



Steady-state analysis and optimization of a queueing-inventory system under joint impact of imperfect production and inspection processes

Serife Ozkar^{*1}, Umay Uzunoglu Kocer²

¹*Department of International Trade and Logistics, Balıkesir University, Balıkesir, Türkiye*

²*Department of Statistics, Dokuz Eylül University, İzmir, Türkiye*

Abstract

During the production process, defective production rates as well as inspection processes become essential, as customers have become increasingly attentive to the quality of the products they receive. This paper aims to analyze a queueing-inventory system in which both the production and inspection processes are imperfect. The novelty of this study lies in the combination of these two imperfections in a single inventory-queueing system. The inspection process involves two kinds of misclassification errors: Type-I occurs when a non-defective item is identified as defective, while Type-II occurs when a defective item is incorrectly labeled as non-defective. Two storage facilities of limited capacity are considered, one allocated for all produced items and the other reserved for items deemed non-defective after inspection, and a production process controlled by an (s, S) -policy. The proposed model is formulated as a continuous-time Markov chain and analyzed by the Matrix Geometric Method. The stability condition is derived, and several key performance measures are obtained. The cost function is then constructed and minimized with respect to production, inspection, and service rates using the Particle Swarm Optimization algorithm. The numerical study shows that the probability of defective production has a considerable impact on all performance measures and that the capacity of the second depot affects both the mean number of customers in the system and the total cost. The results indicate that production, inspection, and service rates must be adjusted jointly according to the probability of defect and the capacity of the depot to achieve the minimum cost.

Mathematics Subject Classification (2020). 90B22, 60K30, 60K25

Keywords. Imperfect inspection, imperfect production, finite capacity depots, optimization, queueing-inventory system

^{*}Corresponding Author.

Email addresses: serife.ozkar@balikesir.edu.tr (S. Ozkar), umay.uzunoglu@deu.edu.tr (U. Uzunoglu Kocer)

Received: 23.02.2026; Accepted: 08.08.2026

1. Introduction

Increasing global competition, fluctuating customer demand, and the increasing complexity of supply chains are forcing businesses to develop more effective inventory management strategies. Queueing inventory systems integrate inventory systems in which customer demand is met by products in stock and orders are placed according to a pre-determined inventory policy, with queueing systems where customer arrivals are random, and service is completed when products are available from stock. Queueing inventory systems integrate the inventory replenishment process and customer demand fulfillment, providing a holistic assessment of the relationship between inventory levels and waiting times. This allows for the simultaneous optimization of both inventory policies and service performance. In recent years, the use of queueing inventory models has gained attention, especially in production inventory systems, to satisfy customer satisfaction. In production processes, defective items may emerge due to factors such as equipment deterioration, variability in process conditions, or the use of suboptimal component quality. The distribution of such nonconforming items to customers not only undermines customer satisfaction, but also poses significant risks to organizational reputation and long-term competitiveness. Hence, the incorporation of an inspection mechanism becomes indispensable to prevent the transport of defective items to customers. On the other hand, when inspection activities rely on human judgment or manual procedures, inspection errors may occur. In the literature, this phenomenon is referred to as an imperfect inspection process, which is typically modeled through two forms of misclassification: Type-I error is classifying a non-defective item as defective, whereas Type-II error is classifying a defective item as non-defective.

Many papers in the literature have focused on the Economic Order Quantity (EOQ) model and the Economic Production Quantity (EPQ) model that include imperfect quality items. One of the important papers that extends a traditional production/inventory model to a model that includes ordered or produced items with imperfect quality is [39]. The authors assumed that poor-quality items are sold as a single batch at a discounted price at the end of the 100% inspection process and also there are no human errors in the inspection process. Konstantaras et al. [22] extends the paper [39] in two directions, in that it assumes that acceptable items are sent to the inventory in batches, not unit-by-unit, and it allows for reworking of the imperfect quality items. According to their approach, there are two ways to deal with imperfect quality items: sell them to a secondary market as a single batch at a lower price, or rework them at some cost and then use them as new ones. Another extension of the EOQ model defined in [39] is provided by Khan et al. [21]. Unlike [39], the authors assume an imperfect inspection process, which means that the inspector can misclassify items in the inspection process, and the probabilities of misclassification errors are assumed to be known. Khan et al. [21] assumes two types of misclassification errors; a non-defective item may be classified as defective (Type-I error) or a defective item may be classified as non-defective (Type-II error).

The EOQ models that include imperfect items have been an interesting research area. There is a vast literature in which EOQ/EPQ models with a 100% inspection process and with a classification error (Type-I and Type-II) in the inspection process. Asadkhani et al. [5] assumed an EOQ model with four different types of imperfect items, such as salvage items, repairable items, scrap items, and rejected items. Asadkhani et al. [5] performed an inspection process 100% in which the inspector can misclassify items as described in [21]. Unlike other papers, Khan et al. [21] paid attention to the ability of the inspector to reduce inspection errors through learning. Chan et al. [9] provided an EPQ model that integrates the following three situations: lower pricing, rework, and rejection. A 100% inspection process was performed to identify the number of good quality items,

poor quality items, and defective items. The reader may refer to [4, 15–18, 24, 28, 29, 36, 40] for detailed information on inspection procedures and related papers in the literature.

Some studies have been conducted in which sampling is performed instead of 100% inspection. Moussawi-Haidar et al. [30] proposed an EOQ-type inventory model where only a sample is inspected instead of all items in the inspection process. If the number of imperfect items is below an "acceptance number", the sample corresponds to the standards and no further inspection is needed. Otherwise, the lot is subject to 100% inspection. The authors numerically proved that the optimal profit obtained by their proposed model is significantly higher than that provided by Salameh and Jaber [39]. Another paper used sampling in the inspection process is [10]. They assumed that a sampling process is performed as soon as the order is received. If the lot holds the requirements of the quality inspection, all items in the lot are considered of perfect quality. Otherwise, either the lot is returned to the supplier, or it is kept in inventory and is taken back during the next order from the supplier.

The papers mentioned so far assume an inspection process that has been utilized by humans. Some papers in the literature have been discussing an inspection process through an automation policy. Dey et al. [11] proposed an automation policy to detect imperfect items. The policy they proposed makes the production process smarter and results in an error-free inspection. The purpose of an automation policy in a smart production system is to inspect the inventory during the production run. The articles [12, 41] used the automated inspection policy suggested by Dey et al. [11].

Among studies on production-inventory models with imperfect quality in a supply chain, Kurdhi et al. [23] examined a vendor-buyer production-inventory model. They assumed 100% inspection process with Type-I and Type-II inspection errors. All defective items are returned to the vendor by the buyer. The articles [35, 38] developed a three-tier supply chain model with the supplier, manufacturer, and retailer, where the production system is imperfect. After the inspection process, the imperfect items are reworked by the manufacturer and the perfect items are sent to the retailer. The retailer also inspects the items in the inspection process. Saha and Chakrabarti [38] assumed the manufacturer has two production units. The manufacturer inspects all the products produced in the imperfect inspection process and sells the perfect ones to the retailer. The retailer inspects the received items due to the manufacturer's erroneous inspection process and returns the imperfect items to the manufacturer after inspection. The manufacturer sends the defective items returned from the retailer and the main production unit to the refurbishment unit for reworking. Recent contributions have further extended this line of research. Yadav et al. [48] examined the effect of in-process inspection on a highly imperfect production system, where an investment in in-process inspection reduces the rate at which defective items are formed, and showed that this policy can lower the total cost by up to 9.3% compared with a system without in-process inspection. Namnabat et al. [31] developed extended EOQ and EPQ models for repairable imperfect quality items, incorporating partial repair success, inspection errors, and finite and infinite withdrawal rates, with preventive maintenance integrated into the EPQ model.

The first studies on queueing-inventory systems were conducted by the papers in [27, 43]. Subsequently, the literature has diversified according to different assumptions of arrival and service process, whether there is interruption of service due to service breakdown or vacation [19, 44], multiple customer classes [26, 33, 34], and whether the inventory is filled by an internal production or an external supplier [6, 7, 37]. An $M/M/1/\infty$ queue with an attached inventory system is examined by Saffari et al. [37], with an infinite capacity waiting room and also with an infinite inventory. The arrivals when the inventory level is zero are rejected, that is, it is assumed that customers are lost. On the other hand, when the inventory reaches zero, customers in the queue are required to wait for the next replenishment. The (r, Q) -policy is used. Baek and Moon [6] studied a model similar to

[37] except the replenishment policy. While Saffari et al. [37] replenished the inventory only by ordering, and Baek and Moon [6] considered an internal production unit. By comparing the results of the proposed model with those of [37], it was shown that the model with an internal production is more effective. The production-inventory model integrated into an $M/M/c/\infty$ queue where the queue capacity is infinite and the inventory capacity is S units was studied by Baek and Moon [7]. Inventory is replenished with the internal production, which is governed by the (s, S) -policy. That is, the production facility is turned off as soon as the inventory level reaches S and is turned on when the inventory level drops to s . The authors utilized the traditional $M/M/1/K$ queue and the regenerative process in their analysis. The papers [6, 7] and [37] are the first studies to demonstrate the independence between the queue length process and the inventory level process for the proposed models. Then, these two processes were analyzed separately. More recently, Vinitha et al. [45] analyzed an integrated branching queueing-inventory system with diverted service, in which each branch operates under its own (s, Q) ordering policy, and derived the stability condition and the rate matrix applying the geometric method of the Neuts matrix.

Although numerous studies have examined issues related to defective items in various settings within classical inventory and supply chain frameworks, we have observed that such problems have received relatively little attention in the queueing-inventory literature. Recently, Aghsami et al. [2] introduced a new model that integrates an $M/M/1$ queueing system with an imperfect production-inventory process and an inspection queue featuring inspection errors. They employed a nonlinear programming approach to optimize system costs. The assumptions of this model are largely consistent with those in [6, 37], except for the incorporation of imperfect production and inspection processes with inspector errors. Aghsami et al. [2] provides an analysis using a methodology similar to that adopted in [6, 7, 37]. The authors first demonstrated that, in the long run, the inspection queue operates independently of the queueing-inventory system; subsequently, they analyzed the queueing-inventory system separately from the inspection queue. In [3], the authors investigated a retailer supplier queueing-inventory framework with imperfect items under a sampling inspection plan. Aghsami et al. [1] developed a two-echelon blood bank network employing a 100% inspection policy. Mehta et al. [25] examined a queueing-inventory model with balking and retrial orbits, in which jobs join a retrial orbit if the server is busy. Both the production process and the 100% inspection process are imperfect, and the inspection stage may lead to Type-I or Type-II misclassification errors. This model closely resembles the model given by Aghsami et al. [2], apart from the inclusion of balking and retrial behaviors, and relies on the same analytical methodology. Consequently, the integrated queueing-inventory model discussed by Mehta et al. [25] comprises three distinct and independent subsystems, each modeled as an $M/M/1/\infty$ queue.

Based on the existing literature, only a limited number of studies have examined queueing-inventory models in which both the production and inspection processes are imperfect. This paper is inspired by the queueing-inventory framework analyzed by Aghsami et al. [2]. The authors assumed that the storage areas for items both post-production and post-inspection are assumed to have infinite capacity. In contrast, our model introduces two finite-capacity depots. The produced items are stored in Depot-1 with capacity S , while the items that successfully pass the inspection process are stored in Depot-2 with capacity K . Moreover, Aghsami et al. [2] assumed that items are produced according to a Poisson process, implying continuous production without any governing control policy. However, the assumption of continuous production and also continuous inspection during production might be impractical, especially when the production rate is high and the proportion of defective items is low. Therefore, we considered that production is governed by a (s, S) -policy. Specifically, when the inventory level in Depot-1 falls to s , the production process is switched *on* and continues until the inventory level reaches S . Once the level

attains S , the production process is turned *off*. In this paper, the proposed queueing-inventory system is represented by a Quasi Birth-and-Death (QBD) structure, and the system is analyzed by using the Matrix Geometric Method (MGM).

To the best of our knowledge, our work introduces the first queueing-inventory model to simultaneously address imperfect production and imperfect inspection, distinguishing itself by explicitly grounding the analysis in a fundamental production-inventory concept: the optimal production-inventory policy. Building on the research gaps identified above, this study is particularly motivated by and closely related to [2] who developed an $M/M/1$ queueing-inventory model incorporating an imperfect production process and an imperfect inspection queue. However, our work differs from [2] in several key aspects. First, while Aghsami et al. [2] assumed unlimited storage capacity for both the pre-inspection and post-inspection items, we consider two finite-capacity depots, which better reflects the physical storage constraints encountered in real production environments. Second, Aghsami et al. [2] modeled production as a continuous Poisson process without any governing control policy, whereas our model regulates production through an (s, S) policy, integrating the inspection process directly into this control mechanism. Third, while Aghsami et al. [2] showed that the inspection queue and the queueing-inventory system are independent in the long run and analyzed them separately, our findings indicate that, under the (s, S) policy considered in this study, the two subsystems are no longer independent and must be treated as a combined system; we address the resulting complexity by formulating the model as a Quasi Birth-and-Death (QBD) process and solving it via the well-established Matrix Geometric Method (MGM). Finally, whereas Aghsami et al. [2] optimized only the production rate through a single-variable nonlinear programming model, our study jointly optimizes the production rate, inspection rate, and service rate using Particle Swarm Optimization (PSO), given the increased complexity of the cost function under the finite-capacity, policy-driven setting. While PSO has been employed to optimize complex cost functions in various QBD-based queueing systems, such as retrial queues with working vacations [49] and queueing-inventory systems with retrial demands [8], its use has typically been restricted to single- or limited-parameter settings. Due to the analytical intractability of the resulting cost function in our model — arising from the joint dependence on three decision variables and the underlying QBD structure — this study extends this line of work by employing PSO to obtain the optimal production, inspection, and service rates simultaneously, going beyond the single-variable analytical optimization approach adopted in [2]. To sum up, the key distinguishing features of this study can be summarized as follows:

- **Finite-capacity storage:** Our study incorporates two finite-capacity depots, offering a more realistic representation of physical storage constraints.
- **Policy-driven production:** While Aghsami et al. [2] modeled production as an uncontrolled continuous Poisson process, our study governs production through an (s, S) policy, with the inspection process embedded directly into this control mechanism.
- **Joint system analysis:** Previous studies, such as [2], showed the independence of the inspection queue and the queueing-inventory system and analyzed them separately, by adopting (s, S) policy, our study shows that the two subsystems must be treated jointly.
- **Multi-parameter optimization:** Our study jointly optimizes the production rate, inspection rate, and service rate using Particle Swarm Optimization (PSO).

