International Journal of Production Research
Vol. 49, No. 7, 1 April 2011, 1827–1850

An analytical model for service level and tardiness in a single machine MTO production system

Klaus Altendorfer* and Herbert Jodlbauer

Department of Operations Management, Upper Austria University of Applied Science, School of Management, 4400 Steyr, Austria

(*Received 23 April 2009; final version received 25 January 2010*)

In this paper an analytical model to calculate service level, FGI and tardiness for a make-to-order (MTO) production system based on the production leadtime, utilisation and WIP is presented. The distribution of customer required leadtime is linked to the already available equations for an M/M/1 production system from queuing theory. Explicit equations for service level, FGI, FGI leadtime and tardiness are presented for an M/M/1 production system within an MTO environment. For a G/G/1 production system an approximation based extension is provided – discussing the influence of variation in the inter-arrival and processing time distribution in this framework. Moreover, the integration of a work ahead window (WAW) work release policy is discussed. Based on a numerical study, a high potential to decrease FGI (up to 97% FGI reduction) when applying such a WAW strategy is found and it is shown that the higher the targeted service level is, the higher the FGI reduction potential. The paper contributes to a better understanding of the relationship between customer required leadtime distribution and the M/M/1 production system. By applying this model a decision maker can base his capacity investment decisions on the service level and expected tardiness for certain levels of FGI and WIP and can additionally define the optimal WAW policy.

Keywords: service level; tardiness; production economics; production management; queuing theory; MTO

1. Introduction

The logistics service a company provides to its customers is one of the most important success factors in a competitive business environment. According to the literature, in order to identify the quality of this logistics service in make-to-order (MTO) production systems two performance indicators are important. The first is service level, meaning percentage of orders or pieces delivered on time. The second is average tardiness, meaning the time a customer has to wait if an order is late.

For production systems, the relationship between utilisation, work in process (WIP) and production leadtime is analytically well defined in the literature. Nonetheless, this relationship lacks a link to the customer's perspective. This analytical link was recently addressed through models discussing the on-time probability depending on the current production system state (see Duenyas and Hopp 1995, Hopp and Roof-Sturgis 2000, Liu and Yuan 2001). Furthermore, empirical and simulation studies have been conducted

*Corresponding author. Email: Klaus.altendorfer@fh-steyr.at

ISSN 0020–7543 print/ISSN 1366–588X online

DOI: 10.1080/00207541003660176
http://www.informaworld.com

concerning the topics of service level and tardiness in production systems as well as the link between utilisation, production leadtime, and WIP (see Jones 1973 or Jodlbauer and Huber 2008).

The lack of analytical models to identify the link between the inherent production system behaviour and the external customer related measures of service level and tardiness for MTO production systems leads to the model developed in this paper. The approach applied in this paper is the combination of the distribution of customer required leadtime with the known equations for an M/M/1 production system from queuing theory. This approach leads to a set of equations defining the service level, the expected FGI leadtime, the expected FGI and the expected tardiness for an M/M/1 production system working under MTO. This set of equations is presented in an explicit form for an exponentially distributed customer required leadtime. The integral form of the equations delivered ensures a broad applicability of this model to further research as well. Additionally, a set of equations for the aforementioned measures is developed for the application of a work ahead window (WAW) policy (see Hopp and Spearman 1996 for WAW). It is proven, that such a WAW policy keeps the Poisson input stream in the production system if customer required leadtime is exponentially distributed. For the G/G/1 case the influence of inter-arrival time as well as processing time variation is discussed based on the production leadtime approximation of Chen and Yao (2001).

For capacity investment decisions this model provides explicit equations to balance the service level, tardiness and FGI against capacity invested and WIP held. Additionally, the influence of the WAW policy for work release on FGI, service level and tardiness is discussed as a managerial tool to reduce FGI.

The remainder of the paper is structured as follows: in Section 2 the relevant literature concerning the research topic is described and related models are introduced. Section 3 provides the analytical development of the model for the M/M/1 production system. In Section 4 a numerical example is provided to show the application of the developed model and discuss the FGI reduction potential of the WAW policy. The conclusion is stated in Section 5.

2. Literature review

In this section, the available literature about the relationship between service level and tardiness in MTO production systems is reviewed. Some literature discussing the logistic characteristic curves is discussed and some simulation studies of production systems supporting the validity of the logistic characteristic curves are presented.

2.1 *Logistic characteristic curves*

The relationship between inventory, production leadtime and utilisation is discussed for a one machine production system in different papers which all use the relationship between inventory, production leadtime and throughput as described by Little (1961).

Karmarkar (1987) stated that the actual leadtime is highly dependent on actual workloads and lot sizes of a one machine production system. Using queuing theory this relationship is described by Medhi (1991), Hopp and Spearman (1996), and Tijms (2003) who deliver equations for the relationship between expected production leadtime and utilisation as well as between expected WIP and utilisation in an M/M/1 production system based on the expected processing time. Furthermore Medhi (1991) proves that in the M/M/1 case the production leadtime is exponentially distributed.

Based on empirical studies as well as through simulation, Wiendahl and Breithaupt (1999), and Wiendahl *et al.* (2005) find a qualitatively similar relationship between utilisation and production leadtime as in queuing theory. A further mathematical model for the calculation of the relationship between inventory, production leadtime and utilisation for a one machine system in a continuous setting is given by Jodlbauer (2008a).

Other approaches to discuss logistic relationships are empirical studies based on real world data or on simulation. The empirical studies approach is conducted, for example, by Wiendahl and Breithaupt (1999), and Wiendahl *et al.* (2005) who describe the relationship between inventory, utilisation and production leadtime (based on empirical data). Jones (1973), and Hopp and Roof-Sturgis (2000) discuss similar relationships applying the simulation approach.

In this paper the equations from queuing theory for utilisation, production leadtime and WIP will be used and referenced as the logistic characteristic curves.

2.2 *Service level and tardiness in production systems*

The calculation of a service level and tardiness reached in a certain production system are two key figures for the evaluation of the logistic performance of such a production system (see Hopp and Spearman 1996 or Jodlbauer and Huber 2008).

Van Nieuwenhuyse *et al.* (2007) present a model to determine the service level for an MTS production system by using queuing theory to determine the distribution of production leadtime and the distribution of demand within the production leadtime and compare it with finished goods inventory (FGI). In Spearman and Zhang (1999) due date setting strategies are discussed for a one product multi machine MTO production system with sequential routing. In the paper of Lutz *et al.* (2003) the use of characteristic curves representing the relationship between mean inventory level and mean tardiness or α-service level for a single machine production system is presented. An assembly line fed by two processing lines is discussed in Liu and Yuan (2001) whereby the service level is defined as the probability that the modelled production leadtime is not greater than a constant customer required leadtime. In Bertsimas and Paschalidis (2001) production policies are derived for an MTS production system with constant production rate and multi items to guarantee predefined α-service levels. A method for predicting maximum production leadtime is presented in Hopp and Roof-Sturgis (2000) which is based on a predetermined safety factor (also called service level in the paper) and on the jobs in the M/M/1 production system for due date setting purposes. In Duenyas and Hopp (1995) the distribution of production leadtime for a certain system state is used to define a due date for the customer, whereby the on time probability is again the probability that the production leadtime is shorter than the promised leadtime to the customer. Jodlbauer (2008b) discusses the use of an operation characteristic to define the MTO ability of production systems and to define the work ahead window needed to reach a certain capacity oriented service level target. The books by Medhi (1991), Chen and Yao (2001), and Tijms (2003) give a broad overview about queuing system basics also including some exact as well as some approximate equations for the relationship between production leadtime, WIP and utilisation for different production system structures. In Zipkin (2000) a broad overview concerning stock replenishment systems and different problem structures is given. Most of the problems (and solution procedures) discussed in Zipkin (2000) consider either service level or customer waiting time or both as part of the cost function or as a constraint whereby the parameters for the replenishment strategy are searched for. Such systems usually correspond to make-to-stock (MTS) production systems with production leadtime values which are distributed independently from the actual workload of the system. The review given in Zipkin (2000) shows that for stock replenishment systems the values for service level and customer waiting time are extensively used to define the optimal order policy.

The literature reviewed in this part, shows that a number of methods for service level and tardiness calculation already exist and that these are mostly based on the current queuing system state (Duenyas and Hopp 1995, Hopp and Roof-Sturgis 2000, Liu and Yuan 2001). The distribution of the customer required leadtime to define the service level and tardiness of an MTO production system is not yet used in queuing models. Therefore, such a combination of the M/M/1 queuing model with the distributed customer required leadtime is presented in this paper.

3. Model development

The model developed in this paper provides a deeper understanding of the relationship between production leadtime (utilisation and WIP) in an M/M/1 production system and the service level as well as tardiness of such a system. An extension of the model is also presented to include the possibility of reducing FGI. This model can be used by decision makers in cases of capacity investment decisions to define the capacity needed for a certain target service level or a certain target average tardiness and to define an optimal WAW policy.

The model development section starts with the description of the model assumptions, followed by a brief introduction of the basic queuing theory model with deterministic customer required leadtime. In Section 3.3 the distribution of customer required leadtime is used to find equations for service level, FGI leadtime, FGI and tardiness in an M/M/1 queuing model. In Section 3.4 equations to integrate the WAW policy for work release are developed and in Section 3.5 an extension of the model for a G/G/1 queuing system is discussed.

3.1 *Model assumptions*

The following assumptions are taken to create the model:

- The customers demand certain due date values once they state their orders and do not change these stated due dates later on. Based on these due dates the customer required leadtime values can be calculated. This stochastic customer required leadime is exponentially distributed.
- The customer required leadtime cannot be influenced by the production system and all customer orders are accepted and released into the system.
- The M/M/1 model from queuing theory is used. This means inter-arrival time and processing time are exponentially distributed and a single machine production system is studied.
- The distribution of customer required leadtime is independent of the distribution of production leadtime.

