

A joint replenishment production-inventory model as an MMAP[K]/PH[K]/1 queue

Ann M. Noblesse^a, Nikki Sonenberg^b, Robert N. Boute^{c,d}, Marc R. Lambrecht^c,
Benny Van Houdt^e

^aDeloitte Consulting, Belgium, ^b Heilbronn Institute for Mathematical Research, University of Bristol, UK, ^cResearch Center for Operations Management, KU Leuven, Belgium,
^dTechnology and Operations Management Area, Vlerick Business School, Belgium,
^eDepartment of Mathematics and Computer Science, University of Antwerp, Belgium.

ARTICLE HISTORY

Compiled January 25, 2022

ABSTRACT

In this paper we analyse a continuous review finite capacity production-inventory system with two products in inventory. With stochastic order quantities and time between orders, the model reflects a supply chain that operates in an environment with high levels of volatility. The inventory is replenished using an independent order-up-to (s, S) policy or a can-order (s, c, S) joint replenishment policy in which the endogenously determined lead times drive the parameters of the replenishment policy. The production facility is modelled as a multi-type MMAP[K]/PH[K]/1 queue in which there are K possible inventory positions when the order is placed and the age process of the busy queue has matrix-exponential distribution. We characterize the system and determine the steady state distribution using matrix analytic methods. Using numerical methods we obtain the inventory parameters that minimize the total ordering and inventory related costs. We present numerical comparisons of independent and joint replenishment policies with varying lead times, order quantities, and cost reductions. We further demonstrate the interplay between the two products in terms of lead times, order quantities and costs.

KEYWORDS

production-inventory; can-order policy; fluid queues; multi-type queues;

1. Introduction

Inventory management has been studied for decades, but volatile market conditions have increased the complexity of modelling and analysis, as supply chain factors such as production rate, demand rate, and lead time become more variable. While much of the research in inventory management deals with single product analysis, companies need to make decisions on many aspects of inventory management, and the replenishment of multiple items is a common operational concern^[6].

We study a continuous review production-inventory system with a single production line, two products in inventory, stochastic order quantities, stochastic time between orders, and where the objective is to determine the parameters of the inventory policy

for each product that minimizes the total ordering and inventory related costs for both products.

We consider two policies of replenishing each product's inventory: an *order-up-to* (s, S) policy in which each product is replenished independently (Arrow *et al.*^[1]); and a *can-order* joint replenishment (s, c, S) policy (Balintfy^[5]), where if the inventory position of one product reaches its reorder point s or below, all other products whose inventory position is below their can-order level c , are included in the same replenishment order and any order placed raises the inventory position of that product to the order-up-to level S . Considerable savings may be realised by such coordinated replenishments (Bastos *et al.*^[8]) and an important factor making this approach cost effective is the joint set-up cost structure. However, the interaction between items makes determination of optimal parameters a difficult task.

The *lead time* is the time between the moment an order is placed and the moment it is received in inventory and thus consists of a waiting time in queue (if the system is busy), a setup time and a production time. The lead time depends on the way orders are placed and the production process, and influences inventory levels and inventory related costs.

Much of the inventory literature treats lead times as a fixed constant or as an *exogenous* variable with a given probability distribution, which means that the time required to deliver an order is assumed to be independent of the size of the current order and independent of the lead time of previous orders (Kulkarni and Yan^[22], Zipkin^[40]). This approach is justified when both production and inventory are decoupled through a large inventory at the production; if the owner of the production system guarantees a fixed delivery date; or if transportation lead times are much longer than production lead times, as in these scenarios the inventory policy does not impact significantly on lead times.

When multiple products are produced on the same production line, the order process of each individual product loading the production queue, impacts the lead times for all products ordered and all inventory levels are affected. Ignoring this impact may lead to suboptimal inventory parameters and higher costs. For example, in the context of a fixed order period (r, S) policy, it was shown that relying on exogenous lead times can result in a severe underestimation of the required inventory on hand (Van Houdt and Pérez^[39]).

In an integrated production-inventory setting, the interaction between the inventory control system and the production system is explicitly modelled: inventory influences production by initiating orders, and production influences inventory by completing and delivering orders to inventory. Accordingly, the lead times are *endogenously* determined by the production system: as the orders of each product load the production line, they determine the lead times, hence influence the time to replenish each product in inventory, and impact the holding and backlogging costs. Thus there is correlation between demand and lead times: high demand depletes inventory more, creating a larger order to replenish inventory, which in turn takes longer to produce. In order to characterize the retailer's inventory behavior in a production-inventory system in an exact way, we include this correlation in our analysis.

Much queueing theory literature assumes the arrival processes of customers are independent processes. The introduction of the Markov arrival process with marked transitions, MMAP $[K]$ (He^[17]), allowed modelling of complicated queues with multiple types of customers and correlated arrival times (Artalejo^[2]). Further, the application of standard matrix analytic methods has allowed the development of efficient algorithms for computing relevant performance measures (for example, He^[15] and Horvath^[18]).

In this paper, the production facility is modelled as a MMAP[K]/PH[K]/1 queue (He^[16], Van Houdt and Blondia^[38]). Queues of this type have a single server, correlated interarrival times and K types of customers, where each customer type may have a different phase-type service requirement. In our case, the customer types correspond to the possible inventory positions when the order is placed.

The state space for the process records information on the type of order, the inventory position when the order was placed, and the current state of production or setup. The *age process* that observes this queue during the busy period keeps track of the age of the order in production, the current production phase and has matrix-exponential distribution (Sengupta^[31,32]). To enable efficient computation of the inventory level distribution, we construct a fluid queue to enable computation of required matrices (Dziel *et al.*^[13]) and from these we can determine the cost-minimizing order quantity of the replenishment orders.

In our model the setup, change-over and production times are stochastic and follow a phase-type distribution^[24]. The advantage of a phase-type distribution is that its Markovian nature allows for an exact analysis and performance evaluation.

The contribution of this work is to construct an exact evaluation for the can-order joint replenishment policy while taking endogenous lead times into account, and hence to study the complex interplay between mutually dependent product inventory levels. We provide illustrations of the interaction between the inventory properties for the two product case, in terms of lead times, order quantities, and cost reductions from a can-order joint replenishment policy as compared with an independent order-up-to policy. The analysis is applicable to more products but leads to state space explosion.

The structure of this paper is as follows: in Section 3 we define the production-inventory model and then, in Section 4, define the associated Markov process characterizing the production-inventory system. From the matrix exponential form of the steady state distribution, we construct the associated fluid process in Section 5. In Section 6 we present results required for the computation of the expected costs, outline the numerical method, and present some individual and joint replenishment outcomes, illustrating the impact of different parameter values. Concluding remarks are given in Section 7.

2. Background

Inventory analyses are generally aimed at minimizing ordering and holding costs while satisfying demand and there is a vast literature addressing differing assumptions on how customer orders and production systems are managed. We mention just a few variations. For example, inventory control policies assuming independent and identically distributed inter-event times have received most attention, but a recent study of production policies to minimise expected holding and backlog costs, while considering *correlated* inter-arrival and processing times, showed that effective production control policies should take correlations in service and demand into account (Dizbin^[12]). We use a model that assumes customer orders are held until the item is produced. Much research seeks ordering policies to maintain the inventory position at a constant level, but a recent analysis illustrates opportunities for performance improvement by using the current inventory level to set a dynamic target for inventory-in-transit, and to place orders that follow this target (Stolyar^[36]). Other work considers situations where a customer may have a ready alternative source and sales are lost (Faaland^[14]), and situations where the supplier has access to an emergency supply

to raise the inventory level (Barron^[7]).

In the context of an inventory managing multiple items, the joint replenishment problem deals with the prospect of saving resources through coordinated replenishments, by considering interdependencies amongst product orders involving a single supplier. Studies of joint replenishment have led to many models, solution methods and applications^[8]. As the complexity of many variations of the joint replenishment problem has been determined to be NP-hard^[11], studies focus on approximate and heuristic methods.

Replenishment models cover both deterministic demand types, for examples see^[8], and stochastic demand types, for example see^[6,21,22]. With the stochastic demand models, there are two main replenishment policies: (i) the periodic replenishment policy that considers a Poisson distributed demand, in which all products follow the same replenishment periodic interval, and each one is replaced to a pre-determined inventory level^[4]. (ii) the continuous review can-order policy, in which if any item reaches a must-order point, a replenishment order is made, and if the other items have inventory levels below a can-order point, they are included in the same order^[5,20,30]. On stochastic demand models based on the can-order policy, solution methods have focused mostly on determining optimal values for re-order and can-order parameters in relation to the selected policy^[8].

Markov processes are valuable as approximate processes, due to their versatility in matching key statistical properties of the demand process^[10,26]. Phase-type distributions^[27] are often used, and in which their matrix structure supports efficient computational implementation^[7]. When general distributions are appropriate for modeling the demand, phase-type distributions can be taken into account in a natural way, as any non-negative continuous distribution of the probability can be approximated with a phase-type distribution^[3].

Prior studies of models incorporating endogeneous lead times include: early work on assemble-to-order production-inventory systems with stochastic lead times, in which one must allow for correlated demands by jointly managing inventories and production capacities across several items – that work used matrix geometric methods for exact analysis to evaluate order-fulfilment performance measures (Song, Hu, Liu^[34]); performance analysis of a single product production-inventory system with periodic demands (Boute *et al.*^[9]); and more recent work studying a single product dual-source inventory system where the lead times at both sources are stochastic and endogeneous, namely sojourn times in a queueing system – where the authors establish a new approach to finding an optimal policy (Song *et al.*^[35]), in part drawing on results on endogeneous stochastic supply networks (Song and Zipkin^[33]).

A key element influencing the complexity and usefulness of any Markov model is the choice of the state space and transition function. In our model, the state representation records information on the type of order, the inventory position when the order was placed, and the current state of production or setup, and the analysis takes advantage of matrix analytic and fluid queue results applicable to the associated age process. An approach similar to ours was used by Noblesse *et al.*^[28] for the single item problem, but the analysis in the current paper allows for systems with batch demand arrivals, variable order quantities, and non-exponential production times. Related approaches in less complex settings than that studied here have been used by Horvath and Van Houdt^[19], Otten *et al.*^[29], Van Houdt^[37].