The rest of the paper is organized as follows: In Section. 2 the model description is presented. In Section. 3, steady-state analysis of the model is performed, and some key performance measures of the system are derived. Illustrative numerical examples,

including an optimization problem, are discussed in Section. 4. Finally, some concluding remarks are given in Section. 5.

2. Model Description

In this study, we discuss a queueing-inventory system where both production and inspection processes are imperfect. There are two finite depots to store the products in the system. That is, the items produced that will be inspected are placed in Depot-1 of capacity S , and the items decided to be non-defective after the inspection process are placed in Depot-2 of capacity K . In practice, storage areas are physically bounded by available warehouse space, shelving capacity, or, in the case of items requiring special conditions, by the capacity of the storage equipment, such as refrigerated units. Assuming unlimited storage capacity, as is commonly done in the queueing-inventory literature, therefore does not reflect the constraints encountered in real production environments. Although capacitated systems are analytically more complicated, we consider two finite-capacity depots precisely because this assumption provides a more realistic representation of industrial settings and increases the applicability of the model. The pictorial representation of the system is given in Figure 1.

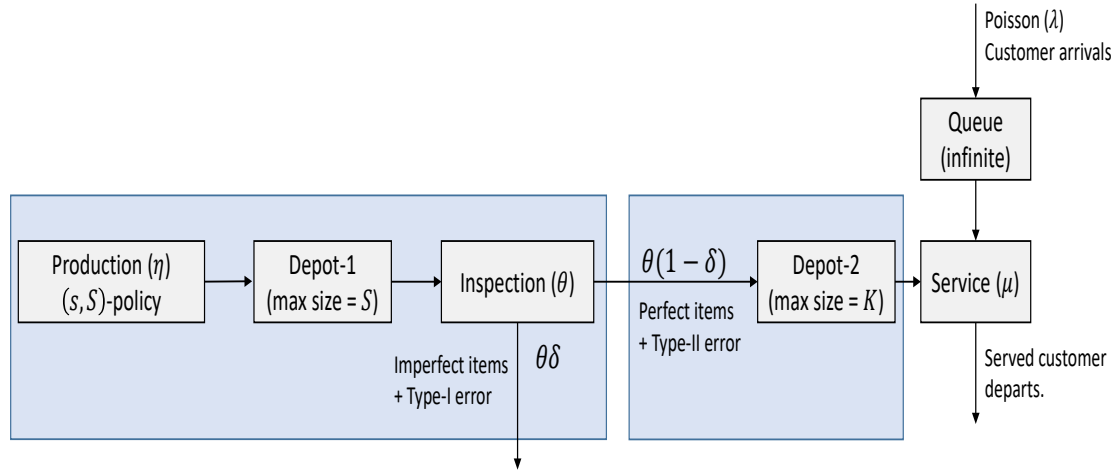


Figure 1. The proposed queueing-inventory system, where both the production and the inspection processes are imperfect.

We assume that customers arrive according to a Poisson process with rate λ . When a customer arrives on the system, if the server is idle and the Depot-2 inventory level is greater than zero, service is initiated immediately when the customer arrives. Having completed the service, the customer leaves the system. Each customer requires exactly one item from the inventory. So, the inventory level in Depot-2 decreases by one unit. Upon arrival, if the server is busy, the arriving customer joins the queue and the queue capacity is assumed to be infinite. Customers are serviced on a First Come First Served (FCFS) basis. The service time of a customer follows an exponential distribution with a parameter μ . The exponential distribution is adopted for the inter-arrival, service, production, and inspection times because it is the natural modeling choice for continuous-time Markov chain (CTMC) formulations, owing to its memoryless property, which ensures that the state of the system at any time depends only on the current state and not on the elapsed time since the last event. When the server is idle and the inventory level in Depot-2 is zero, the customer joins the queue. Additionally, when a customer's service is completed, the inventory level may be depleted to zero. If any customer is waiting in the queue at that moment, they continue to wait patiently. In any case, there is no loss to the customer.

The internal production process is carried out depending on the inventory level in Depot-1. The (s, S) inventory policy is implemented; that is, whenever the inventory level in Depot-1 drops to s , the production process is turned *on*. The production continues until the inventory level reaches S . Therefore, S is the maximum inventory level for Depot-1. When the inventory level hits S , the production is switched *off*. The production time is exponentially distributed with parameter η . We assume that the production process is imperfect and the fraction of defective production is constant. α represents the probability that a produced item is defective.

Continuous production without any governing control policy implies that items are produced and inspected uninterruptedly, which is impractical when the production rate is high and the proportion of defective items is low, as it leads to unnecessary production, inspection effort, and holding costs. In practice, production is typically initiated only when the inventory falls to a predetermined reorder level and is stopped once a target level is reached. The (s, S) policy captures this behavior directly: the production process is switched *on* when the inventory level in Depot-1 drops to s and switched *off* when it reaches S . An important implication of adopting this policy is that the inspection queue and the queueing-inventory system are no longer independent, as they are under the continuous-production assumption; consequently, the two subsystems must be analyzed jointly, which is addressed in this study by formulating the model as a QBD process and solving it via the Matrix Geometric Method.

We also assume that there is an imperfect inspection process as given by Aghsami et al. [2]. There are two types of misclassifications: Type-I is classifying a non-defective item as defective (represented with probability p_1), and Type-II is classifying a defective item as non-defective (represented with probability p_2). Once an inspection is completed, the inspected item may be a truly non-defective item or a defective item that has been identified as non-defective. The probability that an item is identified as defective is δ .

$$\delta = p_1(1 - \alpha) + (1 - p_2)\alpha. \quad (2.1)$$

The inspection time follows an exponential distribution with parameter θ . So, non-defective items are sent to Depot-2 with a rate $\theta(1 - \delta)$, and the defective items are discarded from the system with a rate $\theta\delta$.

Production and inspection are affected by distinct and independent sources of error in practice. Defective production typically results from mechanical wear, material variability, or process drift, while inspection errors arise from limitations in human judgment, sensor precision, or testing methodology, such as detection thresholds or sampling variability. Because these two error sources originate from unrelated physical mechanisms, it is realistic for an imperfect production process to be paired with an imperfect inspection process that cannot fully compensate for it. This dual-imperfection structure is observed in various real-world settings. In electronic component manufacturing, for instance, soldering or assembly defects may occur during production, and automated optical inspection (AOI) systems used to detect them have their own false-positive and false-negative rates due to sensor resolution and algorithmic limitations. Similarly, in blood screening processes, an infected unit can pass undetected due to the serological window period, or a healthy unit can be incorrectly flagged as positive due to the limited sensitivity and specificity of the diagnostic test. The proposed model reflects these practically relevant cases by allowing both α , the probability of production defects, and p_1 and p_2 , the probabilities of inspection misclassification to be specified independently.

The probability of producing a defective item (α) and the misclassification probabilities (p_1, p_2) are treated as fixed parameters over the planning horizon, consistent with the modeling approach adopted in [2]. This assumption is also necessary for analytical

tractability: since these probabilities directly determine the transition rates of the underlying CTMC, thereby affecting the stationary condition required for the CTMC/QBD formulation and, consequently, the stability analysis given by Theorem 3.1. In practice, defect and misclassification rates are typically estimated from historical process data and treated as constant over a given operating period, being periodically re-calibrated as process conditions change, which is consistent with this modeling assumption. To examine the sensitivity of the system to these parameters, extensive numerical experiments are conducted in Section 4, in which α , p_1 , and p_2 are varied across a range of values, allowing their effect on the performance measures and the optimal cost to be assessed without requiring them to be modeled as time-varying quantities.

3. The steady-state analysis

In this section, the steady-state analysis of the queueing-inventory system described in Section 2 is performed.

Let $N(t)$, $I(t)$, $J(t)$, and $M(t)$ denote, respectively, the number of customers in the system, the inventory level of the Depot-1, the inventory level of the Depot-2, and the status of the production process, at time t . The process $\{(N(t), I(t), J(t), M(t)), t \geq 0\}$ is a continuous-time Markov chain (CTMC), and the state space in the lexicographical order is given by

$$\begin{aligned} \Omega = & \{ \{(n, i, j, 1) : 0 \leq i \leq s, 0 \leq j \leq K\} \cup \\ & \{(n, i, j, m) : s+1 \leq i \leq S-1, 0 \leq j \leq K, m = 0, 1\} \cup \\ & \{(n, S, j, 0) : 0 \leq j \leq K\} \}, n \geq 0. \end{aligned}$$

As can be seen from the state-space definition above, each state of the system is represented by a four-dimensional vector (n, i, j, m) , corresponding to the number of customers in the system, the inventory level of Depot-1, the inventory level of Depot-2, and the status of the production process, respectively. Since the process $\{(N(t), I(t), J(t), M(t)), t \geq 0\}$ is a CTMC, its infinitesimal generator can be represented using the Quasi Birth-and-Death (QBD) structure, in which the state space is organized into levels according to the number of customers n , and each level is further partitioned into a set of phases corresponding to the remaining state variables (i, j, m) . Consequently, the generator matrix Q takes a block-tridiagonal form, where each block matrix corresponds to a specific level n , and transitions between blocks represent transitions between consecutive levels (i.e., customer arrivals and departures), while transitions within a block represent changes in the inventory or production status at a specific level.

The level $(n, i, j, 1)$ corresponds to the case when there are n customers in the system, the inventory level of Depot-1 is i , $0 \leq i \leq s$, the inventory level of Depot-2 is j , $0 \leq j \leq K$ and the production is *on*. The level (n, i, j, m) corresponds to the case when there are n customers in the system, the inventory level of Depot-1 is i , $s+1 \leq i \leq S-1$, the inventory level of Depot-2 is j , $0 \leq j \leq K$ and m is the status of the production process (if $m = 0$, the production is *off* if $m = 1$, the production is *on*). The level $(n, S, j, 0)$ corresponds to the case when there are n customers in the system, the inventory level of Depot-1 is S , the inventory level of Depot-2 is j , $0 \leq j \leq K$ and the production is *off*. The transitions rates are listed as follows:

1. State transitions triggered by customer arrivals.
 $(n, i, j, m) \rightarrow (n+1, i, j, m)$ with rate λ for $n \geq 0$, $0 \leq i \leq S$, $0 \leq j \leq K$, $m = 0, 1$.
2. Transitions due to the service of customers.
 $(n, i, j, m) \rightarrow (n-1, i, j-1, m)$ with rate μ for $n \geq 1$, $0 \leq i \leq S$, $1 \leq j \leq K$,

- $m = 0, 1$.
3. State transitions triggered by production of items.
 - a. $(n, i, j, 1) \rightarrow (n, i + 1, j, 1)$ with rate η for $n \geq 0, 0 \leq i \leq S - 2, 0 \leq j \leq K$.
 - b. $(n, S - 1, j, 1) \rightarrow (n, S, j, 0)$ with rate η for $n \geq 0, 0 \leq j \leq K$.
 4. Transitions due to an item being labeled as non-defective as a result of the inspection process.
 - a. $(n, i, j, m) \rightarrow (n, i - 1, j + 1, m)$ with rate $\theta(1 - \delta)$ for $n \geq 0, 1 \leq i \leq S$ ($i \neq s + 1$), $0 \leq j \leq K - 1, m = 0, 1$.
 - b. $(n, s + 1, j, 0) \rightarrow (n, s, j + 1, 1)$ with rate $\theta(1 - \delta)$ for $n \geq 0, 0 \leq j \leq K - 1$.
 5. Transitions due to an item being labeled as defective as a result of the inspection process.
 - a. $(n, i, j, m) \rightarrow (n, i - 1, j, m)$ with rate $\theta\delta$ for $n \geq 0, 1 \leq i \leq S$ ($i \neq s + 1$), $0 \leq j \leq K - 1, m = 0, 1$.
 - b. $(n, s + 1, j, 0) \rightarrow (n, s, j, 1)$ with rate $\theta\delta$ for $n \geq 0, 0 \leq j \leq K - 1$.

