A multi-item capacitated make-to-order production system with considerable demand fluctuations is discussed. The relationship between the available capacity and the inventory needed to meet customer requirements with a pre-defined service level is modeled. Furthermore, the total cost for both capacity and inventory is minimized and it is shown that, assuming negligible change-over times, the double of the surplus inventory cost has to be equal to the excess capacity cost to ensure minimum total cost.

1. Introduction

According to the hierarchical planning approach, see Hax and Meal (1975), strategic and long-term decisions are made by top management whereas operational and short term tasks are performed by lower-level management. In general, capacity investment decisions are considered as strategic and inventory management as operational tasks. Practice in enterprises as well as most models in the literature view capacity investment decisions separately to the inventory needed to meet customer requirements with restricted available capacity.

In this paper, a relationship between the available capacity and the inventory needed to fulfill the customer orders with a required service level is developed. It is shown that there is a considerable impact of the available capacity (investment decision) on the inventory needed. Furthermore, the optimum capital invested in capacity is determined by minimizing the total cost for capacity as well as cost for capital invested in the required inventory. A make-to-order (MTO) production system under dynamic customer demand with considerable fluctuations is considered. Furthermore, the buying behavior of the customers is modeled by the statistical distribution of the customer required delivery lead time. The multi-item capacitated production model is based on a Fixed Order Period lot-sizing policy and production orders are released if the remaining time to due date is shorter than the work ahead window. Two strategies for the management of short term demand peaks are compared. The first strategy involves dedicating some excess capacity to expected future demand peaks while the second requires pre-producing on stock known customer orders with due dates far in the future.

The research objective of the paper is twofold:

1. development of a function with respect to available capacity, describing the inventory needed to meet the customer requirement with a pre-defined service level
2. development of a model to minimize the total costs for capacity as well as costs for capital invested in inventory.

The related literature focuses mainly on capacity and inventory management. A good survey of capacity expansion and capacity management literature is given in Luss (1982) and more recently in Van Mieghem (2003). For a good review of inventory management see Silver et al. (1998). In the literature references below, models which combine the capacity and the inventory are discussed.

Bish et al. (2005) investigated a multi-period two-stage supply chain, comprising two uncapacitated suppliers and two capacitated plants over an infinite horizon. They studied the performance measured in sales, production variability, variability propagating upstream,
component inventory and outbound distribution assuming an MTO environment with lost sales under several capacity allocation policies to manage short-term order variability.

Bradley and Glynn (2002) considered a joint optimization of capacity and inventory decisions in a single-product, single-stage, single server, produce-to-stock and limited capacity manufacturing model with the objective of minimizing the long-run average operating cost due to penalty, holding and capacity costs. This was one of the first research contributions addressing the joint decision of capacity and inventory. Angelus and Porteus (2002) as well as Van Mieghem and Rudi (2002) also addressed the joint (inventory against capacity investment) decision problem.

Raman and Kim (2002) stated that the error induced by ignoring the impact of inventory holding costs can be substantial. In their model with a high gross margin, unpredictable demand and high obsolescence risk, it was demonstrated that the reduction of inventory enabled by a higher flexible capacity invested can lead to a reduction of the sum of capacity and inventory cost.

Van Mieghem (1998) discussed optimal investment decisions on capacity at the strategic level by discussing the trade-off between revenue and capital investment costs. He stated that the greatest increase in revenue can be generated by flexible capacity when demands are negatively correlated.

Rajagopalan and Swaminathan (2001) explored the interaction between production planning and capacity acquisition in a multi-item and multi-period environment with known, varying and long-term growing demand. They discussed the trade-off between early investment (use of excess capacity for smaller lot-sizes to reduce the inventory) and later investment (use of excess capacity for building up inventories to meet the demand growth with postponed machine purchase).

Mincoyvics et al. (2009) discuss a production system with a certain permanent capacity and contingent capacity to meet non-stationary stochastic demand in an MTS production system. They develop a model to economically evaluate the balance between inventory, permanent capacity and contingent capacity. They show that the value of flexibility decreases with an increasing capacity acquisition leadtime.

Zhang et al. (2004) considered a discrete-time capacity expansion problem involving multiple product families, multiple machine types and non-stationary stochastic demand with no finished goods inventory and backorders. Their objective was to minimize the sum of capacity investment and the cost of lost sales.

A framework for the modeling and analysis of MTO, MTS and delay product differentiation (DD) is developed by Gupta and Benjaafar (2004). They discuss a model for minimization of inventory costs subject to a service level constraint in a multi-product production system. Based on a certain accepted customer waiting time on which the service level constraint is based, the inventory holding costs are evaluated for an MTS, MTO and DD production systems. No FGI is held in this model based on the assumption that a customer does not require a certain leadtime but is satisfied whenever the real customer waiting time is shorter than the accepted customer waiting time.

The remainder of this article is organized as follows. In the next section the model, which is based on four steps, is introduced. These steps are customer required capacity determination, describing the buying behavior of the customers by the order characteristic, calculating the inventory needed and finally formulating as well as solving the cost minimization problem. In Section 3 numerical illustrations and comparison of four different settings are presented. Section 4 concludes and all proofs are summarized in the Appendix A.

2. Model

A multi-period, multi-item and single-resource production environment with limited capacity and fluctuating dynamic demand is studied. The single-resource concerned may be an assembly or production line in a plant or a bottleneck machine. The production system is working on a make-to-order basis. This means only known customer orders are released into the production system. Some of the future demand is known while some remains unknown. The proportion of known to unknown orders is described by the order characteristic presented in Section 2.2. The following Fig. 1 provides an overview of the developed model.

As shown in Fig. 1, the customer orders a certain quantity of a certain product with a certain due date.1 All these orders are stored in a customer order buffer. The orders are released from that order buffer to production if the remaining time to the due date is shorter than the pre-defined work ahead window (as marked on the left hand side of Fig. 1). The fixed order period FOP n rule constitutes the applied lot sizing strategy. This means that in the MTO environment, the known customer demand of the next n periods is summarized to one production lot. The work order is defined by the earliest due date rule. The fixed order period FOP n is combined with the release policy applying the condition that the length of n periods is equal to the work ahead window. After the processing of the production lot at the machine, the finished goods are stored in the FGI buffer until their due date is reached. This production on stock is not an MTS strategy since the products are dedicated to certain customer orders. In case of late deliveries the products are not stored. Each production lot will usually contain more than one customer order since it contains the demand of the next n periods of a certain product. For that reason the FGI will usually contain items of most product types.

The order release strategy of using a WAW is often applied in MTO production systems and is comparable with CONWIP, see Hopp and Spearman (1996), (in the case that the WIP-CAP is very high or equivalent the work ahead window is the only control variable) or with Simplified Drum Buffer Rope, see Schragenheim and Dettmer (2001).

The main interest of this article lies in the trade-off between capital invested in capacity and capital invested in inventory. The evaluation of this trade-off is based on a cost objective which takes inventory holding as well as capacity into account. The capacity cost incurred is the depreciation for the capacity employed. One key point in the trade-off is excess capacity, which can be used instead of inventory to meet the customer requirement in the situation of a temporary demand peak.

The suggested approach, especially the concept of customer required capacity as well as the order characteristic to describe the buying behavior of the customers, is based on ideas of the recently introduced customer driven production model by Jodlbauer (2008). The model consists of the parameters and variables shown in Table 1.

