Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

STATE PROBABILITIES FOR AN M/M/1 QUEUING SYSTEM WITH TWO CAPACITY LEVELS

Alexander Hübl

Logistikum
University of Applied Sciences Upper Austria
Wehrgrabengasse 1-3
4400 Steyr, AUSTRIA

Klaus Altendorfer

Department of Operations Management
University of Applied Sciences Upper Austria
Wehrgrabengasse 1-3
4400 Steyr, AUSTRIA

ABSTRACT

Flexible capacity of production system gets in times of short-time working and economic crisis an increasingly important status. Therefore, models to handle flexible capacity of production systems are necessary. In production planning queuing theory is a widely applied modeling approach. Since classical M/M/1 queuing models neglect flexible capacity this work implements two production rates in an M/M/1 queuing model. Whenever the queue length is more than k, the system runs at high speed otherwise low speed is used. The aim of this paper is the calculation of the state probabilities of the Markov chain. The state probabilities are the basis for developing an approximation for the production lead time which is dedicated to further research. Finally, a simulation study for the evaluation of state probabilities for flexible capacity with one and two switching points is conducted.

1 INTRODUCTION

On the one hand the due dates of the customers can be negotiated to create a smoother capacity demand (Corti et al. 2006; Hegedus and Hopp 2001; Hopp and Roof Sturgis 2000; Keskinocak and Tayur 2004). On the other hand the capacity can be adjusted to the fluctuations of the customer demand (Altendorfer et al. 2014; Bradley and Glynn 2002; Defregger and Kuhn 2006; Kok 2000; Li et al. 2009; Mincsovics and Dellaert 2009; van Mieghem and Rudi 2002). The methods discussed in this paper are based on capacity adjustment literature.

2 MODEL DESCRIPTION

A simple M/M/1 queue is assumed where two different production rates are possible. For solving this problem queuing theory is applied for a M/M/1 queue with a WIP and FGI (see Figure 1). In literature only a constant production rate μ is used (Altendorfer and Jodlbauer 2011; Altendorfer and Minner 2011; Buzacott and Shanthikumar 1993; Medhi 1991). Therefore, literature is extended by applying two different production rates μ_L and μ_H. Whenever more than k orders are in the WIP of the production system, it runs at speed μ_H otherwise μ_L is used. (Altendorfer and Jodlbauer 2011) demonstrated how a customer required lead time distribution can be implemented in a M/M/1 queuing model with constant μ for service level, FGI, FGI lead time and tardiness.

Figure 1 shows the basic model where order i is stated by a customer with rate λ. If the machine is idle, the order is released into production, otherwise the order is waiting in the buffer for processing. Only one order can be processed at a time. The production unit consists of a buffer for waiting orders and a

978-1-4673-9743-8/15/$31.00 ©2015 IEEE

processing step that requires a random processing time. Whenever more than k orders are in the WIP of the production system, it runs at speed μ_H otherwise is μ_L used.

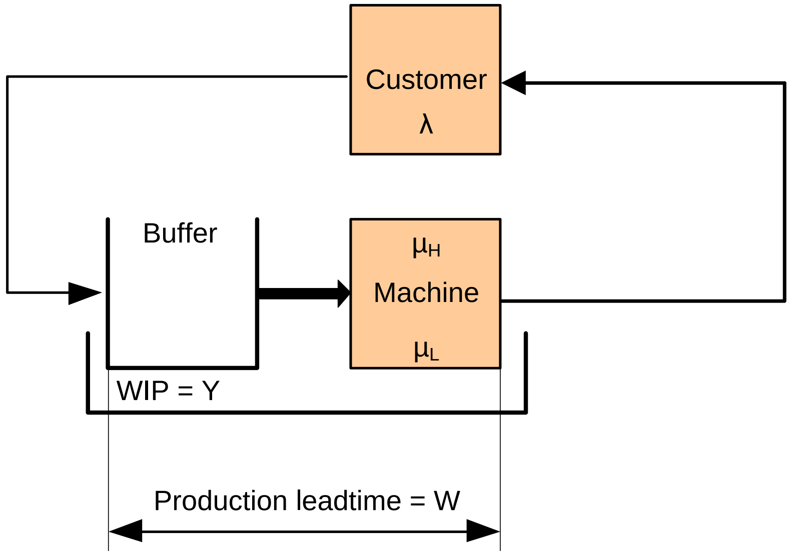

Figure 1: M/M/1 Production system with switching capacity.