We now establish the notation employed in the subsequent analysis. e is a unit column vector; e_j is a unit column vector is of dimension j ; $e_j(i)$ is a column vector with 1 in the i^{th} position and 0 elsewhere; I_k is an identity matrix of order k and $E = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ is a matrix of order 2. The symbol $\otimes$ represents the Kronecker product. The transpose notation is denoted by $'$. Finally, we define $d = (s + 2)(K + 1) + (S - s - 1)(2K + 2)$, which denotes the total number of states within a level (i.e., the dimension of each block matrix), used as the dimension of the block matrices A , B , B_0 , and C introduced below. The infinitesimal generator matrix of the queueing-inventory system has a block-tridiagonal matrix structure and is given by

$$Q = \begin{pmatrix} B_0 & A & & & \\ C & B & A & & \\ & C & B & A & \\ & & \ddots & \ddots & \ddots \end{pmatrix}. \quad (3.1)$$

The matrix A of dimension d represents the arrival of the customers in the system:

$$A = \begin{pmatrix} \lambda I_{K+1} & & & & \\ & \ddots & & & \\ & & \lambda I_{K+1} & & \\ \hline & & & \lambda I_{2K+2} & \\ & & & & \ddots \\ \hline & & & & & \lambda I_{2K+2} \\ \hline & & & & & & \lambda I_{K+1} \end{pmatrix}.$$

The matrix C of dimension d represents the departures of the customers from the system:

$$C = \begin{pmatrix} C_1 & & & & \\ & \ddots & & & \\ & & C_1 & & \\ \hline & & & C_2 & \\ & & & & \ddots \\ \hline & & & & & C_2 \\ \hline & & & & & & C_1 \end{pmatrix},$$

with

$$C_1 = \begin{pmatrix} 0 & & & & \\ \mu & 0 & & & \\ & \mu & 0 & & \\ & & \ddots & \ddots & \\ & & & \mu & 0 \end{pmatrix}, \quad C_2 = \begin{pmatrix} 0 & & & & \\ \mu & 0 & & & \\ & \mu & 0 & & \\ & & \ddots & \ddots & \\ & & & \mu & 0 \end{pmatrix} \otimes I_2.$$

The B_0 is a d -dimensional matrix and has a tridiagonal structure. The upper diagonal and lower diagonal matrices represent the production process of the items and the inspection process of the items, respectively.

$$B_0 = \left(\begin{array}{cccc|cccc|c} B_{01} & \hat{B}_1 & & & & & & & \\ \tilde{B}_1 & B_{02} & \hat{B}_1 & & & & & & \\ & \ddots & \ddots & \ddots & & & & & \\ & & \tilde{B}_1 & B_{02} & \hat{B}_1 & & & & \\ & & & \tilde{B}_1 & B_{02} & \hat{B}_2 & & & \\ \hline & & & & \tilde{B}_2 & B_{03} & \hat{B}_3 & & \\ & & & & & \tilde{B}_3 & B_{03} & \hat{B}_3 & \\ & & & & & & \ddots & \ddots & \ddots \\ & & & & & & & \tilde{B}_3 & B_{03} & \hat{B}_3 \\ & & & & & & & & \tilde{B}_3 & B_{03} & \hat{B}_4 \\ \hline & & & & & & & & & \tilde{B}_4 & B_{04} \end{array} \right),$$

with

$$B_{01} = (-\lambda - \eta)I_{K+1}, \quad B_{02} = \begin{pmatrix} (-\lambda - \eta - \theta)I_K & \\ & (-\lambda - \eta), \end{pmatrix},$$

$$B_{04} = \begin{pmatrix} (-\lambda - \theta)I_K & \\ & -\lambda \end{pmatrix},$$

$$B_{03} = \begin{pmatrix} I_K \otimes \begin{pmatrix} (-\lambda - \theta) & 0 \\ 0 & (-\lambda - \eta - \theta) \end{pmatrix} & \\ & \begin{pmatrix} -\lambda & 0 \\ 0 & (-\lambda - \eta) \end{pmatrix} \end{pmatrix},$$

$$\hat{B}_1 = \eta I_{K+1}, \quad \hat{B}_2 = I_{K+1} \otimes \eta e'_2(2), \quad \hat{B}_3 = I_{K+1} \otimes \eta E, \quad \hat{B}_4 = I_{K+1} \otimes \eta e_2(2),$$

$$\tilde{B}_1 = \begin{pmatrix} \theta\delta & \theta(1-\delta) & & & \\ & \theta\delta & \theta(1-\delta) & & \\ & & \ddots & \ddots & \\ & & & \theta\delta & \theta(1-\delta) \\ & & & & 0 \end{pmatrix},$$

$$\tilde{B}_2 = \begin{pmatrix} \theta\delta & \theta(1-\delta) & & & \\ \theta\delta & \theta(1-\delta) & & & \\ & \theta\delta & \theta(1-\delta) & & \\ & \theta\delta & \theta(1-\delta) & & \\ & & \ddots & \ddots & \\ & & & \theta\delta & \theta(1-\delta) \\ & & & \theta\delta & \theta(1-\delta) \\ & & & & 0 \\ & & & & 0 \end{pmatrix},$$

$$\tilde{B}_3 = \begin{pmatrix} \theta\delta I_2 & \theta(1-\delta)I_2 & & & \\ & \theta\delta I_2 & \theta(1-\delta)I_2 & & \\ & & \ddots & \ddots & \\ & & & \theta\delta I_2 & \theta(1-\delta)I_2 \\ & & & & 0 \\ & & & & 0 \end{pmatrix},$$

$$\tilde{B}_4 = \begin{pmatrix} \theta\delta e'_2(1) & \theta(1-\delta)e'_2(1) & & & \\ & \theta\delta e'_2(1) & \theta(1-\delta)e'_2(1) & & \\ & & \ddots & \ddots & \\ & & & \theta\delta e'_2(1) & \theta(1-\delta)e'_2(1) \\ & & & & 0 \end{pmatrix}.$$

The matrix B also has a d -dimensional and tridiagonal structure. The upper diagonal and lower diagonal matrices are the same ones as in the matrix B_0 . Only the matrices in the main diagonal change.

$$B = \left(\begin{array}{cccc|ccc|c} B_1 & \hat{B}_1 & & & & & & \\ \tilde{B}_1 & B_2 & \hat{B}_1 & & & & & \\ & \ddots & \ddots & \ddots & & & & \\ & & \tilde{B}_1 & B_2 & \hat{B}_1 & & & \\ & & & \tilde{B}_1 & B_2 & & & \\ \hline & & & & \tilde{B}_2 & \hat{B}_2 & & \\ & & & & B_3 & \tilde{B}_3 & & \\ & & & & \tilde{B}_3 & B_3 & \hat{B}_3 & \\ & & & & & \ddots & \ddots & \\ & & & & & & \tilde{B}_3 & B_3 & \hat{B}_3 \\ & & & & & & & \tilde{B}_3 & B_3 & \hat{B}_4 \\ \hline & & & & & & & \tilde{B}_4 & B_4 \end{array} \right),$$

where

$$B_1 = \begin{pmatrix} (-\lambda - \eta) & \\ & (-\lambda - \eta - \mu)I_K \end{pmatrix},$$

$$B_2 = \begin{pmatrix} (-\lambda - \eta - \theta) & & \\ & (-\lambda - \eta - \theta - \mu)I_{K-1} & \\ & & (-\lambda - \eta - \mu) \end{pmatrix},$$

$$B_4 = \begin{pmatrix} (-\lambda - \theta) & & \\ & (-\lambda - \theta - \mu)I_{K-1} & \\ & & (-\lambda - \mu) \end{pmatrix},$$

$$B_3 = \begin{pmatrix} B_{31} & & \\ & B_{32} & \\ & & B_{33} \end{pmatrix} \text{ with}$$

$$B_{31} = \begin{pmatrix} (-\lambda - \theta) & 0 \\ 0 & (-\lambda - \eta - \theta) \end{pmatrix}, \quad B_{33} = \begin{pmatrix} (-\lambda - \mu) & 0 \\ 0 & (-\lambda - \eta - \mu) \end{pmatrix},$$

$$B_{32} = I_{K-1} \otimes \begin{pmatrix} (-\lambda - \theta - \mu) & 0 \\ 0 & (-\lambda - \eta - \theta - \mu) \end{pmatrix}.$$

3.1. Stability condition

Let $\mathbf{z} = (z_0, z_1, \dots, z_s, z_{s+1}, \dots, z_{S-1}, z_S)$ be the steady-state probability vector of the finite generator matrix $D = A + B + C$. That is, $\mathbf{z}$ satisfies

$$\mathbf{z}D = \mathbf{0} \quad \text{and} \quad \mathbf{z}\mathbf{e} = 1. \quad (3.2)$$

The vector $\mathbf{z}_i = [z_{i,0}(1), \dots, z_{i,K}(1)]$, $0 \leq i \leq s$ and the vector $\mathbf{z}_S = [z_{S,0}(0), \dots, z_{S,K}(0)]$ include the probabilities that the production is *on* and the production is *off*, respectively. The dimension of each vector is $K+1$. The vector $\mathbf{z}_i = [z_{i,0}(0), z_{i,0}(1), \dots, z_{i,K}(0), z_{i,K}(1)]$, $s+1 \leq i \leq S-1$, of dimension $2(K+1)$, represents the probabilities that the production is both *on* and *off*. In other words, $z_{i,j}(m)$ denotes the probability that there are i items in Depot-1, there are j items in Depot-2, and the production is *on* for $m = 1$ and is *off* for $m = 0$.

The defined queueing-inventory system has a QBD-type generator matrix given in (3.1) and hence it is stable *if and only if* $\mathbf{z}A\mathbf{e} < \mathbf{z}C\mathbf{e}$ (see, e.g., Neuts [32]). The left side and the right side of the inequality can be given as follows:

$$\begin{aligned} \mathbf{z}A\mathbf{e} &= \lambda \left[\sum_{i=0}^s \sum_{j=0}^K z_{i,j}(1) + \sum_{i=s+1}^{S-1} \sum_{j=0}^K [z_{i,j}(0) + z_{i,j}(1)] + \sum_{j=0}^K z_{S,j}(0) \right], \\ \mathbf{z}C\mathbf{e} &= \mu \left[\sum_{i=0}^s \sum_{j=1}^K z_{i,j}(1) + \sum_{i=s+1}^{S-1} \sum_{j=1}^K [z_{i,j}(0) + z_{i,j}(1)] + \sum_{j=1}^K z_{S,j}(0) \right]. \end{aligned} \quad (3.3)$$

From the equations given in (3.3) and the normalizing condition in (3.2), we get the following Theorem 3.1 for the stability condition.

Theorem 3.1. *The queueing-inventory system with imperfect production and inspection processes under study is stable if and only if the following condition is satisfied:*

$$\lambda < \mu \left[1 - \left[\sum_{i=0}^{S-1} z_{i,0}(1) + \sum_{i=s+1}^S z_{i,0}(0) \right] \right]. \quad (3.4)$$

We define the two probability vectors $\mathbf{z}_i(1)$ and $\mathbf{z}_i(0)$, respectively, corresponding to the situations when the number of items in Depot-1 is i , the production is *on* and the production is *off*. These vectors are given by

$$\begin{aligned} \mathbf{z}_i(1) &= [z_{i,0}(1), z_{i,1}(1), \dots, z_{i,K}(1)], \quad 0 \leq i \leq S-1, \\ \mathbf{z}_i(0) &= [z_{i,0}(0), z_{i,1}(0), \dots, z_{i,K}(0)], \quad s+1 \leq i \leq S. \end{aligned} \quad (3.5)$$

The left-hand side of equation (3.4), λ , represents the customer demand arrival rate, while the right-hand side represents the effective rate at which the system can process this demand, taking into account the fraction of time the server is available (i.e., not blocked by an empty Depot-2). In other words, the condition requires that the long-run customer arrival rate does not exceed the long-run effective service capacity of the system. Equivalently, this condition can be expressed as an intensity rate ρ , computed by dividing λ by the effective service rate. The system is stationary if and only if $\rho < 1$.

Lemma 3.2. *The sum of the probabilities that the production is on equals the following result*

$$\sum_{i=0}^{S-1} \mathbf{z}_i(1)\mathbf{e} = \left(\frac{\theta}{\eta}\right) \left[1 - \mathbf{z}_0\mathbf{e} - \left(\sum_{i=1}^{S-1} z_{i,K}(1) + \sum_{i=s+1}^S z_{i,K}(0) \right) \right]. \quad (3.6)$$

Proof. The steady-state equations in (3.2) can be rewritten as

$$\begin{aligned} -\eta z_{0,0}(1) + \mu z_{0,1}(1) + \theta \delta z_{1,0}(1) &= 0, \\ -(\eta + \mu) z_{0,j-1}(1) + \mu z_{0,j}(1) + \theta(1 - \delta) z_{1,j-2}(1) + \theta \delta z_{1,j-1}(1) &= 0, \quad 2 \leq j \leq K, \\ -(\eta + \mu) z_{0,K}(1) + \theta(1 - \delta) z_{1,K-1}(1) &= 0. \end{aligned} \quad (3.7)$$

For $1 \leq i \leq s-1$:

$$\begin{aligned} \eta z_{i-1,0}(1) - (\eta + \theta)z_{i,0}(1) + \mu z_{i,1}(1) + \theta \delta z_{i+1,0}(1) &= 0, \\ \eta z_{i-1,j-1}(1) - (\eta + \theta + \mu)z_{i,j-1}(1) + \mu z_{i,j}(1) \\ &+ \theta(1 - \delta)z_{i+1,j-2}(1) + \theta \delta z_{i+1,j-1}(1) = 0, \quad 2 \leq j \leq K, \\ \eta z_{i-1,K}(1) - (\eta + \mu)z_{i,K}(1) + \theta(1 - \delta)z_{i+1,K-1}(1) &= 0. \end{aligned} \quad (3.8)$$

For $i = s$:

$$\begin{aligned} \eta z_{s-1,0}(1) - (\eta + \theta)z_{s,0}(1) + \mu z_{s,1}(1) + \theta \delta [z_{s+1,0}(0) + z_{s+1,0}(1)] &= 0, \\ \eta z_{s-1,j-1}(1) - (\eta + \theta + \mu)z_{s,j-1}(1) + \mu z_{s,j}(1) + \theta(1 - \delta)[z_{s+1,j-2}(0) + z_{s+1,j-2}(1)] \\ &+ \theta \delta [z_{s+1,j-1}(0) + z_{s+1,j-1}(1)] = 0, \quad 2 \leq j \leq K, \\ \eta z_{s-1,K}(1) - (\eta + \mu)z_{s,K}(1) + \theta(1 - \delta)[z_{s+1,K-1}(0) + z_{s+1,K-1}(1)] &= 0. \end{aligned} \quad (3.9)$$

For $s+1 \leq i \leq S-2$:

$$\begin{aligned} -\theta z_{i,0}(0) + \mu z_{i,1}(0) + \theta \delta z_{i+1,0}(0) &= 0, \\ \eta z_{i-1,0}(1) - (\eta + \theta)z_{i,0}(1) + \mu z_{i,1}(1) + \theta \delta z_{i+1,0}(1) &= 0, \\ -(\theta + \mu)z_{i,j-1}(0) + \mu z_{i,j}(0) + \theta(1 - \delta)z_{i+1,j-2}(0) + \theta \delta z_{i+1,j-1}(0) &= 0, \quad 2 \leq j \leq K, \\ \eta z_{i-1,j-1}(1) - (\eta + \theta + \mu)z_{i,j-1}(1) + \mu z_{i,j}(1) + \theta(1 - \delta)z_{i+1,j-2}(1) \\ &+ \theta \delta z_{i+1,j-1}(1) = 0, \\ -\mu z_{i,K}(0) + \theta(1 - \delta)z_{i+1,K-1}(0) &= 0, \\ \eta z_{i-1,K}(1) - (\eta + \mu)z_{i,K}(1) + \theta(1 - \delta)z_{i+1,K-1}(1) &= 0. \end{aligned} \quad (3.10)$$