3. Model description

Consider a production-inventory system with $j \in \{1, 2\}$ products, where demands arrive according to a compound Poisson process with rate $\lambda^{(j)}$. The sizes of the batch demands are i.i.d. and follow a general discrete distribution with maximum demand size m_j . Let $d_x^{(j)}$ denote the probability of a demand of size x for product j . If inventory is insufficient to fulfill demand, unmet demand is backlogged.

The inventory of product j is replenished according to a (s_j, c_j, S_j) policy, where $c_j \geq s_j$: if the inventory position of product j reaches its reorder point s_j , product $\ell \neq j$ will also place an order if its inventory position is at or below the can-order level c_ℓ , otherwise it will not join the order. Setting $c_j = s_j$, the (s_j, c_j, S_j) policy reduces to an (s_j, S_j) policy, therefore our analysis holds for both joint and independent replenishment policies. The maximum order size is $o^{(j)} = S_j - s_j + m_j - 1$.

Orders are sent to the production queue and produced on a first-come-first-served basis by a single machine which produces the units of each order sequentially. Each order requires a *major setup time* and a *production time per unit*. If a joint order is placed, only one major setup time is required and an additional *minor change-over time*. Typically, the minor change-over time will be smaller than the major setup time, although the model does not require this assumption.

When both the setup is complete and the last unit of the order is produced, the order is replenished in inventory. If a joint order is placed, the order is only delivered in inventory when the production for both products is completed. We assume the unit production time is identical for both products and follows phase-type distribution $PH(\gamma_p, U_p)$, that is, with subgenerator matrix U_p and initial probability vector γ_p of order n_p . The major setup time and minor change-over time are assumed to both follow phase-type distributions with $PH(\gamma_+, V_+)$ of order n_+ , and $PH(\gamma_-, V_-)$ of order n_- , respectively.

Denote by η_j the expected number of orders placed per unit time of only product j and η_j^c the expected number of joint orders per unit time initiated by product j . An inventory holding cost h_j is incurred per unit per time and Φ_j^+ is the expected number of units in inventory of product j at a random point in time. A backlogging cost p_j is incurred per unit per time and Φ_j^- denotes the expected number of units of product j backlogged at a random point in time. The expected total cost per unit of time for both products is

$$E[C] = \sum_{j=1}^2 \left((\tilde{k} + k_j)(\eta_j + \eta_j^c) + k_j \eta_{3-j}^c + h_j \Phi_j^+ + p_j \Phi_j^- \right), \quad (1)$$

where a major fixed order cost \tilde{k} is incurred per order placed and a minor fixed order cost k_j if product j is included in the order. For ease of reference, a summary of key notation is provided in Table 1.

The objective is to determine the inventory parameters (s_j, c_j, S_j) , $j \in \{1, 2\}$, that minimize $E[C]$. We next describe the Markov process of the underlying model, and key theoretical results are set out in the following two sections, enabling the iterative procedure used for the numerical solutions discussed in Section 6.3.

4. Markov process of the two product system

Below we describe the state representation and transitions (Section 4.1). For the busy process we use a state space construct similar to that in He^[15] and Noblesse *et al.*^[28]. Then we present the transition rate and density matrices (Sections 4.2 and 4.3), and introduce the product utilisation rate, ρ_p (Section 4.4) that is used in the inventory level distribution that is derived in Section 5.3.

4.1. Process description

When the server is idle, we use as state space $\mathcal{I} = \{(i, j) : s_1 + 1 \leq i \leq S_1, s_2 + 1 \leq j \leq S_2\}$ where each element of \mathcal{I} characterizes an inventory position. Clearly \mathcal{I} has dimension $(S_1 - s_1)(S_2 - s_2)$. A Markov process $(I_t : t \in \mathbb{R}^+)$ then records the inventory depletions due to arrivals of demands.

When the server is busy, we represent the system using the Markov process $(X_t, L_t : t \in \mathbb{R}^+)$. The first component $X_t \in \mathbb{R}^+$ represents the *age*, i.e., the elapsed time since the order in service was placed, and the second component $L_t \in \mathcal{P}$ represents the state of the order in service at time t . This busy process describes the transitions in the MMAP[K]/PH[K]/1 queue that models the production facility (He^[17]).

For an order in production, the state space \mathcal{P} is defined as $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$, where the states in \mathcal{P}_j describe the order when initiated by product j , i.e., the product that reached its reorder point, and we set $\mathcal{P}_j = \mathcal{P}_{j,indiv} \cup \mathcal{P}_{j,joint}$, partitioned according to whether the order in production is an individual or joint order.

For an order initiated by product 1, define

$$\mathcal{P}_{1,indiv} = \{(i_1, i_2, z) \mid s_1 - m_1 + 1 \leq i_1 \leq s_1, c_2 + 1 \leq i_2 \leq S_2, 1 \leq z \leq z^{1,indiv}\},$$

where i_j represents the inventory position of product j when the order in production at time t was placed, i.e., at time $t - X_t$, and the third component, z , represents the progress of the production: i.e., it is either in major setup (in a state in $z \in \{1, \dots, n_+\}$), or in production with $r - 1$ remaining items to be produced to complete the order and the server is currently in state $k \in \{1, \dots, n_p\}$. So $z = n_+ + (r - 1)n_p + k$. Accordingly, $z^{1,indiv} = n_+ + o^{(1)}n_p$.

Define

$$\mathcal{P}_{1,joint} = \{(i_1, z) \mid s_1 - m_1 + 1 \leq i_1 \leq s_1, 1 \leq z \leq z^{1,joint}\},$$

where $z^{1,joint} = n_- + n_+ + (o^{(1)} + S_2 - s_2 - 1)n_p$.

In contrast to $\mathcal{P}_{1,indiv}$, here we do not need to keep track of the inventory position i_2 as it will increase to S_2 irrespective of the value of i_2 . We note that the range of z is now larger as the order also involves at most $S_2 - s_2 - 1$ items of product 2 and the server can also be in one of the states of the minor change-over time. If $z \in \{1, \dots, n_-\}$, the server is in state z of the minor change-over time, if $z \in \{n_- + 1, \dots, n_- + n_+\}$ a major setup is ongoing and the remaining z values correspond to the production of an item.

When the inventory position of product 1, exceeds c_1 , the order only includes product 2, while for inventory positions up to c_1 , the order is joint with product

1. So, for an order initiated by product 2, define

$$\mathcal{P}_{2,indiv} = \left\{ (i_1, z) \mid c_1 + 1 \leq i_1 \leq S_1, 1 \leq z \leq z^{2,indiv} \right\},$$

where $z^{2,indiv} = n_+ + o^{(2)}n_p$ and

$$\mathcal{P}_{2,joint} = \left\{ (i_1, z) \mid s_1 + 1 \leq i_1 \leq c_1, 1 \leq z \leq z^{2,joint} \right\},$$

where $z^{2,joint} = n_- + n_+ + (o^{(2)} + S_1 - s_1 - 1)n_p$.

We have defined the model from the perspective of product 1 and thus the sets $\mathcal{P}_{1,indiv}$ and $\mathcal{P}_{2,indiv}$ are structurally different. This representation provides sufficient information to derive the inventory distributions, but enables smaller matrix calculations than if we retained the same structure for $\mathcal{P}_{2,indiv}$ as $\mathcal{P}_{1,indiv}$.

We also note that to obtain a Markov process it is not necessary to keep track of the inventory position, i_1 of product 1 (as the inventory position increases to S_1 irrespective of the value of i_1), but we have included this in the representation as it can then be used in deriving the inventory level distribution of product 1 in Section 5.3, which in turn is used in our iterative procedure to compute the expected cost in Section 6.3.

The process evolves as follows: as the production of an order continues, X_t increases linearly over time and a downward jump in X_t occurs when an order completes service. Three types of transitions can occur from state $(X_t, L_t) = (x, i)$:

- A transition to state (x, j) for $j \neq i \in \mathcal{P}$ when the state of production/ change-over/setup changes but the same order remains in service. Denote by $(A_0)_{i,j}$ as the rate of this transition.
- A transition to a state (y, j) with $y \in [x - u, x)$, for $0 < u < x$ when an order completes service and there is a subsequent order in the queue enters service. When the inter-arrival time of the subsequent order is (at most) u , time units, the order has waited at least $x - u$ time units, illustrated in Figure 1. Denote by $A_{i,j}(u)$ the transition rate, with $dA_{i,j}(u)$ its corresponding density function. The matrix $A(u)$ of rates $A_{i,j}(u)$ takes into account the correlation between j (which represents the new order size) and u (the inter-arrival time between the new order and the previous one).
- A transition to state $(0, j)$ when the order completes service and the queue is empty upon service completion. This occurs with rate $\int_x^\infty dA_{i,j}(u)$.

The matrices A_0 and $dA(u)$ are described in the next sections.

4.2. Transition rate matrix A_0

Define

$$A_0 = \begin{bmatrix} F_{++}^{(1)} & \mathbf{0} \\ \mathbf{0} & F_{++}^{(2)} \end{bmatrix}, \quad (2)$$

where $F_{++}^{(j)}$ is an order $|\mathcal{P}_j|$ matrix that contains the rates when the order was initiated by product j . Then

$$F_{++}^{(1)} = I_{m_1} \otimes \begin{bmatrix} F_{++}^{(1),indiv} & \mathbf{0} \\ \mathbf{0} & F_{++}^{(1),joint} \end{bmatrix}, \quad (3)$$

where I_{m_1} records the inventory position of product 1 the moment the order was placed, the matrices $F_{++}^{(1),joint}$ and $F_{++}^{(1),indiv}$ describe the joint and individual orders initiated by product 1, respectively, and \otimes denotes the Kronecker product. Then,

$$F_{++}^{(1),joint} = \begin{bmatrix} V_- & & & & 0 \\ v_+ \gamma_- & V_+ & & & \\ & u_p \gamma_+ & U_p & & \\ & & \ddots & \ddots & \\ 0 & & & u_p \gamma_p & U_p \end{bmatrix}, \quad (4)$$

is an order $z^{1,joint}$ matrix and describes the transitions in production and setup phases and where the vectors $u_p = -U_p e$ and $v_+ = -V_+ e$ denote the completion rates of the unit production and the major setup time, respectively. Define,

$$F_{++}^{(1),indiv} = I_{S_2 - c_2} \otimes \begin{bmatrix} V_+ & & & & 0 \\ u_p \gamma_+ & U_p & & & \\ & u_p \gamma_p & \ddots & & \\ & & \ddots & \ddots & \\ 0 & & & u_p \gamma_p & U_p \end{bmatrix}, \quad (5)$$

