University of Groningen

Integrating make-to-order and make-to-stock in job shop control (Reprint of International Journal of Production Economics 185, pp 1-10)

Beemsterboer, Bart; Land, Martin; Teunter, Ruud; Bokhorst, Jos

Published in:
International Journal of Production Economics

DOI:
[10.1016/j.ijpe.2017.11.002](https://doi.org/10.1016/j.ijpe.2017.11.002)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Beemsterboer, B., Land, M., Teunter, R., & Bokhorst, J. (2017). Integrating make-to-order and make-to-stock in job shop control (Reprint of International Journal of Production Economics 185, pp 1-10). *International Journal of Production Economics*, 194, 3-12. <https://doi.org/10.1016/j.ijpe.2017.11.002>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Reprint of “Integrating make-to-order and make-to-stock in job shop control”[☆]

Bart Beemsterboer, Martin Land, Ruud Teunter^{*}, Jos Bokhorst

University of Groningen, Dept. Operations, PO Box 800, 9700 AV Groningen, The Netherlands

ARTICLE INFO

Keywords:

Production
Hybrid production
Job shop
Job shop control
Discrete event simulation

ABSTRACT

Demand fluctuations in make-to-order job shops lead to utilisation fluctuations and delivery delays, particularly in periods with high demand. Many job shop production companies therefore include some standardised products in their product mix and use a hybrid make-to-order/ make-to-stock production approach. However, as the control of a make-to-order job shop is usually focused on meeting due dates, which do not apply to make-to-stock items, integrating make-to-stock items in the control of a job shop is not straightforward. We propose four methods of integrating make-to-stock items in the control of a job shop and evaluate these using discrete event simulation in Python. We show that a popular but simple method of always giving priority to MTO items is strongly outperformed by more advanced methods of integrating MTS into job shop control as we have been able to reduce the MTS lost sales by a considerable 60%. We further show that considering up-to-date status information may improve the performance substantially. Loosening the due dates of MTS replenishment orders during periods of low MTS demand enables better use of the production capacity to maximise the delivery performance of MTO instead.

1. Introduction

Offering differentiated products is essential for many production companies. Job shop production is the ultimate form of high variety manufacturing, with orders following different routings along multi-purpose machines. Most products are unique; they are typically produced in a make-to-order fashion. However, operating entirely in a make-to-order fashion implies that demand fluctuations lead to utilisation fluctuations and delivery delays, particularly in periods with high demand (Land et al., 2015). We observe that many job shops in practice produce some standard products to stock in periods with low demand and deliver them from inventory. Combining make-to-order (MTO) and make-to-stock (MTS) is known as hybrid production.

However, including MTS items in a generally MTO oriented job shop leaves production planning and control with a challenge. As MTO is subject to stringent due dates, existing planning and control methods generally focus on achieving a high delivery performance. For MTS, this is not directly applicable because delivering on time boils down to sustaining a sufficient amount of inventory. If dispatching decisions are

commonly based on due dates of MTO items, it is not obvious what dispatching decisions between MTO and MTS items should be based on. A common approach in practice, which is also seen in the literature (Federgruen and Katalan, 1999; Hadj Youssef et al., 2004; Chang and Lu, 2010; Fernandes et al., 2015), is to always give priority to MTO items, which is obviously suboptimal. From a practical point of view, it would be ideal to integrate MTS items into the job shop control mechanism that is applied to MTO. A straightforward approach to do so is to define appropriate operation due dates for both MTO and MTS items and select the most urgent order at the time of dispatching. In this paper we aim to develop such a control mechanism and compare several design options in a simulation study.

Apart from several studies assuming simple unified production systems, few studies have considered a control mechanism for hybrid job shop systems. Most of them relate to very specific cases of job shop production. Chang et al. (2003) and Wu et al. (2008) develop comprehensive, purpose-built planning methodologies for the case of semiconductor foundries. Fernandes et al. (2015) study a hybrid make-to-order/assemble-to-order system where, besides MTO items, the

DOI of original article: <https://doi.org/10.1016/j.ijpe.2016.12.015>.

[☆] This article is a reprint of a previously published article. For citation purposes, please use the original publication details; International Journal of Production Economics, 185, pp. 1–10.

^{*} Corresponding author.

E-mail addresses: b.j.beemsterboer@rug.nl (B. Beemsterboer), m.j.land@rug.nl (M. Land), r.h.teunter@rug.nl (R. Teunter), j.a.c.bokhorst@rug.nl (J. Bokhorst).

semi-finished products of a number of items are made to stock, which are then completed on an MTO basis. The authors focus on order release, and in the final prioritisation at workstations they always give priority to the MTO items. [Schönsleben \(2011\)](#) develops a comprehensive methodology that provides for the full planning and control of a hybrid production job shop. However, the logic of this approach was never tested and will be included in our study.

No one has yet dealt with basic issues of integrating MTS items into an existing control method of a make-to-order job shop. Nevertheless, many hybrid job shop production companies are confronted with this problem as existing job shop control methods are inapplicable to MTS items. Therefore, we address the following research question:

- How can MTS items be integrated into the control of a make-to-order job shop oriented on due dates?

This paper is structured as follows. In [Section 2](#), we discuss the existing theories on planning and control in hybrid production systems. [Section 3](#) deals with the experimental design and proposes four methods for integrating MTS items into the control of the job shop specified. [Section 4](#) evaluates the methods using a discrete event simulation approach. [Section 5](#) summarises our conclusions.

2. Literature review

Literature on planning and control of make-to-order job shops is much more extensive than for hybrid production systems. Early contributions consider either a mathematical optimisation approach (see e.g. [Graves, 1981](#)), or dispatching rules for shop floor control. [Panwalkar and Iskander \(1977\)](#), [Blackstone et al. \(1982\)](#) and [Ramasesh \(1990\)](#) all provide thorough revisions of various types of dispatching methods. [Kanet and Hayya \(1982\)](#) focus on dispatching methods based on operation due dates and compare several methods of obtaining these operation due dates. [Land et al. \(2014\)](#) compare operation due date based dispatching methods in conjunction with order release mechanisms.

