Int J Adv Manuf Technol (2016) 83:449–459
DOI 10.1007/s00170-015-7587-z

ORIGINAL ARTICLE

Production control policies for a multistage serial system under MTO-MTS production environment

Paolo Renna[1]

Received: 28 November 2014 / Accepted: 12 July 2015 / Published online: 25 July 2015

Abstract A manufacturing system composed by several stages in serial system is considered. Each stage can produce several types of products from a semi-finished product. At each stage, a production control strategy is performed to release MTO and MTS orders. The MTS orders try to reduce the lead times and increase the service level for the customers. This research proposes MTS control policies observing the customer demand with higher service level reducing the stock level of the buffers at each stage. A simulation environment based on multi-domain methodology has been developed to test the proposed approach compared to policies proposed in literature. The numerical results are obtained for different levels of customer demand, fluctuations of the product type requested, and the reliability of the production system. The main results show how the proposed approach leads to better results both for service level and reduce the MTS level in all conditions tested.

Keywords Production control · Make to order · Make to stock · Multi domain simulation

1 Introduction

The major part of the research characterizes the production systems as make-to-stock (MTS) or mate-to-order (MTO) management approaches. The MTS systems support production systems characterized by low variety of product type and less expensive products [16]. In MTO systems, the main competitive factor is the reduction of the lead times increasing the customer service level (as the due date). The combination of MTO and MTS (hybrid MTO-MTS) can lead to the better trade-off between the MTO and MTS benefits.

In recent years, companies have shifted their production strategies towards hybrid MTS/MTO environments to achieve the advantages of both pure MTS and pure MTO systems simultaneously, among which lower inventory levels and shorter delivery times are the most notable [8].

Several studies investigated the hybrid MTO-MTS problem [2, 12, 16, 18], but these works focused on the distinction of the products to manufacture in MTO or MTS system or focused on the inventory management. The relevance of this theme is related to the several industrial cases of hybrid MTO-MTS systems as steel industry [4], food production systems [16], chemical industry [15], agricultural machines [9], electronic industry, and other similar markets where many product configurations can be produced from intermediate interchangeable modules [6]. An example of company that adopts this strategy is Dell Computer Corporation [14].

The reference context studied in this research concerns hybrid MTO production systems where the introduction of buffers can improve the customer service level when the MTO system is not able to meet the due date requested.

In this context, three problems need to be solved. The first concerns how the customer orders are satisfied using the MTO system or by a buffer in the production system. The second decision regards how the stock level of the buffers is determined. Finally, the replacement policy of the buffers is the last problem to solve.

The original contribution of this research regards mainly the determination of the stock level of the buffers and the replenishment policy of the buffers. Moreover, the simulation

---

✉ Paolo Renna
paolo.renna@unibas.it

[1] University of Basilicata, Potenza, Italy

environment developed based on multi- domain ExtendSim® gives the test case more realistic than the model investigated in literature introducing demand fluctuations and reliability. The main objectives are the reduction of the lead times in order to improve the customer service level (tardiness performance) with adequate level of work in process (WIP) in the production systems.

The paper is organized as follows. Section 2 presents an overview of the literature; the reference context is described in Section 3. Section 4 describes the proposed approach, while the simulation environment is presented in Section 5. Section 6 provides a discussion of the simulation results. Finally, conclusions and future research paths are drawn in Section 7.

2 Literature review

One of the first studies on hybrid MTS/MTO was presented by Williams [21]. He considered production/inventory systems as deterministic and single product with stochastic demand and capacity constraint using queuing theory. Therefore, the main issue is the stock system because the production system is composed by only one stage.

Among recent studies on this theme, Soman et al. [16] studied a food production company, which is a common MTS/MTO environment. The authors propose a comprehensive hierarchical planning framework that covers the important production management decisions to serve as a starting point for evaluation and further research on the planning system for MTO–MTS situations. Moreover, the authors identified as "Simulation studies might be helpful to study the MTO/MTS decision and the interactions between the products and the capacity under varying demand patterns, set-up times and processing times." Therefore, one of the gaps that the proposed research tries to fill is the development of a simulation environment more realistic.

Some studies focused on only one or two production stages reducing the possibility to extend their approaches in complex production systems. Soman et al. [17] studied a single-stage production system considering an economic-lot scheduling problem in a system including MTS, MTO and hybrid MTS/MTO; they focused on the inventory system. Almehdawe and Jewkes [1] explored a possible way in which MTS–MTO systems could be adapted to take advantage of economies of scale in ordering semi-finished goods. The primary contribution of this work is to show the potential benefit of such batching and to demonstrate that there can be substantial savings to the manufacturer, but little cost to the consumer in terms of additional delays. The main limits regard the production system composed by only two stages that manufacture only one product type. Moreover, the use of queue network to model the production system studied the system only in steady state.

Mathematical and optimization models (as genetic algorithm) have been proposed, but these models are characterized by higher computational time and, in some cases, are too complex to support real industrial applications. Zaerpour et al. [19] proposed a decision-making structure to choose the proper strategy for producing the products. They developed a Fuzzy TOPSIS-Analytic Hierarchical Process to determine partitioning of MTS, MTO, and MTS/MTO products. However, the assumptions of this model are too complex and not applicable in the real-world environments. Ghrayeb et al. [5] developed a hybrid push/pull system of an ATO manufacturing environment. The objective function for the presented hybrid model is to minimize the sum of inventory holding cost and delivery lead-time cost. The model is applied to solve the inventory and late delivery problems in an assemble-to-order manufacturer. A genetic algorithm is used. A discrete event simulation model is used to evaluate the objective function for each chromosome in the genetic algorithm. Compared to the pure push or pure pull production systems, the results showed that the hybrid production system could reduce the company cost significantly.