2.1 MODEL ASSUMPTIONS

The following assumptions are taken into consideration to create the model:

- M/M/1 model is assumed with two levels of $\mu \rightarrow \{\mu_L, \mu_H\}$
- Processing times are exponentially distributed and only one order can be processed at time. Whenever more than k orders are in the WIP μ_H otherwise μ_L is used as production rate and no transient behavior at the switching point is assumed
- Customer orders for single items are stated piecewise from many different customers which supports the exponentially distributed customer order interarrival times with rate λ for single items i
- Customer orders are not changed once they are stated and have be accepted and released into the production system
- Nothing is produced without a customer order (MTO system)
- FIFO is applied as dispatching rule after the customer orders are released
- Machine capacity cannot be stored

2.2 STATE PROBABILITES FOR SWITCHING PROCESSING RATES

As described above, a simplified M/M/1 production system is applied in this paper and whenever the queue length is more than k, the system runs at speed μ_L and at speed μ_H otherwise. For $k = 2$ the following derivation of the state probabilities P_n of the Markov chain holds:

$$
\begin{aligned}
&P_0\lambda = P_1\mu_L \Rightarrow P_1 = P_0\frac{\lambda}{\mu_L} \text{ and}\\
&P_1\lambda + P_1\mu_L = P_2\mu_L + P_0\lambda \Rightarrow P_1\lambda = P_2\mu_L \Rightarrow P_2 = P_1\frac{\lambda}{\mu_L} = P_0\left(\frac{\lambda}{\mu_L}\right)^2 \text{ and}\\
&P_2\lambda + P_2\mu_L = P_3\mu_H + P_1\lambda \Rightarrow P_2\lambda = P_3\mu_H \Rightarrow P_3 = P_2\frac{\lambda}{\mu_H} = P_0\left(\frac{\lambda}{\mu_L}\right)^2\frac{\lambda}{\mu_H}\\
&\Rightarrow P_4 = P_3\frac{\lambda}{\mu_H} = P_0\left(\frac{\lambda}{\mu_L}\right)^2\left(\frac{\lambda}{\mu_H}\right)^2
\end{aligned}
\tag{1}
$$

Which can be generalized to:

$$
P_n = P_0\left(\frac{\lambda}{\mu_L}\right)^{\min(k,n)}\left(\frac{\lambda}{\mu_H}\right)^{\max(0,n-k)}
\tag{2}
$$

And therefore P_0 can be calculated as:

$$
\begin{aligned}
&\sum_{n=0}^{\infty} P_n = 1 \Leftrightarrow \sum_{n=0}^{\infty} P_0\left(\frac{\lambda}{\mu_L}\right)^{\min(k,n)}\left(\frac{\lambda}{\mu_H}\right)^{\max(0,n-k)} = 1 \Leftrightarrow P_0\sum_{n=0}^{k}\left(\frac{\lambda}{\mu_L}\right)^{n} + P_0\left(\frac{\lambda}{\mu_L}\right)^{k}\sum_{n=k+1}^{\infty}\left(\frac{\lambda}{\mu_H}\right)^{n-k} = 1\\
&\Leftrightarrow P_0\frac{\lambda\left(\frac{\lambda}{\mu_L}\right)^k - \mu_L}{\lambda-\mu_L} + P_0\left(\frac{\lambda}{\mu_L}\right)^k\frac{\lambda}{\mu_H-\lambda} = 1 \Leftrightarrow P_0\frac{\mu_L}{\mu_L-\lambda} + P_0\left(\frac{\lambda}{\mu_L}\right)^k\left(\frac{\lambda}{\mu_H-\lambda} - \frac{\lambda}{\mu_L-\lambda}\right) = 1\\
&\Leftrightarrow P_0\left(\frac{\mu_L^{\,k+1}(\mu_H-\lambda)}{\mu_L^{\,k}(\mu_L-\lambda)(\mu_H-\lambda)} - \frac{\lambda^{k+1}(\mu_H-\mu_L)}{\mu_L^{\,k}(\mu_L-\lambda)(\mu_H-\lambda)}\right) = 1\\
&\Leftrightarrow P_0 = \frac{\mu_L^{\,k}(\mu_L-\lambda)(\mu_H-\lambda)}{\mu_L^{\,k+1}(\mu_H-\lambda) - \lambda^{k+1}(\mu_H-\mu_L)}
\end{aligned}
\tag{3}
$$