For $i = S-1$:

$$\begin{aligned} -\theta z_{S-1,0}(0) + \mu z_{S-1,1}(0) + \theta \delta z_{S,0}(0) &= 0, \\ \eta z_{S-2,0}(1) - (\eta + \theta)z_{S-1,0}(1) + \mu z_{S-1,1}(1) &= 0, \\ -(\theta + \mu)z_{S-1,j-1}(0) + \mu z_{S-1,j}(0) + \theta(1 - \delta)z_{S,j-2}(0) + \theta \delta z_{S,j-1}(0) &= 0, \quad 2 \leq j \leq K, \\ \eta z_{S-2,j-1}(1) - (\eta + \theta + \mu)z_{S-1,j-1}(1) + \mu z_{S-1,j}(1) &= 0, \\ -\mu z_{S-1,K}(0) + \theta(1 - \delta)z_{S,K-1}(0) &= 0, \\ \eta z_{S-2,K}(1) - (\eta + \mu)z_{S-1,K}(1) &= 0. \end{aligned} \quad (3.11)$$

For $i = S$:

$$\begin{aligned} \eta z_{S-1,0}(1) - \theta z_{S,0}(0) + \mu z_{S,1}(0) &= 0, \\ \eta z_{S-1,j-1}(1) - (\theta + \mu)z_{S,j-1}(0) + \mu z_{S,j}(0) &= 0, \quad 2 \leq j \leq K, \\ \eta z_{S-1,K}(1) - \mu z_{S,K}(0) &= 0. \end{aligned} \quad (3.12)$$

Firstly, we obtain the following equations from the first two equations in (3.7) are *iteratively* added,

$$\mu z_{0,K}(1) = \eta[z_{0,K-1}(1) + \dots + z_{0,1}(1) + z_{0,0}(1)] - \theta[z_{1,0}(1) + z_{1,1}(1) + \dots + z_{1,K-2}(1)] - \theta \delta z_{1,K-1}(1),$$

and from the last equation in (3.7),

$$\mu z_{0,K}(1) = -\eta z_{0,K}(1) + \theta z_{1,K-1}(1) - \theta \delta z_{1,K-1}(1).$$

Then from the equality of these equations, the result $\eta z_0(1)e = \theta[z_1e - z_{1,K}(1)]$ for $i = 0$ is obtained.

Secondly, from the equality between the equation obtained *iteratively* adding the first two equations and the last equation in (3.8), separately for each value of i , we obtain the result $\eta z_i(1)e = \theta[z_{i+1}e - z_{i+1,K}(1)]$ for $1 \leq i \leq s-1$.

By performing similar operations from the equations in (3.9)-(3.11), the results are given as follows.

$$\begin{aligned}\eta z_i(1)e &= \theta[z_{i+1}e - z_{i+1,K}(1)], \quad 0 \leq i \leq s-1, \\ \eta z_i(1)e &= \theta[z_{i+1}e - (z_{i+1,K}(0) + z_{i+1,K}(1))], \quad s \leq i \leq S-2, \\ \eta z_{S-1}(1)e &= \theta[z_S e - z_{S,K}(0)].\end{aligned}\quad (3.13)$$

By adding the equations in (3.13) for all i values and using the normalization equation in (3.2), we obtain the result in Lemma 3.2. $\square$

Now, we define the probability $\hat{z}_j$ corresponding to the situation where the number of items in Depot-2 is j . This probability is given by

$$\hat{z}_j e = \sum_{i=0}^{S-1} z_{i,j}(1) + \sum_{i=s+1}^S z_{i,j}(0), \quad 0 \leq j \leq K. \quad (3.14)$$

Lemma 3.3. *The probability that the number of items in Depot-2 is not zero equals the following result*

$$(1 - \hat{z}_0 e) = \left(\frac{\theta(1 - \delta)}{\mu} \right) \left[1 - z_0 e - \left(\sum_{i=1}^{S-1} z_{i,K}(1) + \sum_{i=s+1}^S z_{i,K}(0) \right) \right]. \quad (3.15)$$

Proof. The steady-state equations in (3.2) can be rewritten as

For $j = 1$:

$$\begin{aligned}-\eta z_{0,0}(1) + \mu z_{0,1}(1) + \theta \delta z_{1,0}(1) &= 0, \\ \eta z_{i-1,0}(1) - (\eta + \theta) z_{i,0}(1) + \mu z_{i,1}(1) + \theta \delta z_{i+1,0}(1) &= 0, \quad 1 \leq i \leq s-1, \\ \eta z_{s-1,0}(1) - (\eta + \theta) z_{s,0}(1) + \mu z_{s,1}(1) + \theta \delta [z_{s+1,0}(1) + z_{s+1,0}(0)] &= 0, \\ -\theta z_{i,0}(0) + \mu z_{i,1}(0) + \theta \delta z_{i+1,0}(0) &= 0, \quad s+1 \leq i \leq S-2, \\ \eta z_{i-1,0}(1) - (\eta + \theta) z_{i,0}(1) + \mu z_{i,1}(1) + \theta \delta z_{i+1,0}(1) &= 0, \\ -\theta z_{S-1,0}(0) + \mu z_{S-1,1}(0) + \theta \delta z_{S,0}(0) &= 0, \\ \eta z_{S-2,0}(1) - (\eta + \theta) z_{S-1,0}(1) + \mu z_{S-1,1}(1) &= 0, \\ -\theta z_{S,0}(0) + \mu z_{S,1}(0) + \eta z_{S-1,0}(1) &= 0.\end{aligned}\quad (3.16)$$

For $2 \leq j \leq K$:

$$\begin{aligned}-(\eta + \mu) z_{0,j-1}(1) + \mu z_{0,j}(1) + \theta(1 - \delta) z_{1,j-2}(1) + \theta \delta z_{1,j-1}(1) &= 0, \\ \eta z_{i-1,j-1}(1) - (\eta + \theta + \mu) z_{i,j-1}(1) + \mu z_{i,j}(1) + \theta(1 - \delta) z_{i+1,j-2}(1) \\ + \theta \delta z_{i+1,j-1}(1) &= 0, \quad 1 \leq i \leq s-1, \\ \eta z_{s-1,j-1}(1) - (\eta + \theta + \mu) z_{s,j-1}(1) + \mu z_{s,j}(1) + \theta(1 - \delta) [z_{s+1,j-2}(0) + z_{s+1,j-2}(1)] \\ + \theta \delta [z_{s+1,j-1}(0) + z_{s+1,j-1}(1)] &= 0, \\ -(\theta + \mu) z_{i,j-1}(0) + \mu z_{i,j}(0) + \theta(1 - \delta) z_{i+1,j-2}(0) + \theta \delta z_{i+1,j-1}(0) &= 0, \quad s+1 \leq i \leq S-2, \\ \eta z_{i-1,j-1}(1) - (\eta + \theta + \mu) z_{i,j-1}(1) + \mu z_{i,j}(1) + \theta(1 - \delta) z_{i+1,j-2}(1) + \theta \delta z_{i+1,j-1}(1) &= 0, \\ -(\theta + \mu) z_{S-1,j-1}(0) + \mu z_{S-1,j}(0) + \theta(1 - \delta) z_{S,j-2}(0) + \theta \delta z_{S,j-1}(0) &= 0, \\ \eta z_{S-2,j-1}(1) - (\eta + \theta + \mu) z_{S-1,j-1}(1) + \mu z_{S-1,j}(1) &= 0, \\ -(\theta + \mu) z_{S,j-1}(0) + \mu z_{S,j}(0) + \eta z_{S-1,j-1}(1) &= 0.\end{aligned}\quad (3.17)$$

$$\begin{aligned}
& -(\eta + \mu)z_{0,K}(1) + \theta(1 - \delta)z_{1,K-1}(1) = 0, \\
& \eta z_{i-1,K}(1) - (\eta + \mu)z_{i,K}(1) + \theta(1 - \delta)z_{i+1,K-1}(1) = 0, \quad 1 \leq i \leq s-1, \\
& \eta z_{s-1,K}(1) - (\eta + \mu)z_{s,K}(1) + \theta(1 - \delta)[z_{s+1,K-1}(0) + z_{s+1,K-1}(1)] = 0, \\
& -\mu z_{i,K}(0) + \theta(1 - \delta)z_{i+1,K-1}(0) = 0, \quad s+1 \leq i \leq S-2, \\
& \eta z_{i-1,K}(1) - (\eta + \mu)z_{i,K}(1) + \theta(1 - \delta)z_{i+1,K-1}(1) = 0, \\
& -\mu z_{S-1,K}(0) + \theta(1 - \delta)z_{S,K-1}(0) = 0, \\
& \eta z_{S-2,K}(1) - (\eta + \mu)z_{S-1,K}(1) = 0, \\
& -\mu z_{S,K}(0) + \eta z_{S-1,K}(1) = 0.
\end{aligned} \tag{3.18}$$

When the equations in (3.16) are added, we get the equation $\mu \hat{z}_1 \mathbf{e} = \theta(1 - \delta)[\hat{z}_0 \mathbf{e} - z_{0,0}(1)]$.

Similarly, for each value of j , the equations in (3.17) are added and we get the equation $\mu \hat{z}_j \mathbf{e} = \theta(1 - \delta)[\hat{z}_{j-1} \mathbf{e} - z_{0,j-1}(1)]$, $j = 2, \dots, K$. So, the following result can be written.

$$\mu \hat{z}_j \mathbf{e} = \theta(1 - \delta)[\hat{z}_{j-1} \mathbf{e} - z_{0,j-1}(1)], \quad 0 \leq j \leq K. \tag{3.19}$$

By adding the equations in (3.19) for all j values and using the normalization equation $\mathbf{z} \mathbf{e} = 1$ in (3.2), we obtain the result in Lemma 3.3. $\square$

When we look at the equalities given in Lemma 3.2 and Lemma 3.3, we see that the expressions in square brackets are the same. It follows that

$$\mu(1 - \hat{z}_0 \mathbf{e}) = \eta(1 - \delta) \sum_{i=0}^{S-1} z_i(1) \mathbf{e}. \tag{3.20}$$

We would like to emphasize the following points. The right side of the inequality in Theorem 3.1 equals to the left side of the equation in (3.20). So, from this relation, the stability condition in Theorem 3.1 can be rewritten such as the first inequality in Theorem 3.4.

From the similarity between the right side of the inequality in Theorem 3.1 and the left side of the equation in Lemma 3.3, we can also rewrite the stability condition in Theorem 3.1 as given in the second inequality in Theorem 3.4.

Theorem 3.4. *The defined queuing-inventory system with imperfect production and inspection processes is stable if and only if the following condition is satisfied:*

$$\begin{aligned}
& \lambda < \eta(1 - \delta) \sum_{i=0}^{S-1} z_i(1) \mathbf{e}, \\
& \lambda < \theta(1 - \delta) \left[1 - z_0 \mathbf{e} - \left(\sum_{i=1}^{S-1} z_{i,K}(1) + \sum_{i=s+1}^S z_{i,K}(0) \right) \right].
\end{aligned} \tag{3.21}$$

3.2. The steady-state probability vector

Let $\mathbf{y} = (\mathbf{y}(0), \mathbf{y}(1), \mathbf{y}(2), \dots)$ denotes the limiting probability vector of the generator matrix Q in (3.1). That is, $\mathbf{y}$ satisfies

$$\mathbf{y} Q = \mathbf{0} \quad \text{and} \quad \mathbf{y} \mathbf{e} = 1. \tag{3.22}$$

The vector $\mathbf{y}(n)$, $n \geq 0$, of dimension $(2S - s)(K + 1)$ is partitioned into vectors of dimension $(K + 1)$ as

$$\mathbf{y}(n) = [\mathbf{y}_0(n, 1), \mathbf{y}_1(n, 1), \dots, \mathbf{y}_s(n, 1), \mathbf{y}_{s+1}(n, 0), \mathbf{y}_{s+1}(n, 1), \dots, \mathbf{y}_{S-1}(n, 0), \mathbf{y}_{S-1}(n, 1), \mathbf{y}_S(n, 0)].$$

where $\mathbf{y}_i(n, m) = [y_{i,0}(n, m), y_{i,1}(n, m), \dots, y_{i,K}(n, m)]$, $0 \leq i \leq S$, $m = 0, 1$.

$y_{i,j}(n, 1)$, $0 \leq i \leq s$, $0 \leq j \leq K$, denotes the probability that the number of customers in the system is n , the number of items in the inspection is i , the number of items in the inventory is j and the production process is *on*.

$y_{i,j}(n, m)$, $s + 1 \leq i \leq S - 1$, $0 \leq j \leq K$, $m = 0, 1$, denotes the probability that the number of customers in the system is n , the number of items in the inspection is i , the number of items in the inventory is j and the production process is *off* for $m = 0$ (the production process is *on* for $m = 1$).

$y_{S,j}(n, 0)$, $0 \leq j \leq K$, denotes the probability that the number of customers in the system is n , the number of items in the inspection is maximum level S , the number of items in the inventory is j and the production process is *off*.

