s+1))} - \frac{c-s+\frac{c(r-1+2rs)}{2r(s+1)}}{s}$, subject to the stability condition $c(s+1) < s$. We deduce that $\frac{\partial E(N)}{\partial r} = -\frac{c(2s-c(s+1))}{2r^2 s(s+1)(s-c(s+1))} < 0$, which establishes that $E(N)$ decreases with the number of phases $r$. Furthermore, by letting $r$ tend to infinity, analogous to the deterministic production time case, we obtain the expression $E(N) = 1 + \frac{1}{1-c} - \frac{3c}{2s} - \frac{c}{s+1} - \frac{c(1+c^2)}{2(1-c)(s-c(s+1))}$.

The observations made for the effect of the variability of the production time are similar to those for the variability of the matching time. An illustration is provided in the Appendix. In particular, production time variability predominantly influences export-heavy regions, where a decrease in production time variability leads to a reduction in the expected cost. Conversely, in import-heavy regions, instances can be identified where the expected cost increases with lower variability. Furthermore, it is noteworthy that the optimal withholding threshold level is not affected by production time variability.

## 9. Conclusion

We investigated optimal container management for a hinterland consignee, aiming to minimize travel and holding costs in a double-ended queue system. The consignee's decision on whether to store or return a container, determined by a withholding threshold, was explored. Closed-form performance measures were derived for a shipper with single storage facility, using a matrix approach that was extended to a numerical method in the general case, along with a fixed-point approximation for efficient computations. Results highlighted diverse withholding roles based on import, balance, or export focus, and the non-monotonic impact of traffic intensity and service capacity on withholding thresholds. In the dynamic admission control problem, we found the optimal withholding policy to be state-dependent in the single-truck case, with a marginal cost reduction in multi-resource scenarios. Examination of variability in matching and production times showed some cost savings in export areas, with a marginal impact on withholding decisions.

There are several avenues for future research. First, the assumptions made can be modified by considering generally distributed processes for the arrival of and demand for containers or a non-Erlang distribution of matching times. Although changing distributions may lead to different quantitative results, it is unlikely to modify the insights provided by the Markovian analysis. Other costs could be included in the analysis, such as the cost to empty and clean containers upon arrival or the cost to buy a new container when an old one is no longer suitable for use. Finally, instead of considering the individual optimization for the consignee in isolation, we could determine the optimal inventory policy for both the consignee and shipper as a way to minimize their overall costs. This would open discussions of how the benefits of the street turn strategy should be shared between the two participants and how such a collaboration could be achieved in practice. Further, street turn strategies have the potential to reduce unnecessary container movements leading to lower carbon emissions. We could reconsider the optimization question by including environmental objectives to further incentivize matching operations.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ejor.2024.02.035.

## References


Bakshi, N., Flynn, S. E., & Gans, N. (2011). Estimating the operational impact of container inspections at international ports. *Management Science*, *57*(1), 1–20.

BigRentz (2022). How much does a storage container cost? Sizes and types. https://www.bigrentz.com/blog/storage-container-cost.

Boucherie, R., & Van Dijk, N. (2017). *Markov decision processes in practice*: *vol. 248*, Springer.

Conolly, B., Parthasarathy, P., & Selvaraju, N. (2002). Double-ended queues with impatience. *Computers & Operations Research*, *29*(14), 2053–2072.

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

Di Crescenzo, A., Giorno, V., Kumar, B., & Nobile, A. (2012). A double-ended queue with catastrophes and repairs, and a jump-diffusion approximation. *Methodology and Computing in Applied Probability*, *14*(4), 937–954.

Diamant, A., & Baron, O. (2019). Double-sided matching queues: Priority and impatient customers. *Operations Research Letters*, *47*(3), 219–224.

Elalouf, A., Perlman, Y., & Yechiali, U. (2018). A double-ended queueing model for dynamic allocation of live organs based on a best-fit criterion. *Applied Mathematical Modelling*, *60*, 179–191.

Green, L., & Kolesar, P. (1991). The pointwise stationary approximation for queues with nonstationary arrivals. *Management Science*, *37*(1), 84–97.

Gross, D., & Harris, C. (1985). *Fundamentals of queueing theory* (2nd ed.). Wiley series in probability and mathematical statistics.

He, Q., Nie, T., Yang, Y., & Shen, Z. (2021). Beyond repositioning: Crowd-sourcing and geo-fencing for shared-mobility systems. *Production and Operations Management*, *30*(10), 3448–3466.

Jiang, T., Chai, X., Liu, L., Lv, J., & Ammar, S. (2021). Optimal pricing and service capacity management for a matching queue problem with loss-averse customers. *Optimization*, *70*(10), 2169–2192.

Kashyap, B. (1966). The double-ended queue with bulk service and limited waiting space. *Operations Research*, *14*(5), 822–834.

Kendall, D. (1951). Some problems in the theory of queues. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *13*(2), 151–173.

Kim, W., Yoon, K., Mendoza, G., & Sedaghat, M. (2010). Simulation model for extended double-ended queueing. *Computers & Industrial Engineering*, *59*(2), 209–219.