The major part of the studies focused on the location of the order decoupling point or the case in which the products are assigned exclusively to MTO or MTS system. Köber and Heinecke [9] studied a hybrid MTO/MTS system by the position of the customer order decoupling point. The evaluation of the production strategies is based on an industrial case of a global manufacturer of agricultural machinery and is accomplished with the help of system dynamics. Rafiei and Rabbani [13] presented a novel decision support system for order acceptance/rejection in a hybrid make-to-stock/make-to-order production environment. In particular, they proposed a fuzzy ANP structure to locate the customer order decoupling point of every family of coming orders. Lu et al. [10] focused on solving the multistage process push/pull junction point location problem in order to satisfy both high service levels and low inventory levels. A technique for order-preference by similarity-to-ideal solution (TOPSIS) is used to select a suitable option. The optimisation involves evaluation of stochastic performance measures within alternative scenarios among candidate junction-point locations using a discrete event simulation model. A practical thin-film transistor-liquid crystal display (TFT-LCD) process case study is utilized to illustrate the proposed method. After implementing a hybrid push/pull production strategy, simulation results indicate that the inventory level was reduced by over 18 % while the service level remained about the same. For another scenario, a 3.4 % decrease in service level can be paid off by a 46 % decrease in inventory level and 34 % improvement in lead time. Chen et al. [3] considered a production system, which is capable to produce two types of products: the first type of products is make-to-order, while the second type is make-to-stock.

The objective is to find the optimal production and pricing policy that maximizes the total discounted profit over an infinite planning horizon. They formulated the production-managing problem to an optimisation problem, which is then solved by two switch curves. They also integrated the product-pricing decision problem into consideration.

Hemmati and Rabbani [7] presented a decision-making structure to determine the appropriate product delivery strategy for different products in a manufacturing system. The strategies considered include make-to-stock (MTS), make-to-order (MTO), and hybrid MTS/MTO production systems. They used analytic network process that generalizes analytic hierarchical process by considering the interdependencies among factors. Finally, in order to show the applicability of the proposed structure in practice, the structure is implemented to choose the best production policy among three aforementioned strategies in the real industrial case company.

Zhang et al. [20] developed a multi-server queuing model of this system, where a subset of the servers or machines is dynamically switched between MTS and MTO production via a congestion-based switching policy. They developed analytical formulae for quantifying all major performance measures of the system. Numerical results are used to illustrate the general behavior of the dynamic hybrid system and to compare its performance to that of a more conventional static hybrid facility with dedicated MTS and MTO servers. For high levels of traffic intensity, the dynamic system is shown to provide superior customer service for both sales channels with lower finished goods inventory levels.

Morikawa et al. [11] proposed production control policies for a make-to-order manufacturing system composed of several stages under uncertain demand. Eight make-to-stock policies are prepared by combining buffer selection rules, matching acceptance rules, and make-to-stock replenishment rules. Their performance is evaluated by computer calculations, but no simulation environment has been developed to improve the level of detail.

Based on the above literature review, the following limitations can be drawn:

(a) The approaches proposed in literature concern simplified production systems (often one or two stages) investigating the environmental conditions in a steady state.

(b) The great part of the research focused on two main issues: decision-making approach to support the decision on the products to manufacture in MTO or MTS and the customer order decoupling point insertion in the production system.

The research proposed in this paper resulted to the above limitations in the following issues:

(a) The proposed approaches have a low computational complexity, and they can be applied in a wide range of production systems (several production stages).

(b) It is considered a hybrid MTO/MTS without customer order decoupling point where the buffer can support the performance improvements of the MTO system.

(c) The simulation environment developed allows to investigate the proposed approach in several environmental conditions as: mix fluctuations and reliability of the manufacturing resources.

The proposed approach starts from the research proposed by Morikawa et al. [11]. The main differences are the following. The first difference regards the relations between each couple of stages; the research proposed in this paper considers the quantity of items in the buffers of the upstream stage to take the decision for the downstream buffers for each specification of the items. The decisions in Morikawa et al. [11] do not take into account the amount of items in the upstream buffers. Therefore, this research investigates how the decision for the buffer level affects the decision of the buffers in other stages. The simulation model developed allows to investigate the reliability of the manufacturing resources considering different production capacities among the stages and improve the level of detail of the model. The numerical results are obtained conducting an opportune number of replications to assure the confidence interval of the numerical results. For the above reasons, the benchmark used to evaluate the proposed approach is a pure MTO system.

3 Reference context

The production environment is the same tested in Morikawa et al. [11]. In the following, it is briefly reassumed. The production system is characterized by four serial production stages; the items visit all stages. The first three stages work with two buffers: MTSB^i that is the MTS buffer of the stage i and MTOB^i that is the MTO buffer of the stage i. The last stage works only in a MTO and delivery the final products to the customers.

Each stage can produce two variants of the products; therefore, there are 16 variants of the products for the four stages considered. The products can be delivered to the customers before the due date. The customer orders input the orders characterized by four terms: product types, due date, the time to fix the quantity, and estimated order quantity before the fix. The time to fix the order is considered in order to investigate the case of orders that contains planned quantity when arrives. The quantity may vary over time until the order is fixed. In this research, some modifications from the above reference context are made as described in the following.