Further simplifying the state probabilities leads to:

for $n \le k$:

$$
\begin{aligned}
&P_n(n\le k) = P_0\left(\frac{\lambda}{\mu_L}\right)^n = \frac{\mu_L^{\,k}(\mu_L-\lambda)(\mu_H-\lambda)}{\mu_L^{\,k+1}(\mu_H-\lambda) - \lambda^{k+1}(\mu_H-\mu_L)}\left(\frac{\lambda}{\mu_L}\right)^n\\
&= \frac{\mu_L^{\,k-n}\lambda^n(\mu_L-\lambda)(\mu_H-\lambda)}{\mu_L^{\,k+1}(\mu_H-\lambda) - \lambda^{k+1}(\mu_H-\mu_L)}
\end{aligned}
\tag{4}
$$

for $n > k$:

$$P_n(n>k) = P_0\left(\frac{\lambda}{\mu_L}\right)^k\left(\frac{\lambda}{\mu_H}\right)^{n-k} = \frac{\mu_L^{\,k}(\mu_L-\lambda)(\mu_H-\lambda)}{\mu_L^{\,k+1}(\mu_H-\lambda)-\lambda^{k+1}(\mu_H-\mu_L)}\left(\frac{\lambda}{\mu_L}\right)^k\left(\frac{\lambda}{\mu_H}\right)^{n-k}$$
$$= \frac{\lambda^n(\mu_L-\lambda)(\mu_H-\lambda)}{\mu_H^{\,n-k}\left(\mu_L^{\,k+1}(\mu_H-\lambda)-\lambda^{k+1}(\mu_H-\mu_L)\right)} = \left(\frac{\lambda}{\mu_H}\right)^n\frac{\mu_H^{\,k}(\mu_L-\lambda)(\mu_H-\lambda)}{\mu_L^{\,k+1}(\mu_H-\lambda)-\lambda^{k+1}(\mu_H-\mu_L)} \tag{5}$$

Which can be applied to identify P_H the percentage of time the production system is running at flexible overcapacity μ_H:

$$P_H = 1-\sum_{n=0}^{k}P_n = 1-\sum_{n=0}^{k}P_0\left(\frac{\lambda}{\mu_L}\right)^n = 1-P_0\frac{\left(\mu_L^{\,k+1}-\lambda^{k+1}\right)}{\mu_L^{\,k}(\mu_L-\lambda)}$$
$$= 1-\frac{\mu_L^{\,k}(\mu_L-\lambda)(\mu_H-\lambda)}{\mu_L^{\,k+1}(\mu_H-\lambda)-\lambda^{k+1}(\mu_H-\mu_L)}\frac{\left(\mu_L^{\,k+1}-\lambda^{k+1}\right)}{\mu_L^{\,k}(\mu_L-\lambda)} = 1-\frac{(\mu_H-\lambda)\left(\mu_L^{\,k+1}-\lambda^{k+1}\right)}{\mu_L^{\,k+1}(\mu_H-\lambda)-\lambda^{k+1}(\mu_H-\mu_L)} \tag{6}$$
$$= \frac{\lambda^{k+1}(\mu_L-\lambda)}{\mu_L^{\,k+1}(\mu_H-\lambda)-\lambda^{k+1}(\mu_H-\mu_L)}$$

3 SIMULATION STUDY

In the simulation study the state probabilities are analyzed. Therefore, 10 replications for each parameter combination are conducted. The simulation time is set to 100,000 periods. For all results a confidence interval of 99% holds true. The following parameters are set:

- Arrival rate $\lambda \in \{0.8; 0.9\}$
- Switching point $k = 12$
- $\mu_H = \mu + d; \mu_L = \mu - d; d \in \{0; 0.05; 0.1; 0.15\}$

3.1 BASIC SCENARIO

Figure 2 and Figure 3 illustrate the evolution of state probabilities under flexible capacity scenarios. When no flexible capacity is available ($d = 0$) then state probabilities decrease exponentially since only one capacity level is available. For the high utilization scenario (Figure 2) it is shown that increase flexibility in capacity leads to an change in the state probabilities. In this simulation study the switching point is set to $k = 12$. In the states assuming low WIP the probability deceases when more flexible capacity is available compared to the constant capacity scenario. After reaching the switching point the probability decreases and the production system is able to decrease the WIP in the system due to providing more capacity.