---

1 Comparing this model to the model of Gupta and Benjaafar (2004) shows that they use an accepted waiting time of the customer while in this model a certain due date is specified by the customer.
The demand for each product (orders of product \(i\) due in period \(t\)) is independently and identically distributed over time. Demands of different products are also statistically independent. Inventory is measured in work load that is the processing time spent for production of the stored items. Based on the variance of sales for each product it is possible to create considerable demand fluctuations as will be discussed in this section. A seasonal demand behavior cannot be discussed with this model since it uses the steady state assumption which is often used in production logistics (e.g. Hopp and Spearman, 1996 and Jodlbauer, 2008).

### Table 1

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_{it})</td>
<td>Random demand rate of product type (i) with due date in period (t)</td>
<td>(\text{N/T})</td>
</tr>
<tr>
<td>(p_i)</td>
<td>Processing time for the (i)th product type</td>
<td>(\text{T/N})</td>
</tr>
<tr>
<td>(t_i)</td>
<td>Change-over time for the (i)th product type</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>Variance of the sales of product type (i) in one period</td>
<td>(\text{N}^2/\text{T}^2)</td>
</tr>
<tr>
<td>(\Delta)</td>
<td>Duration of one period</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(q_i)</td>
<td>Lot-size of the (i)th product type</td>
<td>(\text{N})</td>
</tr>
<tr>
<td>(k_i)</td>
<td>Random customer required capacity with due date in period (t)</td>
<td>(\text{T/T} = 1)</td>
</tr>
<tr>
<td>(\bar{k}_{it})</td>
<td>Random (n)-periods average customer required capacity with due date in period (t)</td>
<td>(\text{T/T} = 1)</td>
</tr>
<tr>
<td>(\bar{h})</td>
<td>Number of periods</td>
<td>(1)</td>
</tr>
<tr>
<td>(I)</td>
<td>Number of different product types considered</td>
<td>(1)</td>
</tr>
<tr>
<td>(F_{k,x})</td>
<td>Statistical distribution function of (k_{it})</td>
<td>(1)</td>
</tr>
<tr>
<td>(\sigma^2_{k_{it}})</td>
<td>Variance of (k_{it})</td>
<td>(\text{T/T} = 1)</td>
</tr>
<tr>
<td>(\mu_{k_{it}} = \bar{k}_0)</td>
<td>Expected customer required capacity without set-up times</td>
<td>(\text{T/T} = 1)</td>
</tr>
<tr>
<td>(\sigma^2_{k_{it}})</td>
<td>Variance of the average customer required capacity without set-up times</td>
<td>(1)</td>
</tr>
<tr>
<td>(\bar{h}<em>{it} = \bar{k}</em>{it}/\Delta)</td>
<td>Coefficient of variation of the average customer required capacity without set-up times</td>
<td>(1)</td>
</tr>
<tr>
<td>(h)</td>
<td>Work ahead window</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(L_i)</td>
<td>Capacity oriented work ahead window</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(P)</td>
<td>Minimum required lead time</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(s)</td>
<td>Capacity oriented service level</td>
<td>(1)</td>
</tr>
<tr>
<td>(L_i)</td>
<td>Random customer required delivery lead time of the (i)th product type</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(\mu_{li})</td>
<td>Expected customer required delivery lead time of the (i)th product type</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(\sigma^2_{li})</td>
<td>Variance of the customer required delivery lead time of the (i)th product type</td>
<td>(\text{T}^2)</td>
</tr>
<tr>
<td>(F_{L(n)})</td>
<td>Statistical distribution function of the customer required delivery lead time</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(\mu_{L(n)})</td>
<td>Expected customer required delivery lead time with set-ups and FOP</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(\sigma^2_{L(n)})</td>
<td>Variance of the customer required delivery lead time with set-ups and FOP</td>
<td>(\text{T}^2)</td>
</tr>
<tr>
<td>(\bar{L}_0)</td>
<td>Random customer requested capacity delivery lead time without set-ups</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(F_{L(n)})</td>
<td>Statistical distribution function of the customer required delivery lead-time</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(\mu_{L(n)})</td>
<td>Expected customer required delivery lead time</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(\sigma^2_{L(n)})</td>
<td>Variance of the customer required delivery lead time</td>
<td>(\text{T}^2)</td>
</tr>
<tr>
<td>(\bar{Z}_0)</td>
<td>Total set-up time</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(c_h)</td>
<td>Total set-up time</td>
<td>(\text{M})</td>
</tr>
<tr>
<td>(c_K)</td>
<td>Total set-up time</td>
<td>(\text{M})</td>
</tr>
<tr>
<td>(K)</td>
<td>Capacity invested or available capacity</td>
<td>(\text{T/T} = 1)</td>
</tr>
<tr>
<td>(K_{excess})</td>
<td>Excess capacity</td>
<td>(\text{T/T} = 1)</td>
</tr>
<tr>
<td>(K_{setup})</td>
<td>Set-up capacity</td>
<td>(\text{T/T} = 1)</td>
</tr>
<tr>
<td>(\mu_y)</td>
<td>Expected inventory needed</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(Y_{min})</td>
<td>Minimum required inventory to cover (P) and to protect against a very short delivery time lead</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(Y_{min})</td>
<td>Surplus inventory</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(Y_{setup})</td>
<td>Set-up inventory</td>
<td>(\text{T})</td>
</tr>
<tr>
<td>(\xi)</td>
<td>Sum of the capacity and inventory cost</td>
<td>(\text{M})</td>
</tr>
<tr>
<td>(U)</td>
<td>Utilization</td>
<td></td>
</tr>
</tbody>
</table>
2.1. Customer required capacity

The customer required capacity $k_t$ is a random variable describing the capacity needed to fulfill the customer orders with due dates in period $t$, whereby no time offset is considered. This customer required capacity depends on the demand, the processing time, the set-up time as well as the lot-size applied. The customer required capacity $k_t$ with respect to the period $t$ without considering any planned lead time (variable lead times are integrated into the approach in the section “combination of customer required capacity with the capacity order characteristic”) is calculated using

$$k_t = \sum_{i=1}^{l} x_t \left( p_i + \frac{r_i}{q_i} \right)$$  \hspace{1cm} (1)

The parameter $q_i$ denotes the applied lot-size for the $i$th product type, $r_i$ the set-up time and $p_i$ the processing time for the $i$th product type. The value $x_t$ describes the random variable demand rate of product type $i$ in period $t$. The number of different items is denoted by $l$. The average production time needed for one product equals $\left(p_i + \frac{r_i}{q_i}\right)$. The customer required capacity is yielded by multiplying this expression by the sales and summing up over all product types. In general the customer required capacity $k_t$ has a high short-term variability and will sometimes be greater than the available capacity $K$. One possible measure for managing these demand peaks is to temporarily pre-produce (in a period with relatively low customer required capacity) on stock. This temporary on stock pre-production is modeled by a time-average operator. The random variable mean value $k_{t,n}$ of $n$ successive customer required capacities $k_t, k_{t+1}, \ldots, k_{t+n-1}$ describes the required capacity if the implemented work ahead window in the MTO environment is $n$ periods long.

$$k_{t,n} = \sum_{i=1}^{l} \left( p_i + \frac{r_i}{q_i} \right) \frac{1}{n} \sum_{t=1}^{t+n-1} x_t$$  \hspace{1cm} (2)

The interpretation of the average customer required capacity $k_{t,n}$ is as follows: If the available capacity is larger than $k_{t,n}$ and all future customer orders are known and if customer orders are released to the production $n$ periods (= work ahead window) before the due date then all customer requirements will be met. The fixed order period (summarizing the demand of $n$ succeeding periods to one production lot) rule is applied for the lot-sizing.

$$q_i = \Delta \sum_{t=1}^{t+n-1} x_t$$  \hspace{1cm} (3)