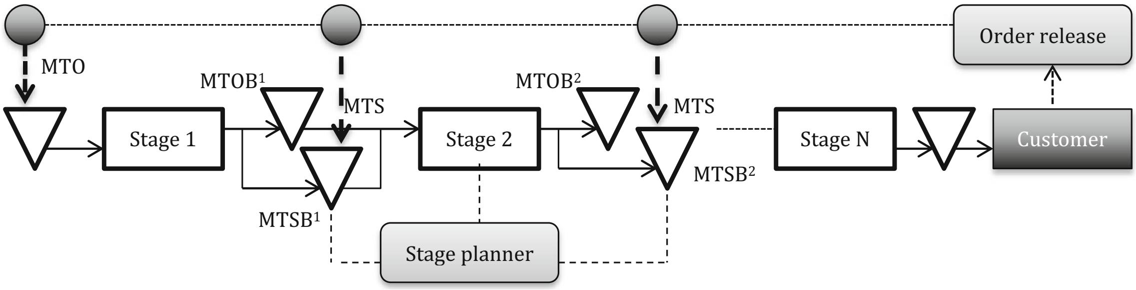

Fig. 1 Production system

Figure 1 shows the model of the production system considered. The architecture of the production system consists of an order release and a stage planner for each stage.

The order release applies the policy to decide if the customer order will be satisfied by MTO or MTS using a buffer over the production stages.

The stage planner controls each stage; in particular, the stage planner collects the information of the upstream and downstream buffers of the generic stage, the capacity available of the stage in order to decide if some MTS orders may be released to replenishment the buffer level.

Each stage is characterized by a queue of the order released; the MTO orders have the priority to the MTS orders.

Each stage is characterized by a production capacity in terms of volume of items that the stage can release after a fixed duration (volume of items/unit of period). The reliability of the production system is considered reducing the capacity of the stage in order to include events that make impossible to utilize the capacity at 100 %. For the above considerations, each item requires the same amount of capacity regardless of order specification at that stage.

The setup activities are included in the production lead time at each stage and defective items are not considered. The capacity of each stage is the same in order to avoid a bottleneck stage. Finally, the required raw material at the first production stage is always available.

4 Proposed approach

The customer orders follow an exponential distribution as described above and each order is characterized by product type, due date, time to fix the order, and the unfixed volume. The volume required is fixed after the time to fix; therefore, the quantity is uncertain until the order is fixed. When the order is fixed, the first decision concerns the release of the order in the production system. The order can be released as a MTO or MTS using the items in the buffers. The decision is evaluated considering the estimated due date of the generic customer order *o*:

$$\mathrm{Tnow} + \sum_{i=1}^{N} t_i - \mathrm{duedate}_o < 0 \qquad (1)$$

where,

Tnow is the time of the decision
t_i is the time to release an order in the production stage i (each t_i the production stage release a volume of items that depend on the production capacity)
N is the number of production stages of the production system
$\mathrm{duedate}_o$ is the due date of the order o.

Expression 1 evaluates if the production in MTO system of the order *o* can be in delay. If the expression 1 is true, the customer order *o* is released in a MTO system because the production in MTO does not lead to delivery the order in delay to the customer. If the expression 1 is false, it is evaluated if the order *o* can be released in MTS system to avoid or reduce the delay. In this case, it is evaluated if the order can be released in a buffer of the *i* stage of the production system.

Figure 2 shows the order release in case of MTS policy. The controller of the production system that applies the order release process evaluates the buffer levels starting from the last stage in MTS that is the *N-1* (the last stage N works only in a MTO system) to the stage 2 (the first stage is connect directly to the inventory of raw materials). The controller of the production system sets the stage to evaluate $k=N-1$ and checks if the buffer for the product typology required is enough to satisfy the volume requested. If the volume is enough for the order, the order is assigned to the buffer of the stage evaluated *k*; otherwise the controller evaluates the before stage $k=k-1$ until the stage 2. If the volume is not enough in all stages evaluated, the order is released in MTO system. This strategy allows to reduce the tardiness of the order.

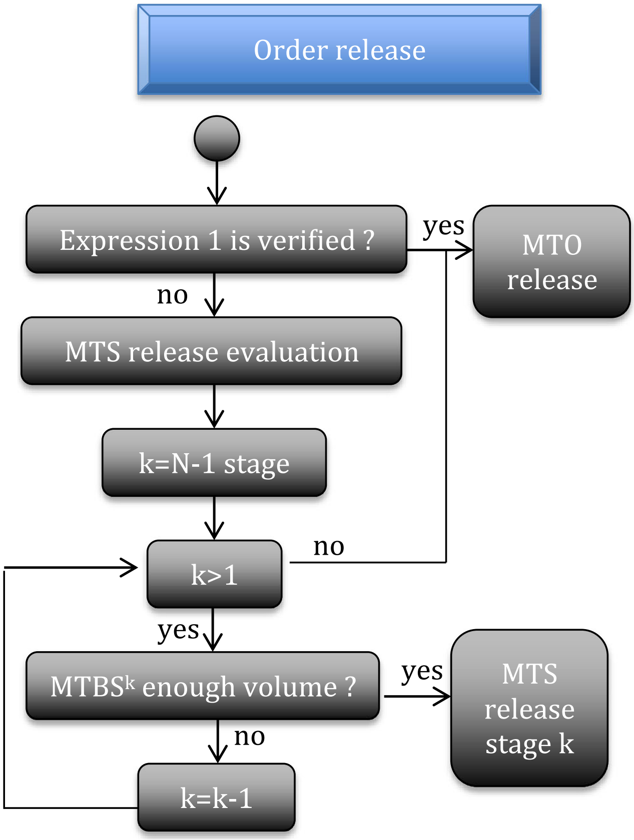

Fig. 2 Order release in MTS system policy

The stage planner performs the operations on the buffer levels of the MTS system of each production stage. The stage planner of each stage performs a periodic review policy (T_p periods) on the capacity available of its stage. The T_p periods corresponds to the time t_i each production stage. Every t_i, each stage produces a volume of items (the volume depends on the capacity of the production stage). In this research, t_i is one period of time (as in [11]). In case of available capacity, the stage planner evaluates how to distribute the MTS order to the different product types according to the policy explained in the following sub-sections.