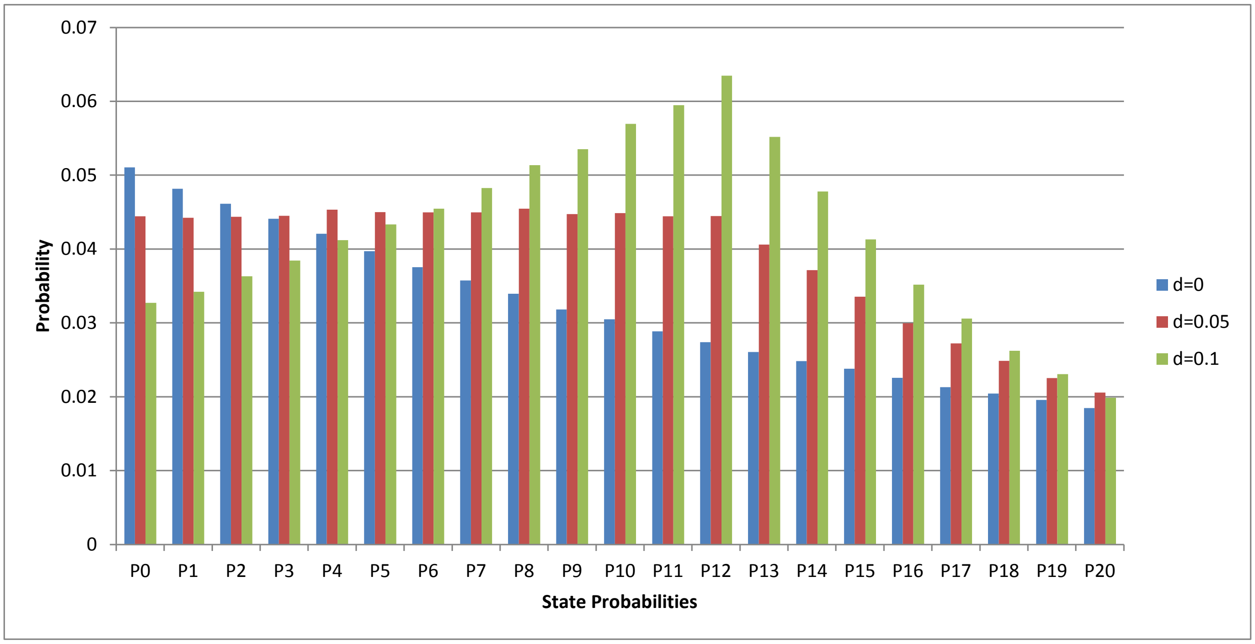

Figure 2: Evolution of state probabilities under flexible capacity for high utilization scenario $\lambda = 0.9$.

In the low capacity scenario flexible lead on all two ($d \in \{0.05; 0.1\}$) to a decrease of state probabilities. For the states assuming low WIP again the probability decrease when more flexible capacity is provided and the probability increases when reaching the switching point compared to the constant capacity scenario.