Relatively few contributions address the control of hybrid production systems and most have a scope different from job shop production. An early contribution is that by [Williams \(1984\)](#), who considers a hybrid production system where the stock process and production process behave independently. The system is not modeled as a job shop, but it has a fixed number of identical machines. When the inventory of one of the MTS items reaches a certain threshold value, a production order is generated. The production process then runs via First Come, First Served (FCFS). [Federgruen and Katalan \(1999\)](#) consider a production system where an MTO item is added to an existing MTS production system. They compare several planning policies: two where priority is given to MTO demands, in which MTS is either preempted or not; and one where the MTO item is added to the existing MTS production sequence. [Tsubone et al. \(2002\)](#) consider a hybrid production system on an aggregate planning level, mainly focusing on capacity allocation. [Hadj Youssef et al. \(2004\)](#) derive optimal stock levels for two control policies of a hybrid production system, namely FCFS and priority for the MTO items. They identify circumstances under which either of the two should be selected. [Chang and Lu \(2010\)](#) analyze a hybrid production system that produces standard MTS products and customised MTO products by performing an additional customisation operation to finished MTS items. They assume that customisation operations always precede production of standard items.

[Soman et al. \(2004\)](#) provide a literature study on hybrid make-to-order/make-to-stock production planning in the food processing industry. After reviewing contributions that address various research questions regarding hybrid production systems, the authors provide a hierarchical planning framework. In a follow-up study, [Soman et al. \(2006\)](#) evaluate the performance of several existing MTS scheduling policies in a hybrid production system to find out whether the proven policies for MTS systems are also suitable for hybrid systems. They

integrate MTO items by using the value of the due date allowance instead of the value for the run-out time, which is done for MTS items, in determining the priority of items to be produced. Continuing this line of research, [Soman et al. \(2007\)](#) apply their hierarchical planning framework to a case study for a food processing company.

A number of contributions take a closer look at detailed production decisions in hybrid production systems. [Carr and Duenyas \(2000\)](#), [Iravani et al. \(2012\)](#) and [Beemsterboer et al. \(2016\)](#) consider stylised models with one MTO and one MTS product, and one unit of production capacity. All use Markov Decision Processes to find optimal policies and study their structure. [Carr and Duenyas \(2000\)](#) and [Iravani et al. \(2012\)](#) consider this from the perspective of contract manufacturers, who produce on an MTS basis for contracted and hence more important customers, and on an MTO basis for others. [Beemsterboer et al. \(2016\)](#) explore the benefits of a planning method that does not take the perspective of either product type, by taking relevant aspects of both product types into account and determining production planning decisions based on both the MTS inventory level and the state of the MTO order book. They conclude that substantial benefits can be obtained when the state of both MTO and MTS is taken into account in the planning policy of a hybrid production system.

All contributions on hybrid production control so far consider the production system as a unified system, i.e. not distinguishing between individual workstations. Hence, none of them addresses the problem of selecting between MTO and MTS items at individual workstations in hybrid production systems. However, this is done by, for instance, [Chang et al. \(2003\)](#) and [Wu et al. \(2008\)](#). Both develop sophisticated planning procedures for the control of a hybrid semiconductor foundry. The dispatching method that [Chang et al.](#) propose distinguishes four priority levels for both MTO and MTS, and classifies items based on their routing and the work in progress in front of the machine. [Wu et al.](#) adapt a control method called Starvation Avoidance ([Glassey and Resende, 1988](#)). They require MTO to be released according to the just-in-time principle, i.e. it should be released on time but preferably not too early. This provides the opportunity to release MTS items to fill remaining capacity. Dispatching at the workstations takes place on the basis of the Critical Ratio (CR) for MTO items. The choice between MTO and MTS is made separately, on the basis of the idea that an MTS item can be produced in case none of the MTO items in the queue becomes operation-delayed (see [Wu et al., 2008](#) for details); otherwise the MTO item with the highest Critical Ratio is produced.

[Fernandes et al. \(2015\)](#) consider the control of a hybrid make-to-order/assemble-to-order system, where some of the products are made to order and some semi-finished products are stocked and finished on order in a second stage. They apply workload control and address two research questions regarding order release: which order release method should be applied, and where in the production system should release be controlled? Regarding the latter, they compare three possibilities: release control only before stage I, release control only before stage II, and release control before both stages. Regarding shop floor control, the authors assume that MTO items have priority above replenishment orders in the first production stage.

[Schönsleben \(2011, p. 731–739\)](#) presents an advanced planning principle for hybrid job shops called Capacity-Oriented Materials Management (Corma). It covers multiple areas in production control, namely order release, shop floor control, and inventory control. The main element of the principle is that, besides a regular release method, MTS replenishment orders are released to the shop floor when generally well utilised work centers are about to starve and these MTS replenishment orders can keep them busy. All MTS replenishment orders receive a due date based on the demand rate, a planned lead time, and a safety stock. This due date is in turn used in the shop floor control, leading to planned starting times for each operation. These planned starting times are recalculated periodically, for instance to make room for newly accepted MTO orders. The planned starting times are also updated in case the demand is higher or lower than expected to speed up or slow down the

MTO replenishment order accordingly.

The control of hybrid production systems has thus been studied from various perspectives. Few studies have addressed the question of selecting between MTO and MTS orders at the individual workstation level in job shop production, and these considered very specific environments or planning frameworks. We observe that the current literature does not provide the basic knowledge how to integrate MTS items in the control of a job shop and make the basic decision which order to give priority when a workstation becomes available. Previous studies generally assume a method of integration in order to focus on other research questions regarding hybrid production planning. We bridge this gap by defining four basic methods to integrate MTS and prioritise orders in a make-to-order job shop. The methods include basic principles that have been assumed in earlier studies. Their performance is compared and assessed using a discrete event simulation approach.