an order $(S_2 - c_2)z^{1,indiv}$ matrix. Since there is no minor setup time, we record the inventory position of product 2 with $I_{S_2 - c_2}$, the moment the order was placed. Then we obtain a zero matrix, except for the $S_2 - c_2$ sub-matrices on the diagonal where each of sub-matrix defines the transitions in production and setup and specifying i_2 . Similarly, define

$$F_{++}^{(2)} = \begin{bmatrix} F_{++}^{(2),indiv} & \mathbf{0} \\ \mathbf{0} & F_{++}^{(2),joint} \end{bmatrix}, \quad (6)$$

where

$$F_{++}^{(2),indiv} = I_{S_1 - c_1} \otimes \begin{bmatrix} V_+ & & & & 0 \\ u_p \gamma_+ & U_p & & & \\ & u_p \gamma_p & \ddots & & \\ & & \ddots & \ddots & \\ 0 & & & u_p \gamma_p & U_p \end{bmatrix}, \quad (7)$$

is of order $(S_1 - c_1)z^{2,indiv}$ and

$$F_{++}^{(2),joint} = I_{c_1 - s_1} \otimes \begin{bmatrix} V_- & & & & 0 \\ v_+ \gamma_- & V_+ & & & \\ & u_p \gamma_+ & U_p & & \\ & & \ddots & \ddots & \\ 0 & & & u_p \gamma_p & U_p \end{bmatrix}, \quad (8)$$

is of order $(c_1 - s_1)z^{2,joint}$. In both joint and individual orders, we keep track of the inventory position of product 1 at the moment the order was placed.

4.3. Density matrix $dA(u)$

The density function of the rate matrix $A(u)$ can be expressed as

$$dA(u) = F_{+-} e^{F_{--} u} F_{-+}, \quad (9)$$

where F_{+-} describes the transitions from the states in \mathcal{P} prior to service completion to the inventory positions of both products immediately after the order was placed. The matrix exponential $e^{F_{--} u}$ gives the density of the inter-arrival time of u , where F_{--} describes the evolution of the inventory position until the subsequent order is placed. The matrix F_{-+} describes the transitions to the initial production state of the subsequent order beginning at $t - X_t + u$.

4.3.1. Production completion of the order in service (F_{+-})

Denote by α_j the vector of the steady state probabilities of the inventory position immediately after a replenishment, then $\alpha_j = (1, 0, \dots, 0)$ with length $(S_j - s_j)$. Let

$$F_{+-} = \begin{bmatrix} F_{+-}^{(1)} \\ F_{+-}^{(2)} \end{bmatrix}, \quad (10)$$

where $F_{+-}^{(j)}$ describes the transitions when the order was initiated by j . Recalling that completion occurs after a major setup time in an individual order, or after a minor change over time of a joint order, we have

$$F_{+-}^{(1)} = e_{m_1} \otimes \begin{bmatrix} \alpha_1 \otimes \left(I_{S_2 - c_2} \otimes \nu_+^{(1)} \right) \\ \nu_-^{(1)} (\alpha_1 \otimes \alpha_2) \end{bmatrix}, \quad (11)$$

with dimensions $|\mathcal{P}| \times (S_1 - s_1)(S_2 - s_2)$ and where $\nu_+^{(j)} = (v_+^\top, 0, \dots, 0)^\top$ is a vector of length $z^{j,indiv}$ and $\nu_-^{(j)} = (v_-^\top, 0, \dots, 0)^\top$ a vector of length $z^{j,joint}$, where v_+^\top denotes the transpose of v_+ . The vector e_{m_1} records the inventory level of product 1 during production as a new order is about to begin. The upper sub-matrix in Equation (11) describes an individual order of product 1, then the inventory position of product 1 is raised to S_1 and we record the inventory position of product 2 with $I_{S_2 - c_2}$. The lower sub-matrix describes a joint order, in which case the inventory positions of both

products are raised to their respective S_j . Similarly,

$$F_{+-}^{(2)} = \begin{bmatrix} I_{S_1-c_1} \otimes \left(\nu_+^{(2)} \alpha_2 \right) \\ \left(e_{c_1-s_1} \otimes \nu_-^{(2)} \right) (\alpha_1 \otimes \alpha_2) \end{bmatrix}, \quad (12)$$

where $I_{S_1-c_1}$ is used to keep track of the inventory position of product 1 when only product 2 is replenished, and $e_{c_1-s_1}$ records the inventory position of product 1 in a joint order.

4.3.2. Inventory position until subsequent order (F_{--})

The matrix F_{--} is of order $(S_1 - s_1)(S_2 - s_2)$ and describes the transition rates in the inventory positions of both products due to demand arrivals until the reorder point of one of the products is reached. Then

$$F_{--} = \begin{bmatrix} -\lambda^{(1)} & \lambda^{(1)} d_1^{(1)} & \dots & \lambda^{(1)} d_{S_1-s_1-1}^{(1)} \\ 0 & -\lambda^{(1)} & \dots & \lambda^{(1)} d_{S_1-s_1-2}^{(1)} \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & -\lambda^{(1)} \end{bmatrix} \otimes I_{S_2-s_2} \\ + I_{S_1-s_1} \otimes \begin{bmatrix} -\lambda^{(2)} & \lambda^{(2)} d_1^{(2)} & \dots & \lambda^{(2)} d_{S_2-s_2-1}^{(2)} \\ 0 & -\lambda^{(2)} & \dots & \lambda^{(2)} d_{S_2-s_2-2}^{(2)} \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & -\lambda^{(2)} \end{bmatrix}. \quad (13)$$

The time u between subsequent orders is then equivalent to the time it takes for the process to decrease from the inventory positions just after an order was placed until absorption in one of the reorder points (or below), thus its density is given by $e^{F_{--}u}$.

4.3.3. Start of production of the subsequent order (F_{-+})

Let $F_{-+} = [F_{-+}^{(1)} \ F_{-+}^{(2)}]$ where $F_{-+}^{(j)}$ records the transitions when the order was initiated by product j and which has dimensions $(S_1 - s_1)(S_2 - s_2) \times |\mathcal{P}_j|$. Then

$$F_{-+}^{(1)} = \sum_{k=0}^{m_1-1} \left(\eta_k^{(1)} \otimes I_{S_2-s_2} \right) \Gamma_k^{(1)}, \quad (14)$$

where

$$\eta_k^{(j)} = \left[\lambda^{(j)} d_{S_j-s_j+k}^{(j)} \quad \lambda^{(j)} d_{S_j-s_j+k-1}^{(j)} \quad \dots \quad \lambda^{(j)} d_{k+1}^{(j)} \quad 0 \quad \dots \quad 0 \right]^\top, \quad (15)$$

is a column vector $\eta_k^{(j)}$ of length $S_j - s_j$ and contains the rates of demand arrivals for product 1 which result in an order of size $S_1 - s_1 + k$. The matrix $\Gamma_k^{(1)}$ determines the state in which production starts if the order size of product 1 equals $S_1 - s_1 + k$ and if the inventory position of product 2 is $S_2 - i_2 + 1$ at the moment when product 1 initiates an order, for $i_2 = 1, \dots, S_2 - s_2$. If only product 1 is replenished, $\Gamma_k^{(1)}$ also records the

inventory position of product 2, i_2 . Thus $\Gamma_k^{(1)}$ is an $(S_2 - s_2) \times |\mathcal{P}_1|$ matrix with all its entries equal to zero, except for n_p entries given by γ_p . For $i_2 = 1, \dots, S_2 - c_2$ the inventory position of product 2 remains above c_2 and the n_p nonzero entries of row i_2 of $\Gamma_k^{(1)}$ are on the columns corresponding to the states $(k, S_2 - i_2 + 1, n_s^+ + z) \in \mathcal{P}_{1,indiv}$ with $z = (S_1 - s_1 + k - 1)n_p + 1$ to $(S_1 - s_1 + k)n_p$. For $i_2 = S_2 - c_2 + 1, \dots, S_2 - s_2$ we have a joint order and the n_p nonzero entries on row i_2 of $\Gamma_k^{(1)}$ are on the columns corresponding to the states $(k, n_s^- + n_s^+ + z)$ of $\mathcal{P}_{1,joint}$ with $z = (S_1 - s_1 + k + i_2 - 2)n_p + 1$ to $(S_1 - s_1 + k + i_2 - 1)n_p$ as the joint order has size $S_1 - s_1 + k + i_2 - 1$.

Similarly, $F_{-+}^{(2)}$ defines the transition rates when product 2 reaches its reorder point,

$$F_{-+}^{(2)} = \sum_{k=0}^{m_2-1} \left(I_{S_1-s_1} \otimes \eta_k^{(2)} \right) \Gamma_k^{(2)}, \quad (16)$$

with $\Gamma_k^{(2)}$ an $(S_1 - s_1) \times |\mathcal{P}_2|$ matrix with all its entries equal to zero, except that each row contains n_p entries equal to γ_p . More specifically, the non-zero entries on row $i_1 \in \{1, \dots, S_1 - c_1\}$ appear on the columns corresponding to the states $(i_1, n_+ + z) \in \mathcal{P}_{2,indiv}$ with $z = (S_2 - s_2 + k - 1)n_p + 1$ to $z = (S_2 - s_2 + k)n_p$. For $i_1 \in \{S_1 - c_1 + 1, \dots, S_1 - s_1\}$ the columns corresponding to the states $(i_1, n_- + n_+ + z)$ of $\mathcal{P}_{2,joint}$ with $z = (S_2 - s_2 + k + i_1 - 2)n_p + 1$ to $(S_2 - s_2 + k + i_1 - 1)n_p$ holds the entries of γ_p .

4.4. Utilisation of the production system

Define ρ_p as the *production utilisation rate*, that is, excluding the setup times, then

$$\rho_p = \sum_{j=1}^2 \left(-\lambda^{(j)} \gamma_p U_p^{-1} e \sum_{i=1}^{m_j} id_i^{(j)} \right), \quad (17)$$

where $\gamma_p (-U_p)^{-1} e$ is the mean unit production time and $\sum_{i=1}^{m_j} id_i^{(j)}$ is the mean demand size for product j . The *overall utilisation* ρ is then determined by the number of orders placed,

$$\rho = \rho_p - (\gamma_+ V_+^{-1} e + \gamma_- V_-^{-1} e) (\eta_1^c + \eta_2^c) - (\gamma_+ V_+^{-1} e) (\eta_1 + \eta_2), \quad (18)$$

where $-\gamma_+ V_+^{-1} e$ and $-\gamma_- V_-^{-1} e$ are the mean major setup and minor change-over times, respectively.