In this equation, $\Delta$ denotes the duration of one period, for instance one day. Without loss of generality $\Delta = 1$ is assumed and the parameter $\Delta$ is neglected in the remainder. This leads to

$$k_{t,n} = \frac{1}{n} \sum_{t=1}^{l} \left( p_i + \frac{r_i}{q_i} \right) \sum_{t=1}^{t+n-1} x_t = \frac{1}{n} \sum_{i=1}^{l} p_i \sum_{t=1}^{t+n-1} x_t + \frac{R}{n}$$  \hspace{1cm} (4)

In Eq. (4) $R$ denotes the sum of all change-over-times.

$$R = \sum_{i=1}^{l} r_i$$  \hspace{1cm} (5)

The variance of the average required capacity $k_{t,n}$ is smaller than the variance of the customer required capacity $k_t$. This characteristic is the basis for calculating the minimum possible number of periods needed to ensure that the average customer required capacity is less than the available capacity with a certain probability. In order to do so, a capacity oriented service-level $s$ is introduced as the probability that the average customer required capacity is less than the available capacity. If the average required capacity is in some periods higher than the available capacity then there will be a problem to meet the customer requested due dates (under the assumption of fixed order period FOP $n$ lot-sizing rule and the proposed release policy with work ahead window equals $n$ periods). On the other hand if the average required capacity is less than the available capacity then the customer requirements will be met. The introduced capacity oriented service-level is equal to the well-known $x$-service-level if the workload (processing as well as set-up time) of every customer order is identical and to the $\beta$-service-level if the workload of every item is identical. For a capacity constrained resource or for machine intensive branches, a capacity oriented service-level is useful because it describes the ratio of capacity which is used for production in time over customer requested capacity.

The probability that the average customer required capacity is less than the available capacity $K$ should be equal to the aimed and predefined capacity oriented service level $s$.

$$F_{k_{t,n}}(s) = s$$  \hspace{1cm} (6)

Eq. (6) is an implicit equation for the number of periods $n$ needed. A greater $n$ correlates with a higher capacity oriented service level and more pre-produced items on stock. Expression $F_{k_{t,n}}(\cdot)$ denotes the statistical distribution function of the $n$-periods average of customer required capacity. In our model the service level $s$ describes the ratio of customer required capacity which is used for products which are delivered in time over available capacity $K$.

The expectation value $\mu_{k,n}$ as well as the variance $\sigma_{k,n}^2$ of the average customer required capacity $k_{t,n}$ assuming independent sales $x_{t,i}$ is determined by

$$\mu_{k,n} = \sum_{i=1}^{l} \left( p_i \mu_{x_i} + \frac{r_i}{n\Delta} \right) = \frac{1}{n} \sum_{i=1}^{l} p_i \mu_{x_i} = \frac{R}{n\Delta} \mu_{x_i} = \mu_{x_i} + \frac{R}{n\Delta}$$  \hspace{1cm} (7)

$$\sigma_{k,n}^2 = \frac{1}{n} \sum_{i=1}^{l} p_i^2 \sigma_{x_i}^2 = \frac{1}{n} \sum_{i=1}^{l} p_i^2 \sigma_{x_i}^2 = \frac{1}{n} \sigma_{k}^2$$  \hspace{1cm} (8)
Whereby $\mu_k$ describes the non time varying expectation value of the customer required capacity without change-over times and $\sigma_k^2$ the variance of the customer required capacity $k$.

The expectation value of the average customer required capacity without change-over times has to be less than the expectation value of the average customer required capacity. If the available capacity is less than the long-term average customer required capacity the number of backorders or lost sales will continuously increase. To prevent this it is assumed that the expectation value of the average customer required capacity has to be less than or equal to the available capacity. Such assumptions are often taken in queuing theory as well (see Hopp and Spearman, 1996) to discuss a stationary setting. Putting all together:

$$\begin{align*}
\mu_k &< K \\
\mu_k &\leq K
\end{align*}$$

(9) (10)

In the case of a normal distribution Eq. (6) is explicitly solvable. In Bish et al. (2005) the applicability of the normal distribution is discussed. In their article the probability of negative values is assumed to be negligible. Furthermore, in this paper only the right tail of the distribution function is used (see Eq. (6) and the fact that $s$ should be near 1).

**Proposition 1.** Considering normally distributed independent sales (or equivalent: $F_{k,s}$ equals the distribution function of the normal distribution with expectation value $\mu_{k,s}$ and variance $\sigma_{k,s}^2$) and assuming $\mu_k < K$ the number of periods needed for pre-production on stock to ensure a capacity oriented service-level $s$ with available capacity $K$ is given by

$$n = \left( \frac{F_{N(0,1),\delta}(s)\sigma_k + \sqrt{\frac{F_{N(0,1),\delta}(s)\sigma_k^2 + 4(K - \mu_k)^2}{2(K - \mu_k)}}}{2} \right)^2$$

(11)

whereby $F_{N(0,1),\delta}(\cdot)$ denotes the quantile (inverse of the standard normal distribution function). Furthermore, the capacity invested (available capacity) can be expressed by

$$K = \mu_k + \frac{R}{\Delta n} + \frac{F_{N(0,1),\delta}(s)\sigma_k}{\sqrt{n}}$$

(12)

**Proof.** See Appendix A. □

The available capacity $K$ is the sum of three terms.

$$K = K_0 + K_{\text{setup}} + K_{\text{excess}}$$

whereby

$$\begin{align*}
K_0 &= \mu_k \\
K_{\text{setup}} &= \frac{R}{\Delta n} \\
K_{\text{excess}} &= \frac{F_{N(0,1),\delta}(s)\sigma_k}{\sqrt{n}}
\end{align*}$$

(13)

The first part $K_0$ is the capacity needed for the production of the mean customer demand without set-up. The set-up capacity $K_{\text{setup}}$ equals the capacity needed to perform the change-overs. The last term of the sum $K_{\text{excess}}$ denotes the excess capacity which is used to meet the customer requirements in demand peak situations.

For practical usage the number of periods needed has to be rounded up to the next integer. This article uses the real value because in the remainder the derivative of expression (11) is needed. In case of negligible change-over times a simplification is achieved.

**Corollary 1.** Considering normally distributed independent sales and assuming no change-over times ($R=0$) as well as $\mu_k < K$, the number of periods needed for pre-production on stock to ensure a service-level $s$ with available capacity $K$ is given by

$$n = \left( \frac{F_{N(0,1),\delta}(s)\sigma_k}{(K - \mu_k)} \right)^2$$

(14)

**Proposition 1** shows that the number of periods needed to manage temporary demand peaks is an increasing function with respect to the variance of the sales, the mean of the sales, the total set-up time as well as of the service level and a decreasing function with respect to the available capacity. The capacity oriented work ahead window which equals the duration of the time period used for averaging is determined by

$$h_k = n\Delta$$

(15)

### 2.2. Order characteristic

In this section the implicit assumption that all future customer orders are known is relaxed by taking into account the customer buying behavior as well as the customer requested delivery lead times. The question of how long before the customer requested due date how much capacity is booked by customer orders is answered. Based on the statistical distribution of the customer requested delivery lead times of the product types the statistical distribution of the capacity delivery lead time can be determined. The random customer required delivery lead time of the $i$th product type $l_i$ denotes the time period between ordering the product by the customer and the customer...
If $Z(n)$ is defined by the workload weighted average (processing as well as set-up time according to the bill of material and FOP $n$) of the customer delivery lead times of all product types.

$$Z(n) = \frac{1}{\sum_{i=1}^{l} p_i \mu_{n_i}} \sum_{i=1}^{l} L_i \left(p_i \mu_{n_i} + \frac{r_i}{n\lambda}\right)$$ (16)