Under the stability condition given in Theorem 3.1 (also Theorem 3.4) the steady-state probability vector $\mathbf{y}$ is obtained (see Neuts [32]) as

$$\mathbf{y}(i) = \mathbf{y}(0)R^i, \quad i \geq 1, \quad (3.23)$$

where the matrix R is the minimal nonnegative solution to the following matrix quadratic equation

$$R^2C + RB + A = \mathbf{0}, \quad (3.24)$$

and the vector $\mathbf{y}(0)$ is obtained by solving

$$\mathbf{y}(0)[B_0 + RC] = \mathbf{0}, \quad (3.25)$$

subject to the normalizing condition

$$\mathbf{y}(0)(I - R)^{-1}\mathbf{e} = 1. \quad (3.26)$$

3.3. Performance measures

Several performance measures characterizing the system are computed and summarized below:

1) The probability that the number of items in Depot-2 equals to zero:

$$P_{D2(zero)} = \sum_{n=1}^{\infty} \left[\sum_{i=0}^{S-1} y_{i,0}(n, 1) + \sum_{i=s+1}^S y_{i,0}(n, 0) \right].$$

2) The probability that the system is idle (i.e., contains no customers):

$$P_{idle} = \mathbf{y}(0)\mathbf{e}.$$

3) The probabilities that the server is idle $P_{i(s)}$ and the inspector is idle $P_{i(i)}$, respectively:

$$P_{i(s)} = \mathbf{y}(0)\mathbf{e} + P_{d2(zero)}.$$

$$P_{i(i)} = \sum_{n=0}^{\infty} \left[\sum_{j=0}^K y_{0,j}(n, 1) + \sum_{i=1}^{S-1} y_{i,K}(n, 1) + \sum_{i=s+1}^S y_{i,K}(n, 0) \right].$$

4) The probabilities that the production is *on* (PP_{ON}) and the inspection is *on* (PI_{ON}):

$$PP_{ON} = \sum_{n=0}^{\infty} \sum_{i=0}^{S-1} \sum_{j=0}^K y_{i,j}(n, 1).$$

$$PI_{ON} = \sum_{n=0}^{\infty} \sum_{j=0}^{K-1} \left[\sum_{i=1}^{S-1} y_{i,j}(n, 1) + \sum_{i=s+1}^S y_{i,j}(n, 0) \right].$$

5) The mean production rate (E_{PR}) and the mean inspection rate (E_{IR}):

$$E_{PR} = \eta PP_{ON} \quad \text{and} \quad E_{IR} = \theta PI_{ON}.$$

We note that $E_{PR} = E_{IR}$ and $\eta(1 - \delta)PP_{ON} = \lambda$.

6) The mean discard rate (E_{DR}) and the mean sending rate to Depot-2 for the inspected item (E_{SRI}):

$$E_{DR} = \theta \delta PI_{ON} \quad \text{and} \quad E_{SRI} = \theta(1 - \delta)PI_{ON} = \lambda.$$

7) The mean rate at which production is switched *on*:

$$ER_{ON} = \theta \sum_{n=0}^{\infty} \sum_{j=0}^{K-1} y_{s+1,j}(n, 0).$$

8) The mean number of customers in the system:

$$EN = \sum_{n=1}^{\infty} n \mathbf{y}(n) \mathbf{e} = \mathbf{y}(0) R(I - R)^{-2} \mathbf{e}.$$

9) The mean number of items in Depot-1:

$$EI_{D1} = \sum_{n=0}^{\infty} \left[\sum_{i=1}^{S-1} i \sum_{j=0}^K y_{i,j}(n, 1) + \sum_{i=s+1}^S i \sum_{j=0}^K y_{i,j}(n, 0) \right].$$

10) The mean number of items in Depot-2:

$$EI_{D2} = \sum_{n=0}^{\infty} \sum_{j=1}^K j \left[\sum_{i=0}^{S-1} y_{i,j}(n, 1) + \sum_{i=s+1}^S y_{i,j}(n, 0) \right].$$

4. Numerical Study

The objective of this section is twofold. First, the behavior of the system performance measures is examined for different values of various input parameters. Some key results from the sensitivity analysis are summarized. Second, a cost function is constructed, and its optimization is performed by using Particle Swarm Optimization (PSO). For the given values of the unit costs, some optimum parameter values that minimize the cost function are obtained. The results of the optimization study are also provided in this section. All numerical calculations, including the computation of the steady-state probabilities, the system performance measures, and the PSO-based cost optimization, were performed using codes developed by the authors in MATLAB.

4.1. Sensitivity analysis of the system performance measures

Numerical results are presented to examine the effects of the parameters on the performance measures under the stability condition. The input parameters used for illustrative purposes are the reorder level $s = 3$ and the capacity of Depot-1 $S = 8$, the Type-I and Type-II misclassification probabilities $p_1 = 0.03$ and $p_2 = 0.05$, the production rate $\eta = 6$, the inspection rate $\theta = 8$, the service rate $\mu = 7$, and the customer arrival rate $\lambda = 3$.

The sensitivity of the system performance measures is examined for different values of the defective production probability $\alpha = 0.01, 0.04, 0.07$ and the capacity of Depot-2, $K = 4, 6, 8, 10$. Throughout this section, the performance measures introduced in Section 3.3 are examined, namely the mean number of customers in the system (EN), the mean number of items in Depot-1 (EI_{D1}) and in Depot-2 (EI_{D2}), the mean rate at which

production is switched on (ER_{ON}), the mean discard rate (E_{DR}), the mean production rate (E_{PR}), and the probabilities that production and inspection are on (PP_{ON} and PI_{ON})

Figure 2 illustrates the effect the arrival rate (λ), service rate (μ), production rate (η) and inspection rate (θ), on the mean number of customers (EN) in the system. The analysis is carried out for two levels of the defective production probability, namely $\alpha = 0.01$ and $\alpha = 0.07$. As shown in Figure 2, EN increases with the arrival rate λ since a higher demand rate leads to greater congestion in the system. In contrast, EN decreases as the service rate μ , the production rate η , or the inspection rate θ increases, because each of these parameters improves the rate at which customers can be served. These changes become steeper and more pronounced when the system operates under a high traffic intensity ρ is high. Figure 2 also allows the effect of the defective production rate α on EN to be assessed across the different parameters. Comparing the cases $\alpha = 0.01$ and $\alpha = 0.07$ reveals that the influence of α becomes particularly pronounced at high values of ρ , since a larger fraction of defective items reduces the effective replenishment rate of Depot-2 and thereby intensifies congestion in an already heavily loaded system.

In addition to this, we can see the effect of the capacity of Depot-2 (K) on the behaviour of EN for the values of $K = 4, 6, 8, 10$ in Figure 2. When ρ is high, it is observed that the capacity K has an important effect on EN . For increasing values of K , the values of EN tend to decrease.

Figure 3 illustrates the effect of the arrival rate (λ), service rate (μ), production rate (η) and inspection rate (θ), on the mean inventory level in Depot-1 (EI_{D1}). The analysis was carried out for different values of the capacity of Depot-2 K , and EI_{D1} was found to decrease as K increases. This is because a larger Depot-2 becomes full less frequently, so that inspected items are transferred out of Depot-1 more readily and accumulate to a lesser extent. Since the effect of K on EI_{D1} exhibits a similar pattern for all system parameters, Figure 3 is presented for a single value of K .

An increase in λ , raises the traffic intensity rate ρ , whereas an increase in η lowers it. Consequently, as can be seen from Figure 3, EI_{D1} takes low values when ρ is high. A higher arrival rate depletes Depot-2 more rapidly, which in turn draws items out of Depot-1 through the inspection process, leaving fewer items in Depot-1. Conversely, a higher production rate means that, once production is switched on, Depot-1 is refilled to the level S more quickly, so that the system spends a greater proportion of time at higher inventory levels in Depot-1. The defective production probability α is also observed to affect EI_{D1} under ρ , since a larger fraction of defective items reduces the effective inflow to Depot-2 and thereby intensifies the depletion of Depot-1.

Figure 3 shows that EI_{D1} decreases as either μ or θ increases, with the decrease being more pronounced for changes in θ than for changes in μ . This is expected, since θ governs the rate at which items leave Depot-1 through inspection and therefore acts directly on the outflow, whereas μ affects Depot-1 only indirectly, through the depletion of Depot-2. For the parameters μ and η , the curves corresponding to different values of the defective production probability α remain almost parallel, so that EI_{D1} decreases as α increases by a nearly constant amount. This is consistent with the previous observation: since an increase in μ lowers the system intensity, these panels correspond to relatively low values of ρ , where the effect of α does not become pronounced.

The behavior of the mean number of items in Depot-2 (EI_{D2}) was examined and presented in Figure 4 for different values of the capacity K as the system parameters vary. For low values of EI_{D2} also takes low values, since the inventory level in Depot-2 is bounded above by its capacity. Figure 4 is therefore presented for a single value of α with the curves corresponding to the different capacity levels.

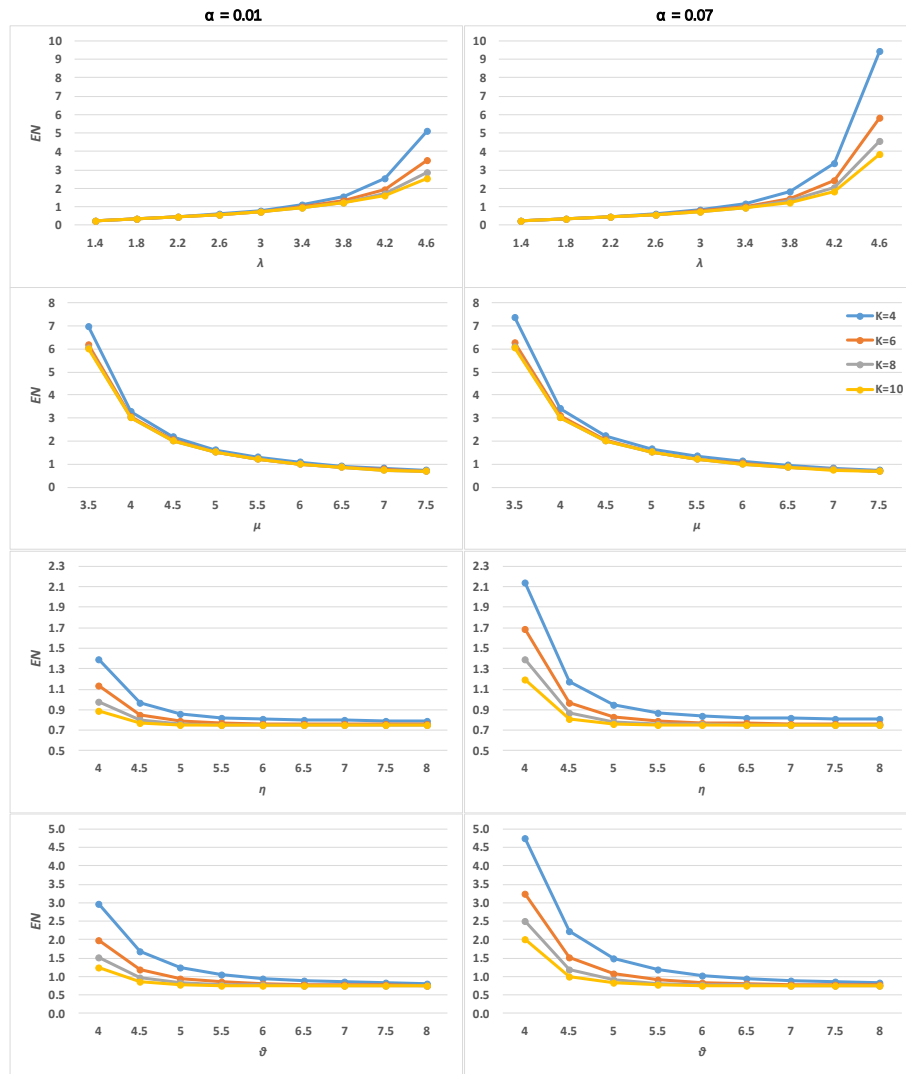


Figure 2. Effect of the parameters on the mean number of customers in the system.

Figure 4 presents the change in EI_{D2} with respect to λ and η . When the system intensity is high, EI_{D2} is expected to be high, and in this case the effect of the capacity K becomes considerably more pronounced, since a larger Depot-2 allows more items to accumulate before the capacity limit is reached. Moreover, as K increases, both the rate of increase and the rate of decrease become sharper: over a given range of λ the curve for $K = 10$ rises more steeply than that for $K = 4$, while over a given range of η , the curve for $K = 10$ falls more steeply than that for $K = 4$. This is because, at low capacities, the inventory level is constrained by the capacity limit and therefore responds less sensitively to changes in the system parameters. Figure 4 also shows that the effect of μ on EI_{D2} is negligible, with only a very slight decrease observed as μ increases. In contrast, the behavior of EI_{D2} with respect to θ depends on the capacity of Depot-2. For low capacity values ($K = 4$ and $K = 6$) EI_{D2} decreases as θ increases, whereas for high capacity values ($K = 8$ and $K = 10$) EI_{D2} first decreases and then increases slowly.

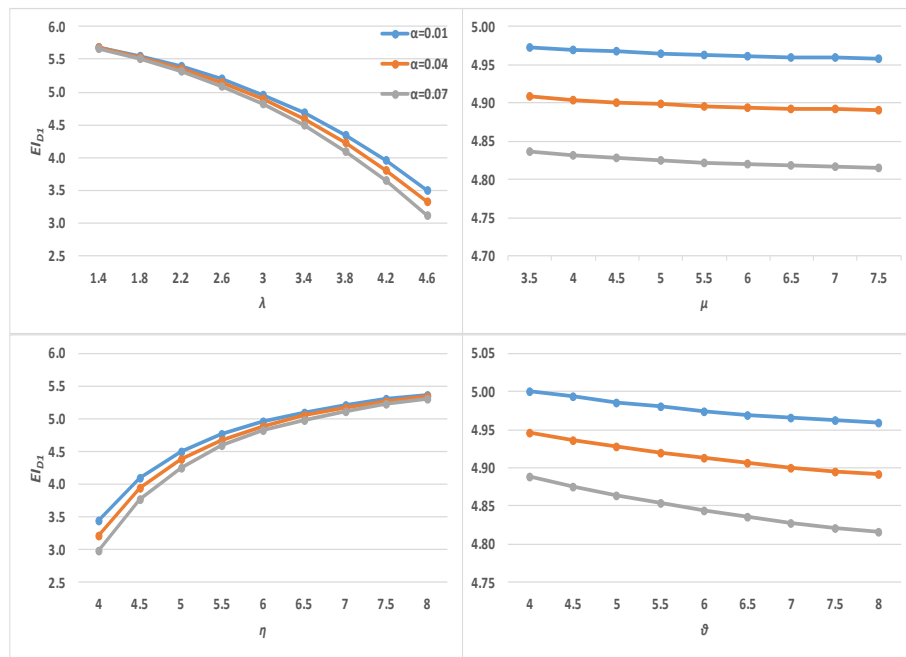


Figure 3. The effect of the system parameters on the mean number of items in Depot-1 for $K=4$.