Koole, G. (2007). Monotonicity in Markov reward and decision chains: Theory and applications. *Foundations and Trends in Stochastic Systems*, *1*(1), 1–76.

Kozan, E. (1997). Comparison of analytical and simulation planning models of seaport container terminals. *Transportation Planning and Technology*, *20*(3), 235–248.

Lee, C., Liu, X., Liu, Y., & Zhang, L. (2021). Optimal control of a time-varying double-ended production queueing model. *Stochastic Systems*, *11*(2), 140–173.

Lee, C., & Song, D. (2017). Ocean container transport in global supply chains: Overview and research opportunities. *Transportation Research, Part B (Methodological)*, *95*, 442–474.

Legros, B., Bouchery, Y., & Fransoo, J. (2019). A time-based policy for empty container management by consignees. *Production and Operations Management*, *28*(6), 1503–1527.

Li, J., Liu, K., Leung, S., & Lai, K. (2004). Empty container management in a port with long-run average criterion. *Mathematical and Computer Modelling*, *40*(1–2), 85–100.

Liu, X., Gong, Q., & Kulkarni, V. (2015). Diffusion models for double-ended queues with renewal arrival processes. *Stochastic Systems*, *5*(1), 1–61.

Liu, H., Li, Q., & Zhang, C. (2020). Matched queues with matching batch pair (m, n). arXiv preprint arXiv:2009.02742.

Liu, X., & Weerasinghe, A. (2021). Admission control for double-ended queues. http://dx.doi.org/10.48550/arXiv.2101.06893, arXiv preprint arXiv:2101.06893.

Meester, L., & Shanthikumar, J. (1990). Concavity of the throughput of tandem queueing systems with finite buffer storage space. *Advances in Applied Probability*, *22*(3), 764–767.

Neuts, M. (1981). *Matrix-geometric solutions in stochastic models: An algorithmic approach*. Mineola: Johns Hopkins University Press.

Nguyen, H., & Phung-Duc, T. (2022). Strategic customer behavior and optimal policies in a passenger–taxi double-ended queueing system with multiple access points and nonzero matching times. *Queueing Systems*, *102*(3–4), 481–508.

Pakulniewicz, M. (2021). European road transport rates rise further; see where it has been highest. Trans.info. https://trans.info/en/european-road-transport-rates-rise-further-see-where-it-has-been-more-severe-250003.

Puterman, M. (1994). *Markov decision processes*. Hoboken, NJ: John Wiley and Sons.

Roy, D., De Koster, R., & Bekker, R. (2020). Modeling and design of container terminal operations. *Operations Research*, *68*(3), 686–715.

Roy, D., van Ommeren, J., de Koster, R., & Gharehgozli, A. (2022). Modeling land-side container terminal queues: Exact analysis and approximations. *Transportation Research, Part B (Methodological)*, *162*, 73–102.

Sasieni, M. (1961). Double queues and impatient customers with an application to inventory theory. *Operations Research*, *9*(6), 771–781.

Shi, Y., & Lian, Z. (2016). Optimization and strategic behavior in a passenger–taxi service system. *European Journal of Operational Research*, *249*(3), 1024–1032.

Shi, Y., Lian, Z., & Shang, W. (2015). Study of a passenger-taxi queueing system with nonzero matching time. In *2015 12th international conference on service systems and service management*. IEEE.

Song, D. (2007). Characterizing optimal empty container reposition policy in periodic-review shuttle service systems. *Journal of the Operational Research Society*, *58*(1), 122–133.

Song, D., & Zhang, Q. (2010). A fluid flow model for empty container repositioning policy with a single port and stochastic demand. *SIAM Journal on Control and Optimization*, *48*(5), 3623–3642.

Stidham, S. (1985). Optimal control of admission to a queueing system. *IEEE Transactions on Automatic Control*, *30*(8), 705–713.

Su, Y., & Li, J. (2023). Admission control of double-sided queues with multiple customer types. *IEEE Transactions on Automatic Control*.

Vis, I., & Roodbergen, K. (2009). Scheduling of container storage and retrieval. *Operations Research*, *57*(2), 456–467.

Wang, Z., Yang, C., Liu, L., & Zhao, Y. (2023). Equilibrium and socially optimal of a double-sided queueing system with two-mass point matching time. *Quality Technology & Quantitative Management*, *20*(1), 89–112.

xChange (2023). Save money on demurrage and detention fees in top 10 ports. https://www.container-xchange.com/blog/demurrage-detention/.

Xie, Y., Liang, X., Ma, L., & Yan, H. (2017). Empty container management and coordination in intermodal transport. *European Journal of Operational Research*, *257*(1), 223–232.

Yu, M., Fransoo, J., & Lee, C. (2018). Detention decisions for empty containers in the hinterland transportation system. *Transportation Research, Part B (Methodological)*, *110*, 188–208.

Zhang, B., Ng, C., & Cheng, T. (2014). Multi-period empty container repositioning with stochastic demand and lost sales. *Journal of the Operational Research Society*, *65*(2), 302–319.