5. Steady state analysis

We consider the form of steady state of the Markov process in Section 5.1 and construct the associated fluid queue in Section 5.2. In Section 5.3 we obtain an expression for the inventory position and the number of demand arrivals since an order was placed and combine the results to determine the inventory level distribution.

5.1. Preliminaries

For $x > 0$, $j \in \mathcal{P}$, denote the steady state density of (X_t, L_t) as

$$\delta_j(x) = \lim_{t \rightarrow \infty} \frac{d}{dx} P[X_t < x, L_t = j], \quad (19)$$

and the corresponding vector $\delta(x) = \{\delta_j(x), j \in \mathcal{P}\}$. If $\rho < 1$ the process (X_t, L_t) has matrix exponential form (Sengupta^[31,32]) and there exists a matrix T of order $|\mathcal{P}|$ such that

$$\delta(x) = \delta(0)e^{Tx}, \quad (20)$$

where the matrix T is the smallest non-negative solution to

$$T = A_0 + \int_0^\infty e^{Tx} dA(x), \quad (21)$$

and $\delta(0) = \theta(-T)$, where θ is the stationary vector of

$$A = A_0 + \int_0^\infty dA(u). \quad (22)$$

See Van Houdt^[37] for details.

Using the matrices A_0 and $dA(x)$, the matrix T can be computed iteratively, $T_{n+1} = A_0 + \int_0^\infty e^{T_n x} dA(x)$, with $T_0 = 0$ (Sengupta^[31]). As this method results in linear convergence and is therefore impractical under high loads, in the next section we present a more efficient approach.

5.2. Reduction to a fluid queue

To employ a method for computing T that converges quadratically we construct a fluid queue (Dzial *et al.*^[13]). In order to obtain a fluid queue, the fluid process must be skip-free in both directions so we replace the immediate downward jumps in X_t by intervals of the appropriate length during which the level decreases linearly. Then there are $|\mathcal{P}|$ phases in which the fluid increases and $|\mathcal{I}|$ artificial phases in which the fluid decreases and the rate matrix of the underlying continuous-time Markov chain is

$$F = \begin{bmatrix} F_{++} & F_{+-} \\ F_{-+} & F_{--} \end{bmatrix}. \quad (23)$$

The matrix $F_{++} = A_0$ contains the rates at which the phase changes while the fluid increases (the same order remains in service), F_{+-} contains the transitions from an increase to a decrease in fluid (a service completion occurs), F_{-+} contains the transitions from a decrease to an increase in fluid (a new order enters service) and F_{--} contains the transition rates while the fluid decreases (demands arrive depleting the inventory, but no product has reached its order point).

When the fluid queue hits level zero the phase continues to evolve according to F_{--} until an event part of F_{-+} occurs and the fluid level becomes positive again.

Censoring the queue only on the periods where the level increases we obtain the Markov process of Section 4, in which the age always increases, unless there is a downward jump. Due to Equation (9), a downward jump to level zero occurs with rate $\int_x^\infty dA_{i,j}(u) = (F_{+-} e^{F_{--}x} (-F_{--})^{(-1)} F_{+-})_{i,j}$, where the part $(-F_{--})^{(-1)} F_{+-}$ captures the evolution of the phase of the fluid queue at zero.

If we take the expression for the steady state of a fluid queue (Latouche^[23]) and observe the queue only when the level is increasing, the steady state has the form given in Equation (20) with

$$T = F_{++} + \Psi F_{-+}, \quad (24)$$

and $\delta(0) = \theta(-T)$, where θ is the stationary vector of $A = A_0 + F_{+-} (-F_{--})^{-1} F_{-+}$ and matrix Ψ , the first return probabilities to the initial level, is the minimal non-negative solution to an algebraic Riccati equation

$$\Psi F_{-+} \Psi + F_{++} \Psi + \Psi F_{--} + F_{+-} = 0. \quad (25)$$

To reduce the computation time of θ , we employ the power method (Van Houdt and Perez^[39]). Further, as F_{++} has block diagonal form, we can apply the algorithm of Meini^[25], which exploits this structure to reduce the memory and time complexity.

Observing the phase process only during the intervals of time in which the fluid level is decreasing, that is, when the production facility is idle, the rate matrix is

$$D = F_{--} + F_{-+} \Psi. \quad (26)$$

from which we obtain the steady state probabilities of the idle production facility, $\hat{\theta}$, such that $\hat{\theta}D = 0$, $\hat{\theta}e = 1$.

5.3. Inventory level distribution of product 1

For $i \geq 0$, denote by ϕ_i the probability the inventory level is $S_1 - i$. To derive this distribution, we separately consider the situations in which the server is busy and where it is idle.

Firstly when the server is busy, define $q_{k,n}$ as the probability that the inventory position of product 1 was $S_1 - k$ at the moment when the order in production was placed and n demand arrivals of product 1 have depleted inventory since. The corresponding vector $q_n = \{q_{k,n}, k = 0, 1, \dots, o^{(1)}\}$ is partitioned such that

$$q_n = [q_{(0:S_1-c_1-1),n} \ q_{(S_1-c_1:S_1-s_1-1),n} \ q_{(S_1-s_1:o^{(1)}),n}], \quad (27)$$

where the entries refer to individual orders placed by product 2, joint orders initiated by product 2, and orders initiated by product 1 (joint or individual), respectively. During the time x this order has spent in the system, the probability that the inventory is further depleted by the arrival of n demands for product 1 is $(1/n!) (\lambda^{(1)} x)^n e^{-\lambda^{(1)} x}$. Let

$$p_n(x) = \int_0^\infty \delta(x) \frac{(\lambda^{(1)} x)^n}{n!} e^{-\lambda^{(1)} x} dx = \delta(0) \left(\lambda^{(1)}\right)^n \left(\lambda^{(1)} I_{|\mathcal{P}|} - T\right)^{-(n+1)}, \quad (28)$$

then $q_n = p_n(x) [(I_{m_1} \otimes e_{|\mathcal{P}_1|/m_1}), (I_{c_1-s_1} \otimes e_{z^{2,joint}}), (I_{S_1-c_1} \otimes e_{z^{2,indiv}})]$.

We can then write

$$\phi_i = (1 - \rho) g_i + \rho f_i, \quad (29)$$

where $1 - \rho$ is the probability of a idle server in which the inventory level of product 1 is at least $s_1 + 1$ and the idle probabilities g_i are given by

$$g_i = \left(\hat{\theta}(I_{S_1-s_1} \otimes e_{S_2-s_2}) \right)_{i+1} 1[0 \leq i \leq S_1 - 1], \quad (30)$$

where $\hat{\theta}$ is derived from the fluid queue construction in Equation (26) and where $1[A] = 1$ if A is true and $1[A] = 0$ otherwise.

When the server is busy, with probability ρ , ϕ_i is obtained by $q_{k,n}$ and the n -fold convolution of the demand size distribution,

$$f_i = \sum_{\substack{k \leq i, \\ n \leq i-k}} q_{k,n} \sum_{\substack{k_1, \dots, k_n > 0, \\ k_1 + \dots + k_n = i-k}} \prod_{s=1}^n d_{k_s}^{(1)}. \quad (31)$$

Then the expected number of units in inventory and expected number of units backlogged are

$$\Phi_1^+ = \sum_{i=0}^{S_1} \phi_i(S_1 - i), \quad \Phi_1^- = \sum_{i=S_1+1}^{\infty} \phi_i(i - S_1). \quad (32)$$

Denote by L the lead time and define $\zeta(x)$ as its steady state density,

$$\zeta(x) = \frac{\delta(x)\hat{\eta}}{\int_0^{\infty} \delta(y)\hat{\eta}dy} = \frac{\theta(-T)e^{Tx}\hat{\eta}}{\theta\hat{\eta}}, \quad (33)$$

where $\hat{\eta} = F_{+-}e$, and with corresponding distribution function

$$Z(x) = 1 - \frac{\theta e^{Tx}\hat{\eta}}{\theta\hat{\eta}}, \quad (34)$$

then the expected lead time of a random order is $E[L] = -\theta T^{-1}\hat{\eta}(\theta\hat{\eta})^{-1}$. Conditioning on the order if product 1 is included, the product dependent expected lead time is

$$E[L^{(j)}] = -\theta T^{-1}\hat{\eta}^{(j)}(\theta\hat{\eta}^{(j)})^{-1}, \quad (35)$$

where

$$\hat{\eta}^{(1)} = \left[\begin{array}{c} F_{+-}^{(1)} \\ I_{S_1-c_1} \otimes \left(\nu_+^{(2)} \alpha_2 \right) \\ \left(e_{c_1-s_1} \otimes \nu_-^{(2)} \right) (\alpha_1 \otimes \alpha_2) \end{array} \right] \circ \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \\ \mathbf{1} \end{bmatrix}, \quad (36)$$

using the expression for $F_{+-}^{(2)}$ in Equation (12), and where \circ denotes the Hadamard product.

6. Expected total cost

For the computation of the expected total cost, we need to characterize the time between orders. In Section 6.1 we consider the time, η_j , between individual orders, and similarly in Section 6.2 for the time, η_j^c , between joint orders. In Section 6.3 we present the iterative procedure employed to derive the policy parameters for both products and give numerical illustrations.