3. Experimental design and specification of integration methods

3.1. Production system

We consider the following production system for our simulation study. We have a production facility that contains six workstations and applies two ways of production. Customised products, with varying routings through the facility and varying processing times at the workstations, are produced on an MTO basis. The facility further produces one type of standardised products on an MTS basis, with a fixed routing and constant processing times.

The configuration closely resembles the job shop of Melnyk and Ragatz (1989), which has served as a standard in job shop research (see also Land et al., 2015). Demands for both types of products follow Poisson processes. Each demand concerns one unit of either MTO or MTS. For MTO items, the number of workstations in the routing of the job is a random number from 1 to 6, each with probability 1/6. Also the workstations included in the routing are drawn randomly, where each workstation has an equal probability of being selected. The routings are then subject to a ‘dominant flow direction’ (see Oosterman et al., 2000). More specifically, the workstations are numbered as 1–6 and are always visited in the order of increasing workstation number. The processing times of each of the operations is taken from a 2-Erlang distribution with a mean of 1 time unit and is assumed to be known from the moment of ordering. Since each MTO job visits on average 3.5 out of 6 workstations, this leads to a mean load contribution of 7/12 time units for each workstation. MTS items require processing at each of the six workstations, again in the order from 1 to 6, each with a fixed processing time of 1. Hence, the mean load contribution of an MTS job is 1 time unit for all six workstations. The demand rates are obtained from the target utilisation, which we set at 90%, and a target load division of 80% MTO and 20% MTS. Using the load contributions of 7/12 for MTO and 1 for MTS, we obtain arrival rates of approximately $\lambda_o = 1.234$ for MTO and of $\lambda_s = 0.18$ for MTS.

MTO jobs must be delivered before a due date, which is obtained as the time of ordering plus a random allowance. This allowance, in turn, is obtained randomly from a uniform distribution with minimum 30 and maximum 40 time units. MTS demands are fulfilled from inventory when available and are otherwise lost. Since lost sales are not processed, this decreases the actual utilisation compared to the target utilisation. MTS stock is controlled by using the well known base stock policy, that keeps a constant base stock level or inventory position (which is equal to the inventory on hand plus the inventory on order). Hence, each MTS demand triggers a replenishment order of size one. We set the base stock level at 13, 20, and 30 to determine its effects in the main analysis. The range of the MTO due date allowance and the MTS base stock levels are obtained from initial simulation runs in which we aimed to achieve both MTO tardiness and MTS lost sale percentages in the realistic range of 0–20%. Table 1 summarises the properties of the production system.

We further assume that the machines are always available and their capacities are fixed, i.e. they cannot be extended in busy periods. We also

do not consider machine changeover times. These simplifying assumptions avoid side effects that distract from our main research interest.

As we are interested in integrating MTS into the control of a make-to-order job shop, we have to define a standard dispatching policy into which we can integrate MTS. We select Operation Due Dates (ODD) as the primary dispatching method and we briefly discuss this dispatching method in the next subsection.

3.2. Primary dispatching rule: operation due dates

Our setting concerns make-to-order aimed at delivery performance, and dispatching rules based on this performance criterion have been shown to perform best (Kanet and Hayya, 1982). We therefore select Operation Due Dates (ODD) as the standard dispatching policy. This method defines ‘operation due dates’ for each operation of a job. When a workstation becomes available, the item with the earliest operation due date is selected for production.

Several approaches to calculating operation due dates have been developed, and allowing a fixed amount of time to each of the operations in the routing of a job was proven to function particularly well in situations with uncontrolled order release (Land et al., 2014), as we apply in this study. Following Land et al. (2015), we allow 5 time units for each operation. This is also in line with the due date allowance lower bound of 30 time units, which is just sufficient for jobs that require processing at all six workstations. More specifically, the operation due date of a job at workstation k , which we denote as D_k , is obtained from its final due date D as follows.

$$D_k = D - 5 \times (W - j_k), \quad (1)$$

where W is the number of workstations in the routing of the job and j_k is the position of workstation k . For example, if the final due date of a job is $D=24$ and its routing 1–3–6, we have $W=3$, $j_1 = 1$, $j_3 = 2$ and $j_6 = 3$ and we obtain operation due dates with (1) as $D_1 = 24 - 5 \times (3 - 1) = 14$, $D_3 = 24 - 5 \times (3 - 2) = 19$ and $D_6 = 24 - 5 \times (3 - 3) = 24$. We leave D_2 , D_4 , D_5 , j_2 , j_4 and j_5 undefined as they are not relevant for this example.

Two of the four integration methods that we will introduce in Section 3.4 integrate MTS by determining a fictitious final due date D and then treating MTS items in the same way as MTO items. As all MTS items have job routing 1–2–3–4–5–6, application of (1) to these MTS items for the concerning integration methods simplifies to

$$D_k = D - 5 \times (6 - k). \quad (2)$$

Note that more complex dispatching methods exist, such as Modified Operation Due Dates (Land et al., 2015). The strength of that method is that in periods of high workload, prioritizing jobs on their operation due date will not lead to satisfying as many due dates as possible, but rather to a large number of violations by a small amount of time. Therefore, in high load periods, this method prioritises jobs based on the Shortest Processing Time. In our setting the postponement of MTS replenishment orders can fulfill the same task of preventing that many MTO orders exceed their due dates in temporary periods of high workload.

3.3. Performance measures

As indicated in Section 1, the planning goals of MTO and MTS production are different, and so are the performance measures associated with them. For MTO, the general goal is a timely delivery of as many jobs as possible. Related performance measures consist of the percentage of tardy jobs, the mean tardiness, the mean lateness, and the standard deviation of the lateness. In our presentation of the results, we focus on the percentage of tardy jobs, though other measures have been recorded to check whether a low percentage tardy was not realised at the cost of a larger mean tardiness. A timely job completion is also one of the goals for MTS production. Here it concerns avoidance of stock outs, so as to avoid

lost sales. We will use the percentage of lost sales as a relevant MTS performance measure.