4.1 Stage planner rule 1

The stage planner of the resources in stage *i* evaluates the capacity available of the production stage and the level of the MTSB buffers of the downstream stage *i+ 1* (buffer level replenishment) and the buffers level of the stage *i* (raw material available for the upstream production stage *i*).

The first step is the computation of the capacity available at stage *i* (cap_av^i [items/period]) after the allocation of the MTO orders order^{MTO} as shown in expression 2:

$$\text{cap\_av}^i = \text{cap}^i - \text{order}^{MTO} \tag{2}$$

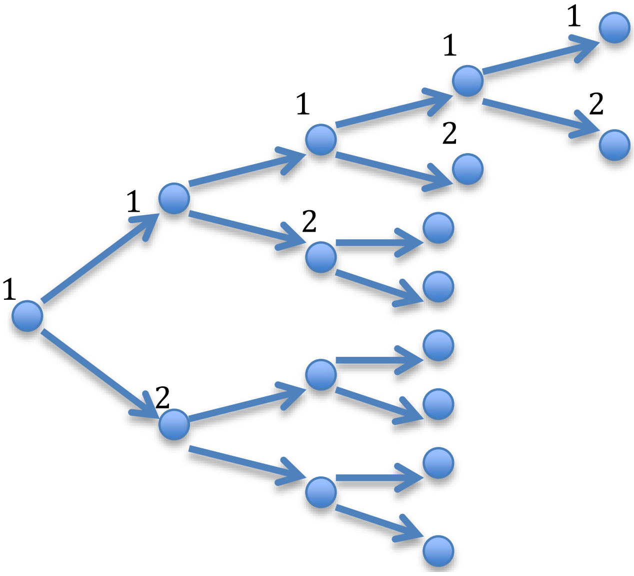

Fig. 3 Production types tree

where,

cap^i is the capacity of the production stage *i* in terms of items/unit of period.

Figure 3 shows the production types tree; the indexes reported in the Fig. 3 are used to explain the MTS release orders of the generic production stage. The first index concerns the upstream buffer of the production stage and the second index the two possible variants of the product after the production stage (downstream buffers) of the generic stage *i* (each stage can produce two variants of the item).

After that, it is computed the quantity requested for the buffers at stage $i+1$ as shown in expressions 3 and 4:

$$\text{delta}^{11} = \text{LBS}^{11} - \text{MTSB}^{11} \tag{3}$$

$$\text{delta}^{12} = \text{LBS}^{12} - \text{MTSB}^{12} \tag{4}$$

where the LBS^{11} and LBS^{12} are the desiderated level of the two downstream buffers of the production stage.

Then, it is evaluated the semi-finished products available to satisfy the quantity requested delta; each semi-finished buffer at stage *i* can satisfy two semi-finished at stage $i+1$.

The expressions 5 and 6 compute the allocation in proportion to the two semi-finished requested at stage *i+ 1* considering the semi-finished available at stage *i* (raw).

$$\text{rich}^{11} = \text{MIN}\left[\text{raw}^{1} \cdot \frac{\text{delta}^{11}}{\text{delta}^{11} + \text{delta}^{12}}, \text{delta}^{11}\right] \tag{5}$$

$$\text{rich}^{12} = \text{MIN}\left[\text{raw}^{1} \cdot \frac{\text{delta}^{12}}{\text{delta}^{11} + \text{delta}^{12}}, \text{delta}^{12}\right] \tag{6}$$

Expressions 7 and 8 distribute the quantity requested in proportion to the capacity available.

$$\mathrm{all}^{11} = \mathrm{MIN}\left[\mathrm{cap\_av}^{1} \cdot \frac{\mathrm{rich}^{11}}{\mathrm{rich}^{11} + \mathrm{rich}^{12}}, \mathrm{rich}^{11}\right] \quad (7)$$

$$\mathrm{all}^{12} = \mathrm{MIN}\left[\mathrm{cap\_av}^{1} \cdot \frac{\mathrm{rich}^{12}}{\mathrm{rich}^{11} + \mathrm{rich}^{12}}, \mathrm{rich}^{12}\right] \quad (8)$$

The values of all^{11} and all^{12} are the order released for the replacement of the buffers level. This rule tries to keep the desiderated level of the buffers. The main drawback is the determination of the desiderated level for each buffer and modifies these values when the customer demands changes.

4.2 Stage planner rule 2

In order to avoid the main drawback of the rule described above, the stock level of the MTS buffers is dynamically determined. The controller of the production system observes the unfixed customer demand. The controller forecasts the level of the buffers evaluating if the production system should satisfy the order in MTO or it is necessary the MTS support (considering the due date). This is a forecast because the order will be released in the production system, only when the order is fixed. In case of MTS, the stage i of the buffer that could satisfy the order able to met the due date is determined. The controller computes the average ($\mathrm{averageMTSB}^{i}$) and standard deviation ($\mathrm{standard\_deviationMTSB}^{i}$) for each buffer at each stage of the past customer unfixed orders observed.