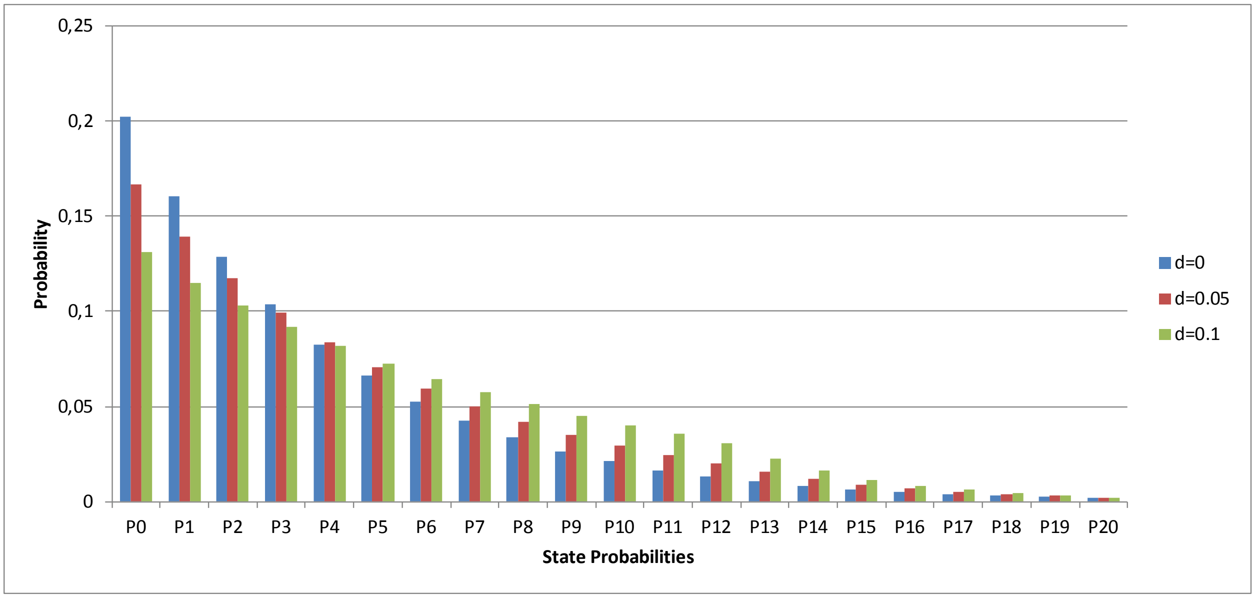

Figure 3: Evolution of state probabilities under flexible capacity for low utilization scenario $\lambda = 0.8$.

3.2 TWO SWITCHING POINTS

In order to get an more realistic situation two switching points are also implemented in the simulation study. If more than k_H orders are in the system the production rate from μ_L to μ_H is increased. The production rate is decreased from μ_H to μ_L when less or equal than k_L orders are in the system. For testing these cases the following relationships hold true: $k_L = k - i$; $k_H = k + i$; $i \in \{0; 1; 2; 3; 4\}$. If i is two then the production system produces at high capacity when the WIP is greater than 13 and the system switches back to low capacity when the WIP reaches 11 units. For the "two switching points" situation also both utilization scenarios are tested for evaluating the state probabilities and are presented in Figure 4 and Figure 5.

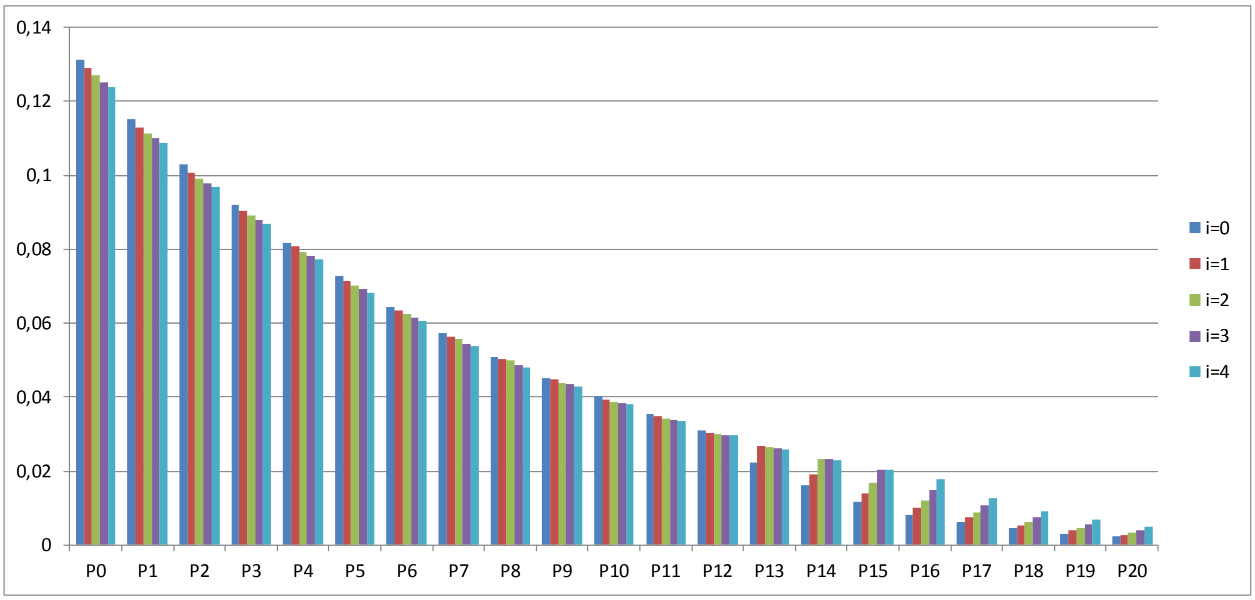

Figure 4. State probabilities for the low utilization scenario for different i at $d = 0.1$.