As we are specifically interested in a situation where MTO and MTS are produced in the same facility, the performances of both types should be taken into account. There is a natural trade-off between the MTO and the MTS performance. Giving more priority to MTO in order to obtain a better due date performance implies less priority for MTS, leading to an increased probability of stock outs, and vice versa. Hence, one policy might outperform another in terms of MTO performance, whereas the second outperforms the first one on MTS performance. To have a meaningful comparison, we will consider the two performances simultaneously. We will therefore use parameterised policies, with a parameter that controls the focus between MTO and MTS. The trade-offs will then be shown using *performance frontiers*, depicting the MTO and MTS performance in two-dimensional graphs.

3.4. Integration methods

We propose four methods for integrating MTS into the control of our job shop setting. We compare the performance of these methods in Section 4.

The methods either determine a final due date D for the MTS items and subsequently obtain operation due dates through (2), or obtain operation due dates for MTS items right away. Note however that the integration method will only affect decisions between MTO and MTS. As the operation due dates of MTO items are unaffected, dispatching decisions between different units of MTO are made in exactly the same way for any integration method. Moreover, dispatching between different units of MTS is not meaningful as they are identical by definition. In our simulation, we handle them using First Come, First Served.

In describing the integration method, we distinguish two kinds of adaptations: (1) actions that take place when MTS is reordered, and (2) additional actions at dispatching moments on the shop floor, on top of the dispatching via ODD.

Arguably the most straightforward way of integrating MTS is by defining a fictitious due date for each MTS replenishment order and treating them in the same way as MTO items, yielding the following integration method.

Fixed MTS Due Dates

Actions at the time of reordering MTS

All MTS replenishment orders receive a fictitious due date. If we denote the time of ordering as t and if we define a fixed due date allowance as δ , the fixed due date of a replenishment order D is given as

$$D = t + \delta. \quad (3)$$

After obtaining D , operation due dates D_k are obtained through (2). The due date allowance δ is the same for all MTS replenishments and is used as the tuning parameter of the integration method.

Actions when dispatching takes place

There are no additional actions when dispatching takes place. MTO and MTS items are considered simultaneously for shop floor dispatching decisions via ODD.

Note that satisfying the MTS due date is not an objective on its own. MTS replenishment orders that are completed after the due date should not be considered as ‘tardy jobs’, as the due date is fictitious and the stock level may have been sufficient, even with the replenishment arriving later than planned (or the other way around, stockouts may occur even though the replenishment arrives before its due date). The performance of MTS is determined by the number of lost sales only and the method parameter δ controls the fraction of demand that is lost. A tighter due date allowance δ increases the focus on MTS, leading to a lower probability of lost sales, but at the same time it decreases the focus on MTO and it thereby inevitably leads to a higher probability of tardiness.

As noted by Beemsterboer et al. (2016), control decisions in hybrid production systems between MTO and MTS should be based on the state of both types of items. Regardless of the integration method we choose,

the MTO status is fully taken care of as we integrate MTS in an existing MTO job shop control method. In our case, the ODD method is closely related to the MTO objective of avoiding tardiness. However, the integration method Fixed MTS Due Dates does not satisfy the need to determine decisions that are also based on the MTS status. For this, we need to take, for instance, the actual inventory level into account in the shop floor control, which more closely relates to the MTS performance than a fixed fictitious due date. Furthermore, using a due date that is fictitious allows us to redefine it whenever we desire. We therefore propose the following two dispatching methods that, besides the MTO status, also take an actualised MTS status into account in dispatching decisions.

Dynamic MTS Due Dates

Actions at the time of reordering MTS

None; MTS replenishment orders are sent to the shop floor without calculating additional properties.

Actions when dispatching takes place

We calculate a due date as the expected unit consumption time (based on the MTS units in stock and those downstream) plus a (possibly negative) additional amount of time. More specifically, if we denote the MTS inventory level as I and the number of MTS units in production and downstream of the considered unit as P , and if we define α as a parameter representing an additional amount of time, then the fictitious due date of the MTS replenishment order D is given by

$$D = t + \frac{I + P}{\lambda_s} + \alpha. \quad (4)$$

After this, MTS operation due dates are obtained via (2) and then ODD dispatching takes place in which the MTO and MTS items are considered simultaneously.

As for the previous method, higher values of α correspond with a smaller focus on MTS and vice versa. For $\alpha = 0$, the method is based on the pure expected unit consumption time. We can equally well use negative values of α , in case MTS has a higher priority than MTO.

The control part of the Corma method by Schönsleben (2011) resembles the above integration method. However, we deviate in several aspects. For instance, Corma assumes using a safety stock for the MTS items. By contrast, we work with an in essence ‘neutral’ parameter α that may be negative (in which case it is a ‘safety time’, equivalent to a safety stock) as well as positive (in which case the method gives in general more priority to MTO than to MTS). Furthermore, we do not adopt a periodic recalculation of the due dates, but recalculate at all dispatching instances.

A modified version of the above integration method that takes the slack per remaining operation for the MTS items into account is the following.

Dynamic MTS Slack Per Remaining Operation

Actions at the time of reordering MTS

None.

Actions when dispatching takes place

For MTS items, we calculate the operation due date allowance as the slack per remaining operation plus a (possibly negative) additional amount of time. More specifically, if we denote the remaining processing time of the job as T , the remaining number of operations as N , a parameter representing the additional time as β , and D , I and P as above, the fictitious operation due date of the MTS replenishment order is obtained as

$$D_k = t + \frac{I + P}{\lambda_s} - \frac{T}{N} + \beta. \quad (5)$$

Then ODD dispatching takes place in which MTO and MTS are considered simultaneously.