The level of the buffers of the MTS system in a generic stage is computed by expressions 9 and 10.

$$\begin{aligned}\mathrm{LBS}^{11} = {} & \mathrm{averageMTSB}^{11} + 3 \\ & \cdot \mathrm{standard\_deviationMTSB}^{11} - \mathrm{MTSB}^{11} \\ & + \mathrm{MinL}^{11}\end{aligned} \quad (9)$$

$$\begin{aligned}\mathrm{LBS}^{12} = {} & \mathrm{averageMTSB}^{12} + 3 \\ & \cdot \mathrm{standard\_deviationMTSB}^{12} - \mathrm{MTSB}^{12} \\ & + \mathrm{MinL}^{12}\end{aligned} \quad (10)$$

The three times of the standard deviation allows to obtain a 97.7 % of the probability to forecast the real value of the buffer levels considering a Gaussian distribution. The Gaussian distribution is used because several sources of distortion have to be considered as follows: as the demand, the capacity of the stage, the buffer levels of the stages, the number of production stages, etc.

The value of the MinL is added to avoid that the buffers have zero components. The allocation of the capacity follows the stage planner rule 1 with the value LBS computes as described in this sub-section.

Because the LBS change dynamically, the value of delta *(*expressions 3 and 4*)* cannot be negative; therefore, in this case expressions 3 and 4 are modified as follows:

$$\mathrm{delta}^{11} = \mathrm{MAX}\left[\mathrm{LBS}^{11} - \mathrm{MTSB}^{11}; 0\right] \quad (11)$$

$$\mathrm{delta}^{12} = \mathrm{MAX}\left[\mathrm{LBS}^{12} - \mathrm{MTBS}^{12}; 0\right] \quad (12)$$

This rule can adapt the level of the buffers when the customer demand changes in terms of mix, due date request (for example, rush orders), and the time to fix the orders. The value of MinL is like a safety stock to determine. Moreover, the level of the buffers is different for each production stage.

5 Simulation environment

The objective of the simulation experiments is to measure the performance of the proposed approach benchmarked to a pure MTO system in different environment conditions.

The simulation environment developed is based on the multi-domain software ExtendSim® by Imagine That Inc. ExtendSim® is a simulation program for modeling discrete event, continuous, agent-based, and discrete rate processes. There are four ExtendSim® packages: CP for continuous processes; OR (operations research) which adds discrete event; AT (advanced technology) which adds discrete rate, a number of advanced modeling features, and Stat Fit for statistical distribution fitting; and Suite which adds 3D animation.

The combination of agent-based, information, and OR allows to develop a complete simulation environment that is able to connect to simple industrial information as access and excel used in small and medium enterprises.

The length of the simulation 100,000 periods and the arrival of the orders followed a Poisson distribution with a four values of mean (see Table 1). The quantity of items (q) requested per order are extracted from a discrete uniform (DU) distribution $q = \mathrm{DU}[6,15]$. Then, it is determined the time to fix the order is extracted from a uniform distribution $u = [1,5]$. The due date assigned to the order is taken from a uniform distribution [1,5] adding up the time of the arrival and the time to fix the order. The fixed quantity of the order is taken from a discrete uniform distribution $DU[q-u,q+u]$. The product type of the order is taken form a discrete uniform distribution $DU[1,16]$ (as reported in [11]).

The simulation advances at discrete period of one (t_i and T_p); at the beginning of each period, the decisions about the buffers level are made.

The capacity of the production stages (volume of items/period) are the same and there are considered two levels as shown in Table 1. Therefore, the production time is the same for all item specifications.

Table 1 Production system data

	A1	A2	A3	A4
Arrival parameter	2	4	8	12
	C1	C2		
Capacity [products/unit time]	10	20		
	Stage 2-3-4	Stage 2-3-4		
Buffer level	40-20-10	60-30-15		
	MinL1	MinL2		
Buffer MinL	5	10		

The approach with fixed desired inventory level is considered for two cases as shown in Table 1. The proposed approach is tested for two levels of minimum inventory MinL as shown in Table 1. Two additional experiments are considered for all classes considered as follows:

- It is considered the demand of product type that over the 1000 periods is only the product type 1. This emulates the extreme case of unbalanced product type orders in order to investigate how the production system reacts to mix change.
- The second additional case concerns the reliability of the stages; this is considered by a reduction of the capacity taken from a uniform distribution [0,0.25]. This means a reduction between 0 and 25 % of the capacity of each stage.

For each experiment class, a number of replications able to assure a 1.5 % confidence interval and 95 % of confidence level for each performance measure have been conducted.

In summary, 64 classes simulated compose the experimental plan for the three models.

The performance measures investigated are the following:

- The average tardiness of the orders
- The average inventory of the MTS system
- The average number of orders satisfied by MTS system

6 Numerical results

Table 2 reports the tardiness and the inventory level for the case base (steady state). The tardiness is reported as a percentage difference compared to the pure MTO system. Moreover, two cases of buffer level of the stage planner rule 1 are reported as follows: the first concerns the buffer level for the three stages 40-20-10 and the second 60-30-15.

The proposed approach leads always to the better tardiness reduction. The greater difference is obtained when the capacity of the production stages is low (C1) and the inter-arrival of the customer orders is low (A4). In case of high capacity (C2), the proposed approach has the same tardiness of the case 60-30-15.

The increasing of capacity of the production stages allows to improve the tardiness performance for all cases tested. The approaches with fixed level of buffers reduce drastically the improvement when the congestion is higher (see A4).

The proposed approach allows to obtain the above tardiness performance reducing drastically the average inventory in the buffers. The average reduction of the inventory is about 27 %. It can be noticed that the inventory level has a low variation between the two capacity tested, except for the case of high congestion (A4). The lower capacity does not allow the production stages to replacement the products in the buffers. The standard deviation highlights how the proposed approach is more robust when the inter-arrival parameter changes and the capacity is lower.