6.1. Expected number of individual orders per unit time

The time between two subsequent individual orders has distribution $PH(\hat{\pi}_j, \hat{F}_{--}^{(j)})$ of order $(S_1 - s_1)(S_2 - s_2)$ and the time to place the subsequent order is equivalent to the time to reach absorption. The expected time between subsequent individual orders of product j is given by $\eta_j^{-1} = \hat{\pi}_j \left(-\hat{F}_{--}^{(j)}\right)^{-1} e$. Define

$$D^{(j)} = \begin{bmatrix} \sum_{k=0}^{m_j-1} \lambda^{(j)} d_{S_j-s_j+k}^{(j)} \\ \vdots \\ \sum_{k=0}^{m_j-1} \lambda^{(j)} d_{1+k}^{(j)} \end{bmatrix}. \quad (37)$$

The matrix F_{--} records the transitions in inventory position until the subsequent replenishment, thus only a decrease in inventory positions is possible. However in $\hat{F}_{--}^{(j)}$, an increase in inventory positions is possible when product $i \neq j$ places an order. Then for $j = 1$,

$$\hat{F}_{--}^{(1)} = F_{--} + D^{(1)}\alpha_1 \otimes \begin{bmatrix} 0 \\ e_{c_2-s_2}\alpha_2 \end{bmatrix} + \begin{bmatrix} I_{S_1-c_1} & 0 \\ e_{c_1-s_1}\alpha_1 \end{bmatrix} \otimes D^{(2)}\alpha_2, \quad (38)$$

where the second term refers to an increase in inventory positions due to a joint order initiated by product 1, and the third term refers to a (joint or individual) order placed for product 2. The vector $\hat{\pi}_1$ contains the probabilities of starting in a given state following an individual order of product 1 and is the stationary distribution of

$$\left(-\hat{F}_{--}^{(1)}\right)^{-1} \left(D^{(1)}\alpha_1 \otimes \begin{bmatrix} I_{S_2-c_2} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \right),$$

as an individual order for product 1 can only occur if the inventory position of product 2 is at least $c_2 + 1$. Similarly,

$$\hat{F}_{--}^{(2)} = F_{--} + D^{(1)}\alpha_1 \otimes \begin{bmatrix} I_{S_2-c_2} & 0 \\ e_{c_2-s_2}\alpha_2 \end{bmatrix} + \begin{bmatrix} 0 \\ e_{c_1-s_1}\alpha_1 \end{bmatrix} \otimes D^{(2)}\alpha_2, \quad (39)$$

and $\hat{\pi}_2$ is the stationary distribution of

$$\left(-\hat{F}_{--}^{(2)}\right)^{-1} \left(\begin{bmatrix} I_{S_1-c_1} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \otimes D^{(2)}\alpha_2 \right).$$

6.2. Expected number of joint orders per unit time

The time between two subsequent joint orders initiated by product 1 has distribution $PH(\alpha_1 \otimes \alpha_2, \hat{F}_{--}^{(12)})$, then $(\eta_1^c)^{-1} = (\alpha_1 \otimes \alpha_2) \left(-\hat{F}_{--}^{(12)}\right)^{-1} e$. As the inventory position of both products after placing a joint order is (S_1, S_2) , the initial state vector is $\alpha_1 \otimes \alpha_2$ and

$$\hat{F}_{--}^{(12)} = F_{--} + D^{(1)}\alpha_1 \otimes \begin{bmatrix} I_{S_2-c_2} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} I_{S_1-c_1} & 0 \\ e_{c_1-s_1}\alpha_1 & \end{bmatrix} \otimes D^{(2)}\alpha_2, \quad (40)$$

where the second term accounts for the individual orders placed of product 1, and the third term refers for the orders placed by product 2. The time between two subsequent joint orders initiated by product 2 is $PH(\alpha_1 \otimes \alpha_2, \hat{F}_{--}^{(21)})$ with $(\eta_2^c)^{-1} = (\alpha_1 \otimes \alpha_2) \left(-\hat{F}_{--}^{(21)}\right)^{-1} e$, where

$$\hat{F}_{--}^{(21)} = F_{--} + D^{(1)}\alpha_1 \otimes \begin{bmatrix} I_{S_2-c_2} & 0 \\ e_{c_2-s_2}\alpha_2 & \end{bmatrix} + \begin{bmatrix} I_{S_1-c_1} & 0 \\ 0 & 0 \end{bmatrix} \otimes D^{(2)}\alpha_2. \quad (41)$$

6.3. Iterative solution and numerical examples

To obtain the inventory parameters that minimize $E[C]$ in Equation (1), we use an iterative procedure to obtain successive approximations for both products using neighbor search and steepest descent algorithms. Given a set of initial inventory parameters for product 1, we determine the optimal inventory parameters for product 2. Then given this inventory policy for product 2, we determine the inventory parameters for product 1 that minimize $E[C]$ and repeat until convergence is reached. We note that to compute the cost for both products simultaneously would involve matrices with significantly larger dimensions and hence be less tractable.

We present results for six experiments for different values of the ordering costs and minor change-over times, as these drive the order quantities, $(S_j - s_j)$ and $(S_j - c_j)$. The parameter values are chosen to highlight the impact of changing the key replenishment drivers and costs, being the ordering costs and setup times. The holding and backlogging costs represent a service level of 90%. Parameter values common to all experiments are listed in Table 2 and the different values are listed in Table 3. The unit production time, major setup and minor change-over times are each exponentially distributed and the demand size distribution is zero-truncated binomial.

We demonstrate the iterative procedure for deriving parameters for the case of an independent replenishment (s_j, S_j) policy and discuss the interplay between the quantities of interest. We then illustrate the cost reduction by using a joint replenishment policy compared to independent policy.

Obtaining cost minimizing parameters for individual replenishment. For Experiment 1, we outline the solution for (s_j^*, S_j^*) . The iterations are listed in Table 4 with an initial guess of $S_1 - s_1 = 5$, from which we obtain $(s_1^*, S_1^*) = (16, 38)$ and $(s_2^*, S_2^*) = (11, 27)$.

To understand how the quantities of interest change throughout the iteration, Figures 2-4 illustrate how ρ , $E[C]$ and s_j evolve.

Iteration 1. For the initial order quantity $S_1 - s_1 = 5$, the small size leads to multiple setups, congested production, long lead times and high costs. The solid line Figure 2(a) illustrates how ρ decreases as the order quantity of product 2 increases. The cost-minimizing reorder point s_2 as a function of the order quantity is illustrated in Figure 3(a) and Figure 4(a) illustrates the impact on $E[C]$. Due to the small order of product 1, a large order of product 2 is optimal, as this tempers the utilisation rate and lead times.

Iteration 2. Given $S_2 - s_2 = 52$, the solid lines of Figures 2,3,4(b) illustrate the impact of the order quantity of product 1 on the utilisation rate ρ , optimal reorder point s_1 and expected total costs $E[C]$, respectively. Increasing the order quantity of product 1 from 5 to 19, leads to more favorable lead times (due to a reduction in setups) and lower total system costs.

Iteration 3. Given $S_1 - s_1 = 19$, the dashed lines of Figures 2,3,4(a) illustrate how $S_2 - s_2$ impacts ρ , s_2^* and $E[C]$, respectively. Compared to the scenario in Iteration 1, the order quantity of 1 is larger, thus the order quantity of 2 can be reduced to 16 units while $E[C]$ is also reduced.

Iterations 4-5. Given $S_2 - s_2 = 16$, the dashed lines in Figures 2,3,4(b) illustrate quantities for iteration 4. We obtain $(s_1, S_1) = (16, 38)$ with $E[C] = 60.74$ and $\rho = 0.75$. Proceeding to the next iteration with $S_1 - s_1 = 22$, we find $(s_2, S_2) = (11, 27)$ thus the procedure ends and the optimal inventory parameters are $(s_1^*, S_1^*) = (16, 38)$ and $(s_2^*, S_2^*) = (11, 27)$.

The results for the all experiments are listed in Table 5. As expected, the minor change-over time has no impact as no joint orders are placed. Different values of \tilde{k} and k_j have no impact as long as $\tilde{k} + k_j$ remains the same for product j , which explains why the result of experiments 1 – 4 are the same for the (s_j, S_j) policy. In experiments 5 and 6 $\tilde{k} + k_j = 0$ and the optimal order quantities are only slightly smaller compared to experiments 1 – 4 due to the impact of the order quantities on setup times and lead times.

Comparison of independent and joint replenishments. Following a similar iterative procedure, we present the solutions for (s_j, c_j, S_j) in Table 5. In Table 6 we give the cost reductions of (s, c, S) compared to the (s, S) policy, listed alongside the ratios of the fixed ordering costs \tilde{k}/k_j and the mean setup time vs. the mean change-over time. As expected, $E[C]$ is lower in the joint replenishment scenarios.

In Experiment 6, due to the small change-over time, we find the highest gains with a cost reduction of 10.73%, in which the benefit is mostly seen from the reduction in setup time by placing joint orders. We see the benefit of joint replenishment is not only influenced by the ratio of \tilde{k}/k_j , but also by the ratio of the joint major setup time and the changeover time. Comparing the individual and joint policies we find that the mean number of orders placed reduces in our set of experiments by more than 14%.

We demonstrate that this cost discrepancy increases in the ratio of \tilde{k} and k_j and in the ratio of the major setup time and the minor change-over time. The impact of a can order policy on lead times is unclear. Although the number of setups reduces,

lead times may not necessarily go down due to a batching time, i.e., waiting time after production until the joint order is completed.

7. Concluding remarks

The analysis of sufficiently detailed production-inventory models is challenging when accounting for the number of variable quantities in the supply-chain process. We consider a continuous review finite capacity production-inventory system with two products in inventory. As the order process of each individual product impacts the utilisation rate of the production system and thus the lead times for all products, the inventory levels are mutually dependent.

The contribution of this work is to construct an exact evaluation for the can-order joint replenishment policy while taking endogenous lead times into account and hence, using numerical examples, to study the complex interplay between product inventory levels, adding to the relatively limited studies of multi-item inventory analysis. The analysis in the current paper allows for systems with batch demand arrivals, variable order quantities, and non-exponential production times. Modelling the production facility as a multi-type queue, with correlated arrivals and type dependent service, the matrix exponential form of the age process of the busy queue is used to compute the inventory level distribution. For a given set of input values, our analysis generates parameters to specify a minimal cost can-order policy, a well-known heuristic for the joint replenishment problem.

We use numerical methods to obtain inventory parameters that minimize total ordering and holding costs and provide examples of the interaction between the inventory properties for the two product case, in terms of lead times, order quantities, and cost reductions from a can-order joint replenishment policy as compared with an independent order-up-to policy. We demonstrate scenarios where there is decreased cost of a can-order (s, c, S) joint replenishment policy as compared with an independent order-up-to (s, S) policy. Our analysis also can apply to a setting with multiple retailers of a single product that is produced on the same production line (Zipkin^[41]) and joint replenishment in transport applications (Padilla Tinoco *et al.*^[30]). A limitation is the computation time, which grows rapidly with the number of products, the maximum demand size or the order quantity. The analysis is applicable to more products but further work includes establishing optimality of the solution obtained numerically, and development of heuristics to manage combinatorial state space explosion. The analysis can be extended to a system with MAP arrival processes, however this will increase the dimension of the matrices involved (by a factor equal to the product of the number of states of both MAPs).