Again, higher values of the method parameter (here β) imply that the method focuses more on MTS, and vice versa.

We could simplify (5) if we apply it to our setting, as we have six workstations that are all contained in the MTS routings with a processing time of one time unit, holding that $S = N = 7 - k$. However, we prefer to present the integration method in a more general formulation here.

Table 1
Production system properties.

Main	Production system	Job shop
	Number of workstations	6
MTO	Demand distribution	Poisson with $\lambda_o \approx 1.234$
	# workstations to be visited	1 to 6, each with prob. 1/6
	Routing	Dominant flow direction; 1→6
	Processing times	Erlang-2 with mean 1 time unit
	Due date allowance distribution	Uniform (30, 40)
MTS	Demand distribution	Poisson with $\lambda_s = 0.18$
	# workstations to be visited	6
	Routing	1 - 2 - 3 - 4 - 5 - 6
	Processing times	1 on all workstations
	Replenishment policy	Base stock
	Base stock levels	13,20,30

Table 2
Parameter values of the integration methods.

Method (tuning parameter)	Base stock level	Tuning parameter values
Fixed	13	0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140
MTS Due	20	30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 180, 200
Dates (δ)	30	70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 200, 220, 240
Dynamic	13	-60, -55, -50, -45, -30, -25, -20, -15, -10, -5, -2, -1, 0, 2, 5, 10, 20
MTS Due	20	-60, -55, -50, -45, -30, -25, -20, -15, -10, -5, -2, -1, 0, 2, 5, 10, 20
Dates (α)	30	-40, -30, -25, -20, -15, -10, -5, -2, 0, 2, 5, 10, 20
Dynamic	13	-50, -45, -40, -35, -30, -25, -20, -15, -10, -5, -2, -1, 0, 2, 5, 10, 15, 20
MTS S/RO	20	-40, -35, -30, -25, -20, -15, -10, -5, -2, -1, 0, 2, 5, 10, 15, 20
(β)	30	-40, -30, -25, -20, -15, -10, -5, -2, -1, 0, 2, 5, 10, 20
Rolling	13	0, 1, 2, 4, 6, 8, 10, 12, 15, 20, 25
MTS ODDs	20	0, 2, 4, 6, 8, 10, 12, 15, 20, 30, 40
(γ)	30	0, 2, 4, 6, 8, 10, 12, 15, 20

The previous integration methods are mainly built on the rationale that MTS items should be completed on time so as to avoid lost sales. On the shop floor, the methods do not distinguish between MTO and MTS apart from looking at their operation due dates. As in job shops the MTO items are generally more important, we could integrate MTS by regarding it as a secondary item and only allowing it to be dispatched when the MTO items in the queue offer a sufficient slack with respect to their operation due dates. We can postpone MTS items as long as there are MTO items in the queue with an operation due date that is not further than some allowance γ away from now, as the following integration method describes.

Rolling MTS Operation Due Dates

Actions at the time of reordering MTS

None.

Actions when dispatching takes place

We define D_k as the operation due date at workstation k and γ as parameter representing a fixed MTS operation due date allowance. Then, the MTS items receive an operation due date given as

$$D_k = t + \gamma, \quad \text{for any } k. \quad (6)$$

After this, ODD dispatching takes place in which MTO and MTS are considered simultaneously.

The parameter γ controls the focus of the method on MTO and MTS in the same way as the parameters of the other methods. However, in this case negative values or a zero value of γ are not very meaningful as these would hold that MTS items would have priority above MTO items even when these are already too late for the concerning operation.

We finally remark that the assumption of organizing dispatching via the ODD method, which assigns an equal amount of time to each of the operations, is not required for any of the four integration methods. The first two methods, Fixed and Dynamic MTS Due Dates, obtain final due dates D for MTS items, which could be used in any dispatching method that is based on these rather than on the operation due dates. For instance, dispatching by giving priority to the job with the least slack per remaining operation would yield a meaningful alternative. There are also deviations of the general ODD method, such as Modified Operation Due Dates (Land et al., 2015) or an ODD variant that assigns operation due dates by equally dividing the slack over all operations in the routing of the job.

3.5. Reference methods

We introduce the following two reference integration methods. The first always prioritises MTO and is commonly used in the literature, for instance by Chang and Lu (2010) and Fernandes et al. (2015) who use this concept to integrate MTS into their planning framework. Hadj Youssef et al. (2004) and Federgruen and Katalan (1999) use the method for a comparison with different methods.

MTO Priority

Actions at the time of reordering MTS

None.

Actions when dispatching takes place

MTO items have priority over MTS items. Selection between MTO items is based on ODD.

The second method gives priority to MTS and has not been proposed or used by anyone, to the best of our knowledge, but we include this final integration method for completeness.

MTS Priority

Actions at the time of reordering MTS

None.

Actions when dispatching takes place

MTS items have priority over MTO items. Selection between MTO items is based on ODD.

3.6. Experimental setup

The experimental factors in the main analysis are the integration method (Fixed MTS Due Dates, Dynamic MTS Due Dates, Dynamic MTS Slack Per Remaining Operation, Rolling MTS Operation Due Dates, MTO Priority, and MTS Priority), the base stock level (13, 20, 30), and the tuning parameter of the integration method, except for the reference methods. The parameter ranges for δ , α , β , and γ are chosen such that it shows as much as possible of the range of the performance that the method can achieve, ranging from near 0% MTO orders tardy to 0% MTS lost sales when the method would allow for it. The ranges are obtained through initial simulation runs and specified in Table 2.