The orders satisfied in MTS system (MTS orders) show how the proposed approach leads to lower orders satisfied in MTS system when the inter-arrival parameter is lower; this highlights how the proposed approach improve tardiness performance allocating the products with a better distribution among the buffers of the production stages.

Table 3 reports the same performance measures of Table 2 in case of mix changes. In this case, the benefits of the proposed approach are greater than the case base. The proposed approach is more able to capture the mix fluctuations of the customer demand. In case of mix change, the MinL1 is better than the cases with fixed level of buffers reducing the average inventory level. In this environmental condition, the proposed approach is better also in the case with high capacity (C2).

Table 4 reports the same performance measures of Table 2 in case of reliability of the production stages. As the reader can notice, all hybrid approaches leads to worst results when the congestion level is high (A4) and capacity is low (C2). In the other cases, the results confirm the above considerations. This supports the robustness of the approach proposed.

Figure 4 shows the tardiness performance considering the effects of inter-arrival, capacity, production system's condition, and the average over all simulations conducted.

When the congestion level is high (see Fig. 4a, A4), the approach with higher stock level of the buffers leads to worst tardiness performance. The proposed approach with MinL1 is able to improve the tardiness performance compared to the MTO system. This means that in case of higher customer demand, the stock level of the buffers needs to be lower. The same trend of the tardiness performance can be observed when the production system is characterized by failures (see Fig. 4c, reliability).

Therefore, the proposed approach can be more competitive if the minimum level of stock of the buffers can

Table 2 Average lateness—case base

	C1				C2			
	40-20-10	60-30-15	MinL2	MinL1	40-20-10	60-30-15	MinL2	MinL1
Tardiness								
A1	−76.40 %	−87.58 %	−88.82 %	−78.88 %	−88.43 %	−95.87 %	−96.69 %	−89.26 %
A2	−72.73 %	−84.85 %	−86.67 %	−75.76 %	−86.07 %	−94.26 %	−95.08 %	−85.25 %
A3	−61.71 %	−74.86 %	−77.14 %	−64.00 %	−80.00 %	−89.60 %	−90.40 %	−78.40 %
A4	−22.38 %	−0.95 %	−44.29 %	−33.33 %	−67.69 %	−79.23 %	−79.2354	−63.85 %
Av.	−58.31 %	−62.06 %	−74.23 %	−62.99 %	−80.55 %	−89.74 %	−90.35 %	−79.19 %
Dev.	24.75 %	41.11 %	20.60 %	20.79 %	9.28 %	7.49 %	7.88 %	11.17 %
Inventory level								
A1	236.58	355.92	261.64	191.59	246.51	356.28	261.48	192.39
A2	234.85	354.01	259.14	189.73	235.28	354.45	259.41	190.18
A3	229.17	347.51	252.69	183.98	230.53	348.84	253.76	184.97
A4	212.72	325.6	234.46	169.09	221.14	337.28	241.80	174.73
Av.	228.33	345.76	251.98	183.60	233.37	349.21	254.11	185.57
Dev.	10.88	13.91	12.28	10.20	10.55	8.56	8.83	7.87
MTS orders								
A1	6654.20	6665.40	6645.60	6462.30	6662.54	6665.59	6659.44	6522.67
A2	9978.16	10,014.60	9980.73	9589.11	9992.14	10,002.34	9982.70	9639.36
A3	20,047.86	20,000.25	19,858.41	18,743.85	19,960.15	19,996.67	19,933.77	18,881.10
A4	39,924.53	39,987.82	39,232.63	35,549.08	39,989.84	39,997.20	39,442.29	36,445.18

Table 3 Average lateness—mix changes

	C1				C2			
	40-20-10	60-30-15	MinL2	MinL1	40-20-10	60-30-15	MinL2	MinL1
Tardiness								
A1	−72.67 %	−79.50 %	−90.06 %	−88.82 %	−84.30 %	−87.60 %	−97.52 %	−95.87 %
A2	−69.09 %	−75.15 %	−88.48 %	−86.06 %	−80.33 %	−86.07 %	−95.90 %	−94.26 %
A3	−58.86 %	−55.43 %	−82.29 %	−78.86 %	−73.60 %	−71.20 %	−92.00 %	−88.80 %
A4	−40.00 %	−33.33 %	−66.19 %	−60.48 %	−62.31 %	−60.00 %	−80.00 %	−76.92 %
Av.	−60.15 %	−60.85 %	−81.76 %	−78.55 %	−75.13 %	−76.22 %	−91.36 %	−88.96 %
Dev.	14.66 %	21.13 %	10.90 %	12.76 %	9.62 %	13.10 %	7.92 %	8.58 %
Inventory level								
A1	232.56	349.16	277.85	207.65	233.89	349.44	269.15	207.80
A2	230.23	344.96	275.84	205.43	229.73	347.30	266.72	205.51
A3	223.25	326.96	269.25	198.90	226.53	333.41	270.30	198.07
A4	210.65	311.85	251.81	185.41	218.39	322.09	257.38	189.00
Av.	221.38	327.92	265.63	196.58	227.14	338.06	265.89	200.10
Dev.	9.92	16.58	12.42	10.21	6.56	12.80	5.86	8.48
MTS orders								
A1	6650.28	6689.41	6656.03	6624.15	3989.84	3997.20	3942.29	3645.18
A2	9987.29	10,002.75	10,000.70	9929.23	9995.94	10,000.52	10,001.07	9925.48
A3	19,950.40	19,888.25	19,898.19	19,462.69	19,920.10	19,996.90	19,875	19,621.18
A4	3823.30	39,700.68	37,919.41	35,218.00	39,327.67	39,969.97	38,203.70	37,052.92