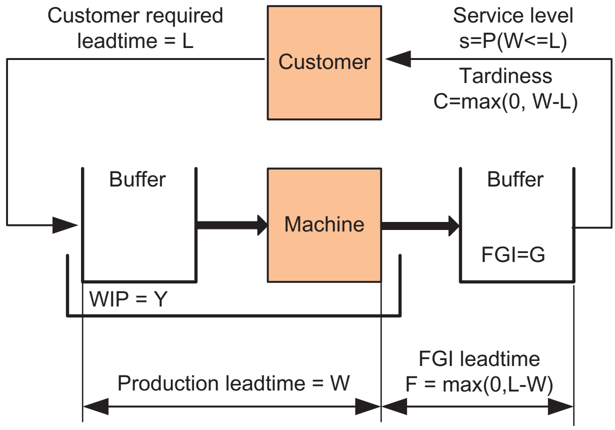

Figure 1. Production system.

- The studied production system is an MTO system, nothing is produced without a customer order.
- The queuing discipline is first in first out (FIFO).
- Machine capacity cannot be stored.

The one machine production system shown in Figure 1 with a WIP and an FGI is discussed in this paper.

The following variables are defined:

L	random variable for the customer required leadtime;
W	random variable of production leadtime needed for one order from arrival at the production system until its completion;
F	random variable of FGI leadtime; whenever an order is finished earlier than the due date it is stored in the FGI for this duration;
C	random variable of tardiness; whenever an order is finished later than the due date it is tardy for this duration;
Y	random variable of WIP in items in front of the machine and in the machine
G	random variable of FGI in items after the machine;
s	service level reached in the production system.

3.2 *Queuing model with deterministic customer required leadtime*

In this part some available queuing theory findings are applied to define the basic equations for the developed approach. The following basic relationships hold within an M/M/1 queuing system (see Tijms 2003):

$$\begin{aligned} \rho &= \frac{\mu}{\lambda} \\ E[W] &= \int_0^\infty f_W(\tau)\tau \, \mathrm{d}\tau = \frac{1}{\mu(1-\rho)} \\ E[Y] &= \lambda E[W] = \frac{1}{1-\rho}. \end{aligned} \tag{1}$$

Whereby:

ρ	utilisation of the machine (probability that machine is busy);
λ	order arrival rate to the system;
μ	processing rate of the machine;
$f_W(\cdot)$	probability distribution function (PDF) of the production leadtime.

When an order has a certain customer required leadtime, the on time probability of this order is equal to the probability that the production leadtime is shorter than this value for customer required leadtime (see Duenyas and Hopp 1995, Liu and Yuan 2001):

$$s = \mathrm{P}(W \leq \hat{L}) = F_W(\hat{L}) = 1 - e^{-(\mu(1-\rho)\hat{L})}. \tag{2}$$

Whereby $F_W(\cdot)$ is the cumulative distribution function (CDF) of the random variable W for production leadtime and $\hat{L}$ is the deterministic customer required leadtime. For Equation (2) to hold true, the assumption that the queuing discipline is FIFO is essential. Each order in the production system which has a production leadtime being smaller than the deterministic customer required leadtime has the FGI leadtime $F = \hat{L} - W$ and zero otherwise, so $F = [\hat{L} - W]^+$ holds. Based on the customer required leadtime and on the exponentially distributed production leadtime the following equation holds for the expected value of the FGI leadtime:

$$\begin{aligned} E[F] &= \int_0^{\hat{L}} f_W(\tau)(\hat{L} - \tau)\mathrm{d}\tau \\ &= \int_0^{\hat{L}} \mu(1-\rho)e^{-(\mu(1-\rho)\tau)}(\hat{L} - \tau)\mathrm{d}\tau \\ &= \frac{\hat{L}\mu(1-\rho) + e^{-(\mu(1-\rho)\hat{L})} - 1}{\mu(1-\rho)}. \end{aligned} \tag{3}$$

Based on Little's law (Little 1961) the expected value of FGI can then be calculated as:

$$E[G] = E[F]\lambda. \tag{4}$$

The same structure as for FGI leadtime also applies for tardiness so $C = [W - \hat{L}]^+$ and (see Duenyas and Hopp 1995 for the integral form):

$$\begin{aligned} E[C] &= \int_{\hat{L}}^{\infty} f_W(\tau)(\tau - \hat{L})\mathrm{d}\tau \\ &= \mu(1-\rho)\int_{\hat{L}}^{\infty} e^{-(\mu(1-\rho)\tau)}(\tau - \hat{L})\mathrm{d}\tau \\ &= \frac{e^{-(\mu(1-\rho)\hat{L})}}{\mu(1-\rho)}. \end{aligned} \tag{5}$$

Equations (1) to (5) from queuing theory will in the next part be linked to the distribution of customer required leadtime to create a model for the relationship between utilisation, production leadtime, service level, FGI, FGI leadtime and tardiness.

3.3 *Distribution of customer required leadtime*

A distribution of customer required leadtime is implemented in this section. If the customer required leadtime is not a deterministic number, but follows a certain distribution, for each possible value the customer required leadtime can take, the service level for that customer required leadtime value can be calculated according to Equation (2). By weighting each of these resulting service levels according to the probability of occurrence of the respective customer required leadtime value the following expected service level results (with an exponentially distributed customer required leadime):

$$s = \int_0^\infty F_W(\tau) f_L(\tau) \mathrm{d}\tau \underset{k=\mu(1-\rho)}{=} \int_0^\infty \left(1 - e^{-(k\tau)}\right) \beta e^{-(\beta\tau)} \mathrm{d}\tau \\ = 1 - \frac{\beta}{(k+\beta)}. \tag{6}$$

Whereby $1/k$ is the mean of the production leadtime and $f_L(\cdot)$ is the PDF of the random variable L for customer required leadtime with parameter β. For proof see Appendix 1.

Calculating the $\lim_{\rho\to 1}$ shows that the service level approaches zero independently of the distribution parameter β of customer required leadtime. For $\lim_{\rho\to 0}$ the service level reaches a value below 100% $(1 - (\beta/(\mu+\beta)))$ depending on the relation between processing rate μ and the parameter of the customer required leadtime distribution β. The result for $\lim_{\rho\to 0}$ can intuitively be argued by the fact that each order needs a certain time to be processed which depends on μ and its on time probability is at $\lim_{\rho\to 0}$ exactly equal to the probability that the processing time is shorter than the customer required leadtime.

Weighting the expected FGI leadtime values from Equation (3) with their probability of occurrence leads to:

$$E[F] = \int_0^\infty \int_0^\theta f_W(\tau)(\theta - \tau) \mathrm{d}\tau f_L(\theta) \mathrm{d}\theta \\ = \int_0^\infty \frac{\theta k + e^{-(k\theta)} - 1}{k} f_L(\theta) \mathrm{d}\theta \\ = \frac{1}{\beta} - \frac{1}{k+\beta}. \tag{7}$$

For proof see Appendix 1.

For $\lim_{\rho\to 1}$ the expected value of FGI leadtime approaches zero. From service level equal to zero at $\lim_{\rho\to 1}$ it is intuitively clear that nothing can be in the FGI when the utilisation reaches 100%. For $\lim_{\rho\to 0}$ a certain maximum expected value of FGI leadtime can be calculated as $(\mu/\beta(\mu+\beta))$. This value decreases in β which means the maximum expected FGI leadtime value increases when the expected customer required leadtime increases. The maximum expected FGI leadtime increases in μ which means that a higher processing rate leads to a higher maximum expected value for the FGI leadtime.

The expected tardiness from Equation (8) is calculated analogous to the expected FGI leadtime in Equation (7):

$$\begin{aligned} E[C] &= \int_0^\infty \int_\theta^\infty f_W(\tau)(\tau-\theta)\mathrm{d}\tau f_L(\theta)\mathrm{d}\theta \\ &= \int_0^\infty \frac{e^{-(k\theta)}}{k} f_L(\theta)\mathrm{d}\theta \\ &= \int_0^\infty \frac{e^{-(k\theta)}}{k} \beta e^{-(\beta\theta)}\mathrm{d}\theta \\ &= \frac{\beta}{k[k+\beta]}. \end{aligned} \tag{8}$$

For $\lim_{\rho\to 1}$ the expected value of tardiness approaches ∞. The reason is that the production leadtime approaches ∞ and so the tardiness has to have the same behaviour. In the case of $\lim_{\rho\to 0}$ a certain minimum value exists for the expected tardiness reached $(\beta/\mu(\mu+\beta))$ similarly to the expected FGI leadtime. This value decreases in μ which means the higher the processing rate is, the lower the minimum expected tardiness. Furthermore, it increases in β which means the higher the expected customer required leadtime, the lower the minimum expected tardiness is.

For the calculation of the FGI in pieces Equation (4) still holds in the case of distributed customer required leadtime values.

3.4 *Integration of a WAW work release policy*

The integration of a customer required leadtime distribution is the basis for the possibility to influence the FGI by integrating a WAW policy. Such a policy is described in Jodlbauer (2008b), and Jodlbauer and Altendorfer (2010), for example, and states that only orders which have a due date within a certain WAW are released to the production system. All the other customer orders are stored in front of the production system and are then released when their due date reaches the WAW. This leads, in the currently developed model, to the behaviour that no order released to the production queue has a remaining customer required leadtime value longer than this WAW w.

When the customer required leadtime is stochastic, the WAW policy leads to the situation that all the orders with a customer required leadtime being greater than the WAW are transferred to an order list (or WAW buffer) when they arrive at the production system. Their release to the production system is triggered when their remaining customer required leadtime becomes smaller than the WAW. This means all the orders having $L > w$ have a customer required leadtime of w. Based on Equation (6) for the service level, the extension for integrating a WAW policy is developed by constraining the value used for calculating the expected service level from Equation (6) with an upper bound being the WAW w:

$$\begin{aligned} s &= \int_0^w F_W(\tau) f_L(\tau)\mathrm{d}\tau + \int_w^\infty F_W(w) f_L(\tau)\mathrm{d}\tau \\ &= 1 - \frac{ke^{-([k+\beta]w)}+\beta}{[k+\beta]}. \end{aligned} \tag{9}$$

For proof see Appendix 1.

For $\lim_{\rho\to 1}$ the service level approaches zero, which is the same behaviour as in Equation (6). When calculating $\lim_{\rho\to 0}$ the following maximum service level can be calculated:

$$1 - \frac{\mu e^{-([\mu+\beta]w)} + \beta}{(\mu + \beta)},$$

which is smaller than the service level without WAW.