In the next section, we carry out a sensitivity analysis to determine the robustness of the methods with respect to various parameter settings. These parameters are the target utilisation, the MTO due date allowance, the operation due date parameter, the processing time variance for MTO jobs, the ratio of MTO and MTS within the demand, and the relative size of MTS products in terms of processing time. The simulations are executed in the discrete event simulation package SimPy, in Python. Each experiment consists of 100 replications of 10,000 time units, with an additional warm-up period before each replication. The warm-up period

Fig. 1. Performance frontiers of Fixed MTS Due Dates for base stock levels 13, 20 and 30.

is 3,000 time units, which was shown to be more than sufficient by a graphical inspection of the results.

Claims that we make in this section regarding differences in the performance of the methods are based on significance of a paired t-test. This approach complies with the use of common random numbers (Kelton and Law, 2000) as used in this study.

4. Simulation results

4.1. Influence of the base stock level and tuning parameters

We first take a closer look at the impact of the base stock level and the tuning parameter for the first method, Fixed MTS Due Dates.

Fig. 1 presents the results in the form of performance frontiers. Each curve shows combinations of the percentage of tardy MTO jobs and the percentage of MTS lost sales that result from the same base stock level by varying the tuning parameter. It is clear that also the base stock level can be used as a control parameter to influence MTO tardiness and MTS lost sales. Lower percentages for both performance measures at the same time can be realised by increasing the base stock level. Obviously, the holding costs will increase for higher base stock levels and a balance must be found between the holding costs and the probability of lost sales. Since we leave the base stock value fixed for each graph, the holding costs remain equal and effects on MTS performance are fully reflected in the lost sales measure.

For instance, in order to bring the percentage of MTO orders tardy below 2%, we have to incur around 15% MTS lost sales with a base stock level of 13, around 7% with a base stock level of 20, and around 2% with a base stock level of 30. When changing the parameter δ (moving along the curve), we are able to achieve a near-perfect performance on either the MTO tardiness or the amount of MTS lost sales, although a near perfect score for one product type implies a weak performance regarding the other.

4.2. Influence of the integration method

Fig. 2 shows the performance frontiers (one for each considered base stock level) obtained from comparing all integration methods. First of all, we observe that also Dynamic MTS Due Dates and Dynamic MTS Slack Per Remaining Operation are able to achieve a near-perfect performance on one of the two measures, but only at the expense of a poor performance on the other. Moreover, both perform better than Fixed MTS Due Dates across almost the entire frontier. For any fixed value of the MTO tardiness, these methods achieve fewer MTS lost sales. Equivalently, for any amount of MTS lost sales they reach a lower tardiness.

As could be expected, a higher base stock level can be used in better

Fig. 2. Relative performance of methods for each of the considered base stock levels.

MTO/MTS performance combinations for all methods. But since the three graphs show that in general the base stock level does not affect the relative performance of the methods, we will focus on the results for a base stock level 20 in our discussion below.

The amounts by which the Dynamic MTS Due Dates and the Dynamic MTS Slack Per Remaining Operation outperform Fixed MTS Due Dates vary across the frontier. For instance, if we would be willing to accept about 2% MTO tardiness under a base stock level of 20, we can reach between 6% and 7% MTS lost sales with Fixed MTS Due Dates, but we only have to accept about 3% with Dynamic MTS Due Dates and even only 2.5% with Dynamic MTS Slack Per Remaining Operation. If we reverse the perspective, reaching 2% MTS lost sales requires accepting an MTO tardiness of again between 6% and 7% for Fixed MTS Due Dates, and only about 4% for Dynamic MTS Due Dates and about 3% for Dynamic MTS Slack Per Remaining Operation.

Comparing the two dynamic methods, we observe that Dynamic MTS Slack Per Remaining Operation outperforms Dynamic MTS Due Dates at

Fig. 3. Frequency distributions of observed finished goods inventory level.

the right hand side of the curves, i.e. when giving priority to MTO performance. For low values of MTS lost sales, Dynamic MTS Slack Per Remaining Operation results in higher MTO tardiness than Dynamic MTS Due Dates.

As could be expected, the two reference integration methods, MTO and MTS Priority, both achieve a near-perfect performance on the product type they give priority to, but in all cases at the cost of a very bad performance of the other. For example the popular method of giving priority to MTO regardless of the state of the system incurs a substantial amount of 15% MTS lost sales at a base stock level of 20. At this base stock level, three of the four proposed integration methods, the two dynamic methods and Rolling MTS Operation Due Dates, all achieve a similar MTO tardiness performance when incurring about 6% MTS lost sales. The strong performance of Rolling MTS Operation Due Dates at the right hand side of the curves provides interesting opportunities for practice. For companies focusing on MTO performance this method is rather straightforward, as it simply prioritises MTO items as long as their operation due date is within a certain threshold value. However, the method results in extreme MTO tardiness when the MTS lost sales should be decreased below a certain point. Concluding, the results strongly advocate using a more advanced integration method for hybrid job shops that focus on MTS performance.

4.3. Examining the performance differences

In order to better understand the performance differences, we examined and compared two experiments in more detail. Given a base stock level of 20, we found similar amounts of MTS lost sales for the instance of Fixed MTS Due Dates with $\delta = 90$ (2.3%) and the instance of Dynamic MTS Slack Per Remaining Operation with $\beta = -2$ (2.2%).

Fig. 4. MTS inventory level and MTO lateness realisation over time for Fixed MTS Due Dates with $\delta = 90$ and Dynamic MTS Slack Per Remaining Operation with $\beta = -2$.

However, Dynamic MTS Slack Per Remaining Operation significantly outperforms Fixed MTS Due Dates on MTO tardiness with 2.7% against 5.8%. We examine the frequency distributions of the finished goods inventory level observed throughout the simulation in Fig. 3 in order to explain the difference.