Table 4 Average lateness—reliability

	C1				C2			
	40-20-10	60-30-15	MinL2	MinL1	40-20-10	60-30-15	MinL2	MinL1
Tardiness								
A1	−73.94 %	−84.04 %	−85.64 %	−75.53 %	−87.20 %	−95.20 %	−96.00 %	−86.40 %
A2	−69.95 %	−80.31 %	−82.38 %	−70.98 %	−84.92 %	−93.65 %	−94.44 %	−84.13 %
A3	−54.88 %	−67.91 %	−70.70 %	−57.67 %	−78.63 %	−87.79 %	−89.31 %	−76.34 %
A4	90.82 %	805.25 %	931.80 %	34.43 %	−64.08 %	−76.06 %	−77.46 %	−61.27 %
Av.	−26.99 %	143.25 %	173.27 %	−42.44 %	−78.71 %	−88.17 %	−89.31 %	−77.03 %
Dev.								
Inventory level								
A1	236.42	355.83	261.00	191.38	236.77	356.28	261.40	191.87
A2	234.50	356.95	258.86	189.52	235.24	354.41	259.51	190.14
A3	228.01	346.32	251.53	183.17	230.46	348.69	253.50	180.55
A4	194.00	253.97	157.41	156.23	220.39	336.46	241.20	174.02
Av.	218.84	319.08	222.60	176.31	230.72	348.96	253.90	184.15
Dev.	21.75	56.64	56.58	17.67	7.39	8.94	9.11	8.39
MTS orders								
A1	6675.25	6671.41	6659.42	6430.78	6663.27	6644.37	6663	6460.04
A2	9999.13	100,005	10,000.4	9602.97	10,006.89	9976.41	9978.98	9626.66
A3	20,060	19,956.4	19,887.4	18,713.56	19,668.95	20,002.92	20,019.1	18,900.15
A4	39,744.54	39,950.07	37,095.11	33,876.64	39,867.63	40,041.83	39,578.67	36,339.6

be modified from MinL2 to MinL1 in case of demand peak or reduction of the reliability of the production system.

The above considerations are more relevant when the capacity of each production stage is low (see Fig. 4b, C1); as the reader can notice, the lower capacity of the production stages

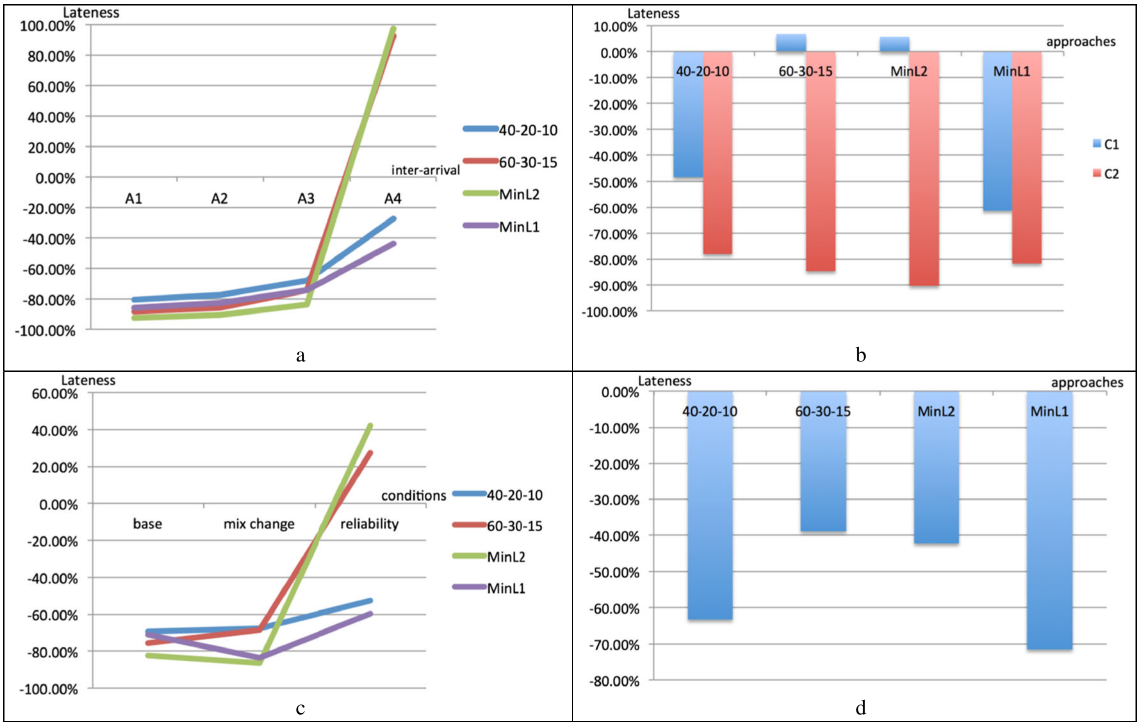

Fig. 4 Lateness evaluation. **a** Inter-arrival. **b** Capacity. **c** Production system condition. **d** Average overall condition

create problems for the approaches with higher buffer stock level. If the capacity is suitable for the stock level of the buffers, the proposed approach MinL1 leads always to the better tardiness performance with inventory reduction.

Figure 4d shows the average tardiness performance over all conditions tested; the proposed approach MinL1 leads to better tardiness performance considering reduction of capacity, peak of customer demand, and reliability of the production system.