Acknowledgements

This research was supported by the Research Foundation-Flanders (FWO-Vlaanderen), grant no. G069016N.

References

- [1] K. Arrow, T. Harris, and J. Marschak. Optimal inventory policy. *Econometrica: Journal of the Econometric Society*, pages 250–272, 1951.
- [2] J. Artalejo, A. Gómez-Corral, and Q. He. Markovian arrivals in stochastic modelling: a survey and some new results. *Statistics and Operations Research Transactions (SORT)*, pages 101–156, 2010.
- [3] S. Asmussen and O. Kella. A multi-dimensional martingale for Markov additive processes and its applications. *Advances in Applied Probability*, 32(2):376–393, 2000.
- [4] D. Atkins and P. Iyogun. Periodic versus can-order policies for coordinated multi-item inventory systems. *Management Science*, 34(6):791–796, 1988.
- [5] J. Balintfy. On a basic class of multi-item inventory problems. *Management Science*, 10(2):287–297, 1964.
- [6] Y. Barron. Critical level policy for a production-inventory model with lost sales. *International Journal of Production Research*, 57(6):1685–1705, 2019.
- [7] Y. Barron. The continuous (S, s, Se) inventory model with dual sourcing and emergency orders. *European Journal of Operational Research*, 2021.
- [8] L. Bastos, M. Mendes, D. Nunes, A. Melo, and M. Carneiro. A systematic literature review on the joint replenishment problem solutions: 2006-2015. *Production*, 27, 2017. .
- [9] R. Boute, M. Lambrecht, and B. Van Houdt. Performance evaluation of a production/inventory system with periodic review and endogenous lead times. *Naval Research Logistics (NRL)*, 54(4):462–473, 2007.
- [10] S. Chakravarthy and B. Rao. Queuing-inventory models with MAP demands and random replenishment opportunities. *Mathematics*, 9(10):1092, 2021.
- [11] T. Cohen-Hillel and L. Yedidsion. The periodic joint replenishment problem is strongly NP-hard. *Mathematics of Operations Research*, 43(4):1269–1289, 2018. .
- [12] N. Dizbin and B. Tan. Optimal control of production-inventory systems with correlated demand inter-arrival and processing times. *International Journal of Production Economics*, 228:107692, 2020.
- [13] T. Dzial, L. Breuer, A. da Silva Soares, G. Latouche, and M. Remiche. Fluid queues to solve jump processes. *Performance Evaluation*, 62(1-4):132–146, 2005.
- [14] B. Faaland, M. McKay, and T. Schmitt. A fixed rate production problem with poisson demand and lost sales penalties. *Production and Operations Management*, 28(3):516–534, 2019.
- [15] Q. He. Age process, workload process, sojourn times, and waiting times in a discrete time SM[K]/PH[K]/1/FCFS queue. *Queueing Systems*, 49(3-4):363–403, 2005.
- [16] Q. He and A. Alfa. The MMAP[K]/PH[K]/1 queues with a last-come-first-served preemptive service discipline. *Queueing systems*, 29(2):269–291, 1998.
- [17] Q. He and M. Neuts. Markov chains with marked transitions. *Stochastic processes and their applications*, 74(1):37–52, 1998.
- [18] G. Horvath. Efficient analysis of the MMAP[K]/PH[K]/1 priority queue. *European Journal of Operational Research*, 246, 03 2015. .
- [19] G. Horvath and B. Van Houdt. A multi-layer fluid queue with boundary phase transitions and its application to the analysis of multi-type queues with general customer impatience. In *Proceedings of the 2012 Ninth International Conference on Quantitative Evaluation of Systems*, pages 23–32. IEEE Computer Society, 2012.
- [20] E. Kayaş, T. Bilgiç, and D. Karabulut. A note on the can-order policy for the two-item stochastic joint-replenishment problem. *IIE Transactions*, 40(1):84–92, 2008.
- [21] A. Krishnamurthy and D. Seethapathy. Production inventory models for a multi-product batch production system. In *Rapid Modelling and Quick Response*, pages 47–59. Springer, 2010.
- [22] V. Kulkarni and K. Yan. Production-inventory systems in stochastic environment and stochastic lead times. *Queueing Systems*, 70(3):207–231, 2012.
- [23] G. Latouche. Structured Markov chains in applied probability and numerical analysis. In

- MAM 2006: Markov Anniversary Meeting*, pages 69–78. C&M Online Media, Inc., 2006.
- [24] G. Latouche and V. Ramaswami. *Introduction to matrix analytic methods in stochastic modeling*. ASA-SIAM series on statistics and applied probability. Society for Industrial and Applied Mathematics, 1999.
- [25] B. Meini. On the numerical solution of a structured nonsymmetric algebraic riccati equation. *Performance Evaluation*, 70(9):682–690, 2013.
- [26] W. Nasr. Inventory systems with stochastic and batch demand: computational approaches. *Annals of Operations Research*, pages 1–25, 2021.
- [27] M. F. Neuts. *Matrix-geometric solutions for stochastic models*. John Hopkins University Press, Baltimore, 1994.
- [28] A. Noblesse, R. Boute, M. Lambrecht, and B. Van Houdt. Lot sizing and lead time decisions in production/inventory systems. *International Journal of Production Economics*, 155:351–360, 2014.
- [29] S. Otten, R. Krenzler, and H. Daduna. Separable models for interconnected production-inventory systems. *Stochastic Models*, 36(1):48–93, 2020.
- [30] S. Padilla Tinoco, S. Creemers, and R. Boute. Collaborative shipping under different cost-sharing agreements. *European Journal of Operational Research*, 263(3):827–837, 2017.
- [31] B. Sengupta. Markov processes whose steady state distribution is matrix-exponential with an application to the GI/PH/1 queue. *Advances in Applied Probability*, pages 159–180, 1989.
- [32] B. Sengupta. The semi-markovian queue: theory and applications. *Stochastic Models*, 6(3):383–413, 1990.
- [33] J. Song and P. Zipkin. Inventories with multiple supply sources and networks of queues with overflow bypasses. *Management Science*, 55(3):362–372, 2009.
- [34] J. Song, S. Xu, and B. Liu. Order-fulfillment performance measures in an assemble-to-order system with stochastic leadtimes. *Operations Research*, 47(1):131–149, 1999.
- [35] J. Song, L. Xiao, H. Zhang, and P. Zipkin. Optimal policies for a dual-sourcing inventory problem with endogenous stochastic lead times. *Operations Research*, 65(2):379–395, 2017.
- [36] A. Stolyar and Q. Wang. Exploiting random lead times for significant inventory cost savings. *Operations Research*, 2021. URL <https://doi.org/10.1287/opre.2021.2129>.
- [37] B. Van Houdt. A matrix geometric representation for the queue length distribution of multitype semi-markovian queues. *Performance Evaluation*, 69(7-8):299–314, 2012.
- [38] B. Van Houdt and C. Blondia. The waiting time distribution of a type k customer in a discrete-time MMAP[K]/PH[K]/ c ($c = 1, 2$) queue using QBDs. *Stochastic Models*, 20(1):55–69, 2004.
- [39] B. Van Houdt and J. Pérez. Impact of dampening demand variability in a production/inventory system with multiple retailers. In *Matrix-Analytic Methods in Stochastic Models*, pages 227–250. Springer, 2013.
- [40] P. Zipkin. Stochastic leadtimes in continuous-time inventory models. *Naval Research Logistics Quarterly*, 33(4):763–774, November 1986.
- [41] P. Zipkin. Performance analysis of a multi-item production-inventory system under alternative policies. *Management Science*, 41(4):690–703, 1995.

Notation	Definition
$\lambda^{(j)}$	demand arrival rate for product j
m_j	maximum batch demand size
$d_x^{(j)}$	probability of a demand of size x
$o^{(j)}$	maximum order size
$PH(\gamma_p, U_p)$	unit production time of order n_p
$PH(\gamma_+, V_+)$	major setup time of order n_+
$PH(\gamma_-, V_-)$	minor change-over time of order n_-
η_j	expected number of orders placed per unit time of only j
η_j^c	expected number of joint orders per unit time placed by j
Φ_j^+	expected number of units in inventory of j
Φ_j^-	expected number of units of j
h_j	inventory holding cost per unit per time
p_j	backlogging cost per unit per time
\tilde{k}	major fixed order cost per order
k_j	minor fixed order cost if j is included in the order

Table 1.: Summary of key notation

Parameter	Product 1	Product 2
$\lambda^{(j)}$ (per hour)	1	1
Maximum demand size m_j	8	6
Mean demand size	4	3
Holding cost h_j (per hour)	1	2
Backlogging cost p_j (per hour)	9	18
Mean unit production time (hours)	1/12	1/12
Mean major setup (hours)	1/2	1/2

Table 2.: Parameter values common to all numerical experiments

Experiment	\tilde{k}	k_1	k_2	Mean minor change-over (hours)
1	8	2	0	1/4
2	8	2	0	1/100
3	0	10	8	1/4
4	0	10	8	1/100
5	0	0	0	1/4
6	0	0	0	1/100

Table 3.: Ordering costs and setup times for each experiment

Iteration	Input	Output	Status
1	$S_1 - s_1 = 5$	$(s_2, S_2) = (26, 78)$	$\rho = 0.91, E[C] = 252.82$
2	$S_2 - s_2 = 52$	$(s_1, S_1) = (20, 39)$	$\rho = 0.71, E[C] = 102.01$
3	$S_1 - s_1 = 19$	$(s_2, S_2) = (11, 27)$	$\rho = 0.77, E[C] = 61.55$
4	$S_2 - s_2 = 16$	$(s_1, S_1) = (16, 38)$	$\rho = 0.75, E[C] = 60.74$
5	$S_1 - s_1 = 22$	$(s_2, S_2) = (11, 27)$	$\rho = 0.75, E[C] = 60.74$

Table 4.: Parameters obtained for the iterative procedure in Experiment 1 for (s_j^*, S_j^*)