The customer required capacity lead time $Z(n)$ describes the time period between when a capacity is booked by a customer order and the customer requested due date. The random variable $Z(n)$ depends on the number of periods used for the lot-sizing as well as production order release policy. The expectation value and the variance of the customer required capacity lead time is determined by:

$$\mu_{Z(n)} = \frac{1}{\sum_{i=1}^{l} p_i \mu_{n_i}} \sum_{i=1}^{l} \mu_{n_i} \left(p_i \mu_{n_i} + \frac{r_i}{n\lambda}\right)$$

$$\sigma_{Z(n)}^2 = \frac{1}{\sum_{i=1}^{l} p_i \mu_{n_i}} \sum_{i=1}^{l} \sigma_{n_i}^2 \left(p_i \mu_{n_i} + \frac{r_i}{n\lambda}\right)^2$$ (17)

The mean delivery time for the $i$th product type is denoted by $\mu_{n_i}$ and the variance of the delivery lead time of the $i$th product type by $\sigma_{n_i}^2$. The buying behavior of the customers, e.g. their requested delivery lead-time is illustrated by the capacity order characteristic. The capacity order characteristic $O_n(\tau)$ describes how long before the due date how much capacity is booked by customer orders.

$$O_n(\tau) = 1 - F_{Z(n)}(\tau)$$ (18)

The expression $F_{Z(n)}(\tau)$ denotes the statistical distribution function of the capacity lead-time. For vanishing change-over times (17) and (18) simplify to

$$\mu_{Z_0} = \frac{1}{\sum_{i=1}^{l} p_i \mu_{n_i}} \sum_{i=1}^{l} \mu_{n_i} p_i \mu_{n_i}$$

$$\sigma_{Z_0}^2 = \frac{1}{\sum_{i=1}^{l} p_i \mu_{n_i}} \sum_{i=1}^{l} \sigma_{n_i}^2 \left(p_i \mu_{n_i}\right)^2$$

$$O(\tau) = 1 - F_{Z_0}(\tau)$$ (19)

whereby

$$Z_0 = \frac{1}{\sum_{i=1}^{l} p_i \mu_{n_i}} \sum_{i=1}^{l} L_i p_i \mu_{n_i}$$

If the change-over times are neglected, the expectation value as well as the variance of the capacity lead time and the capacity order characteristic are independent of the number of periods needed for pre-production to meet the customer requirement in temporary demand peaks.

The following Lemma 1 states some important relationships between the capacity delivery lead time and the capacity order characteristic.

**Lemma 1.** If $F_{Z(n)}(\tau) = 0$ for $\tau \leq 0$ (ordering after due date is not allowed), then for the capacity order characteristic the following holds true:

$$\int_{0}^{\infty} O_n(\tau) d\tau = \mu_{Z(n)}$$ (20)

$$\int_{0}^{\infty} f_{Z(n)}(\tau + h) d\tau = 1 - F_{Z(n)}(h) = O_n(h)$$ (21)

$$\int_{0}^{h} f_{Z(n)}(\tau - h) d\tau = F_{Z(n)}(h) = 1 - O_n(h)$$ (22)

**Proof.** See Appendix A. □

The first equation of Lemma 1 (Eq. (20)) states that the area under the capacity order characteristic curve is equal to the expectation value of the customer required capacity delivery lead time. The following Fig. 2 illustrates the interpretation of Eqs. (21) and (22).

In Fig. 2, the horizontal axis shows the due date and the vertical axis the customer order date. The dotted line illustrates the statistical distribution function of the customer required capacity delivery lead time. This density function equals zero for due dates less than the order date, because it is not allowed to order after the due date. Consequently, there are only valid pairs of due dates and order dates on the right-upper side of the diagonal. The remaining lead time in Fig. 2 denotes the time period from current date until the due date. According to Fig. 2, the left side of Eq. (21) (meaning the integral over the statistical distribution function of the customer required capacity delivery lead time) can be interpreted as the proportion of capacity already booked by customer orders with remaining lead time which equals $h$. Furthermore, the left side of Eq. (22) (meaning the integral over the statistical distribution function of the customer required capacity delivery lead time) describes the proportion of capacity which will be booked by customer orders in the near future with the due date in $h$ time. Consequently, the capacity order characteristic $O_n(h) = 1 - F_{Z(n)}(h)$ has two important interpretations:
(1) By definition it equals the probability that the customer required capacity delivery lead time is greater than $h$

(2) It equals the proportion of booked capacity by known customer orders with remaining lead time $h$

For $1 - O_\alpha(h) = F_{Z(\alpha)}(h)$ the following holds true:

(3) By definition it equals the probability that the customer required capacity delivery lead time is shorter than $h$

(4) It equals the proportion of not yet booked capacity with remaining lead time $h$ (but some orders are expected to be allocated to this capacity).

### 2.3. Combination of customer required capacity and the capacity order characteristic

The next step in the model development is the combination between the customer required capacity and the capacity order characteristic. Fig. 3 illustrates this combination.

The area $A$ in Fig. 3 describes the capacity which will be allocated to customer orders which are expected in the near future and with due dates within $P + h_k$ time. The capacity oriented work ahead window $h_k$ is defined by $h_k = n \Delta$ ($n$ is defined by Eq. (11) – $n$ equals the number of periods needed for pre-production to ensure the required service-level), the time period $P$ denotes the minimum possible remaining lead time that is the sum of all required processing, change-over and transportation time (but excluding waiting or queuing times caused by inventory) to finalize the product and to deliver it to the customer. The expected but not yet booked mean customer required capacity with due dates nearer than $P + h_k$ equals

$$
\frac{A}{P + h_k} - \frac{P + h_k}{\Delta} \cdot h_{k,n} = A \frac{1}{\Delta} h_{k,n}
$$

(23)

Because of the fact that the customer orders which will be allocated to $A$ are not yet known, the product type to be ordered and to be produced is unknown too. To avoid an anonymous production to stock, products which are well defined by known customer orders but with due
dates farther in the future are pre-produced. The capacity needed for this pre-production is illustrated by the area \( B \). The booked mean customer required capacity with remaining capacity delivery lead time between \( h \) and \( P + h_k \) equals

\[
\frac{B}{h - P - h_k} \cdot \frac{h - P - h_k}{\Delta} \mu_{k,n} = \frac{B}{\Delta} \mu_{k,n}
\]  

(24)

To be able to manage all expected short term customer orders the area \( B \) (proportion of required capacity by known customer orders with due dates long in the future) should be equal to the area \( A \) (proportion of required capacity by unknown but expected customer orders with short due dates). The so called work ahead window \( h \) is defined by the condition \( A \) should be equal to \( B \). If all known customer orders which have a remaining lead time shorter than \( h \) (or equivalently which have a due date nearer than \( h \)) are released to production the customer oriented service level \( s \) with the available capacity \( K \) can be reached. For the work ahead window \( h \) the following implicit equation based on \( A = B \) can be derived (see Appendix A) and can be easily understood by interpreting Fig. 3:

\[
\int_0^h 1 - F_{Z(n)}(t) dt = P + h_k
\]  

(25)

Because of (20) it is clear that Eq. (25) has a solution if and only if the time period \( P + h_k \) is shorter than the mean customer required capacity delivery lead time \( \mu_{Z(n)} \). If no finite work ahead window \( h \) is found to fulfill Eq. (25), it is not possible to install a pure MTO system because not enough customer orders are available. There are several strategies to enable an MTO system. Some of these alternatives may be to increase available capacity, to reduce demand fluctuations or to move the capacity order characteristic to the right (which means to induce customers to accept longer delivery lead times). In further research work it is planned to discuss this in more detail. This article is concerned exclusively with MTO systems so it is assumed that the next inequality (26) holds true:

\[
\mu_{Z(n)} > P + h_k = P + n(K)\Delta
\]  

(26)

In inequality (26) as well as in the remainder the number of periods \( n \) is formulated as a function with respect to the available capacity \( K \), to illustrate the dependency of \( n \) on \( K \) according to Eq. (11).