The underlying assumption is that this WAW policy does not disturb the Poisson input stream into the production system. This is necessary to ensure that the production leadtime of the production system is still exponentially distributed. The necessary property is stated in Proposition 1.

Proposition 1: *For every Poisson arrival process the application of the WAW policy leads to a Poisson input stream into the production system if the customer required leadtime is exponentially distributed.*

Proof: See Appendix 2.

The two equations for the expected FGI leadtime and the expected tardiness are created by using Equations (7) and (8) for two different cases. The first case is that the customer required leadtime is smaller than the WAW and the second case is that the customer required leadtime is larger than the WAW. Whenever $L \leq w$ holds true, Equation (7) also holds in the case of conducting a WAW policy. For the situation where $L > w$, the realised customer required leadtime in the production system is w. The first integral of Equation (7) integrates over all possible customer required leadtime values, so this integral has to be split up into the two parts. Adding up these two parts of the expected value for the expected FGI leadtime leads to the following equation:

$$\begin{aligned} E[F] &= \int_0^w \int_0^\theta f_W(\tau)(\theta - \tau)\mathrm{d}\tau f_L(\theta)\mathrm{d}\theta + \int_w^\infty \int_0^w f_W(\tau)(w - \tau)\mathrm{d}\tau f_L(\theta)\mathrm{d}\theta \\ &= \frac{\beta e^{-([k+\beta]w)} - (k + \beta)e^{-(\beta w)} + k}{\beta(k + \beta)}. \end{aligned} \tag{10}$$

For proof see Appendix 1.

Calculating the $\lim_{\rho\to 1}$ shows that the FGI leadtime approaches zero (see Appendix 1), which is the same case as for Equation (7). For $\lim_{\rho\to 0}$ the FGI leadtime leads to the following maximum value:

$$\frac{\beta e^{-([\mu+\beta]w)} - (\mu + \beta)e^{-(\beta w)} + \mu}{\beta(\mu + \beta)},$$

which is smaller than $\mu/\beta(\mu + \beta)$ (for proof see Appendix 1) for the case without a WAW policy.

The expected tardiness can be stated analogous to the expected FGI leadtime, so the first integral of Equation (8) also has to be split up into two parts and then added to the expected tardiness. This leads to the following equation:

$$\begin{aligned} E[C] &= \int_0^w \int_\theta^\infty f_W(\tau)(\tau - \theta)\mathrm{d}\tau f_L(\theta)\mathrm{d}\theta + \int_w^\infty \int_w^\infty f_W(\tau)(\tau - w)\mathrm{d}\tau f_L(\theta)\mathrm{d}\theta \\ &= \frac{k e^{-([k+\beta]w)} + \beta}{k[k + \beta]}. \end{aligned} \tag{11}$$

For proof see Appendix 1.

Calculating the $\lim_{\rho\to 1}$ shows that the tardiness approaches ∞. For $\lim_{\rho\to 0}$ a minimum tardiness level can be calculated:

$$\frac{\mu e^{-([\mu+\beta]w)}+\beta}{\mu(\mu+\beta)},$$

which is higher than in the case without WAW (for proof see Appendix 1).

Equation (4) still holds for the calculation of the FGI in pieces in the case of distributed customer required leadtime values and the application of a WAW policy.

Proposition 2: *Whenever the WAW policy is applied to an M/M/1 production system facing exponentially distributed customer required leadtime, the expected FGI leadtime and the service level with the WAW policy are lower than without the WAW policy and the expected tardiness with the WAW policy is higher than without the WAW policy.*

Proof: See Appendix 2.

Proposition 2 shows that in a system with a WAW policy no improvement of FGI can be achieved without reducing the service level (or equivalently the utilisation as shown later on). Nevertheless the balance between service level and FGI or utilisation and FGI can be discussed.

Based on Equation (4), the following equation can be stated for the expected value of FGI:

$$E[G]=\frac{\rho\mu\left(e^{-([k+\beta]w)}-1\right)}{k+\beta}-\frac{\rho\mu\left(e^{-(\beta w)}-1\right)}{\beta}. \tag{12}$$

For proof see Appendix 1.

Calculating the $\lim_{\rho\to 1}$ and the $\lim_{\rho\to 0}$ for Equation (12) delivers the expected FGI value of zero. This means in both extreme cases for the utilisation there is no FGI in pieces available. The comparison to the $\lim_{\rho\to 0}$ of the FGI leadtime shows, that even with the maximum expected FGI leadtime at utilisation zero there is no expected FGI in pieces available. The reason for that is that at utilisation zero, the input rate has to be zero and for that reason no products are produced.

Proposition 3: *For any M/M/1 production system facing exponentially distributed customer required leadtime with deterministic WAW, the utilisation leading to the maximum expected FGI is indirectly defined by Equation* (13):

$$\frac{\mu e^{-([\mu(1-\rho)+\beta]w)}(1+w\rho\mu)-\mu}{\mu(1-\rho)+\beta}+\frac{\rho\mu^2\left(e^{-([\mu(1-\rho)+\beta]w)}-1\right)}{(\mu(1-\rho)+\beta)^2}=\frac{\mu\left(e^{-(\beta w)}-1\right)}{\beta}. \tag{13}$$

Proof: See Appendix 2.

Equation (13) can only be solved numerically. The utilisation value found with this equation gives the production manager the information whether an increase in utilisation still leads to an increase in average FGI or if it already leads to a decrease, which is valuable information concerning FGI costs and FGI storage.

Proposition 4: *The application of a WAW policy within an M/M/1 production system facing exponentially distributed customer required leadtime leads to a utilisation and an FGI* *reduction as stated in Equations* (14) *and* (15), *respectively, in comparison to an M/M/1 production system without WAW policy for equal service level values between the two systems.*

$$\Delta\rho = \frac{1-\rho_w}{1-s} e^{-([\mu(1-\rho_w)+\beta]w)} \tag{14}$$

$$\Delta E[G] = \frac{\rho_w \mu \left(e^{-([k_w+\beta]w)} - 1\right)}{k_w+\beta} - \frac{\rho_w \mu \left(e^{-(\beta w)} - 1\right)}{\beta} - \frac{(\rho_w + \Delta\rho)\mu k}{k+\beta}. \tag{15}$$

Whereby $\Delta\rho$ *is the utilisation loss and* $\Delta E[G]$ *is the FGI reduction when in the production system applying the WAW policy the utilisation* ρ_w *is reached. The parameters* $1/k$ *and* $1/k_w$ *indicate the mean production leadtime in the production system without WAW policy and with WAW policy, respectively.*

Proof: See Appendix 2.

Based on Proposition 4, the management can discuss the balance between utilisation loss and FGI cost reduction and define an optimal strategy based on utilisation loss costs and FGI costs. The trajectory shown in Figure 2 can be defined for each combination of μ, β, and ρ_w.

A detailed numerical example discussing the FGI reduction potential is presented in Section 4.2.

3.5 *Extension for G/G/1 production system*

To identify the influence of other distributions for inter-arrival and processing time, the influence of increasing the mean of the exponential distribution for production leadtime on service level, FGI, FGI leadtime and tardiness is studied.

Proposition 5: *Increasing the mean of the exponential distribution for production leadtime in an M/M/1 production system facing exponentially distributed customer required leadtime without changing the distribution of customer required leadtime and without changing the input rate leads to a reduction of service level, expected FGI, and expected FGI leadtime as well as to an increase in expected tardiness.*

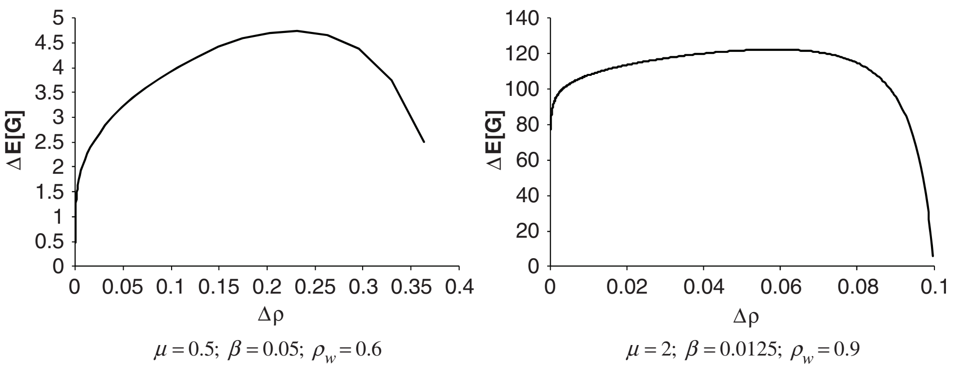

Figure 2. Trajectory utilisation loss versus FGI reduction.

Proof: See Appendix 2.

Based on Chen and Yao (2001) the production leadtime for a G/G/1 queue can be approximated with an exponential distribution with mean:

$$E[W] = \frac{\lambda\left(c_a^2 + c_s^2\right)}{2(1-\rho)}. \tag{16}$$

Whereby c_a^2 and c_s^2 are the coefficients of variation of inter-arrival and processing time distribution, respectively. Based on this approximation, the mean of the production leadtime increases whenever the coefficient of variation of the inter-arrival time distribution or the processing time distribution increases. In combination with Proposition 5, this shows that the results for service level, expected FGI, expected FGI leadtime and expected tardiness depend, for a more general system, on the second moment of these two distributions. Applying Proposition 5 shows that the lower the variation of processing time and inter-arrival time, the better the results for service level, expected FGI, expected FGI leadtime and expected tardiness will be.

4. Numerical example

In this section, firstly (Section 4.1) a comparison between three cases shows the typical characteristic of the relationships between utilisation, expected WIP, expected FGI, expected FGI leadtime, expected tardiness and service level. The advantage of implementing a WAW is discussed in this section based on the calculated curves and some general insights from the shape of the curves are given. Section 4.2 presents the result of an extensive experiment set to evaluate the FGI reduction potential based on the implementation of a WAW work release policy.

4.1 *Shape of the logistic characteristic curves*

In this section a numerical example is conducted with three different sets of parameters. The results for these three different sets of parameters for the M/M/1 queuing model are compared as developed in the paper according to the logistic relationship between utilisation, WIP, production leadtime, service level, FGI, FGI leadtime and tardiness for an exponentially distributed customer required leadtime. The parameters given in Table 1 are set for the numerical example.