Both instances achieve a rather similar occurrence probability of inventory level 0, which is a direct result of our choice to examine two instances with similar MTS lost sale percentages. However, the differences for higher inventory levels are striking. Despite similar lost sale levels, the finished goods inventory levels are centered around eight items for Fixed MTS Due Dates and around four items for Dynamic MTS Slack Per Remaining Operation. This indicates that Dynamic MTS Slack Per Remaining Operation is much better able to react to relatively low inventory levels, and resupply stock just in time. This flexibility allows the method to focus on MTO when this is needed. By contrast, Fixed MTS Due Dates treats the MTS due dates as if these were strict, even when the inventory is at a very high or low level. As a result, the method is less flexible in prioritizing MTO orders when these are close to their due dates, which explains the performance difference for the same percent age lost sales.

To further improve our understanding, we follow Land et al. (2015) and focus on temporary high load periods. In these periods all of the MTO tardiness and most of the MTS lost sales are realised. Hence, the performance difference between integration methods is determined by how they handle these high load periods.

Fig. 4 shows the MTS finished goods inventory level and realisations of the MTO lateness during the first 2000 time units of the first simulation

Table 3

Twelve varied parameter values in the sensitivity study.

	Target utilisation	DD allowance	ODD para-meter	MTO process time	MTS share	MTS product size
Main value	90%	30–40	5	Erlang-2	20%	1
Tested	88%	20–30	3	Erlang-3	10%	0.5
sensitivity values	92%	40–50	7	Erlang-1	30%	2

run after the warm-up period, again for Fixed MTS Due Dates and Dynamic MTS Slack Per Remaining Operation. The x-axis reflects the simulation time and the y-axis specifies both the inventory level for MTS in numbers of items (the black curve) and the MTO lateness in time units (the gray curve). For the MTO orders the lateness of each completed order is set against its time of completion and the resulting points are connected. The depicted first part of the run contains a long high load period from about $t=3600$ to $t=4600$ where many MTO orders get tardy. Here the performance difference between the two integration methods is the most apparent. We observe that especially during the high load period the inventory level held by Dynamic MTS Slack Per Remaining Operation is much lower than that held by Fixed MTS Due Dates. For

example in the circled period, around time 3800, Fixed MTS Due Dates builds up an inventory of 14 items successively resulting in several tardy orders. In the same period, Dynamic MTS Slack Per Remaining Operation does not exceed an inventory level of 6, while avoiding tardy orders. More generally, MTS stock buildups during periods of MTO lateness are avoided by Dynamic MTS Slack Per Remaining Operation.

4.4. Sensitivity study

In order to study the sensitivity of our results to the system parameters, we repeated our simulation for twelve different settings. We have selected six crucial system parameters of which we are interested in the

Fig. 5. Performance frontiers resulting from the sensitivity study (first three parameters).

influence on the results. These parameters are the target utilisation, the range of the MTO due date allowance, the range of the operation due date parameter, the variation of the processing time of MTO jobs, the ratio of MTO and MTS within the demand, and the size of MTS products in terms of processing time. We remark that the MTS product size was varied by adjusting the fixed processing times; the MTS demand rate was adapted correspondingly to ensure that the expected MTS workload is unchanged, and the inventory position was changed accordingly so that it stayed the same in terms of the MTS workload. So, a larger product size implies more intermittent and lumpy MTS demand (for processing time). We further remark that we do not vary the number of machines as preliminary results showed that this does not alter the main findings, and because doing so would not be straightforward as many system parameters have to be changed along with the number of machines to maintain comparability in for instance system utilisation.

In addition to the value used in the main analysis, we select a smaller and a larger value for each of the system parameters. All sensitivity analyses are performed under the mid-level base stock value of 20. When the target utilisation was varied we kept the MTO/MTS ratio fixed and

when the MTO/MTS ratio was varied we kept the target utilisation fixed by choosing appropriate values of the demand rates λ_o and λ_s . Table 3 gives an overview of the parameters varied in this sensitivity study.

Figs. 5 and 6 show the results of the sensitivity study (the reference methods are not shown when they perform outside the scope of the other methods). With the exception of the operation due date parameter, varied parameters do have a notable absolute performance effect. The sensitivity analysis reveals that the conclusions from the main analysis are mostly confirmed by the sensitivity results. The relative performance of the methods is unchanged except for the Dynamic MTS Slack Per Remaining Operation at low MTS lost sales. In all cases, Fixed MTS Due Dates is outperformed by Dynamic MTS Due Dates in terms of the percentage of MTO orders tardy, and also by Dynamic MTS Slack Per Remaining Operation and Rolling MTS Operation Due Dates when MTS lost sales are sufficiently high. Dynamic MTS Slack Per Remaining Operation performs the best for low values of the MTO tardiness.

The distances between the curves in Figs. 5 and 6 indicate the relative benefit that can be obtained by using a dynamic method compared to Fixed MTS Due Dates. In comparing these distances between left and

Fig. 6. Performance frontiers resulting from the sensitivity study (latter three parameters).

right panels of the figures, one should keep in mind that the ranges on the axis are different for most varied parameters. We observe a relatively larger benefit in cases with a higher target utilisation, a lower MTO due date allowance, a higher MTO processing time variation, and a higher MTS product size (and so a more lumpy MTS demand). So, the relative benefit is higher when capacity is tighter or when capacity requirements are more variable. Apparently, the more sophisticated dynamic method is able to cope better with more difficult situations. The ODD parameter and the MTS share have less affect on the relative benefit.

4.5. Implications for practice

Although we have identified Dynamic MTS Slack Per Remaining Operation as the best performing method in most settings, it is also the most advanced method, which comes at a price in terms of ease of use. Dispatching decisions require evaluating the expected consumption time of the first MTS unit in the queue each time a job has to be selected. However, our results show that if this would be too complex to implement or be too time consuming, one of the other methods can be used without large performance losses.

5. Conclusion

Integrating make-to-stock items into the control of a make-to-order job shop with a due date oriented control approach is not straightforward. A simple method is to define a (fixed) due date allowance for MTS items. Even if the focus of a company would be on the extremes of either realizing 0% MTS lost sales or 0% MTO tardy, it already performs better than always giving priority to MTS or MTO items.