2.4. Determination of the required inventory

The random variable lead time is defined as the time it takes from production release until shipment. Thus the lead time is the sum of the production lead time and the queuing time caused by the finished goods inventory (in case of pre-production). Based on the interpretation of the work ahead window that a customer order is released to production, if the remaining time to due date is shorter than the work ahead window, the determination of the expected lead time is performed.

There are three cases to differentiate as shown in Fig. 4.

Case 1: Customer orders with very long customer requested delivery lead times (delivery lead time is longer than the work ahead window \( h \)) are released exactly at \( h \) time before due date to production. So for these customer orders the lead time is equal to the work ahead window.

Case 2: Customer orders with very short customer required delivery lead times (delivery lead time is shorter than the minimum required lead time \( P \)) have at least a lead time of \( P \) (faster than \( P \) is not possible).

Case 3: All other customer orders with customer requested delivery lead times shorter than \( h \) but longer than \( P \) are released to production at the time when the customer orders are received and therefore the lead time equals the customer required delivery lead time.

By applying Little’s law, an equation for the expected inventory (work in process plus finished goods inventory) needed with respect to the available capacity is yielded.
**Proposition 2.** Considering normally distributed independent sales and assuming \( \mu_k < K \) as well as \( \mu_{z,n} > P + h_k \) the mean inventory needed to ensure a service-level \( s \) with available capacity \( K \) is given by

\[
\mu_y = \left( P + \int_0^P F_{Z|(0) K}(t) \, dt + n(K)\Delta \right) \mu_{y,0,K}
\]  

(27)

whereby \( n(K) \) is defined by (11).

**Proof.** See Appendix A. \( \square \)

The expected inventory needed consists of four terms of a sum

\[
\mu_y = \left( P + \int_0^P F_{Z|(0) K}(t) \, dt \right) \mu_k + \left( P + \int_0^P F_{Z|(0) K}(t) \, dt \right) \frac{R}{\Delta n(K)} + n(K)\Delta \mu_k + n(K)\Delta - \frac{R}{\Delta n(K)}
\]  

(28)

Thus the following identity holds true:

\[
\mu_y = Y_0 + Y_{\text{setup}} + Y_{\text{surplus}} + R
\]

whereby

\[
Y_0 = \left( P + \int_0^P F_{Z|(0) K}(t) \, dt \right) \mu_k
\]

(29)

\[
Y_{\text{setup}} = \left( P + \int_0^P F_{Z|(0) K}(t) \, dt \right) \frac{R}{\Delta n(K)}
\]

(30)

\[
Y_{\text{surplus}} = n(K)\Delta \mu_k
\]

The first part of \( Y_0 \) is the inventory which is needed to cover the minimum necessary lead time and the second part of \( Y_0 \) is needed to protect against stock-outs for customer orders with a very short delivery lead time. For both parts of \( Y_0 \) setup times are not taken into account. The set-up inventory \( Y_{\text{setup}} \) is caused by the required set-ups during the minimum necessary lead time. The surplus inventory \( Y_{\text{surplus}} \) equals the pre-produced stock to meet the customer requirements in demand peak situation without taking the set-up times into account. The value \( R \) denotes the sum of all set-up times.

In the case of vanishing set-up times (27) simplifies to

\[
\mu_y = \left( P + \int_0^P F_{Z|(0) K}(t) \, dt + n(K)\Delta \right) \mu_k
\]

(30)

2.5. **Determination of the optimal pair of capacity invested and inventory needed without change-over times**

The objective is to find the optimal capacity investment and the optimal capital which should be invested in inventory to reduce the overall costs. In the model, the capacity invested is the decision variable. According to Eq. (27) the inventory needed is a function with respect to the capacity invested to ensure that the customer requirements can be met with the service level \( s \). The customer requirements are defined by the parameters: mean demand, variance of demand and the statistical distribution of the delivery lead time. The presented optimization problem consists of two stages. In the first stage the management decision for a certain service level has to be taken. In the second stage the capacity investment is optimized based on that management decision.

We start with the simpler case where there are no change-over times. Under the assumption of vanishing set-up time it is possible to present the solution to the minimization problem in a closed form.

**Proposition 3.** Considering normally distributed independent sales and assuming \( R = 0 \), \( \mu_k < K \) as well as \( \mu_{z,n} > P + h_k \), the solution of the non linear cost minimization problem

\[
c(K) = \mu_y(K)c_y + kc_K \rightarrow \min
\]

is

\[
K_{\text{opt}} = \mu_k + \frac{2\mu_k c_H\Delta \left( F_{N(0),1}(s)\sigma_k \right)^2}{c_K}
\]

(32)

\[
\mu_y(K_{\text{opt}}) = \left( P + \int_0^P F_{Z_0}(t) \, dt \right) \mu_k + \left( F_{N(0),1}(s)\sigma_k c_K \right)^2 \frac{\Delta \mu_k}{2 \mu_k c_H \Delta}
\]

(33)

\[
n(K_{\text{opt}}) = \frac{F_{N(0),1}(s)\sigma_k c_K}{2 \mu_k c_H \Delta}
\]

(34)

\[
c(K_{\text{opt}}) = \mu_k c_k + \left( 2\mu_k c_H\Delta \left( F_{N(0),1}(s)\sigma_k c_K \right)^2 \right)^\frac{1}{2} + \left( P + \int_0^P F_{Z_0}(t) \, dt \right) \mu_k c_H + \left( \frac{\mu_k c_H \left( F_{N(0),1}(s)\sigma_k c_K \right)^2}{2 \Delta} \right)^\frac{1}{2} \Delta
\]

(35)

whereby the capacity cost for one capacity unit is denoted by \( c_k \) and the holding cost for one unit stored for a year by \( c_H \). The cost function \( c(K) \) denotes the cost of machine investment as well as the capital costs of holding inventory for a fixed time horizon of one year.

**Proof.** See Appendix A. \( \square \)
The second term of the sum on the right side of Eq. (32) or (33) can be interpreted as the optimal excess capacity and optimal surplus inventory, respectively. Both the excess capacity as well as the surplus inventory are used to manage demand peaks. Proposition 3 and the following Corollary 2 show the cost minimum trade-off between excess capacity and surplus inventory in the case of no change-overs.

**Corollary 2.** Under the condition of Proposition 3, for the optimal choice of capacity invested and inventory needed, the following identity holds true: The double of the surplus inventory cost is equal to the cost of the excess capacity

\[ K_{\text{excess,opt}} C_k = 2Y_{\text{surplus,opt}} C_h \]  

whereby

\[ K_{\text{excess,opt}} = \left( \frac{2\mu C_h \Delta \left( F_{\text{N}(0,1)}^{-1}(s) \sigma_k \right)^2}{\mu_k \Delta} \right)^\frac{1}{2} \]

\[ Y_{\text{surplus,opt}} = \left( \frac{F_{\text{N}(0,1)}^{-1}(s) \sigma_k C_k}{2\mu C_h \Delta} \right)^\frac{1}{2} \mu \Delta \]

**Proof.** See Appendix A. □

Corollary 2 gives an important managerial insight: When the change-over times are negligible the optimal trade-off between the capacity costs and the inventory costs is that the double of the surplus inventory costs should be equal to the costs for excess capacity.