In the numerical example the production rate μ is held constant and the input rate λ is varied to find different points of the logistic characteristic curves. All resulting figures in this section are shown with the expected WIP on the x-axis. This is the same format as the

Table 1. Parameters for curve generation.

Parameter	Case I	Case II	Case III	Unit
w	∞	∞	20	periods
β	0.05	0.2	0.05	pcs/periods
μ	0.5	0.5	0.5	pcs/periods

traditional logistic characteristic curves presented by Wiendahl and Breithaupt (1999), Wiendahl *et al.* (2005), Jodlbauer (2008a) and also partly in Hopp and Spearman (1996).

The logistic characteristic curves as shown in Figure 3 can be calculated for the three cases. The service level calculated without an additional WAW is the maximum reachable service level if each order is released to the production system without any delay and the FIFO dispatching rule is applied (see Proposition 2). The reduction of service level with increasing WIP, as shown in Figure 3(a), is a result of the increasing expected production leadtime (and increasing λ). The maximum service level, maximum FGI leadtime and minimum tardiness values for utilisation zero as well as the asymptotical behaviour of the curves for utilisation approaching 100% is shown in Figure 3.

The result that with a smaller average value for the customer required leadtime the service level decreases faster, the FGI leadtime decreases faster and the tardiness increases faster when WIP increases is consistent with intuition.

In Cases I and II there is the maximum possible service level if each job is released to the system without any delay shown. The integration of a WAW can reduce the FGI inventory and for this reason the costs. The effect of having a WAW w of 20 periods is shown as Case III in Figure 3.

The comparison of Cases I and III in Figure 3(a) and (d) shows the influence of having a policy reducing the longest customer required leadtime values on service level and tardiness. Such a WAW policy leads to a reduction of service level and an increase of

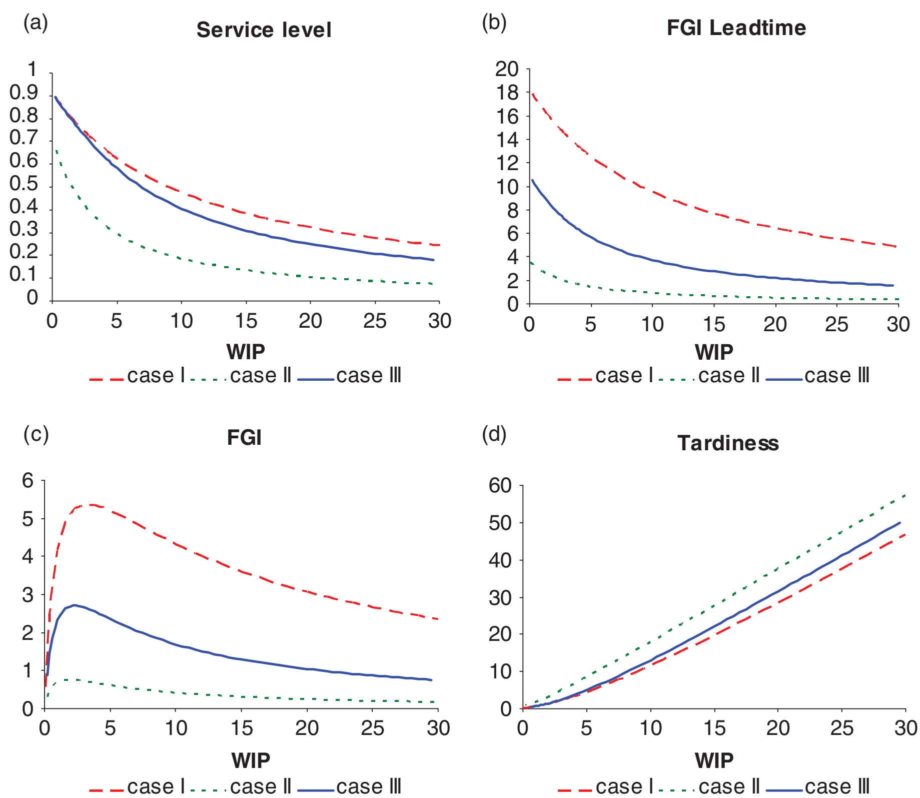

Figure 3. Logistic characteristic curves.

Table 2. Parameters for numerical test cases.

Parameter	Tested parameter values
Service level s (6 values)	0.5, 0.8, 0.9, 0.95, 0.98, 0.99
Production rate μ (7 values)	0.5, 1, 2, 4, 8, 16, 32
Customer required leadtime parameter β (8 values)	0.0125, 0.025, 0.05, 0.1, 0.2, 0.4, 0.8, 1.2
Work ahead window w (143 values)	0.5, 1, 2, 4, ..., 98, 100, 110, 120, ..., 990, 1000, ∞
Utilisation tolerance (3 values)	1%, 0.1%, 0.01%

expected tardiness. Figure 3(b) and (c) show the expected FGI leadtime and the expected FGI in items. The comparison with a WAW of 20 periods, which is exactly the expected value of the exponential distribution of customer required leadtime, shows that approximately half of the FGI in pieces can be reduced with this policy. The service level as well as the expected tardiness at low WIP values is only influenced marginally by this policy.This leads to the managerial insight, that especially the orders with long customer required leadtime values should not be released immediately. A detailed numerical study about the FGI reduction potential of such a WAW policy is given in the next section.

4.2 *FGI reduction potential with WAW policy*

To identify the potential of reducing expected FGI when a WAW work release policy is applied, a broad spectrum of test cases has been generated. All possible combinations of the parameter settings in Table 2 have been tested, whereby for each test case the input rate λ was numerically searched which leads to the specified service level.

Based on Proposition 2 there cannot be a potential for reducing expected FGI by applying a WAW work release policy without either reducing service level, increasing tardiness or reducing utilisation. For this comparison, the utilisation was the factor allowed to be reduced for the sake of keeping the service level constraint. This means the potential to reduce FGI has always to be seen in the light of a slight utilisation loss which also has to be accepted. Three limits for the utilisation loss are tested in this numerical example. There is 1%, 0.1% and 0.01% lower utilisation accepted compared to the test case without WAW policy.

In this numerical study, firstly, for all possible combinations of parameters from Table 2, the input rate to reach the targeted service level was calculated. As shown in the result (Figure 4) the higher the service level target is, the lower the number of test cases where this value can be reached. In the second step, the utilisation reached without WAW is compared to the utilisation values reached with WAW for the same combination of s, μ and β, and the lowest value for w out of the test set still fulfilling the utilisation constraint is determined. For this WAW value the FGI in pieces is compared to the FGI in pieces without WAW and the potential to decrease FGI is calculated as a percentage of the FGI without WAW. The structured results ordered concerning the production system specifications as a combination of s, μ and β, can be found in Appendix 4. The FGI reduction potential of service level of 50% and 99% with 1% and 0.01% utilisation reduction constraint are shown in Figure 4.

The results compared in Figure 4 show that the range of reducing FGI based on the numerical example tested is between 0% and 97% depending on the production system.

Utilisation tolerance 1%

s 50%		μ 0,5	1	2	4	8	16	32	Avg
0	0,0125	60%	72%	82%	89%	95%	96%	97%	84%
	0,025	46%	59%	72%	82%	89%	95%	95%	77%
	0,05	34%	46%	57%	70%	79%	89%	95%	67%
	0,1	24%	34%	46%	53%	70%	79%	89%	56%
	0,2	12%	24%	34%	46%	46%	61%	79%	43%
	0,4	4%	8%	17%	34%	33%	33%	61%	27%
	0,8	---	2%	8%	8%	34%	33%	33%	20%
	1,2	---	---	2%	17%	17%	17%	46%	20%
Avg		30%	35%	40%	50%	58%	63%	74%	**50%**

s 99%		μ 0,5	1	2	4	8	16	32	Avg
0	0,0125	---	---	91%	91%	94%	94%	96%	93%
	0,025	---	---	---	91%	91%	91%	91%	91%
	0,05	---	---	---	---	91%	91%	91%	91%
	0,1	---	---	---	---	---	91%	91%	91%
	0,2	---	---	---	---	---	---	91%	91%
	0,4	---	---	---	---	---	---	---	
	0,8	---	---	---	---	---	---	---	
	1,2	---	---	---	---	---	---	---	
Avg				91%	91%	92%	92%	92%	**92%**

Utilisation tolerance 0.01%

s 50%		μ 0,5	1	2	4	8	16	32	Avg
0	0,0125	8%	11%	15%	21%	29%	40%	52%	25%
	0,025	6%	7%	9%	16%	22%	31%	41%	19%
	0,05	4%	6%	8%	11%	15%	22%	31%	14%
	0,1	2%	4%	5%	8%	11%	14%	20%	9%
	0,2	2%	2%	4%	5%	8%	11%	11%	6%
	0,4	0%	2%	2%	4%	4%	8%	8%	4%
	0,8	---	0%	2%	2%	2%	2%	8%	2%
	1,2	---	---	0%	2%	2%	2%	2%	1%
Avg		4%	5%	6%	9%	12%	16%	21%	**10%**

s 99%		μ 0,5	1	2	4	8	16	32	Avg
0	0,0125	---	---	87%	87%	89%	89%	89%	88%
	0,025	---	---	---	87%	87%	87%	87%	87%
	0,05	---	---	---	---	83%	83%	83%	83%
	0,1	---	---	---	---	---	83%	83%	83%
	0,2	---	---	---	---	---	---	83%	83%
	0,4	---	---	---	---	---	---	---	
	0,8	---	---	---	---	---	---	---	
	1,2	---	---	---	---	---	---	---	
Avg				87%	87%	86%	85%	85%	**86%**

Figure 4. Results of numerical study.
Note: - - - means that the service level target cannot be reached for this parameter combination of μ and β.

Looking at the results in detail shows appositive correlation between service level and FGI reduction potential with such a WAW policy. Furthermore, the lower the possibility to reduce utilisation is, the lower the potential to reduce FGI, which is an intuitive result. Nevertheless, the numerical study shows that even if nearly no flexibility to reduce utilisation is given (case with 0.01% tolerance) the FGI for high service levels can still be reduced by more than 80% when a WAW rule is applied.