Optimal replenishment parameters of product 1									
Exp.	(s_1^*, S_1^*)	$E[C]$	η_1	$E[L]$	(s_1^*, c_1^*, S_1^*)	$E[C]$	η_1	η_1^c	$E[L]$
1	(16,38)	24.77	0.17	3.37	(18,24,38)	24.16	0.07	0.08	3.72
2	(16,38)	24.77	0.17	3.37	(16,28,33)	22.40	0.02	0.13	3.16
3	(16,38)	24.77	0.17	3.37	(18,22,39)	24.64	0.09	0.06	3.72
4	(16,38)	24.77	0.17	3.37	(15,25,33)	23.22	0.03	0.11	3.23
5	(17,38)	23.08	0.18	3.33	(18,23,37)	22.53	0.08	0.08	3.67
6	(17,38)	23.08	0.18	3.33	(15,29,32)	20.76	0.01	0.13	3.10

Optimal replenishment parameters of product 2									
Exp.	(s_2^*, S_2^*)	$E[C]$	η_2	$E[L]$	(s_2^*, c_2^*, S_2^*)	$E[C]$	η_2	η_2^c	$E[L]$
1	(11,27)	35.97	0.17	3.01	(12,19,27)	34.68	0.08	0.05	3.49
2	(11,27)	35.97	0.17	3.01	(11,22,24)	31.92	0.03	0.11	3.11
3	(11,27)	35.97	0.17	3.01	(12,18,27)	35.13	0.11	0.03	3.36
4	(11,27)	35.97	0.17	3.01	(11,20,24)	32.88	0.05	0.10	3.12
5	(12,27)	34.50	0.18	2.98	(12,19,27)	33.64	0.09	0.04	3.44
6	(12,27)	34.50	0.18	2.98	(11,22,23)	30.64	0.02	0.14	3.08

Table 5.: Optimal replenishment parameters for each experiment

Exp.	$\frac{\tilde{k}}{k_j}$	$\frac{E[\text{Setup}]}{E[\text{Change-over}]}$	% Cost reduction
1	8	2	3.13%
2	8	50	10.57%
3	0	2	1.60%
4	0	50	7.64%
5	∞	2	2.45%
6	∞	50	10.73%

Table 6.: Ratio of ordering costs and setup times

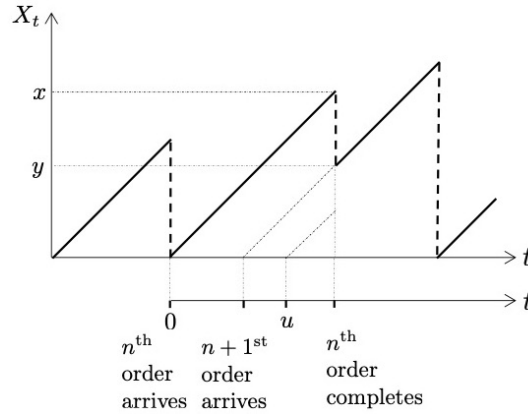


Figure 1.: Sample path of X_t . Upon production completion of the n^{th} order at age x , the subsequent order in service has already spent at least $x - u$ time units in the system.

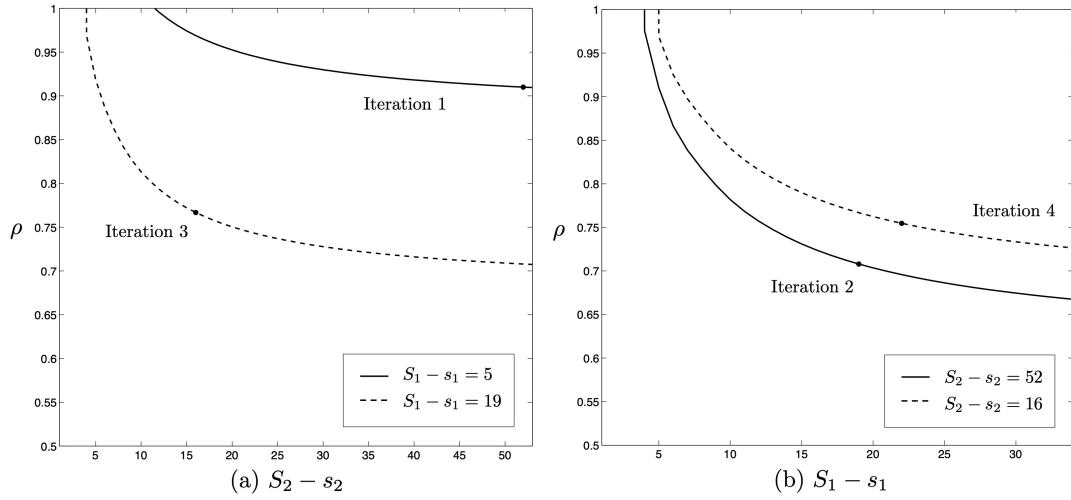


Figure 2.: Order quantity ($S_j - s_j$) vs. utilisation rate ρ .

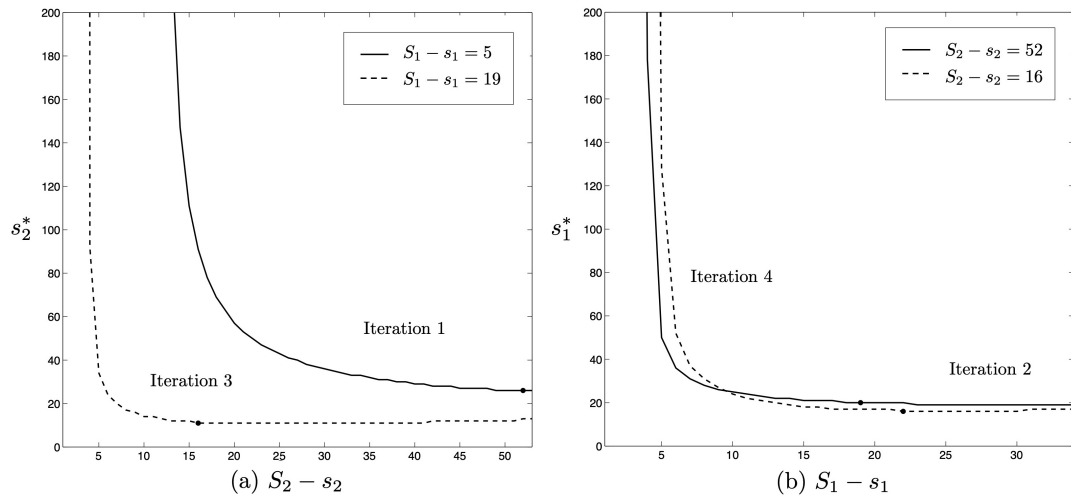


Figure 3.: Order quantity ($S_j - s_j$) vs. reorder points (s_j^*).

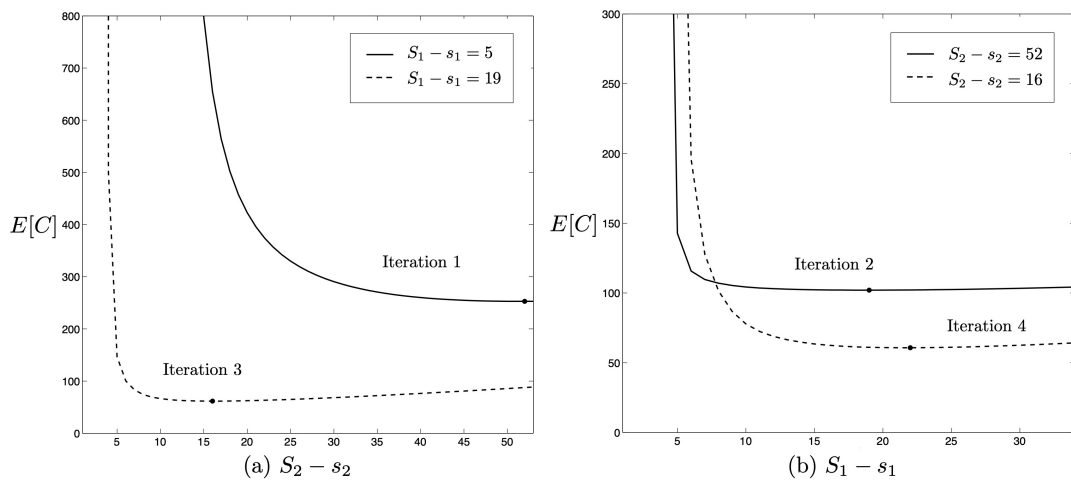


Figure 4.: Order quantity ($S_j - s_j$) vs. expected total cost $E[C]$.

List of Figures

1	Sample path of X_t	22
2	Order quantity $(S_j - s_j)$ vs. utilisation rate ρ	23
3	Order quantity $(S_j - s_j)$ vs. reorder points (s_j^*)	23
4	Order quantity $(S_j - s_j)$ vs. expected total cost $E[C]$	23

Collated Author's Response

We would like to thank the reviewers and the editor for their constructive and detailed feedback and the opportunity to resubmit. We have performed a thorough revision of the manuscript, improving the presentation of the paper. We address each specific comment from the editor and each reviewer in the tables below, and also provide the latexdiff comparison from the the previous submission.