7 Conclusions and future development

A hybrid MTO/MTS multi-stage serial production system has been investigated. Each production stage can produce two specifications of the orders obtaining 16 product typology for the four stages considered. In each production stage, a buffer to improve the tardiness performance of the production system is introduced. This paper proposes a policy to manage the buffers' level in order to minimize the average tardiness of orders with lower stock level of the buffers. A simulation environment based on multi-dimensional software (ExtendSim®) is developed to overcome the limitations of the modeling proposed in literature as the decision of the generic stage is made considering the volume of items available in the upstream buffers, mix fluctuations, and reliability of the production stages. The numerical results are obtained by the evaluation of the proposed approach compared with approaches proposed in literature. The original contributions of this research are the following:

- The proposed approach based on the adaptable level of the buffers (each buffer has different desiderated level) leads to the better performance improving the tardiness performance reducing drastically the average stock level of the buffers. The proposed approach can be controlled by two parameters: minimum level of buffer stock level and the multiplied coefficient of the standard deviation (see expression 9). These two parameters allow to regulate the trade-off between the tardiness and the inventory costs following the management implications.
- The stock level of the buffers needs to be adequate to the capacity of the production stages. If the policy adopted leads to higher level of stock than the production stages can support, the tardiness performance get worse in some cases as higher congestion or low reliability of the production systems.
- At managerial level, the simulation environment developed by multi-domain software can support the design of the policy in hybrid MTO/MTS system in order to reduce the risk and improve the performance of the production systems. These objectives can be pursued if the simulation environment is more realistic considering the effects of the production environment and the relation among the production stages (the main limit of the model proposed in literature).

Future development paths of the proposed research can be as follows: set the parameters of the proposed approach evaluating the costs of inventory and tardiness; a policy that considers the satisfaction of the orders not only by one buffer, but two buffers of two consecutive stages can satisfy the orders (in particular, when the quantity required is high); and the development of a fuzzy tool to decide the dynamic stock level of the buffers in order to include several characteristics.

References

1. Almehdawe E, Jewkes E (2013) Performance analysis and optimization of hybrid manufacturing systems under a batch ordering policy. Int J Prod Econ 144(1):200–208
2. Carr S, Duenyas I (2000) Optimal admission control and sequencing in a make-to-stock make-to-order production system. Oper Res 48(5):709–720
3. Chen X, Tai AH, Yang Y (2014) Optimal production and pricing policies in a combined make-toorder/ make-to-stock system. Int J Prod Res. doi:10.1080/00207543.2014.932930
4. Denton B, Gupta D, Jawahir K (2003) Managing increasing product variety at integrated steel mills. Interfaces 33:41–53
5. Ghrayeb O, Phojan N, Tan BA (2009) A hybrid push/pull system in assemble-to-order manufacturing environment. J Intell Manuf 20: 379–387
6. Gupta D, Weerawat W (2006) Supplier–manufacturer coordination in capacitated two-stage supply chains. Eur J Oper Res 175:67–89
7. Hemmati S, Rabbani M (2010) Make-to-order/make-to-stock partitioning decision using the analytic network process. Int J Adv Manuf Technol 48(5–8):801–813
8. Kalantari M, Rabbani M, Ebadian M (2011) A decision support system for order acceptance/rejection in hybrid MTS/ MTO production systems. Appl Math Model 35(3):1363–1377
9. Köber J, Heinecke G (2012) Hybrid production strategy between make-to-order and make-to- stock – a case study at a manufacturer of agricultural machinery with volatile and seasonal demand. Proc CIRP 3:453–458
10. Lu JC, Yang T, Su C-T (2012) Analysing optimum push/pull junction point location using multiple criteria decision-making for multistage stochastic production system. Int J Prod Res 50(19):5523–5537
11. Morikawa K, Takahashi K, Hirotani D (2014) Make-to-stock policies for a multi stage serial system under a make-to-order production environment. Int J Prod Econ 147:30–37
12. Perona M, Saccani N, Zanoni S (2009) Combining make-to-order and make-to-stock inventory policies: an empirical application to a manufacturing SME. Prod Plan Control 20(7):559–575
13. Rafiei H, Rabbani M (2012) Capacity coordination in hybrid MTS/ MTO production environment. Int J Prod Res 50(3):773–789
14. Serwer A (2002) Dell does domination. Fortune Mag 145(2):70–75
15. Sharda B, Akiya N (2012) Selecting make-to-stock and postponement policies for different products in a chemical plant: a case study using discrete event simulation. Int J Prod Econ 136(1):161–171

16. Soman CA, van Donk DP, Gaalman G (2004) Combined make-to-order and make-to-stock in a food production system. Int J Prod Econ 90:223–235
17. Soman CA, van Donk DP, Gaalman G (2006) Comparison of dynamic scheduling policies for hybrid make-to-order and make-to-stock production systems with stochastic demand. Int J Prod Econ 104(2):441–453
18. Sox CR, Thomas LJ, McClain JO (1997) Coordinating production and inventory to improve service. Manag Sci 43(9):1189–1197
19. Zaerpour N, Rabbani M, Gharegozli AH, Tavakkoli-Moghaddam R (2009) A comprehensive decision making structure for partitioning of make-to-order, make-to-stock and hybrid products. Soft Comput 13(11):1035–1054
20. Zhang ZG, Kim I, Springer M, Cai G, Yu Y (2013) Dynamic pooling of make-to-stock and make-to-order operations. Int J Prod Econ 144:44–56
21. Williams TM (1984) Special products and uncertainty in production/inventory systems. Eur J Oper Res 15(1):46–54