The result of this numerical study leads to an interesting managerial insight since the competition between production companies is more and more forced into the field of logistic performance and it is clearly shown that implementing such a WAW policy helps to keep costs for high service levels down.

5. Conclusion

The current paper introduced an analytical model for integrating the logistic key figures service level, FGI, FGI leadtime and tardiness into the logistic characteristic curves for WIP, utilisation and production leadtime for an M/M/1 production system. Analytically exact equations for service level, expected FGI leadtime, expected FGI and expected tardiness are presented based on an exponentially distributed customer required leadtime for an MTO production system. Furthermore, the effect of the WAW work release policy on these logistic figures can be determined with the model. Especially in the case of capacity investment decisions, the service level and tardiness values reached can be balanced against the capacity invested and the costs for holding FGI.

The implementation of the WAW work release policy enables the discussion of reducing FGI. The numerical example discussed in this paper leads to the managerial insight that a company using the WAW policy for order release to the production system can save up to 97% of the FGI in comparison to releasing all orders directly to the production system. Especially for high service level targets the FGI reduction potential of such a WAW policy is quite high even if only a marginal reduction of utilisation is allowed.

For a G/G/1 production system it is shown, that an increase in the coefficient of variation of inter-arrival or processing time negatively influences the system performance measured as service level, expected FGI leadtime, expected FGI and expected tardiness.

As a further step in research concerning the logistic behaviour of MTO production systems, the extension of this model to multi-machine settings either in an analytically exact way or as an approximation is proposed. Furthermore, the influence of dispatching rules could be discussed and a detailed sensitivity analysis concerning the influence of the distribution shapes for inter-arrival time, processing time and customer required leadtime on the results could be conducted.

Acknowledgement

We would like to thank Stefan Minner and Richard Hartl for their support and comments on the article. Their input led to a considerable increase in the quality of this publication.

Notes

1. For a Poisson process with rate λ and events of type 1 and type 2, the assignment of type 1 and type 2 is given with probability p and $1-p$ independent of all the other events, the two resulting streams of events for types 1 and 2 are again Poisson streams with rate $\varphi = p\lambda$ and $\psi = (1-p)\lambda$, respectively (see Tijms 2003).
2. The merge of two Poisson streams of events with rates ψ and φ lead to a Poisson stream of events with rate $\lambda = \varphi + \psi$ (see Tijms 2003).

References

Bertsimas, D. and Paschalidis, I.C., 2001. Probabilistic service level guarantees in make-to-stock manufacturing systems. *Operations Research*, 49 (1), 119–133.

Chen, H. and Yao, D.D., 2001. *Fundamentals of queueing networks: performance, asymptotics, and optimization*. New York: Springer.

Duenyas, I. and Hopp, W.J., 1995. Quoting customer lead times. *Management Science*, 41 (1), 43–57.

Hopp, W.J. and Spearman, M.L., 1996. *Factory physics: foundation of manufacturing management. 1*. Chicago, IL: Irwin.

Hopp, W.J. and Roof-Sturgis, M.L., 2000. Quoting manufacturing due dates subject to a service level constraint. *IIE Transactions*, 32 (9), 771–784.

Jodlbauer, H., 2008a. A time-continuous analytic production model for service level, work in process, lead time and utilization. *International Journal of Production Research*, 46 (7), 1723–1744.

Jodlbauer, H., 2008b. Customer driven production planning. *International Journal of Production Economics*, 111 (2), 793–801.

Jodlbauer, H. and Huber, A., 2008. Service-level performance of MRP, kanban, CONWIP and DBR due to parameter stability and environmental robustness. *International Journal of Production Research*, 46 (8), 2179–2195.

Jodlbauer, H. and Altendorfer, K., 2010. Trade-off between capacity invested and inventory needed. *European Journal of Operational Research*, 203 (1), 118–133.

Jones, C.H., 1973. An economic evaluation of job shop dispatching rules. *Management Science*, 20 (3), 293–307.

Karmarkar, U.S., 1987. Lot sizes, lead times and in-process inventories. *Management Science*, 33 (3), 409–418.
Little, J.D.C., 1961. A proof for the queuing formula L = lW. *Operations Research*, 9 (3), 383–387.
Liu, L. and Yuan, X.-M., 2001. Throughput, flow times, and service level in an unreliable assembly system. *European Journal of Operational Research*, 135 (3), 602–615.
Lutz, S., Löedding, H., and Wiendahl, H.-P., 2003. Logistics-oriented inventory analysis. *International Journal of Production Economics*, 85 (2), 217–231.
Medhi, J., 1991. *Stochastic models in queuing theory*. Boston, MA: Academic Press.
Spearman, M.L. and Zhang, R.Q., 1999. Optimal lead time policies. *Management Science*, 45 (2), 290–295.
Tijms, H.C., 2003. *A first course in stochastic models*. Amsterdam: Wiley.
Van Nieuwenhuyse, I., *et al*., 2007. Buffer sizing in multi-product multi-reactor batch processes: impact of allocation and campaign sizing policies. *European Journal of Operational Research*, 179 (2), 424–443.
Wiendahl, H.-P. and Breithaupt, J.-W., 1999. Modelling and controlling the dynamics of production systems. *Production Planning & Control*, 10 (4), 389–401.
Wiendahl, H.-H., Von Cieminski, G., and Wiendahl, H.-P., 2005. Stumbling blocks of PPC: towards the holistic configuration of PPC systems. *Production Planning & Control*, 16 (7), 634–651.
Zipkin, P.H., 2000. *Foundations of inventory management*. New York: McGraw-Hill Higher Education.

Appendix 1. Proof of equations

Proof of Equation (6): The probability of an order being delivered on time can be visualised in the production leadtime W and customer required leadtime L space shown in Figure 5. The shaded area indicates when an order is delivered on time. Assuming that the distributions of W and L are independent of each other (which hold true for FIFO dispatching discipline) the following equation

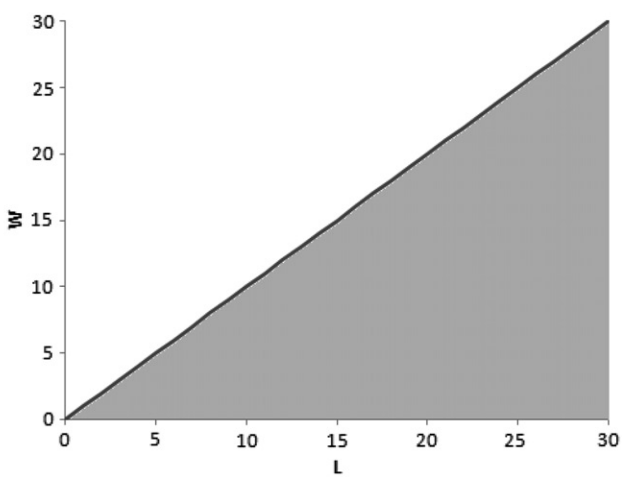

Figure 5. Joint probability space of W and L.

for the service level holds:

$$
\begin{aligned}
s &= \int_0^\infty \int_\tau^\infty f_W(\tau) f_L(\theta)\mathrm{d}\theta\,\mathrm{d}\tau = \int_0^\infty f_W(\tau)\int_\tau^\infty f_L(\theta)\mathrm{d}\theta\,\mathrm{d}\tau \\
&= \int_0^\infty f_W(\tau)(1-F_L(\tau))\mathrm{d}\tau = 1 - \int_0^\infty f_W(\tau)F_L(\tau)\mathrm{d}\tau \\
&\quad \text{with } F_L(0)=0;\quad F_L(\infty)=1;\quad F_W(0)=0;\quad F_W(\infty)=1; \\
&= F_W(\tau)F_L(\tau)\Big|_0^\infty - \int_0^\infty f_W(\tau)F_L(\tau)\mathrm{d}\tau = \int_0^\infty F_W(\tau) f_L(\tau)\mathrm{d}\tau.
\end{aligned}
\tag{A1}
$$

Proof of Equation (7):

$$
\begin{aligned}
E[F] &= \int_0^\infty \int_0^\theta f_W(\tau)(\theta-\tau)\mathrm{d}\tau f_L(\theta)\mathrm{d}\theta = \int_0^\infty \frac{\theta k + e^{-(k\theta)}-1}{k}\beta e^{-(\beta\theta)}\,\mathrm{d}\theta \\
&= \frac{\beta}{k}\left[\int_0^\infty \theta k e^{-(\beta\theta)}\,\mathrm{d}\theta + \int_0^\infty e^{-([k+\beta]\theta)}\,\mathrm{d}\theta - \int_0^\infty e^{-(\beta\theta)}\,\mathrm{d}\theta\right] \\
&= \frac{\beta}{k}\left[\int_0^\infty \theta k e^{-(\beta\theta)}\,\mathrm{d}\theta + \frac{1}{[k+\beta]} - \frac{1}{\beta}\right] = \frac{\beta}{k}\left[\frac{k}{\beta^2} + \frac{1}{[k+\beta]} - \frac{1}{\beta}\right] \\
&= \frac{1}{\beta} - \frac{1}{k+\beta}.
\end{aligned}
\tag{A2}
$$

Proof of Equation (9):

$$
\begin{aligned}
s &= \int_0^w F_W(\tau) f_L(\tau)\mathrm{d}\tau + \int_w^\infty F_W(w) f_L(\tau)\mathrm{d}\tau \\
&= \int_0^w \left(1-e^{-(k\tau)}\right)\beta e^{-(\beta\tau)}\,\mathrm{d}\tau + \int_w^\infty \left(1-e^{-(kw)}\right)\beta e^{-(\beta\tau)}\,\mathrm{d}\tau \\
&= \beta\left[\frac{e^{-(\beta\tau)}}{-\beta}\bigg|_0^w - \frac{e^{-([k+\beta]\tau)}}{-[k+\beta]}\bigg|_0^w\right] + \left(1-e^{-(kw)}\right)\beta\frac{e^{-(\beta\tau)}}{-\beta}\bigg|_w^\infty \\
&= 1 + \frac{\beta e^{-([k+\beta]w)}-\beta}{[k+\beta]} - \frac{[k+\beta]e^{-([k+\beta]w)}}{[k+\beta]} \\
&= 1 - \frac{k e^{-([k+\beta]w)}+\beta}{[k+\beta]}.
\end{aligned}
\tag{A3}
$$