Editor comments	Response
<p>R1 recommends acceptance, subject to some minor revision. The referee mentions that the paper is not easy to follow, but does not have any good idea how to improve it, and therefore leave it as is. R2 is also positive about the technical contribution of the paper. He/she too complains about the paper's writing and exposition. This referee recommends a major revision to improve the readability, clarity, and exposition, and provides some detailed suggestions. R2 feels that the paper will be acceptable after the paper's clarity and exposition has improved to satisfactory level. I would like to thank the referees for their valuable feedback. I concur with the referees assessment of the paper. Given that both reports are consistent in nature, and the remaining weakness is on writing and exposition, I recommend a minor revision. But please take this revision seriously, your contribution can be appreciated and impactful only when the readers can understand it.</p>	<p>The Introduction (Section 1) has been restructured including added clarity on specific contributions of the paper. Section 2 is new and includes new literature (pointed to by Reviewer 2). Particular effort has gone into emphasising the nature of the computation, including adding more signposts as to how and why the model is constructed in this fashion. The language relating to the iterative process has been edited throughout to improve clarity. The Model description in Section 4.1 and Inventory level distribution in Section 5.3 have also been improved.</p>

Reviewer 1 comments	Response
<p>1. It seems to me that the paper is mathematically correct. 2. The main contribution of the paper is the introduction of a Markov process for an inventory model with two types of products. The introduction of the Markov process, and a Markov modulated fluid flow process, is highly technical and creative. Using the stationary distribution of the Markov process, a number of quantities for the inventory model are obtained, which makes it possible to find a good inventory control policy. 3. It is not easy for me to follow the details for the introduction of the Markov process. On the other hand, I am not able to suggest a better way to do it. Thus, the presentation of the paper, especially about the Markov process, is acceptable to me. In summary, the methodology utilized in this paper is innovative and the analysis is mathematically correct. I recommend the paper be accepted for publication in Stochastic Models.</p>	
<p>Minor editorial comments. 1) Page 7, lines 17-18: Elaborate English.</p>	<p>We have added a sentence on the minor change-over time.</p>
<p>2) Page 7, line 43 and line 47: Why is "where" used? Is it better to use "and"?</p>	<p>Corrected</p>
Continued on next page	

Reviewer 1 comments	Response
3) Page 7, equation (1): It seems that the variable “ l ” can be replaced by “ $3 - j$ ”.	We have made this suggested change.
4) Page 8, lines 17 - 24: I do not fully understand the meaning of this sentence. I think it is related to the introduction of the Markov process. For example, the sets $P_{1,indiv}$ and $P_{2,indiv}$ are structurally different.	This paragraph has been rewritten to improve the explanation and a further comment has been added in Section 4.1 as to why they are structurally different.
5) Page 8, line 46: Is it better to define the “state space Γ ” by $(i, j), s_1 + 1 \leq i \leq S_1, s_2 + 1 \leq j \leq S_2$?	This improvement has been made.
6) Page 10, lines 36 – 43: It seems to me that this comment is true for the joint order case (for both i_1 and i_2). $P_{1,indiv}$ has i_1 and i_2 , but $P_{2,indiv}$ does not have i_2 . (I think it is related to page 8, lines 17-24.)	We have expanded the statement to improve the clarity.
7) Page 14, in equation (11): Is “ e_{m_1} ” defined?	The vector e_{m_1} was defined, but the paragraph has been rearranged to improve clarity.
8) Page 14, line 43: Is the notation “ T ” for matrix transpose defined?	Definition added.
9) Page 19, Section 5.2 [old reference Section 4.1] : It is not clear how the border $x = 0$ (zero fluid level) is defined. How does the underlying Markov chain changes when the fluid level enters or leaves zero.	We have added further detail of the boundary behaviour on page 13.
10) Page 21, equation (28): Please make sure that the elements in q_n are arranged in consistent with that in equation (26).	The ordering is consistent.
11) Page 21, equation (30): Is there a short justification for it?	Section 5.3 has been rearranged to improve the clarity of the construction of the inventory level distribution.
12) Page 22, equation (35): Please check the correctness of the column vector with elements $\{1, 0, 1\}$ on the right hand side.	This has been checked.
13) Page 23, line 13, the title of Subsection 5.1: Should “per unit time” be added to it?	This modification has been made.
14) Page 24, line 45, the title of Subsection 5.2: Should “per unit time” be added to it?	This modification has been made.
15) Page 25, line 30, the title of Section 6: It might be better to add “and numerical examples” to it.	We agree with the reviewer and have modified the title of Section 6.3.
16) Page 29: Any comments on whether or not the analysis can be extended to a system with MAP arrival processes for the two types of products?	We have added a comment on the extension to MAP arrival processes in the conclusion - noting that this will increase the dimension of the matrices involved (by a factor equal to the product of the number of states of both MAPs).

Continued on next page

Reviewer 1 comments	Response
<p>17) Page 29: Any comments on whether or not the analysis can be extended to a system with different unit production times?</p>	<p>We note that in principle this could be done, noting that the number of phases in the set $\mathcal{P}_{2,joint}$ increases quite significantly as we need to keep track of both order sizes, instead of just the sum of the two. So the size of for instance the $F_{++,joint}$ matrix would involve the product of $o^{(1)}$ and $(S_2 - s_2 - 1)$ instead of its sum.</p>

Reviewer 2 comments	Response
<p>The paper analyzes a continuous-review finite capacity production-inventory system with two products in inventory. Two replenishment policies are considered: the order-up-to (s, S) policy and the can-order (s, c, S) policy. The main tool is the matrix geometric method. Their aim is to derive the inventory parameters that minimize the total ordering and inventory related costs. Numerical examples, comparison and insights are also provided. The model seems to be important, practical and can be applied in many cases. However, some serious changes should be made before acceptance. The Introduction and the description Sections are poorly written, and thus, it makes the manuscript difficult to understand. The introduction isn't clear, cumbersome, lacks necessary details, explanations and recent studies. The analysis should be restructured. Numerical examples, observations and insights should be expanded. Also, the language needs some improvements; I recommend the use of a technical editor. I recommend a major revision; Following, I detail my comments.</p>	
<p>Introduction. Motivation, Examples, Contribution. The authors should motivate their model and add real-world examples of the model. There are no real-world examples or applications of the two policies. In fact, the policies are shortly described rather than extendedly. The contribution of the paper is not clear. Although there is a list of good papers, they are not described, and their contribution is not clear. As a result, the contribution and the uniqueness of the paper is not clear. In addition, no main results, observations, conclusions and insights are provided in the Introduction.</p>	<p>Section 1 has been reworked to emphasise the motivation and contribution, and added comments on real world interest.</p>
Continued on next page	

Reviewer 2 comments	Response
<p>Literature review. Although there is a list of good papers, most of them are old (just 3 papers are from 2019). The literature review is lacking relevant and recent papers. The paper deals with the matrix geometric method and assume phase-type distributions, thus, I suggest adding relevant papers. In the following, I added few studies dealing with phase-type distributions, matrix geometric approach and (s, S) policy. I believe the authors will find them or part of them (and the examples, applications, motivations and references therein) helpful and relevant.</p>	<p>Thank you to the reviewer for their helpful suggestions. The literature has been updated, specifically the inclusion of References^[7,10,12,14,26] that the reviewer pointed to. Further, we have also added References^[4,8,11]</p>
<p>Model description. Some of the assumptions of the model are not clear or not well motivated. I suggest presenting a list of assumptions and add some motivations or examples. For example, one assumption is that if a joint order is placed, the order is only delivered in inventory when the production for both products is complete (P.7, L.23). Why? Please add some examples and motivation; or if a joint order is placed, only one major setup is required Again, why is the production process influenced by the decision of the managers? I recommend adding real-world examples.</p>	<p>Some motivations for model description and assumptions have been added on page 4.</p>
<p>The evolution of production and inventory processes under different situations (individual and joint order) is unclear. The presentation is unfocused and cumbersome; this makes the policy very hard to understand. Can the authors provide typical sampling paths of several scenarios of the processes (the inventory, the production) for better explanation? Figure 1 is one example. I recommend providing some more.</p>	<p>We have added several signposts in Sections 3 and 4, rewritten Section 5.3 to add further clarity to the production-inventory model and process.</p>
<p>There are many parameters in the paper. I recommend adding a table summarizing all the notations, variables, parameters and costs used.</p>	<p>We have added Table 1 that contains the key notation for the model inputs.</p>
<p>I recommend adding Preliminaries Section with a short description of the phase-type distribution, Kronecker sum and Kronecker product.</p>	<p>We believe that definitions of phase-type distribution, Kronecker sum and Kronecker product are assumed knowledge of the reader, but to aid we have added further introductory references in the introduction.</p>
<p>As I see, matrices (3) and (6) are similar except of their indexes. The same with (5) and (7), and (4) and (8). So, why not to use shared definitions, e.g., $F(j)_{++}$, $j=1,2$, and thus makes this section shorter, clearer and sharper?</p>	<p>We understand the reviewers concern, but since matrices (3) and (6) & (5) and (7) are constructed differently (as (3) and (8) have additional terms) - we believe reducing these equations will not add to the clarity.</p>
<p>Here, too, I recommend adding a summarize table of all the matrices and their interpretations</p>	<p>We have not summarised the matrices from the fluid queue construction as elements of these are standard in the fluid queue literature.</p>

Continued on next page

Reviewer 2 comments	Response
The model does not include a production cost, although for the up-to-level (s, c, S) policy, the number of produced items is random. Thus, the impact of a production cost per item can be significant.	The model includes production time per unit $PH(\gamma_p, U_p)$, thus the concept is included in the total cost.
Preliminaries Section. Many results are taken from previous studies (Sengupta[21,22], Latouche [16] , and many others). I recommend adding Preliminaries Section with a short description of these results (as far as it works with the results and the course of the manuscript), and also a short description of the phase-type distribution, Kronecker sum and Kronecker product.	We have added further preliminary references and explanations in Section 1. Further signposts in Sections 3 and 4, a new Section 5.1, restructured Section 6.
Examples. Can the authors provide some motivation for the selected values in Tables 1 and 2? = The set of experiments (6) is neither satisfying nor detailed. Consider the large number of parameters, I recommend expanding the case studies to include more values, more results and more insights.	We have added the following statement on the choice of parameter values: “to highlight the impact of changing the key replenishment drivers and costs, being the ordering costs and setup times. The holding and backlogging costs represent a service level of 90%.”
Concluding remarks. Please add some future research and extensions.	We have added detail on the extension to a system with MAP arrival processes.
Minor remarks: Eq. (1) includes the terms η_i^c and η_j^c . What is η_i^c ? Maybe a typo? Why both are included?	Each η^c refers to different product, the notation has been modified to make this clearer.
2. Usually, the notation TC is used for the total cost, and C is used to denote the cycle length. If possible, and in order to avoid confusion and to keep consistency, please use TC for the total cost.	We have chosen to keep total cost as C , as we believe this is not uncommon notation.
Summarizing, I believe that the paper is interesting, relevant and fits to the journal’s scope. However, the main lack is its presentation. I suggest the authors reorganize the paper, add explanations, typical sample paths, summary of notations, definitions and parameters, assumptions, motivations, practical and real-world examples and applications. Please expand the literature review and the numerical studies and shorten the presentation of the matrices used. Please make the manuscript consist and comprehensive. By doing so, I recommend that the paper will be acceptable for publication.	We have undertaken a major restructure of the presentation in Sections 1 and 2 (new) - including new recent references and real world motivation. We have added several signposts throughout Sections 3,4,5 to guide the reader as to reasons the results are set out in the current order. We have added a summary table of key notation. The purpose of the numerical examples is illustrative, to show the iterative procedure is feasible and to display some of the tradeoffs, hence we regard the number of examples as sufficient.