Furthermore, the results of more advanced methods show that considering up-to-date status information may further improve the performance substantially. Loosening the due dates of MTS replenishment orders during periods of low MTS demand enables a better use of production capacity to maximise the delivery performance of MTO instead. When inventory levels drop faster than foreseen, lost sales can be avoided by tightening MTS due dates. The more advanced integration methods perform better across a variety of settings.

The results of a method that only prioritises MTS when there are no urgent MTO orders in the queue of a workstation imply that dispatching could also work completely without fictitious MTS due dates. However, this approach is only effective when the main focus of the company is on MTO delivery performance.

This study focused on integration of MTS into job shop dispatching methods. However, the research on planning and control of job shops takes many different approaches and scopes, providing numerous research opportunities on the integration of MTS. For instance, the concept of order release partly centralises the decision whether to prioritise MTO or MTS instead of leaving this to the shop floor control. Moreover, it avoids an early commitment of material resources to orders. Early release of MTS orders in temporary periods of low load may provide interesting opportunities. Another opportunity relates to MTS lot sizing

decisions. The production lot sizes for MTO items are given by the customer orders, but for MTS they can be chosen. In case the machines require setups, producing in larger lots may save production capacity, but this would reduce the responsiveness to incoming MTO orders that may have a tight due date. Future research could examine this trade-off in hybrid job shops.

References

- Beemsterboer, B., Land, M., Teunter, R., 2016. Hybrid mto-mts production planning: an explorative study. *Eur. J. Oper. Res.* 248, 453–461.
- Blackstone, J.H., Phillips, D.T., Hogg, G.L., 1982. A state-of-the-art survey of dispatching rules for manufacturing job shop operations. *Int. J. Prod. Res.* 20, 27–45.
- Carr, S., Duenyas, I., 2000. Optimal admission control and sequencing in a make-to-stock/make-to-order production system. *Oper. Res.* 48, 709–720.
- Chang, K.-H., Lu, Y.-S., 2010. Queueing analysis on a single-station make-to-stock/make-to-order inventory-production system. *Appl. Math. Model.* 34, 978–991.
- Chang, S.-H., Pai, P.-F., Yuan, K.-J., Wang, B.-C., Li, R.-K., 2003. Heuristic pac model for hybrid mto and mts production environment. *Int. J. Prod. Econ.* 85, 347–358.
- Federgruen, A., Katalan, Z., 1999. The impact of adding a make-to-order item to a make-to-stock production system. *Manag. Sci.* 45, 980–994.
- Fernandes, N.O., Silva, C., Carmo-Silva, S., 2015. Order release in the hybrid mto-fto production. *Int. J. Prod. Econ.* 170, 513–520.
- Glassey, C.R., Resende, M.G., 1988. Closed-loop job release control for vlsi circuit manufacturing. *IEEE Trans. Semicond. Manuf.* 1, 36–46.
- Graves, S.C., 1981. A review of production scheduling. *Oper. Res.* 29, 646–675.
- Hadj Youssef, K., Van Delft, C., Dallery, Y., 2004. Efficient scheduling rules in a combined make-to-stock and make-to-order manufacturing system. *Ann. Oper. Res.* 126, 103–134.
- Iravani, S.M., Liu, T., Simchi-Levi, D., 2012. Optimal production and admission policies in make-to-stock/make-to-order manufacturing systems. *Prod. Oper. Manag.* 21, 224–235.
- Kanet, J.J., Hayya, J.C., 1982. Priority dispatching with operation due dates in a job shop. *J. Oper. Manag.* 2, 167–175.
- Kelton, W.D., Law, A.M., 2000. *Simulation Modeling and Analysis*. McGraw Hill, Boston.
- Land, M., Stevenson, M., Thürer, M., 2014. Integrating load-based order release and priority dispatching. *Int. J. Prod. Res.* 52, 1059–1073.
- Land, M.J., Stevenson, M., Thürer, M., Gaalman, G.J., 2015. Job shop control: in search of the key to delivery improvements. *Int. J. Prod. Econ.* 168, 257–266.
- Melnik, S.A., Ragatz, G.L., 1989. Order review/release: research issues and perspectives. *Int. J. Prod. Res.* 27, 1081–1096.
- Oosterman, B., Land, M., Gaalman, G., 2000. The influence of shop characteristics on workload control. *Int. J. Prod. Econ.* 68, 107–119.
- Panwalkar, S.S., Iskander, W., 1977. A survey of scheduling rules. *Oper. Res.* 25, 45–61.
- Ramasesh, R., 1990. Dynamic job shop scheduling: a survey of simulation research. *Omega* 18, 43–57.
- Schönslieben, P., 2011. *Integral Logistics Management: Operations and Supply Chain Management within and Across Companies*. CRC Press.
- Soman, C.A., Van Donk, D.P., Gaalman, G., 2004. Combined make-to-order and make-to-stock in a food production system. *Int. J. Prod. Econ.* 90, 223–235.
- Soman, C.A., van Donk, D.P., 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, 441–453.
- Soman, C.A., Van Donk, D.P., Gaalman, G.J., 2007. Capacitated planning and scheduling for combined make-to-order and make-to-stock production in the food industry: an illustrative case study. *Int. J. Prod. Econ.* 108, 191–199.
- Tsubone, H., Ishikawa, Y., Yamamoto, H., 2002. Production planning system for a combination of make-to-stock and make-to-order products. *Int. J. Prod. Res.* 40, 4835–4851.
- Williams, T., 1984. Special products and uncertainty in production/inventory systems. *Eur. J. Oper. Res.* 15, 46–54.
- Wu, M.-C., Jiang, J.-H., Chang, W.-J., 2008. Scheduling a hybrid mto/mts semiconductor fab with machine-dedication features. *Int. J. Prod. Econ.* 112, 416–426.