Proof of Equation (10):

$$
\begin{aligned}
E[F] &= \int_0^w \int_0^\theta f_W(\tau)(\theta-\tau)\mathrm{d}\tau f_L(\theta)\mathrm{d}\theta + \int_w^\infty \int_0^w f_W(\tau)(w-\tau)\mathrm{d}\tau f_L(\theta)\mathrm{d}\theta \\
&= \int_0^w \frac{\theta k + e^{-(k\theta)}-1}{k}\beta e^{-(\beta\theta)}\,\mathrm{d}\theta + \int_w^\infty \frac{wk + e^{-(kw)}-1}{k}\beta e^{-(\beta\theta)}\,\mathrm{d}\theta \\
&= \frac{\beta}{k}\left[\int_0^w \theta k e^{-(\beta\theta)}\,\mathrm{d}\theta + \int_0^w e^{-([k+\beta]\theta)}\,\mathrm{d}\theta - \int_0^w e^{-(\beta\theta)}\,\mathrm{d}\theta\right] + \beta\frac{wk+e^{-(kw)}-1}{k}\int_w^\infty e^{-(\beta\theta)}\,\mathrm{d}\theta \\
&= \frac{\beta}{k}\left[wk\frac{e^{-(\beta w)}}{-\beta} - \frac{k\left(e^{-(\beta w)}-1\right)}{\beta^2} - \frac{e^{-([k+\beta]w)}-1}{[k+\beta]} + \frac{e^{-(\beta w)}-1}{\beta}\right] + \frac{wke^{-(\beta w)} + e^{-([k+\beta]w)} - e^{-(\beta w)}}{k} \\
&= \frac{ke^{-([k+\beta]w)}+\beta}{k[k+\beta]} - \frac{k\left(e^{-(\beta w)}-1\right)+\beta}{k\beta} \\
&= \frac{\beta e^{-([k+\beta]w)} - (k+\beta)e^{-(\beta w)} + k}{\beta(k+\beta)}
\end{aligned}
\tag{A4}
$$

Limes calculation $\lim_{\rho\to 1}$:

$$\lim_{\rho\to 1}\left(\frac{\beta e^{-([\mu(1-\rho)+\beta]w)}-(\mu(1-\rho)+\beta)e^{-(\beta w)}+\mu(1-\rho)}{\beta(\mu(1-\rho)+\beta)}\right)=\frac{\beta e^{-(\beta w)}-\beta e^{-(\beta w)}}{2\beta}=0. \tag{A5}$$

Limes calculation $\lim_{\rho\to 0}$:

$$\begin{aligned}&\lim_{\rho\to 0}\left(\frac{\beta e^{-([\mu(1-\rho)+\beta]w)}-(\mu(1-\rho)+\beta)e^{-(\beta w)}+\mu(1-\rho)}{\beta(\mu(1-\rho)+\beta)}\right)\\&=\frac{\beta e^{-([\mu+\beta]w)}-(\mu+\beta)e^{-(\beta w)}+\mu}{\beta(\mu+\beta)}\\&=\frac{\beta e^{-([\mu+\beta]w)}-(\mu+\beta)e^{-(\beta w)}}{\beta(\mu+\beta)}+\frac{\mu}{\beta(\mu+\beta)}\quad \text{with } e^{-([\mu+\beta]w)}<e^{-(\beta w)}\\&\Rightarrow \beta e^{-([\mu+\beta]w)}-\beta e^{-(\beta w)}<0\\&\Rightarrow \frac{\beta e^{-([\mu+\beta]w)}-\beta e^{-(\beta w)}-\mu e^{-(\beta w)}}{\beta(\mu+\beta)}+\frac{\mu}{\beta(\mu+\beta)}<\frac{\mu}{\beta(\mu+\beta)}.\end{aligned} \tag{A6}$$

Proof of Equation (11):

$$\begin{aligned}E[C]&=\int_0^w\int_\theta^\infty f_W(\tau)(\tau-\theta)\mathrm{d}\tau f_L(\theta)\mathrm{d}\theta+\int_w^\infty\int_w^\infty f_W(\tau)(\tau-w)\mathrm{d}\tau f_L(\theta)\mathrm{d}\theta\\&=\int_0^w\frac{e^{-(k\theta)}}{k}\beta e^{-(\beta\theta)}\mathrm{d}\theta+\int_w^\infty\frac{e^{-(kw)}}{k}\beta e^{-(\beta\theta)}\mathrm{d}\theta=\frac{ke^{-([k+\beta]w)}+\beta}{k[k+\beta]}.\end{aligned} \tag{A7}$$

Limes calculation $\lim_{\rho\to 0}$:

$$\begin{aligned}&\lim_{\rho\to 0}\left(\frac{\mu(1-\rho)e^{-([\mu(1-\rho)+\beta]w)}+\beta}{\mu(1-\rho)[\mu(1-\rho)+\beta]}\right)=\frac{\mu e^{-([\mu+\beta]w)}+\beta}{\mu[\mu+\beta]}\quad \text{with } \mu e^{-([\mu+\beta]w)}>0\\&\Rightarrow \frac{\mu e^{-([\mu+\beta]w)}+\beta}{\mu(\mu+\beta)}>\frac{\beta}{\mu(\mu+\beta)}.\end{aligned} \tag{A8}$$

Proof of Equation (12):

$$\begin{aligned}E[G]=\lambda E[F]\underset{\lambda=\rho\mu}{=}&\frac{\rho\mu k e^{-([k+\beta]w)}}{k[k+\beta]}+\frac{\rho\mu\beta}{k[k+\beta]}-\frac{\rho\mu k\left(e^{-(\beta w)}-1\right)}{k\beta}-\frac{\rho\mu\beta}{k\beta}\\=&\frac{\rho\mu e^{-([k+\beta]w)}}{[k+\beta]}+\frac{\mu\rho\beta}{k[k+\beta]}-\frac{\rho\mu\left(e^{-(\beta w)}-1\right)}{\beta}-\frac{\rho\mu}{k}\\=&\frac{\rho\mu\left(e^{-([k+\beta]w)}-1\right)}{k+\beta}-\frac{\rho\mu\left(e^{-(\beta w)}-1\right)}{\beta}.\end{aligned} \tag{A9}$$

Appendix 2. Proof of propositions

Proof of Proposition 1: The comparison of a stochastic customer required leadtime value L to a deterministic WAW value w leads to a random assignment of orders to type 1 and type 2 orders. Type 1 orders have $L \le w$ with probability p and type 2 orders have $L > w$ and have probability $(1-p)$. Based on the splitting property (see Tijms 2003)[1] of the input stream being a Poisson stream with rate λ the two resulting streams of events are again Poisson streams with rates $\varphi = p\lambda$ and $\psi = (1-p)\lambda$ respectively. The Poisson stream 1 ($L \le w$) with rate $\varphi = p\lambda$ directly feeds the M/M/1 production system. The Poisson stream 2 with rate $\psi = (1-p)\lambda$ feeds the buffer of units waiting to be released into the system. This Poisson stream 2 has the following waiting time distribution of items:

$$P(l = L - w \mid L > w) = \beta e^{-\beta(l-w)}e^{-\beta w} = \beta e^{-\beta l}. \tag{A10}$$

Equation (A10) shows that the waiting time for the single items in the WAW buffer is again exponentially distributed with rate β. So this WAW buffer can be transformed to an M/M/∞ queuing system for which the output process is equal to the Poisson input process with rate $\psi = (1-p)\lambda$ (see Tijms 2003). This output process of the WAW buffer feeds the M/M/1 production system. Based on the merging property (see Tijms 2003)[2], the input process for the M/M/1 production system is Poisson with rate $\lambda = \varphi + \psi$. □

Proof of Proposition 2: The service level with a WAW policy is lower than without a WAW policy:

$$\text{with } ke^{-([k+\beta]w)} > 0 \Rightarrow 1 - \frac{ke^{-([k+\beta]w)} + \beta}{[k+\beta]} < 1 - \frac{\beta}{[k+\beta]}. \tag{A11}$$

Expected FGI leadtime with a WAW policy is always lower than without a WAW policy:

$$\begin{aligned} E[F] &\underset{(10)}{=} \frac{\beta e^{-([k+\beta]w)} - (k+\beta)e^{-(\beta w)} + k}{\beta(k+\beta)} \\ &= \frac{\beta e^{-([k+\beta]w)} - \beta e^{-(\beta w)} - ke^{-(\beta w)}}{\beta(k+\beta)} + \frac{k}{\beta(k+\beta)} \\ &\quad \text{with } e^{-([k+\beta]w)} < e^{-(\beta w)} \Rightarrow \beta e^{-([k+\beta]w)} - \beta e^{-(\beta w)} < 0 \\ &\Rightarrow \frac{\beta e^{-([k+\beta]w)} - \beta e^{-(\beta w)} - ke^{-(\beta w)}}{\beta(k+\beta)} + \frac{1}{\beta} - \frac{1}{k+\beta} < \frac{1}{\beta} - \frac{1}{k+\beta}. \end{aligned} \tag{A12}$$

Expected tardiness with a WAW policy is always higher than without a WAW policy:

$$\frac{ke^{-([k+\beta]w)} + \beta}{k[k+\beta]} \quad \text{with } ke^{-([k+\beta]w)} > 0 \Rightarrow \frac{ke^{-([k+\beta]w)} + \beta}{k[k+\beta]} > \frac{\beta}{k[k+\beta]}. \tag{A13}$$

□

Proof of Proposition 3:

$$\begin{aligned} E[G] &= \frac{\rho\mu\left(e^{-([\mu(1-\rho)+\beta]w)} - 1\right)}{\mu(1-\rho)+\beta} - \frac{\rho\mu\left(e^{-(\beta w)} - 1\right)}{\beta} \\ \frac{dE[G]}{\mathrm{d}\rho} &= \frac{\mu\left(e^{-([\mu(1-\rho)+\beta]w)} - 1\right)}{\mu(1-\rho)+\beta} + \frac{w\rho\mu^2 e^{-([\mu(1-\rho)+\beta]w)}}{\mu(1-\rho)+\beta} + \frac{\rho\mu^2\left(e^{-([\mu(1-\rho)+\beta]w)} - 1\right)}{(\mu(1-\rho)+\beta)^2} \\ &\quad - \frac{\mu\left(e^{-(\beta w)} - 1\right)}{\beta} = 0 \Leftrightarrow \frac{\mu e^{-([\mu(1-\rho)+\beta]w)}(1 + w\rho\mu) - \mu}{\mu(1-\rho)+\beta} \\ &\quad + \frac{\rho\mu^2\left(e^{-([\mu(1-\rho)+\beta]w)} - 1\right)}{(\mu(1-\rho)+\beta)^2} = \frac{\mu\left(e^{-(\beta w)} - 1\right)}{\beta} \end{aligned} \tag{A14}$$