For the low and high utilization scenario one can observe for state probabilities up to 12 that for setting up two switching points ($i > 0$) the state probability decreases the greater i gets. For state in which the observed scenario reaches k_H leads to the highest probability. Also for the "two switching points" situation, the above observed influence of utilization can be determined on the evolution of state probabilities.

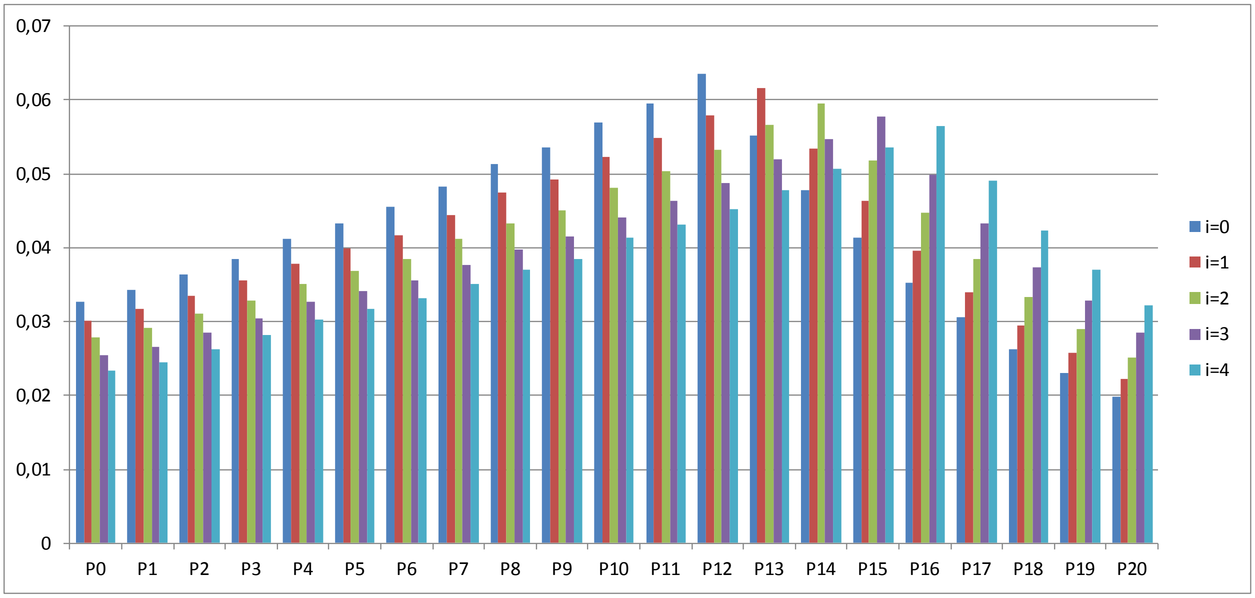

Figure 5. State probabilities for the high utilization scenario for different i at $d = 0.1$.

4 CONCLUSION

In this paper, the state probabilities of a M/M/1 queuing system with two production rates and one switching point when reaching a certain WIP are calculated analytically. In a simulation study the evaluation of the state probabilities is observed by providing flexible capacity of the production system. Moreover, a "two switching point" production system is introduced where the system uses high capacity when the WIP reaches k_H. Unlike previously, the high capacity remains upright after the WIP undercuts k_H. Only when the WIP reaches the lower switching point k_L the low capacity of the production system is applied. This switching behavior of capacity leads to lead modified state probabilities of the production system. The simulation study delivers more insights into the evolution of state probabilities of an queuing system.

The analytic development of the state probabilities is the basis for developing analytical relationships for production and FGI lead time and performance indicators such as service level s and tardiness C of this production system in further research. Finally, the optimal switching point k can be calculated based on an optimization problem in further research.