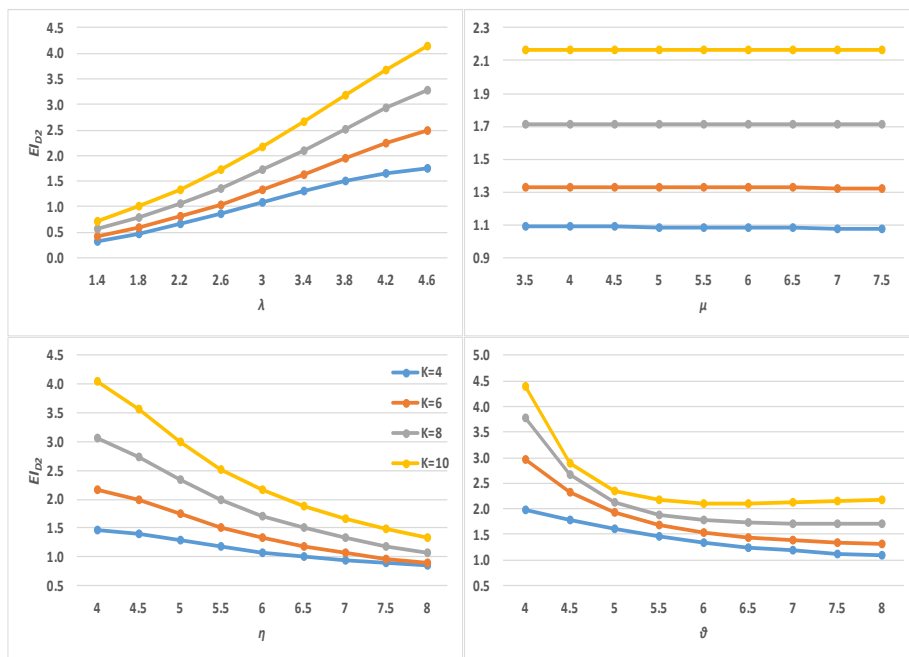


Figure 4. The effect of the system parameters on the inventory level in the Depot-2 for $\alpha=0.07$.

The impact of the mean rate at which production is switched *on* (ER_{ON}) is shown in Figure 5. This effect was also investigated for different values of the capacity K and ER_{ON} was found to decrease as K increases, since a larger Depot-2 becomes full less frequently and therefore allows Depot-1 to be depleted more steadily, so that the reorder level s is

reached less often. Figure 5 is presented for the case $K = 4$. The change in ER_{ON} with increasing λ values exhibits concave behavior. ER_{ON} increases with λ at low values of ρ , but decreases at high values of ρ . In contrast, ER_{ON} increases monotonically with η . For both of these parameters, the defective production probability α has a pronounced effect on ER_{ON} when ρ is high. Figure 5 further shows that ER_{ON} decreases as either μ or θ increases. For these two parameters, larger values of α consistently lead to smaller values of ER_{ON} .

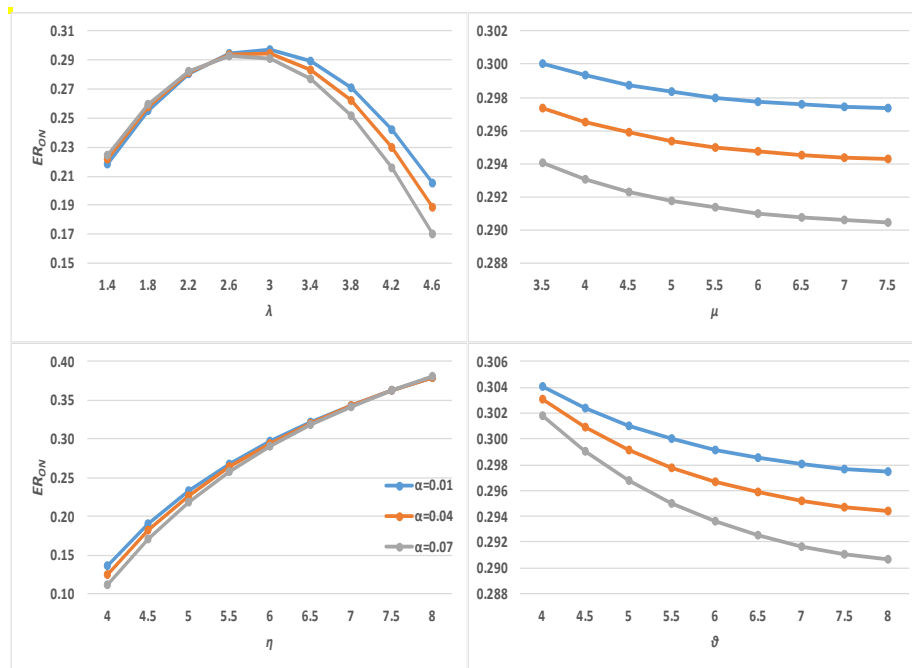


Figure 5. The effect of the system parameters on the mean rate at which production is switched on for $K = 4$.

The effect of the capacity K on the mean discard rate (E_{DR}) and the mean production rate (E_{PR}) was also investigated as the system parameters both measures were found to be independent of K . Figure 6 therefore shows the change in E_{DR} and E_{PR} with increasing values of λ and μ , for $K = 4$; the results obtained when η and θ vary coincide with those obtained for μ . As can be seen from Figure 6, both E_{DR} and E_{PR} remain unchanged as μ increases, whereas they increase linearly with λ . This is consistent with the relation $E_{SRI} = \theta(1 - \delta)PI_{ON} = \lambda$ established in Section 3.3. In steady state, the rate at which non-defective items are sent to Depot-2 must equal the customer arrival rate, so that the production rate, and consequently the discard rate, are determined by λ alone and are unaffected by the service, production, or inspection rates.

Figure 6 further shows that while the effect of α on the mean production rate E_{PR} is quite small, its effect on the mean discard rate E_{DR} is substantial. As α increases, E_{DR} increases as expected, as a larger fraction of the products are identified as defective and discarded during inspection. The service rate, in contrast, is only slightly affected, as it must still meet the same demand rate λ , requiring only a marginal increase to compensate for additional items removed from the system.

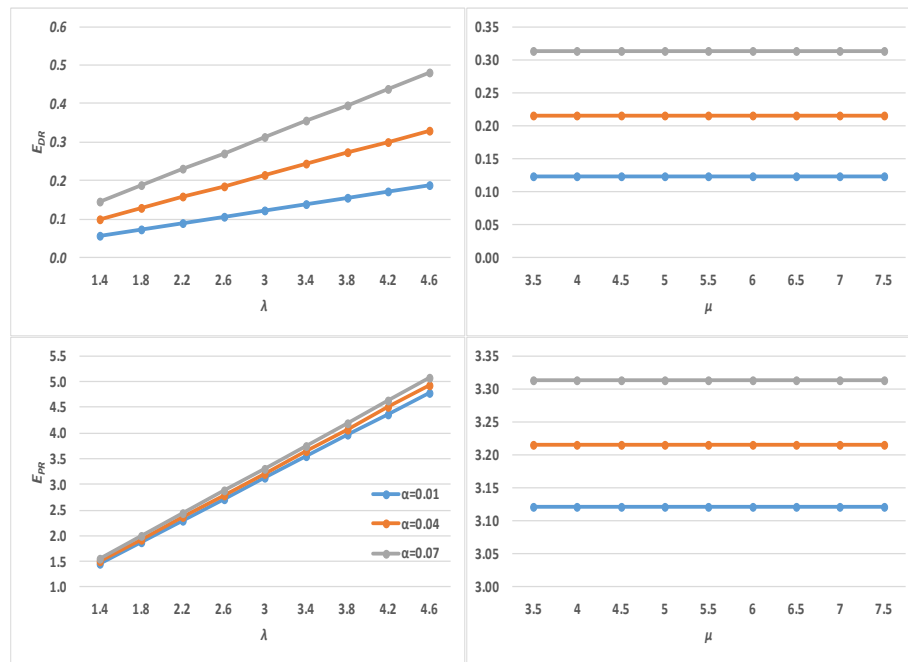


Figure 6. The effect of λ and μ on the mean discard rate and the mean production rate for $K = 4$.

The behavior of the probability that production is *on* (PP_{ON}) and the probability that inspection is *on* (PI_{ON}) was examined for different values of capacity K as the system parameters vary, and both probabilities were found to be independent of K . Therefore, the results are presented for the constant value $K = 4$ in Figure 7 and Figure 8, respectively. When the system intensity is high, production is expected to be turned *on* more often. Figure 7 confirms this relation. PP_{ON} increases with λ and decreases with η , since a higher demand rate depletes Depot-1 more rapidly and thus triggers production more frequently, whereas a higher production rate allows the target level S to be reached more quickly and therefore shortens the periods during which production remains *on*. Moreover, PP_{ON} is not affected by changes in μ and θ , which follows from the relation $\eta(1 - \delta)PP_{ON} = \lambda$ established in Section 3.3, indicating that PP_{ON} is determined solely by λ , η and δ .

Similarly, inspection is expected to be turned *on* more often when the system intensity is high. As shown in Figure 8, PI_{ON} increases with λ and decreases with θ , the latter because a higher inspection rate allows each item to be processed more quickly, leaving the inspector idle for longer periods. As in the case of PP_{ON} the changes in μ and η have no effect on PI_{ON} , which is consistent with the relation $\theta(1 - \delta)PI_{ON} = \lambda$ given in Section 3.3.

As shown in the presented figures, the probability of defective production α has an important effect on performance measures. Our analysis further indicates that the Type-I misclassification probability p_1 has a more significant effect on performance measures than the Type-II misclassification probability p_2 . To illustrate the effect of p_1 , the performance measures are computed for specific values of the system parameters and reported in Table 1, where $\alpha = 0.01$, $p_2 = 0.05$, $s = 3$, $S = 8$, $K = 6$, with computations performed for two values each of the arrival rate, service rate, production rate, and inspection rate.

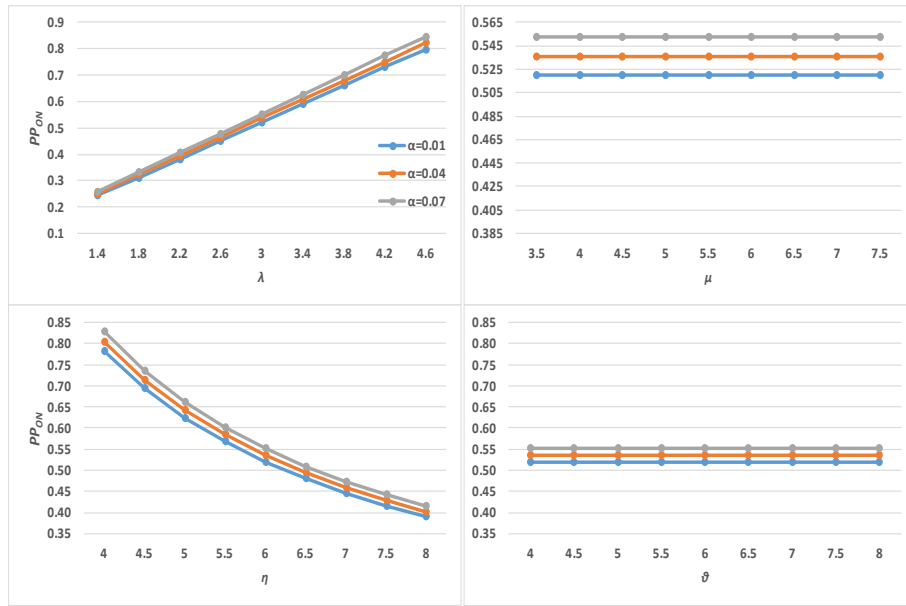


Figure 7. The probabilities that the production is *on* for the case of $K = 4$.

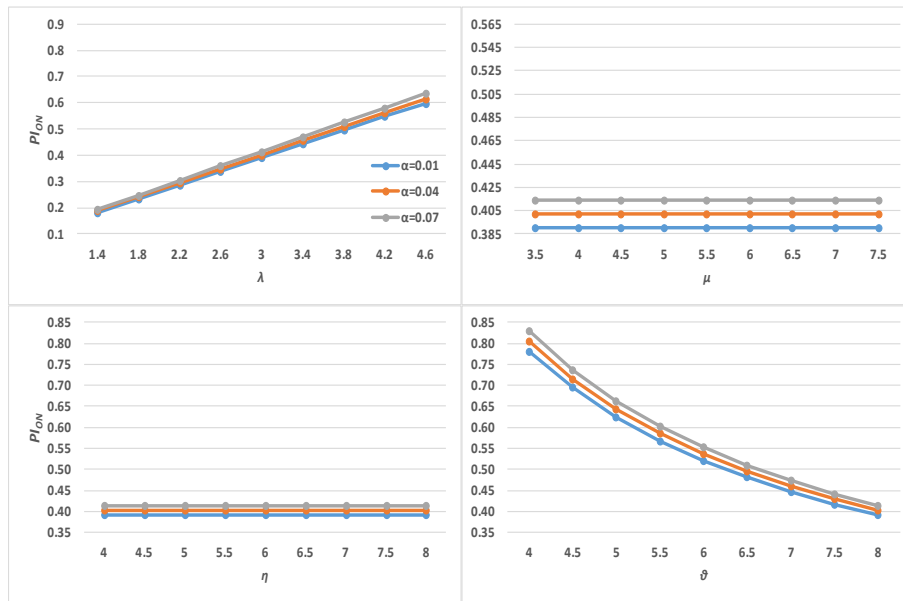


Figure 8. The probabilities that the inspection is *on* for the case of $K = 4$.