□

Proof of Proposition 4: Based on Equations (6) and (9) for the service level in a system without and with a WAW policy the following can be stated provided the service level in both systems is equal:

$$\begin{aligned} s &= 1 - \frac{\beta}{(k+\beta)} \Leftrightarrow k(1-s) - \beta s = 0 \text{ and} \\ s_w &= 1 - \frac{k_w e^{-([k_w+\beta]w)} + \beta}{k_w + \beta} \Leftrightarrow k_w(1-s_w) - k_w e^{-([k_w+\beta]w)} - \beta s_w = 0 \\ &\quad \text{with } s = s_w \Rightarrow k(1-s) - \beta s = k_w(1-s) - k_w e^{-([k_w+\beta]w)} - \beta s \\ &\Leftrightarrow (k-k_w)(1-s) = -k_w e^{-([k_w+\beta]w)} \underset{\substack{k=\mu(1-\rho) \\ k_w=\mu(1-\rho_w)}}{\Leftrightarrow} \rho = \rho_w + \frac{1-\rho_w}{1-s} e^{-([\mu(1-\rho_w)+\beta]w)}. \end{aligned} \tag{A15}$$

Whereby the index w indicates the system applying the WAW policy. The utilisation loss can be calculated as:

$$\Delta\rho = \rho - \rho_w = \frac{1-\rho_w}{1-s} e^{-([\mu(1-\rho_w)+\beta]w)}. \tag{A16}$$

Based on Equations (12) and (4), the following can be stated for the FGI reduction:

$$\begin{aligned}\Delta E[G] &= E[G_w] - E[G] \\ &= \frac{\rho_w \mu\left(e^{-([k_w+\beta]w)}-1\right)}{k_w+\beta} - \frac{\rho_w\mu\left(e^{-(\beta w)}-1\right)}{\beta} - \frac{k\mu\rho}{k+\beta} \\ &= \frac{\rho_w \mu\left(e^{-([k_w+\beta]w)}-1\right)}{k_w+\beta} - \frac{\rho_w\mu\left(e^{-(\beta w)}-1\right)}{\beta} - \frac{(\rho_w+\Delta\rho)\mu k}{k+\beta}.\end{aligned} \tag{A17}$$

□

Proof of Proposition 5: Service level increases with increasing k (decreases with increasing mean production leadtime $1/k$):

$$\begin{aligned}& s = 1 - \frac{ke^{-([k+\beta]w)}+\beta}{k+\beta} \\ & \frac{ds}{dk} = \frac{kwe^{-([k+\beta]w)} - e^{-([k+\beta]w)}}{k+\beta} + \frac{ke^{-([k+\beta]w)}+\beta}{(k+\beta)^2} \overset{!}{>} 0 \\ & \quad \Rightarrow k^2we^{-([k+\beta]w)} + \beta kwe^{-([k+\beta]w)} + \beta \overset{!}{>} \beta e^{-([k+\beta]w)} \\ & \quad \text{is fulfilled with } \beta > \beta e^{-([k+\beta]w)};\ k^2we^{-([k+\beta]w)} + \beta kwe^{-([k+\beta]w)} > 0.\end{aligned} \tag{A18}$$

Expected FGI leadtime decreases with increasing k (increases with increasing mean production leadtime $1/k$):

$$\begin{aligned}& E[F] = \frac{\beta e^{-([k+\beta]w)} - (k+\beta)e^{-(\beta w)} + k}{\beta(k+\beta)} \\ & \frac{dE[F]}{dk} = \frac{1 - e^{-(\beta w)} - w\beta e^{-([k+\beta]w)}}{\beta(k+\beta)} - \frac{k + \beta e^{-([k+\beta]w)} - (k+\beta)e^{-(\beta w)}}{\beta(k+\beta)^2} \overset{!}{>} 0 \\ & \quad \Rightarrow 1 - (k+\beta)we^{-([k+\beta]w)} - e^{-([k+\beta]w)} \overset{!}{>} 0 \Rightarrow e^{(k+\beta)w} - 1 \overset{!}{>} (k+\beta)w \\ & \quad \text{is fulfilled with } e^x - 1 > x.\end{aligned} \tag{A19}$$

Expected tardiness decreases with increasing k (increases with increasing mean production leadtime $1/k$):

$$\begin{aligned}& E[C] = \frac{ke^{-([k+\beta]w)}+\beta}{k[k+\beta]} \\ & \frac{dE[C]}{dk} = \frac{e^{-([k+\beta]w)} - kwe^{-([k+\beta]w)}}{k(k+\beta)} - \frac{ke^{-([k+\beta]w)}+\beta}{k(k+\beta)^2} - \frac{ke^{-([k+\beta]w)}+\beta}{k^2(k+\beta)} \overset{!}{<} 0 \\ & \quad \Rightarrow \frac{(k+\beta)ke^{-([k+\beta]w)} - (k+\beta)k^2we^{-([k+\beta]w)}}{k^2(k+\beta)^2} - \frac{k^2e^{-([k+\beta]w)} + k\beta}{k^2(k+\beta)^2} \\ & \quad\quad - \frac{(k+\beta)ke^{-([k+\beta]w)} + (k+\beta)\beta}{k^2(k+\beta)^2} \overset{!}{<} 0 \\ & \quad \Rightarrow -k^3we^{-([k+\beta]w)} - \beta k^2we^{-([k+\beta]w)} - k^2e^{-([k+\beta]w)} - 2k\beta - \beta^2 \overset{!}{<} 0 \\ & \quad \text{is fulfilled with } k, w, \beta > 0.\end{aligned} \tag{A20}$$

Expected FGI increases with increasing k (decreases with increasing mean production leadtime $1/k$):

$$\begin{aligned} E[G] &= \frac{\rho\mu\left(e^{-([k+\beta]w)}-1\right)}{k+\beta} - \frac{\rho\mu\left(e^{-(\beta w)}-1\right)}{\beta} \\ \frac{dE[G]}{dk} &= -\frac{\rho\mu\left(e^{-([k+\beta]w)}-1\right)}{(k+\beta)^2} - \frac{w\rho\mu e^{-([k+\beta]w)}}{k+\beta} \overset{!}{>} 0 \\ &\Rightarrow -\rho\mu\left(e^{-([k+\beta]w)}-1\right) - (k+\beta)w\rho\mu e^{-([k+\beta]w)} \overset{!}{>} 0 \\ &\Rightarrow e^{([k+\beta]w)} - 1 \overset{!}{>} (k+\beta)w \text{ is fulfilled with } e^x - 1 > x. \end{aligned} \tag{A21}$$

Appendix 3. List of variables

Symbol	Description	Unit
W	random variable of production leadtime	periods
$F_W(\cdot)$	cumulative distribution function of the random variable W for production leadtime	1
$f_W(\cdot)$	distribution function of the random variable W for production leadtime	1
s	service level reached in the production system	1
F	random variable of FGI leadtime	periods
ρ	utilisation of the machine	1
Y	random variable of WIP (work in process) in front of the machine and in the machine	pcs
C	random variable of tardiness	periods
L	random variable of customer required leadtime	periods
$\hat{L}$	deterministic customer required leadtime	periods
$F_L(\cdot)$	cumulative distribution function of the random variable L for customer required leadtime	1
$f_L(\cdot)$	distribution function of the random variable L for customer required leadtime	1
μ	processing rate of the machine	pcs/period
λ	rate of arrival of jobs to the system	pcs/period
β	parameter of the customer required leadtime distribution	1/period
w	work ahead window for work release to the production system	periods

Appendix 4. Results of numerical study

Figures 6–8 show the results for all the tested production system settings.

s		μ							
50%		0,5	1	2	4	8	16	32	Avg
β	0,0125	60%	72%	82%	89%	95%	96%	97%	84%
	0,025	46%	59%	72%	82%	89%	95%	95%	77%
	0,05	34%	46%	57%	70%	79%	89%	95%	67%
	0,1	24%	34%	46%	53%	70%	79%	89%	56%
	0,2	12%	24%	34%	46%	46%	61%	79%	43%
	0,4	4%	8%	17%	34%	33%	33%	61%	27%
	0,8	---	2%	8%	8%	34%	33%	33%	20%
	1,2	---	---	2%	17%	17%	17%	46%	20%
Avg		30%	35%	40%	50%	58%	63%	74%	**50%**

s		μ							
80%		0,5	1	2	4	8	16	32	Avg
β	0,0125	56%	65%	73%	82%	88%	92%	95%	79%
	0,025	48%	55%	65%	73%	82%	88%	91%	72%
	0,05	38%	46%	55%	65%	70%	82%	88%	63%
	0,1	26%	38%	46%	55%	65%	65%	76%	53%
	0,2	---	26%	38%	38%	55%	55%	54%	44%
	0,4	---	---	26%	25%	25%	55%	55%	37%
	0,8	---	---	---	26%	25%	25%	55%	33%
	1,2	---	---	---	---	11%	38%	37%	29%
Avg		42%	46%	50%	52%	53%	62%	69%	**53%**

s		μ							
90%		0,5	1	2	4	8	16	32	Avg
β	0,0125	63%	69%	75%	80%	87%	92%	95%	80%
	0,025	55%	61%	67%	73%	80%	87%	90%	73%
	0,05	41%	55%	61%	67%	73%	80%	87%	66%
	0,1	---	41%	50%	61%	61%	73%	73%	60%
	0,2	---	---	34%	50%	50%	50%	73%	51%
	0,4	---	---	---	22%	50%	50%	50%	43%
	0,8	---	---	---	---	22%	50%	50%	41%
	1,2	---	---	---	---	---	33%	33%	33%
Avg		53%	57%	57%	59%	60%	64%	69%	**60%**