The next corollary discusses the optimal utilization, which is defined as the ratio of average customer required capacity to available capacity.

**Corollary 3.** Under the condition of Proposition 3 for the optimal choice of capacity invested and inventory needed the following cost optimal utilization is gained

\[ U_{\text{opt}} = \frac{\mu_k}{K} = \frac{1}{2 \left( \frac{F_{\text{N}(0,1)}^{-1}(s) \sigma_k C_k}{\mu C_h \Delta} \right)^\frac{1}{2} + 1} \]

**Proof.** See Appendix A. □

The optimal utilization is a decreasing function in service level, coefficient of variation of capacity demand and unit holding cost. Or in other words, more excess capacity is needed for better service level, higher coefficient of variation of capacity demand or for greater unit holding cost to ensure cost optimality. Furthermore, the utilization is an increasing function with respect to unit capacity cost. In case of greater unit capacity cost, it is advantageous to reduce excess capacity. To summarize: The optimal utilization is a decreasing function with respect to the coefficient of variation of the capacity demand, to the service level and to the ratio unit holding cost over unit capacity cost, whereby the influence of the first two (power of 2/3) is higher than of the last one (power of 1/3).

In the next section the more complicated case with set-up times is discussed.

### 2.6. Determination of the optimal pair capacity invested and inventory needed including change-over times

For the case with set-ups, a quadratic equation is developed to find the optimal trade-off between capacity invested and inventory needed. Because of the complexity of the closed form solution, we solve it numerically, whereas a generalization of Corollary 2 is explicitly gained.

In practical usage, the minimum required lead time \( P \) (processing time plus set-up time) will be considerably less than the expected customer requested capacity delivery lead time (only a very small number of customer orders have a delivery lead time shorter than the minimum required lead time). Furthermore, the influence of the number of periods which are summarized to one production lot on the statistical distribution of the capacity delivery lead time is very small (the weighting in (16) change only marginally), so all together we assume that

\[ \int_0^P F_{\text{Z}(0)}(t) \, dt \approx \int_0^P F_{\text{Z}(0)}(t) \, dt < < P \]

holds true. Eq. (39) is technically necessary for the proof and has no impact in practical usage because the left side of (39) vanishes in real application.

**Proposition 4.** Considering normally distributed independent sales and assuming (39), \( \mu_k < K \) as well as \( \mu_{z_0} > P + h_k \) then for the solution of the non linear cost minimization problem

\[ c(K) = \mu_k (K) C_h + K C_k \rightarrow \min \]  

the following equation holds true:

\[ -2\mu C_h \mu^2 (K_{\text{opt}}) + F_{\text{N}(0,1)}^{-1}(s) \sigma_k C_k \sqrt{\mu (K_{\text{opt}})} + 2 \frac{\mu}{\Delta} \left( \int_0^P F_{\text{Z}(0)(K)}(t) \, dt \right) C_h + C_k = 0 \]

For the optimal pair \( K_{\text{opt}} \) and \( \mu_k (K_{\text{opt}}) \) the double of the surplus inventory cost is equal to the excess capacity cost, plus double the set-up inventory cost, plus double the set-up capacity cost.

\[ 2Y_{\text{surplus}} C_h = K_{\text{excess}} C_k + 2Y_{\text{setup}} C_h + 2K_{\text{setup}} C_k \]
Proof. See Appendix A. □

Eq. (41) is a quadratic equation for the square root of the unknown number of periods for the work ahead window. Because of the fact that the closed form solution of the quadratic equation is a very complex expression and no additional insights are expected from the closed form solution, Eq. (41) is solved numerically by standard routines and taking into account that the optimal work ahead window has to be greater than the work ahead window of the corresponding problem without set-up time. After determining the optimal \( n(\text{K}_{\text{opt}}) \) by solving this quadratic equation the inventory needed can be calculated by applying Eq. (27) and the optimal capacity to be invested in by applying Eq. (12).

In the case of no set-up times the general case simplifies to the case without change-overs: Eq. (42) simplifies to (36) and (41) to

\[
-2\Delta \mu_1 c_H n^2(K) + F_{N(0,1)}(s)\sigma_k c_k \sqrt{n(K)} = 0 \tag{43}
\]

By solving (43) the Eq. (34) for the optimal number of periods is gained.

3. Numerical illustration of the results

In this section, four different settings are compared. The basic setting is referenced by A and is defined by the following data (see Table 2).

The statistical behavior of the customer required capacity as well as of the capacity delivery lead time is described by a normal distribution. The three comparison settings are defined by changing only one parameter. For setting B, the service level is increased to 0.95, for C the mean customer required capacity is increased to 580 and for D the total set-up time is equal to 100. In Table 3 and Figs. 5–7 these four settings are presented and compared.

The horizontal axis in Figs. 5–7 denotes the capacity \( K \). The inventory functions of the left hand curves are determined by Eq. (27) and the cost functions of the right hand curves are calculated by

\[
c(K) = \mu_1(K)c_H + Kc_k \rightarrow \min \tag{44}
\]

The utilization in Table 3 is defined by

\[
U = \frac{H_{K,A}}{K} \tag{45}
\]

The figures demonstrate that the inventory needed with respect to the capacity invested is a decreasing and convex function and that the total cost function with respect to the capacity invested is a convex function. The inventory curves show that there is a vertical asymptotic at \( K_0 + K_{\text{setup}} \) and a horizontal asymptotic with constant value \( Y_0 + Y_{\text{setup}} \). An increase in the service level, mean customer required capacity as

---

Table 2

<table>
<thead>
<tr>
<th>Definition of the basic setting A.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>Service level</td>
</tr>
<tr>
<td>Mean capacity demand</td>
</tr>
<tr>
<td>Deviation of capacity demand</td>
</tr>
<tr>
<td>Minimum required lead time</td>
</tr>
<tr>
<td>Mean capacity delivery lead time</td>
</tr>
<tr>
<td>Deviation of capacity delivery lead time</td>
</tr>
<tr>
<td>Total set-up time</td>
</tr>
<tr>
<td>Unit holding cost</td>
</tr>
<tr>
<td>Unit capacity cost</td>
</tr>
</tbody>
</table>

Table 3

<table>
<thead>
<tr>
<th>Comparison of the four settings.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>( K_{\text{opt}} )</td>
</tr>
<tr>
<td>( K )</td>
</tr>
<tr>
<td>( Y_{\text{setup}, \text{opt}} )</td>
</tr>
<tr>
<td>( \mu_1(K_{\text{opt}}) )</td>
</tr>
<tr>
<td>( c_0 )</td>
</tr>
<tr>
<td>( Y_{\text{setup}, \text{opt}} )</td>
</tr>
<tr>
<td>( Y_{\text{setup}, \text{opt}} )</td>
</tr>
<tr>
<td>( H(K_{\text{opt}}) )</td>
</tr>
<tr>
<td>( c(K_{\text{opt}}) )</td>
</tr>
<tr>
<td>( K_{\text{setup}, \text{opt}} + K_{\text{setup}, \text{opt}} )</td>
</tr>
<tr>
<td>( 2(Y_{\text{setup}, \text{opt}} - Y_{\text{setup}, \text{opt}})K )</td>
</tr>
<tr>
<td>( U_{\text{opt}} )</td>
</tr>
</tbody>
</table>
well as in the total set-up time causes a higher inventory needed if the same capacity is invested. Furthermore, for this case a higher optimal capacity invested is caused, a higher optimal inventory needed and higher optimal costs are achieved. In all four settings the relationship between excess capacity, set-up capacity, surplus inventory and set-up inventory (42) (or its simpler form (36) for the case of no change-over times) is confirmed for the optimal choice of capacity invested and inventory needed (see Table 3).