The computations are performed for two values of arrival rate, service rate, production rate, and inspection rate. Table 1 shows that p_1 has a considerable effect on all performance measures, and this effect is more pronounced when the system intensity rate is high. The main findings of the sensitivity analysis are summarized in Table 2, which reports the direction of the effect of each system parameter on the performance measures considered in this section. Several general patterns emerge from this summary. The arrival rate λ is the only parameter that influences every performance measure, reflecting its role as the driver of demand in the system. All other rates adjust to accommodate the demand imposed by λ . In contrast, the service rate μ affects only the measures related to customers and to Depot-1, leaving the discard and production rates unchanged, since in steady state these

are determined by the demand rate rather than by the speed of service. The production rate η and the inspection rate θ act in opposite directions on the two depots: an increase in η raises the inventory level in Depot-1 while lowering it in Depot-2, whereas an increase in θ reduces the inventory in Depot-1 by accelerating the flow of items through inspection.

The probability of defective production α and the capacity of Depot-2 K both influence a wide range of measures, with α affecting the discard rate the most strongly and K affecting the level of inventory in Depot-2. Finally, two measures exhibit non-monotonic behavior, namely ER_{ON} with respect to λ and EI_{D2} with respect to θ , indicating that the effect of these parameters depends on the operating region of the system. In general, the table shows that no single parameter governs the system in isolation, which motivates the joint optimization of the production, inspection, and service rates carried out in the following section.

Table 1. The effect of p_1 on the performance measures.

	p_1	$0.41 < \rho < 0.62$				$0.61 < \rho < 0.89$			
		$\lambda = 2.2$	$\mu = 5.5$	$\eta = 7$	$\theta = 8$	$\lambda = 4.6$	$\mu = 3.5$	$\eta = 4$	$\theta = 6$
EN	0.01	0.4589	1.2142	0.7565	0.7596	3.1275	6.1388	1.0313	0.7970
	0.03	0.4590	1.2173	0.7577	0.7618	3.5100	6.1637	1.1305	0.8060
	0.05	0.4592	1.2210	0.7592	0.7646	4.0557	6.1940	1.2695	0.8173
	0.07	0.4594	1.2258	0.7610	0.7682	4.8651	6.2311	1.4675	0.8314
EI_{D1}	0.01	5.4077	4.9999	5.2390	4.9993	3.5766	5.0022	3.5687	5.0093
	0.03	5.3866	4.9555	5.2092	4.9549	3.4500	4.9583	3.4258	4.9679
	0.05	5.3638	4.9076	5.1769	4.9069	3.3144	4.9109	3.2711	4.9235
	0.07	5.3390	4.8559	5.1421	4.8551	3.1692	4.8597	3.1036	4.8759
EI_{D2}	0.01	0.6883	1.1402	0.9077	1.1398	2.3229	1.1397	2.0927	1.3014
	0.03	0.7158	1.1864	0.9447	1.1857	2.3740	1.1860	2.1294	1.3569
	0.05	0.7449	1.2347	0.9837	1.2338	2.4197	1.2346	2.1568	1.4154
	0.07	0.7756	1.2853	1.0250	1.2841	2.4585	1.2856	2.1720	1.4772
ER_{ON}	0.01	0.2797	0.2989	0.3432	0.2988	0.2115	0.2993	0.1438	0.2998
	0.03	0.2806	0.2971	0.3431	0.2970	0.2005	0.2975	0.1363	0.2982
	0.05	0.2814	0.2950	0.3427	0.2949	0.1886	0.2955	0.1281	0.2965
	0.07	0.2820	0.2925	0.3421	0.2924	0.1759	0.2931	0.1193	0.2944
E_{DR}	0.01	0.0435	0.0594	0.0594	0.0594	0.0910	0.0594	0.0594	0.0594
	0.03	0.0898	0.1224	0.1224	0.1224	0.1877	0.1224	0.1224	0.1224
	0.05	0.1379	0.1881	0.1881	0.1881	0.2884	0.1881	0.1881	0.1881
	0.07	0.1882	0.2566	0.2566	0.2566	0.3935	0.2566	0.2566	0.2566
E_{PR}	0.01	2.2435	3.0594	3.0594	3.0594	4.6910	3.0594	3.0594	3.0594
	0.03	2.2898	3.1224	3.1224	3.1224	4.7877	3.1224	3.1224	3.1224
	0.05	2.3379	3.1881	3.1881	3.1881	4.8884	3.1881	3.1881	3.1881
	0.07	2.3882	3.2566	3.2566	3.2566	4.9935	3.2566	3.2566	3.2566
PP_{ON}	0.01	0.3739	0.5099	0.4371	0.5099	0.7818	0.5099	0.7648	0.5099
	0.03	0.3816	0.5204	0.4461	0.5204	0.7979	0.5204	0.7806	0.5204
	0.05	0.3897	0.5313	0.4554	0.5313	0.8147	0.5313	0.7970	0.5313
	0.07	0.3980	0.5428	0.4652	0.5428	0.8322	0.5428	0.8142	0.5428
PI_{ON}	0.01	0.2804	0.3824	0.3824	0.3824	0.5864	0.3824	0.3824	0.5099
	0.03	0.2862	0.3903	0.3903	0.3903	0.5985	0.3903	0.3903	0.5204
	0.05	0.2922	0.3985	0.3985	0.3985	0.6111	0.3985	0.3985	0.5313
	0.07	0.2985	0.4071	0.4071	0.4071	0.6242	0.4071	0.4071	0.5428

Table 2. Summary of the direction of the parameter effects on the performance measures.

Parameter	$\uparrow$	EN	EI_{D1}	EI_{D2}	ER_{ON}	EDR	EPR	PP_{ON}	PI_{ON}
λ		$\uparrow$	$\downarrow$	$\uparrow$	$\downarrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$
μ		$\downarrow$	$\downarrow$	$\approx$	$\downarrow$	—	—	—	—
η		$\downarrow$	$\uparrow$	$\downarrow$	$\uparrow$	—	—	$\downarrow$	—
θ		$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	—	—	—	$\downarrow$
α		$\uparrow$	$\downarrow$	$\uparrow$	$\downarrow$	$\uparrow$	$\approx$	—	—
K		$\downarrow$	$\downarrow$	$\uparrow$	$\downarrow$	—	—	—	—

$\uparrow$: increases, $\downarrow$: decreases, $\downarrow$: non-monotonic,
 $\approx$: negligible effect, —: no effect

4.2. Cost optimization

In this section, the total expected cost function per unit time is constructed based on the system performance measures obtained in Section 3.3. First, let the cost elements be defined as follows:

- C_1 : cost per unit time of producing a unit
- C_2 : cost per unit time of inspection and holding one unit in Depot-1
- C_3 : cost per unit time of holding one unit in Depot-2
- C_4 : cost per unit time of discarding a nondefective product
- C_5 : cost per unit time of sending a defective product to Depot-2
- C_6 : waiting cost per unit time per customer in the system
- C_7 : cost per unit time of providing a service

The main objective is to determine the optimal production rate η , optimal inspection rate θ , and optimal service rate μ so that the expected cost function is minimized. Based on the unit cost elements defined above and by using the performance measures, the expected total cost function can be formulated as

$$TC(\eta, \theta, \mu) = C_1 E_{PR} + C_2 EI_{D1} + C_3 EI_{D2} + C_4 EI_{D1} [p_1(1 - \alpha)] + C_5 EI_{D2} p_2 \alpha + C_6 EN + C_7 \mu \quad (4.1)$$

Since the expected cost function is highly complex, it is difficult to prove its convexity and also obtain explicit expressions for the parameters using classical optimization methods. Therefore, Particle Swarm Optimization (PSO) is used to find the optimal parameter values.

4.2.1. Particle Swarm Optimization Method. The PSO algorithm was introduced by Kennedy and Eberhard in 1995 [20]. In the PSO algorithm, particles flow through a multidimensional search space and the position of each particle is determined by its own experience (cognitive) and that of its neighbouring particles (social) experiences. The position of a particle is updated by adding a velocity to its current position [14]. Let $x_i(t)$ represent the value of the i^{th} particle in iteration t . Then the value $x_i(t + 1)$ is given by

$$x_i(t + 1) = x_i(t) + v_i(t + 1) \quad (4.2)$$

The velocity vector $v_i(t)$ drives the optimization process and contains both the particle's own experiential knowledge (cognitive component) and the experience it obtains socially from its neighbors (social component). The velocity is computed as below [42]:

$$v_i(t + 1) = wv_i(t) + c_1 U_1(p_{best} - x_i(t)) + c_2 U_2(g_{best} - x_i(t)) \quad (4.3)$$

The performance of PSO depends on some control parameters such as inertia weight, acceleration constants, swarm size, number of iterations, etc. Inertia weight (w) is proposed by Shi and Eberhart [42] and controls the momentum of a particle. U_1 and U_2 are uniform random variables. c_1 and c_2 are acceleration constants, and Eberhart and Shi [13] showed that these coefficients can be taken 1.494 in their studies. Various studies have shown that PSO is capable of finding optimal solutions with swarms of 20 to 60 particles in size [13, 42]. PSO has been successfully applied for optimizing highly complex cost functions in complex queueing problems [46, 47, 49, 50]. In our study, the PSO method is implemented to optimize the expected cost function given the equation in (4.1). The procedure implemented is described below:

Algorithm 1 PSO Algorithm

```

1: Initialization: Generate the initial swarm
2: Check the stability condition  $\rho < 1$ 
3: Initialize pbest, gbest, velocity vector
4: Evaluate the cost function and find the initial cost value
5: while (iter  $\leq$  maxiter) or ( $\epsilon > 10^{-5}$ ) do
6:   velocity update using eq.(4.3)
7:   position update using eq. (4.2)
8:   check the stability condition  $\rho < 1$ 
9:   if  $\rho \geq 1$  then
10:     the corresponding cost is assigned a high value
11:   end if
12:   update pbest and gbest
13:   update minimum cost value
14:   find the difference between the two latest min cost
15:   if difference  $< 10^{-5}$  then
16:      $k = k + 1$ 
17:   end if
18:   if  $k \geq 3$  &  $\epsilon =$  difference then
19:     end if
20: end while

```

4.2.2. Results. In the numerical study, the PSO algorithm was implemented with a swarm size of 50 particles and a maximum number of iterations of $maxiter = 50$. The inertia weight was kept constant at $w = 0.5$, and the acceleration constants were set to $c_1 = c_2 = 1.494$, as suggested by Eberhart and Shi [13]. The algorithm ends when the maximum number of iterations is reached or when the improvement in the minimum cost value falls below the tolerance of 10^{-5} for three consecutive iterations, as described in *Algorithm 1*. The decision variables η , θ , and μ were searched within the bounds [1, 11]. Since not every parameter combination within these bounds necessarily yields a stable system, the stability condition given in Theorem 3.1 was verified for each candidate solution generated during the search. Whenever a candidate solution violates the stability condition, a sufficiently high cost value is assigned to that solution, so that it is effectively excluded from the search and cannot be selected as a personal or global best (see steps 8 – 11 of *Algorithm 1*). The unit costs used for the optimization study for illustrative purposes are given below:

$$C_1 = 50, C_2 = 25, C_3 = 15, C_4 = 100, C_5 = 250, C_6 = 40, C_7 = 20.$$

The optimal parameter values and also the minimum cost are obtained for $K = 5$ and $K = 9$, and the results for $\lambda = 1$ and $\lambda = 1.4$ are given in Table 3 and Table 4, respectively.

Table 3. Optimal parameter values and total costs when $\lambda = 1$.

K	α	S	η^*	θ^*	μ^*	TC^*	ρ
5	0.01	5	1.3632	11.0000	2.5051	196.3002	0.7568
		6	1.3464	7.7550	2.4619	199.3318	0.7661
		7	1.3410	5.8001	2.3285	205.0861	0.7694
		8	1.2987	9.4705	2.3895	210.6275	0.7936
	0.1	5	1.4899	8.3196	2.5424	198.0384	0.7582
		6	1.4824	11.0000	2.5085	200.8652	0.7614
		7	1.5072	8.0587	2.5787	207.0237	0.7486
		8	1.4568	7.9423	2.7295	213.5903	0.7740
9	0.01	5	1.3304	3.4858	2.6697	224.1929	0.7801
		6	1.2752	5.3712	2.4459	222.4970	0.8082
		7	1.2968	5.2621	2.6491	225.0223	0.7945
		8	1.2709	7.8797	2.4836	227.7968	0.8106
	0.1	5	1.3994	7.0394	2.5750	223.1257	0.8060
		6	1.4057	5.3378	2.4340	223.6029	0.8026
		7	1.4197	3.2939	2.6026	227.5683	0.7971
		8	1.3876	5.5752	2.5039	229.9751	0.8126

Table 4. Optimal parameter values and total costs when $\lambda = 1.4$.

K	α	S	η^*	θ^*	μ^*	TC^*	ρ
5	0.01	5	1.9115	7.1750	3.2419	218.9269	0.7584
		6	1.9026	9.4818	3.1649	220.0875	0.7602
		7	1.9062	7.4609	3.1494	225.0871	0.7585
		8	1.8463	11.0000	3.1966	230.2664	0.7818
	0.1	5	2.1016	10.7278	3.2564	219.9151	0.7540
		6	2.0818	11.0000	3.1912	222.0786	0.7604
		7	2.0598	10.9526	3.2847	226.9918	0.7674
		8	2.0330	11.0000	3.1890	233.0795	0.7773
9	0.01	5	1.8446	4.7930	3.0650	244.6230	0.7882
		6	1.7921	11.0000	3.1461	242.3760	0.8050
		7	1.8111	5.0682	3.1353	244.7024	0.7975
		8	1.7732	11.0000	3.1474	247.3239	0.8134
	0.1	5	1.9734	7.2664	3.1578	244.4532	0.8014
		6	1.9525	11.0000	3.2258	243.5074	0.8085
		7	1.9840	5.4414	2.9882	246.9681	0.7970
		8	1.9411	6.7474	3.2668	250.0373	0.8133

In the PSO Algorithm, for each candidate solution satisfying the stability condition, the steady-state probabilities, the system performance measures, and the corresponding expected total cost were computed. This procedure ensures that all the values of the optimal parameters reported in Tables 3 and Table 4 correspond to stable configurations

of the system, as also confirmed by the system intensity values ($\rho < 1$) reported in the last column of these tables.