s		μ							
95%		0,5	1	2	4	8	16	32	Avg
β	0,0125	73%	76%	80%	84%	87%	91%	93%	83%
	0,025	61%	71%	74%	78%	82%	85%	89%	77%
	0,05	---	58%	71%	71%	78%	78%	85%	73%
	0,1	---	---	58%	71%	71%	71%	71%	68%
	0,2	---	---	---	47%	71%	71%	71%	65%
	0,4	---	---	---	---	47%	71%	71%	63%
	0,8	---	---	---	---	---	47%	71%	59%
	1,2	---	---	---	---	---	---	58%	58%
Avg		67%	68%	71%	70%	73%	73%	76%	**71%**

s		μ							
98%		0,5	1	2	4	8	16	32	Avg
β	0,0125	---	84%	86%	88%	90%	92%	94%	89%
	0,025	---	---	84%	84%	88%	88%	92%	87%
	0,05	---	---	---	84%	84%	84%	84%	84%
	0,1	---	---	---	---	84%	84%	84%	84%
	0,2	---	---	---	---	---	84%	84%	84%
	0,4	---	---	---	---	---	---	84%	84%
	0,8	---	---	---	---	---	---	---	
	1,2	---	---	---	---	---	---	---	
Avg			84%	85%	85%	86%	86%	87%	**85%**

s		μ							
99%		0,5	1	2	4	8	16	32	Avg
β	0,0125	---	---	91%	91%	94%	94%	96%	93%
	0,025	---	---	---	91%	91%	91%	91%	91%
	0,05	---	---	---	---	91%	91%	91%	91%
	0,1	---	---	---	---	---	91%	91%	91%
	0,2	---	---	---	---	---	---	91%	91%
	0,4	---	---	---	---	---	---	---	
	0,8	---	---	---	---	---	---	---	
	1,2	---	---	---	---	---	---	---	
Avg				91%	91%	92%	92%	92%	**92%**

Figure 6. Results of numerical study with utilisation reduction limit 1%.

s		μ							
50%		0,5	1	2	4	8	16	32	Avg
β	0,0125	24%	33%	44%	56%	68%	79%	88%	56%
	0,025	17%	25%	33%	44%	55%	67%	79%	46%
	0,05	13%	17%	24%	33%	42%	53%	65%	35%
	0,1	8%	12%	17%	24%	33%	39%	53%	26%
	0,2	5%	8%	12%	17%	24%	33%	33%	19%
	0,4	2%	4%	8%	8%	17%	17%	33%	13%
	0,8	---	2%	2%	8%	8%	8%	8%	6%
	1,2	---	---	2%	2%	2%	2%	17%	5%
Avg		11%	14%	18%	24%	31%	37%	47%	**26%**

s		μ							
80%		0,5	1	2	4	8	16	32	Avg
β	0,0125	32%	40%	47%	54%	62%	70%	79%	55%
	0,025	29%	34%	39%	47%	54%	62%	70%	48%
	0,05	23%	28%	34%	37%	45%	54%	59%	40%
	0,1	17%	21%	25%	31%	37%	45%	54%	33%
	0,2	---	17%	17%	25%	25%	37%	37%	27%
	0,4	---	---	11%	11%	25%	25%	25%	20%
	0,8	---	---	---	5%	5%	25%	25%	15%
	1,2	---	---	---	---	11%	11%	11%	11%
Avg		25%	28%	29%	30%	33%	41%	45%	**33%**

s		μ							
90%		0,5	1	2	4	8	16	32	Avg
β	0,0125	50%	54%	58%	62%	68%	73%	78%	63%
	0,025	43%	50%	52%	58%	61%	67%	73%	58%
	0,05	34%	41%	50%	50%	55%	61%	67%	51%
	0,1	---	34%	41%	50%	50%	50%	61%	47%
	0,2	---	---	34%	33%	50%	50%	50%	43%
	0,4	---	---	---	22%	22%	50%	50%	36%
	0,8	---	---	---	---	22%	22%	50%	32%
	1,2	---	---	---	---	---	33%	33%	33%
Avg		42%	45%	47%	46%	47%	51%	58%	**48%**

s		μ							
95%		0,5	1	2	4	8	16	32	Avg
β	0,0125	64%	67%	71%	72%	76%	80%	82%	73%
	0,025	55%	64%	67%	71%	71%	74%	78%	68%
	0,05	---	52%	64%	64%	71%	71%	71%	65%
	0,1	---	---	47%	58%	58%	71%	71%	61%
	0,2	---	---	---	47%	47%	47%	71%	53%
	0,4	---	---	---	---	47%	47%	47%	47%
	0,8	---	---	---	---	---	47%	47%	47%
	1,2	---	---	---	---	---	---	58%	58%
Avg		59%	61%	62%	62%	62%	62%	65%	**62%**

s		μ							
98%		0,5	1	2	4	8	16	32	Avg
β	0,0125	---	79%	84%	84%	86%	88%	88%	85%
	0,025	---	---	79%	84%	84%	84%	88%	84%
	0,05	---	---	---	76%	84%	84%	84%	82%
	0,1	---	---	---	---	68%	84%	84%	79%
	0,2	---	---	---	---	---	68%	84%	76%
	0,4	---	---	---	---	---	---	68%	68%
	0,8	---	---	---	---	---	---	---	
	1,2	---	---	---	---	---	---	---	
Avg			79%	82%	81%	80%	81%	82%	**81%**

s		μ							
99%		0,5	1	2	4	8	16	32	Avg
β	0,0125	---	---	89%	89%	91%	91%	91%	90%
	0,025	---	---	---	87%	87%	91%	91%	89%
	0,05	---	---	---	---	83%	83%	91%	86%
	0,1	---	---	---	---	---	83%	83%	83%
	0,2	---	---	---	---	---	---	83%	83%
	0,4	---	---	---	---	---	---	---	
	0,8	---	---	---	---	---	---	---	
	1,2	---	---	---	---	---	---	---	
Avg				89%	88%	87%	87%	88%	**88%**

Figure 7. Results of numerical study with utilisation reduction limit 0.1%.

s		μ							
50%		0,5	1	2	4	8	16	32	**Avg**
β	0,0125	8%	11%	15%	21%	29%	40%	52%	25%
	0,025	6%	7%	9%	16%	22%	31%	41%	19%
	0,05	4%	6%	8%	11%	15%	22%	31%	14%
	0,1	2%	4%	5%	8%	11%	14%	20%	9%
	0,2	2%	2%	4%	5%	8%	11%	11%	6%
	0,4	0%	2%	2%	4%	4%	8%	8%	4%
	0,8	---	0%	2%	2%	2%	2%	8%	2%
	1,2	---	---	0%	2%	2%	2%	2%	1%
Avg		4%	5%	6%	9%	12%	16%	21%	**10%**

s		μ							
80%		0,5	1	2	4	8	16	32	**Avg**
β	0,0125	22%	25%	28%	31%	39%	45%	52%	35%
	0,025	19%	22%	25%	29%	34%	39%	45%	30%
	0,05	15%	19%	21%	25%	28%	34%	37%	26%
	0,1	9%	14%	17%	21%	25%	25%	31%	20%
	0,2	---	8%	11%	17%	17%	25%	25%	17%
	0,4	---	---	5%	11%	11%	11%	25%	13%
	0,8	---	---	---	5%	5%	5%	5%	5%
	1,2	---	---	---	---	11%	11%	11%	11%
Avg		16%	17%	18%	20%	21%	25%	29%	**21%**

s		μ							
90%		0,5	1	2	4	8	16	32	**Avg**
β	0,0125	39%	43%	46%	50%	54%	58%	62%	50%
	0,025	35%	39%	43%	45%	50%	52%	58%	46%
	0,05	27%	33%	37%	41%	45%	50%	50%	41%
	0,1	---	27%	33%	33%	41%	41%	50%	38%
	0,2	---	---	22%	33%	33%	33%	33%	31%
	0,4	---	---	---	22%	22%	22%	22%	22%
	0,8	---	---	---	---	22%	22%	22%	22%
	1,2	---	---	---	---	---	10%	33%	22%
Avg		34%	36%	36%	38%	38%	36%	41%	**37%**

s		μ							
95%		0,5	1	2	4	8	16	32	**Avg**
β	0,0125	56%	59%	62%	65%	67%	71%	72%	65%
	0,025	47%	55%	58%	61%	64%	67%	71%	60%
	0,05	---	47%	52%	58%	58%	64%	64%	57%
	0,1	---	---	47%	47%	58%	58%	58%	54%
	0,2	---	---	---	47%	47%	47%	47%	47%
	0,4	---	---	---	---	47%	47%	47%	47%
	0,8	---	---	---	---	---	47%	47%	47%
	1,2	---	---	---	---	---	---	32%	32%
Avg		52%	54%	55%	56%	57%	57%	55%	**55%**

s		μ							
98%		0,5	1	2	4	8	16	32	**Avg**
β	0,0125	---	78%	79%	79%	81%	84%	84%	81%
	0,025	---	---	76%	79%	79%	79%	84%	80%
	0,05	---	---	---	76%	76%	76%	76%	76%
	0,1	---	---	---	---	68%	68%	68%	68%
	0,2	---	---	---	---	---	68%	68%	68%
	0,4	---	---	---	---	---	---	68%	68%
	0,8	---	---	---	---	---	---	---	
	1,2	---	---	---	---	---	---	---	
Avg			78%	78%	78%	76%	75%	75%	**77%**

s		μ							
99%		0,5	1	2	4	8	16	32	**Avg**
β	0,0125	---	---	87%	87%	89%	89%	89%	88%
	0,025	---	---	---	87%	87%	87%	87%	87%
	0,05	---	---	---	---	83%	83%	83%	83%
	0,1	---	---	---	---	---	83%	83%	83%
	0,2	---	---	---	---	---	---	83%	83%
	0,4	---	---	---	---	---	---	---	
	0,8	---	---	---	---	---	---	---	
	1,2	---	---	---	---	---	---	---	
Avg				87%	87%	86%	85%	85%	**86%**

Figure 8. Results of numerical study with utilisation reduction limit 0.01%.