**Fig. 5.** Illustration of the influence of service-level.

**Fig. 6.** Illustration of the influence of mean customer required capacity.

**Fig. 7.** Illustration of the influence of set-up time.
The findings of this paper and the numerical study confirm intuition and deliver a function to explicitly calculate the intuitive assumptions. Concerning the application of this model, the four cases studied could be related to a production system controlled with a CONWIP method. For the CONWIP application, the parameter WIP-cap would be neglected. For this CONWIP production system, the optimal work ahead window and capacity investment can be calculated and the cases A to D could be discussed as what if cases for the production system.

The next test discusses the impact of the ratio unit holding cost to unit capacity cost on optimal utilization. For this purpose the setting A is compared with setting E, whereby all parameters for setting E are equal to the setting A except the unit costs. The unit holding cost equals 100 and the unit capacity cost equals 1 for parameter setting E. Consequently A can be interpreted as an environment with dominant unit capacity cost (e.g. steel production) whereas for E the unit inventory cost is the most important driver (e.g. diamond grinding). Fig. 8 illustrates the cost with respect to utilization.

As expected, in case A (low ratio unit holding cost over unit capacity cost) the optimal utilization is near 1, which means full utilization. In case E (high ratio unit holding cost over unit capacity cost) the optimum utilization is considerably smaller. For practical usage, the different shapes of the two curves may be more interesting and more important. In case A there is a considerable increase in the cost if the utilization is a little higher than optimal. In case E, the cost with respect to utilization is nearly a constant function in the interval 0.2 to 0.5.

4. Conclusions

In this paper a multi-item MTO production system with stationary stochastic demand is considered. The relationship between available capacity and the inventory needed to meet customer requirements with a pre-defined service level is studied. The market is described by the statistical distribution of the demand and by the statistical distribution of the customer required delivery lead time. It is assumed that customer orders are released to production if the remaining time until their due date is shorter than the work ahead window. The Fixed Order Period lot-sizing rule is applied. The first goal of the paper is to develop a function describing the inventory needed with respect to capacity invested. The second goal is to minimize the cost for the capital invested in capacity as well as in inventory.

It is shown that the inventory needed is a decreasing and convex function with respect to the invested capacity. Furthermore, an increase in service level, mean customer demand as well as set-up time necessitates higher inventory needed for the same capacity invested. The cost function with respect to the capacity invested is convex and for the global minimum capacity invested the following characteristic is gained: Double the surplus inventory cost, minus double the setup inventory cost equals the excess capacity cost plus the setup capacity cost. In the case of vanishing change-over times this characteristic simplifies to: Double the surplus inventory cost equals the excess capacity cost.

The model illustrates that there is an impact of the utilization on the cost for available capacity and inventory needed. Especially in the case of a small ratio of unit holding cost over unit capacity cost, the cost is considerably increased if the utilization is only a little higher or smaller than optimal.

The model can be used to support a capacity investment decision by including the expected capital which has to be invested in the inventory in the total cost, whereby the inventory needed is determined by Eq. (27). The characteristic of the global minimum can be used to analyze the relationship between capacity and inventory. If the excess capacity cost is much greater than double the surplus inventory cost, divesting the capacity will improve the cost structure, and if the excess capacity cost is much smaller than double the surplus inventory cost, an increase in available capacity will improve the cost structure. Furthermore, Eq. (26) shows how much capacity is needed to be able to install or to run a pure MTO system.

In further research the model should be extended to Make to Stock systems. In this paper only costs for capacity and inventory were considered. An extension to integrate penalty costs for capacity not delivered on time could be discussed in further research. Furthermore, an extension of the pure cost objective to revenue, lost sales and backorders depending on service level may be aimed. Further development of the model should take into account that the cost of investment is not incurred at the same time as the inventory cost. Additionally, the result stability concerning costs and service level when different values for the stochastic variables occur could be tested in further work.
Appendix A

A.1. Proof of Proposition 1

\[ F_{N_1, \nu + \frac{\nu^2}{\nu + 1}}(K) = s \]
\[ \iff F_{N_{(1), k}}(\frac{K - \mu_k}{\sqrt{k} \sigma_k}) = s \]
\[ \iff K - \mu_k - \frac{R}{\Delta} = F_{N_{(1), 1}}(s) \]
\[ \iff n(K - \mu_k) - \sqrt{n}F_{N_{(0, 1), 1}}(s)\sigma_k - \frac{R}{\Delta} = 0 \]
\[ \iff n = \left( F_{N_{(0, 1), 1}}(s)\sigma_k + \sqrt{(F_{N_{(0, 1), 1}}(s)\sigma_k)^2 + 4(K - \mu_k) \frac{R}{\Delta}} \right) \]
\[ n(K - \mu_k) - \sqrt{n}F_{N_{(0, 1), 1}}(s)\sigma_k - \frac{R}{\Delta} = 0 \]
\[ \iff K = \mu_k + \frac{R}{\Delta n} + \frac{F_{N_{(0, 1), 1}}(s)\sigma_k}{\sqrt{n}} \]

A.2. Proof of Lemma 1

\[ \mu_{Z(a)} = \int_0^\infty F_{Z(0)}(\tau)d\tau = \lim_{a \to -\infty} \left( \int_0^a F_{Z(0)}(\tau)d\tau - \int_0^a F_{Z(0)}(\tau)d\tau \right) = \lim_{a \to -\infty} \left( \int_0^a F_{Z(0)}(\tau) - F_{Z(0)}(\tau)d\tau \right) \]
\[ = \int_0^\infty 1 - F_{Z(0)}(\tau)d\tau = \int_0^\infty O(\tau)d\tau \]
\[ \times \int_0^h F_{Z(0)}(\tau + h)d\tau = \int_0^h F_{Z(0)}(\tau)d\tau = 1 - F_{Z(0)}(h) \]
\[ \times \int_0^h F_{Z(0)}(h - \tau)d\tau = - \int_0^h F_{Z(0)}(\tau)d\tau = \int_0^h F_{Z(0)}(\tau)d\tau = F_{Z(0)}(h) \]

Derivation of the implicit Eq. (25) for the work ahead window \( h \)

\[ A = B \]
\[ \iff \]
\[ \int_0^{P + h} F_{Z(0)}(t)dt = \int_0^h 1 - F_{Z(0)}(t)dt \]
\[ \iff \]
\[ \int_0^h F_{Z(0)}(t)dt = h - (P + h_k) \]
\[ \iff \]
\[ \int_0^h 1 - F_{Z(0)}(t)dt = P + h_k \]