From Tables 3 and 4, it can be concluded that the cost is minimized when the system intensity rate is about 0.80. As the capacity K increased from 5 to 9, the system utilization also increased. Moreover, as can also be seen from the table, another important factor that affects system utilization is α . For higher values of α , we have a higher intensity rate. From Table 3, for $K = 9$ and $\alpha = 0.01$, optimal S is found at 6. When λ increases to 1.4, from Table 4, the optimal value of S is obtained again as 6 for both values of $\alpha = 0.01$ and $\alpha = 0.1$ when $K = 9$.

5. Conclusion

This study examines a queueing-inventory system in which both the production and inspection processes are imperfect. Two finite-capacity depots are used to store the items: produced items awaiting inspection are placed in Depot-1 of capacity S , while items identified as non-defective during the inspection process are transferred to Depot-2 of capacity K . The production process is governed by an (s, S) -policy, which links the replenishment decision directly to the inventory level in Depot-1. The proposed system is represented as a QBD process and analyzed using the geometric matrix method. The stability condition is established theoretically, and several key performance measures are derived. An extensive numerical study is then conducted to examine the effects of the system parameters on these performance measures, followed by a cost optimization carried out by means of the PSO algorithm.

The numerical results reveal that the probability of defective production α has a considerable impact on all performance measures, and this impact becomes more pronounced when the intensity rate of the system is high. The capacity of Depot-2 (K) emerges as another influential parameter, particularly for the mean number of customers in the system and the mean number of items in Depot-2. Regarding cost optimization, both K and α have been found to substantially affect the optimal value of the total cost. These findings indicate that achieving the minimum cost requires that the production rate, the inspection rate and the service rate be adjusted jointly, according to the values of α and K . These results can also be related to those reported in the related literature. The independence between the inspection system and the queueing-inventory system established in [2] under continuous production no longer holds once production is governed by an (s, S) -policy, which constitutes one of the main structural differences between the two models. Moreover, while both studies identify the probability of defects as a parameter with a substantial effect on system performance, the finite depot capacities considered here introduce an additional dimension. The capacity of Depot-2 is shown to influence both the mean number of customers in the system and the optimal total cost, an effect that cannot arise under the unlimited capacity assumption.

This study introduces a queueing-inventory system that incorporates imperfect production and imperfect inspection under finite storage capacities and a production control policy. Given that the occurrence of defective items is an inherent aspect of real-world production systems, the proposed model is expected to have broad applicability in practice. The comprehensive theoretical analysis, supported by the numerical results, offers valuable insights into the efficient operation of such systems. It should be noted that this study is subject to some limitations, which also provide directions for further research. While necessary for the Continuous Time Markov Chain (CTMC) formulation, the assumed exponential distributions for inter-arrival, service, production, and inspection times may not capture the variability observed in production environments where these times are less or more variable than the exponential case. Similarly, the probabilities of defects and misclassification are treated as constant across the planning horizon and therefore may not reflect

potential variations due to equipment breakdown or operator experience. The model also assumes a single server and a single inspector, both of which are perfectly reliable. Items identified as defective are also assumed to be discarded rather than reworked. Relaxing these assumptions could set many goals for future work. The model could be extended by considering phase-type distributions for service, production, or inspection times, by adding multiple servers or inspectors, or by allowing for server and inspector failures. Applying the model to a real production or sorting system using parameter estimates derived from field data represents a valuable aspect for future research.

Acknowledgements

The authors would like to thank the Editor and the anonymous reviewers for their valuable comments and constructive suggestions, which have substantially improved the quality and presentation of this paper.

Author contributions. All the co-authors have contributed equally in all aspects of the preparation of this submission.

Conflict of interest statement. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding. The authors received no funding for this research.

Data availability. No data was used for the research described in the article.

References

- [1] A. Aghsami, Y. Samimi and A. Aghaei, *A combined continuous-time Markov chain and queueing-inventory model for a blood transfusion network considering ABO/Rh substitution priority and unreliable screening laboratory*, Expert Systems with Applications **215**, 119360, 2023.
- [2] A. Aghsami, Y. Samimi and A. Aghaei, *A novel Markovian queueing-inventory model with imperfect production and inspection processes: A hospital case study*, Computers & Industrial Engineering **162**, 2021.
- [3] A. Aghsami, Y. Samimi and A. Aghaei, *An integrated Markovian queueing-inventory model in a single retailer-single supplier problem with imperfect quality and destructive testing acceptance sampling*, Advances in Industrial Engineering **55**(4), 367-401, 2021.
- [4] A. Ait El Cadi, A. Gharbi, K. Dhouib and A. Artiba, *Joint production, maintenance, and quality control in manufacturing systems with imperfect inspection*, Journal of Manufacturing Systems **77**, 848-858, 2024.
- [5] J. Asadkhani, H. Mokhtari and S. Tahmasebpour, *Optimal lot sizing under learning effect in inspection errors with different types of imperfect quality items*, Operational Research **22**, 2631-2665, 2022.
- [6] J.W. Baek and S.K. Moon, *The M/M/1 queue with a production-inventory system and lost sales*, Appl. Math. Comput. **233**, 534-544, 2014.
- [7] J.W. Baek and S.K. Moon, *A production-inventory system with a Markovian service queue and lost sales*, Journal of the Korean Statistical Society **45**, 14-24, 2016.
- [8] X. Cai and Y. Qin, *Profit maximization in a retrial queueing-inventory system: a hybrid algorithm*, Appl. Syst. Innov. **9**(87), 2026.
- [9] W.M. Chan, R.N. Ibrahim and P.B. Lochert, *A new EPQ model: integrating lower pricing, rework and reject situations*, Production Planning & Control **14**(7), 588-595, 2003.

- [10] N. Cheikhrouhou, B. Sarkar, B. Ganguly, A.I. Malik, R. Batista and Y.H. Lee, *Optimization of sample size and order size in an inventory model with quality inspection and return of defective items*, Ann. Oper. Res. **271**, 445-467, 2018.
- [11] B.K. Dey, S. Pareek, M. Tayyab and B. Sarkar, *Autonomation policy to control work-in-process inventory in a smart production system*, International Journal of Production Research **59**(4), 1258-1280, 2021.
- [12] B.K. Dey and H. Seok, *Intelligent inventory management with autonomation and service strategy*, J. Intell. Manuf. **35**, 307-330, 2024.
- [13] R.C. Eberhart and Y. Shi, *Comparing inertia weights and constriction factors in particle swarm optimization*, Proceedings of the 2000 Congress on Evolutionary Computation. CEC00 (Cat. No.00TH8512), La Jolla, CA, USA, 84-88, 2000.
- [14] A.P. Engelbrecht, *Fundamentals of Computational Swarm Intelligence*, Wiley, ISBN: 978-0-470-09191-3, 2006.
- [15] G. Genta, M. Galetto and F. Franceschini, *Inspection procedures in manufacturing processes: recent studies and research perspectives*, International Journal of Production Research **58**(15), 4767-4788, 2020.
- [16] J.T. Hsu and L.F. Hsu, *An EOQ model with imperfect quality items, inspection errors, shortage backordering, and sales returns*, Int. J. Production Economics **143**, 162-170, 2013a.
- [17] J.T. Hsu and L.F. Hsu, *Two EPQ models with imperfect production processes, inspection errors, planned backorders, and sales returns*, Computers & Industrial Engineering **64**, 389-402, 2013b.
- [18] M.Y. Jaber, S. Zanoni and L.E. Zavanella, *Economic order quantity models for imperfect items with buy and repair options*, Int. J. Production Economics **155**, 126-131, 2014.
- [19] K. Jeganathan, V.A. Koffer, K. Lakshmann, K. Loganathan, M. Abbas and A. Shilpa, *Replacement of failed items in a two commodity retrieval queueing-inventory system with multi-component demand and vacation interruption*, Heliyon **10**(2), e24024, 2024.
- [20] J. Kennedy and R. Eberhart, *Particle swarm optimization*, Proceedings of the IEEE International Conference on Neural Networks **4**, 1942-1948, 1995.
- [21] M. Khan, M.Y. Jaber and M. Bonney, *An economic order quantity (EOQ) for items with imperfect quality and inspection errors*, Int. J. Production Economics **133**, 113-118, 2011.
- [22] I. Konstantaras, S.K. Goyal and S. Papachristos, *Economic ordering policy for an item with imperfect quality subject to the in-house inspection*, International Journal of Systems Science **38**(6), 473-482, 2007.
- [23] N.A. Kurdhi, M. Marchamah and N.A. Respatiwan, *A two-echelon supply chain inventory model with shortage backlogging, inspection errors and uniform demand under imperfect quality items*, Int. J. Procurement Management **11**(2), 135-152, 2018.
- [24] A.K. Manna, J.K. Dey and S.K. Mondal, *Imperfect production inventory model with production rate dependent defective rate and advertisement dependent demand*, Computers & Industrial Engineering **104**, 9-22, 2017.
- [25] P. Mehta, M. Jain and S. Dhibar, *Retrial inventory-queueing model with inspection processes and imperfect production*, In: Sharma D.K., Jain M. (eds) Data analytics and artificial intelligence for inventory and supply chain management. Inventory optimization. Springer, Singapore, 2022.
- [26] A.Z. Melikov, R.R. Mirzayev and S.S. Nair, *Numerical Study of a Queueing-Inventory System with Two Supply Sources and Destructive Customers*, J. Comput. Syst. Sci. Int. **61**, 581-598, 2022.
- [27] A. Melikov and A.A. Molchanov, *Stock optimization in transport/storage*, Cybernetics and Systems Analysis **3**, 179-182, 1992.

- [28] L. Moussawi-Haidar, M. Salameh and W. Nasr, *Production lot sizing with quality screening and rework*, Appl. Math. Model **40**, 3242-3256, 2016.
- [29] L. Moussawi-Haidar, M. Salameh and W. Nasr, *Effect of deterioration on the instantaneous replenishment model with imperfect quality items*, Appl. Math. Model **38**, 5956-5966, 2014.
- [30] L. Moussawi-Haidar, M. Salameh and W. Nasr, *An instantaneous replenishment model under the effect of a sampling policy for defective items*, Appl. Math. Model **37**(3), 719-727, 2013.
- [31] S. Namnabat, H. Mokhtari and A. Fallahi, *Extended inventory planning models for repairable imperfect quality items under withdrawal policies and inspection errors*, J. Ind. Prod. Eng. **42**(6), 625-650, 2025.
- [32] M.F. Neuts, *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach*, Baltimore, MD: The Johns Hopkins University Press [1994 version is Dover Edition], 1981.
- [33] S. Ozkar, *Two-commodity queueing-inventory system with phase-type distribution of service times*, Ann. Oper. Res. **331**, 711-737, 2023.
- [34] S. Ozkar and U. Uzunoglu Kocer, *Two-commodity queueing-inventory system with two classes of customers*, Opsearch **58**, 234-256, 2021.
- [35] S. Pal and G.S. Mahapatra, *A manufacturing-oriented supply chain model for imperfect quality with inspection errors, stochastic demand under rework and shortages*, Computers & Industrial Engineering **106**, 299-314, 2017.
- [36] N. Rizky, I.D. Wangsa, W.A. Jauhari and H.M. Wee, *Managing a sustainable integrated inventory model for imperfect production process with type one and type two errors*, Clean Techn. Environ. Policy **23**, 2697-2712, 2021.
- [37] M. Saffari, A. Asmussen and R. Haji, *The M/M/1 queue with inventory, lost sale, and general lead times*, Queueing Syst. **75**, 65-77, 2013.
- [38] S. Saha and T. Chakrabarti, *A supply chain model under return policy considering refurbishment, learning effect and inspection error*, Croatian Operational Research Review **11**, 53-66, 2020.
- [39] M.K. Salameh and M.Y. Jaber, *Economic production quantity model for items with imperfect quality*, Int. J. Production Economics **64**, 59-64, 2000.
- [40] B. Sarkar and S. Saren, *Product inspection policy for an imperfect production system with inspection errors and warranty cost*, Eur. J. Oper. Res. **248**, 263-271, 2016.
- [41] B.K. Sett, B.K. Dey and B. Sarkar, *Automated inspection policy for smart factory: An improved approach*, Mathematics **8**, 1815, 2020.
- [42] Y. Shi and R.C. Eberhart, *Parameter selection in particle swarm optimization* In: Porto V.W., Saravanan N., Waagen D., Eiben A.E. (eds) Evolutionary Programming VII. EP 1998. Lecture Notes in Computer Science, 1447. Springer, Berlin, Heidelberg, 1998.
- [43] K. Sigman and D. Simchi-Levi, *Light Traffic Heuristic for an M/G/1 queue with limited inventory*, Ann. Oper. Res. **40**, 371-380, 1992.
- [44] U. Uzunoglu Kocer and S. Ozkar, *A production queueing-inventory system with two-customer and a server subject to breakdown*, Ann. Oper. Res. **331**, 1089-1117, 2023.
- [45] V. Vinitha, S. Amutha, G. P. Joshi, M. Kameswari, and N. Anbazhagan, *Integrated branching queueing-inventory system with diverted service*, Opsearch **63**, 754-785, 2026.
- [46] C.H. Wu, Y.Y. Dai and D.Y. Yang, *Reliability and cost-effectiveness optimization of G-systems with preventive maintenance strategy and unreliable repairman*, Appl. Math. Model **150**, 116387, 2026.
- [47] C.H. Wu, W.C. Lee, J.C. Ke and T.H. Liu, *Optimization analysis of an unreliable multi-server queue with a controllable repair policy*, Computers and Operations Research **49**, 83-96, 2014.

- [48] S. Yadav, S. Pareek, M. Sarkar, J. Ma and Y. Ahn, *Effect of in-process inspection on highly imperfect production system considering environmental deliberations*, Mathematics **13**, 1074, 2025.
- [49] D.Y. Yang and C.H. Wu, *Cost-minimization analysis of a working vacation queue with N-policy and server breakdowns*, Computers and Industrial Engineering **82**, 151-158, 2015.
- [50] M. Yua, Y. Tanga and Y. Weib, *Production fluid-type inventory model associated with a single vacation queue arising from additive manufacturing in the food sector*, Appl. Math. Model **128**, 242-259, 2024.