ACKNOWLEDGMENTS

The work described in this paper was done within the COMET Project Heuristic Optimization in Production and Logistics (HOPL), #843532 funded by the Austrian Research Promotion Agency (FFG).

5 REFERENCES

Altendorfer, K., A. Hübl, and H. Jodlbauer. 2014. "Periodical capacity setting methods for make-to-order multi-machine production systems." *International Journal of Production Research* 52:4768–4784.

Altendorfer, K., and H. Jodlbauer. 2011. "An analytical Model For Service Level And Tardiness In A Single Machine MTO Production System." *International Journal of Production Research* 49:1827–1850.

Altendorfer, K., and S. Minner. 2011. "Simultaneous optimization Of Capacity And Planned Lead Time In A Two-Stage Production System With Different Customer Due Dates." *European Journal of Operational Research* 213:134–146.

Balakrishnan, N., J. W. Patterson, and V. Sridharan. 1996. "Rationing Capacity Between Two Product Classes." *Decision Sciences* 27:185–214.

Balakrishnan, N., J. W. Patterson, and V. Sridharan. 1999. "Robustness of capacity Rationing Policies." *European Journal of Operational Research* 115:328–338.

Bradley, J. R., and P. W. Glynn. 2002. "Managing Capacity and Inventory Jointly in Manufacturing Systems." *Management Science* 48:273–288.

Buyukkaramikli, N. C., J. Bertrand, and H. P. G. van Ooijen. 2013. "Periodic capacity Management Under A Lead-Time Performance Constraint." *OR Spectrum* 35:221–249.

Buzacott, J. A., and J. G. Shanthikumar. 1993, *Stochastic Models Of Manufacturing Systems*, Prentice-Hall: Englewood Cliffs NJ.

Corti, D., A. Pozzetti, and M. Zorzini. 2006. "A Capacity-Driven Approach To Establish Reliable Due Dates In A MTO Environment." *International Journal of Production Economics* 104:536–554.

Defregger, F., and H. Kuhn. 2006. "Revenue management for a make-to-order company with limited inventory capacity." *OR Spectrum* 29:137–156.

Hegedus, M. G., and W. J. Hopp. 2001. "Due date Setting With Supply Constraints In Systems Using MRP." *Computers & Industrial Engineering* 39:293–305.

Hopp, W. J., and M. L. Roof Sturgis. 2000. "Quoting Manufacturing Due Dates Subject To A Service Level Constraint." *IIE Transactions* 32:771–784.

Keskinocak, P., and S. Tayur. 2004. "Due-Date Management Policies." In *Handbook of quantitative Supply Chain Analysis*, edited by D. Simchi-Levi, et al., 485–553, New York NY: Springer Science + Business Media.

Kok, T. G. de. 2000. "Capacity allocation And Outsourcing In A Process Industry." *International Journal of Production Economics* 68:229–239.

Li, H., L. Hendry, and R. Teunter. 2009. "A Strategic Capacity Allocation Model For A Complex Supply Chain: Formulation And Solution Approach Comparison." *International Journal of Production Economics* 121:505–518.

Medhi, J. 1991, *Stochastic Models In Queueing Theory*, 1st ed., Academic Press: Boston u.a.

Mincsovics, G. Z., and N. P. Dellaert. 2009. "Workload-dependent Capacity Control In Production-To-Order Systems." *IIE Transactions* 41:853–865.

van Mieghem, J. A., and N. Rudi. 2002. "Newsvendor Networks: Inventory Management and Capacity Investment with Discretionary Activities." *Manufacturing & Service Operations Management* 4:313–335.

Yang, W., and R. Y. K. Fung. 2014. "An Available-To-Promise Decision Support System For A Multi-Site Make-To-Order Production System." *International Journal of Production Research* 52:4253–4266.

AUTHOR BIOGRAPHIES

Alexander Hübl is leading the research group logistic optimization at the University of Applied Sciences Upper Austria, Austria. His research interests include discrete-event simulation, agent-based simulation, queuing theory and their applications in logistics and operations management. His email address is alexander.huebl@fh-steyr.at.

Klaus Altendorfer is a Professor of the Department of Operations Management at the University of Applied Sciences Upper Austria, Austria. He holds a Ph.D. in logistics and operations management from University of Vienna, Austria. His e-mail address is klaus.altendorfer@fh-steyr.at.