A.3. Proof of Proposition 2

\[ \mu_{\text{lead time}} = \int_0^\infty \max(P, \min(h, t))f_{Z(0)}(t)dt = P \int_0^P f_{Z(0)}(t)dt + \int_P^h f_{Z(0)}(t)dt + h \int_h^\infty f_{Z(0)}(t)dt \]
\[ = PF_{Z(0)}(P) + hF_{Z(0)}(h) - PF_{Z(0)}(P) - \int_P^h F_{Z(0)}(t)dt + h - hF_{Z(0)}(h) \]
\[ = h - \int_P^h F_{Z(0)}(t)dt = \int_0^h 1 - F_{Z(0)}(t)dt + \int_P^h F_{Z(0)}(t)dt = P + h_k + \int_0^P F_{Z(0)}(t)dt \]
\[ \Rightarrow \text{(by applying Little's Law)} \]
\[ \mu_F = \left( P + h_k + \int_0^P F_{Z(0)}(t)dt \right) \mu_{k,n} \]
A.4. Proof of Proposition 3

\[ c(K) = \mu_t(K)c_H + Kc_K \]
\[ c'(K) = \mu_t'(K)c_H + c_K \]
\[ \mu_t(K) = \left( P + \int_0^p F_{Z_1}(t)dt + n(K)\Delta \right) \frac{\mu_k}{\gamma} = \left( P + \int_0^p F_{Z_1}(t)dt + \frac{F_{N(0,1)}^{-1}(s)\sigma_k}{(K - \mu_k)^2} \Delta \right) \mu_k \]
\[ \mu_t'(K) = -2\mu_k\Delta \frac{\left( F_{N(0,1)}^{-1}(s)\sigma_k \right)^2}{(K - \mu_k)^2} + c_K = 0 \]
\[ \Rightarrow K_{opt} = \left( 2\mu_kC_H \frac{\left( F_{N(0,1)}^{-1}(s)\sigma_k \right)^2}{\gamma} \right)^{\frac{1}{\gamma}} + \mu_k \]
\[ \mu_t(K_{opt}) = \left( P + \int_0^p F_{Z_1}(t)dt + \frac{F_{N(0,1)}^{-1}(s)\sigma_k}{2\mu_kC_H \frac{\left( F_{N(0,1)}^{-1}(s)\sigma_k \right)^2}{\gamma} \Delta} \right) \mu_k = \left( P + \int_0^p F_{Z_1}(t)dt + \frac{F_{N(0,1)}^{-1}(s)\sigma_k c_K}{2\mu_kC_H \frac{\left( F_{N(0,1)}^{-1}(s)\sigma_k c_K \right)^2}{\gamma} \Delta} \right) \mu_k \]
\[ c(K_{opt}) = \mu_t(K_{opt})c_H + K_{opt}c_K \]
\[ = \mu_kc_H + \left( 2\mu_kC_H \frac{\left( F_{N(0,1)}^{-1}(s)\sigma_k c_K \right)^2}{\gamma} \right)^{\frac{1}{\gamma}} + \left( P + \int_0^p F_{Z_1}(t)dt \right) \mu_k + \left( \mu_kC_H \frac{F_{N(0,1)}^{-1}(s)\sigma_k c_K}{2\Delta} \right)^{\frac{1}{\gamma}} \Delta \]

A.5. Proof of Corollary 2

\[ K_{access, opt}c_K \]
\[ = \left( 2\mu_kC_H \frac{\left( F_{N(0,1)}^{-1}(s)\sigma_k c_K \right)^2}{\gamma} \right)^{\frac{1}{\gamma}} = 2 \]

A.6. Proof of Corollary 3

\[ U_{opt} = \frac{\mu_k}{K_{opt}} \]
\[ = \frac{\mu_k}{\left( 2\mu_kC_H \frac{\left( F_{N(0,1)}^{-1}(s)\sigma_k c_K \right)^2}{\gamma} \right)^{\frac{1}{\gamma}} + \mu_k} \frac{1}{\left( \frac{F_{N(0,1)}^{-1}(s)\sigma_k c_K}{2\Delta} \right)^{\frac{1}{\gamma}} + 1} \]

A.7. Proof of Proposition 4

Before proving Proposition 4, a lemma discussing the first derivative with respect to the capacity of the mean average customer required capacity and the capacity oriented work ahead window is formulated.

Lemma 2. For the first derivative of the capacity oriented work ahead window the following differential equation holds true:

\[ n'(K) = -\frac{2n^2(K)}{F_{N(0,1)}^{-1}(s)\sigma_k \sqrt{n(K)} + 2\frac{\mu_k}{\gamma}} \]

and for the first derivative of the mean average customer required capacity with respect to the capacity K the equation

\[ \frac{\partial}{\partial K} \mu_{n(K)} = -\frac{Rn'(K)}{n'(K)} \frac{2\frac{\mu_k}{\gamma}}{F_{N(0,1)}^{-1}(s)\sigma_k \sqrt{n(K)} + 2\frac{\mu_k}{\gamma}} \]

is gained.
A.8. Proof of Lemma 2

\[ n(K)(K - \mu_k) - \sqrt{n(K)F_{N(0,1)}^{-1}(s)\sigma_K} - \frac{R}{\Delta} = 0 \] (see proof of Proposition 1)

\[ n'(K)(K - \mu_k) + n(K) - \frac{n'(K)F_{N(0,1)}^{-1}(s)\sigma_K}{2\sqrt{n(K)}} = 0 \]

\[ n'(K) \left( K - \mu_k - \frac{F_{N(0,1)}^{-1}(s)\sigma_K}{2\sqrt{n(K)}} \right) + n(K) = 0 \]

\[ \Rightarrow \quad n'(K) \left( F_{N(0,1)}^{-1}(s)\sigma_K + \frac{R}{n(K)\Delta} \right) + n(K) = 0 \]

\[ K - \mu_k = \frac{F_{N(0,1)}^{-1}(s)\sigma_K}{\sqrt{n(K)}} + \frac{R}{n(K)\Delta} \text{ from (12)} \]

\[ n'(K) = -\frac{2n^2(K)}{\sqrt{n(K)F_{N(0,1)}^{-1}(s)\sigma_K} + 2 \frac{R}{\Delta}} \]

\[ \mu_{k,u}(K) = \mu_k + \frac{R}{n(K)\Delta} \]

\[ \mu_{k,d}(K) = -\frac{2n^2(K)}{\sqrt{n(K)F_{N(0,1)}^{-1}(s)\sigma_K} + 2 \frac{R}{\Delta}} \]

A.9. Proof of Proposition 4

\[ c(K) = \mu_y(K)c_H + Kc_k \]

\[ \Rightarrow \quad \left( P + \int_0^\rho F_{z_0}(t)dt + n(K)\Delta \right) \mu_{x,y,K}c_H + Kc_k \]

\[ c'(K) = n'(K)\Delta \mu_{x,y,K}c_H + \left( P + \int_0^\rho F_{z_0}(t)dt + n(K)\Delta \right) \mu_{x,y,K}c_H + c_k = 0 \]

\[ -2n^2(K)\Delta \left( \mu_k + \frac{R}{n(K)\Delta} \right)c_H + \left( P + \int_0^\rho F_{z_0}(t)dt + n(K)\Delta \right) 2\frac{R}{\Delta}c_H + c_k \sqrt{n(K)F_{N(0,1)}^{-1}(s)\sigma_K} + 2 \frac{R}{\Delta}c_k = 0 \]

\[ -2n^2(K)\Delta \mu_{x,y,H}c_H + \sqrt{n(K)F_{N(0,1)}^{-1}(s)\sigma_K} + 2 \frac{R}{\Delta} \left( P + \int_0^\rho F_{z_0}(t)dt \right) c_H + 2 \frac{R}{\Delta}c_k = 0 \]

\[ \Rightarrow \quad n(K - \mu_k) - \sqrt{nF_{N(0,1)}^{-1}(s)\sigma_K} - \frac{R}{\Delta} = 0 \]

\[ 2n(K)\Delta \mu_{x,y,H}c_H + \left( n(K)(K - \mu_k) - \frac{R}{\Delta} \right)c_k + 2 \frac{R}{\Delta} \left( P + \int_0^\rho F_{z_0}(t)dt \right) c_H + 2 \frac{R}{\Delta}c_k = 0 \]

\[ \frac{2n(K)\Delta\mu_{x,y,H}c_H}{\frac{R}{\Delta}} = \frac{K - \mu_k - \frac{R}{\Delta}}{c_k} + 2 \frac{R}{\Delta} \left( P + \int_0^\rho F_{z_0}(t)dt \right) c_H + 2 \frac{R}{\Delta}c_k \]

References


