Evaluation of Complex Manufacturing Systems in the Context of Aggregated Operating Curves

Ulf Bergmann, Matthias Heinicke, Gerd Wagenhaus, Sascha Schmidt

Evaluation of Complex Manufacturing Systems in the Context of Aggregated Operating Curves

Ulf Bergmann[1], Matthias Heinicke[2], Gerd Wagenhaus[1], and Sascha Schmidt[1]

[1] Otto-von-Guericke University, 39106 Magdeburg, Germany
ulf.bergmann@ovgu.de
[2] Targus Management Consulting AG, Dechenstraße 7, 40878 Ratingen, Germany

Abstract. For the right configuration of production systems it is crucial to know the optimum operating point by way of a functional correlation between work in process and output rate or throughput time respectively.
As a matter of fact, executives and managers are quite familiar with this correlation which, however, rarely leads to its formal-mathematical application. Among others, this is due to the specific prerequisites arising from the application of general laws from queuing theory and operating curve theory.
Therefore, this paper will discuss our approach for a practical determination and description of suitable reference value systems using aggregated operating curves. The main focus will be directed towards the system's behaviour of complex production systems in versatile batch production, as it is frequently encountered in mechanical and plant engineering.

Keywords: Aggregated Operating Curve, Simulation, Production System Control, Modeling Procedure.

1 Introduction

In order to adequately evaluate the performance of operating systems it is necessary to have significant parameters. Especially useful in this context are models and parameters of both the queuing theory and the curve theory. In the course of evaluating and configurating production structures, an increasing need for aggregated system description is noticeable, which goes beyond familiar bottleneck analyses focusing on one operating system.

Based on the general principles of operating curve theory, this paper will present a practical approach for the aggregated determination and application of a funnel model's parameters. Finally, the verification of the approach will be shown by reviewing a complex production system.

2 Theoretical Aspects

The evaluation of the system's behaviour is usually based on input-output modeling (I/O). With these kind of model significant parameters like work in process (WIP), output rate and throughput time can be ultimately derived from the formal I/O presentation of production systems. The interdependencies between these three parameters are known as Little's Law and, in its modified version, as Funnel Model (see **Fig. 1**).

2.1 Little's Law

The methodological groundwork of Little's Law is the queuing theory. A so-called queue is defined by a number of units (jobs) waiting for their processing at a single workstation. The arrival rate indicates how many units (jobs) are arriving the system on average in a certain period of time. The service rate indicates how many units (jobs) can be served by the system on average in a certain period of time. The realised service rate, then, yields the system's output in terms of quantity (number of jobs) [1].

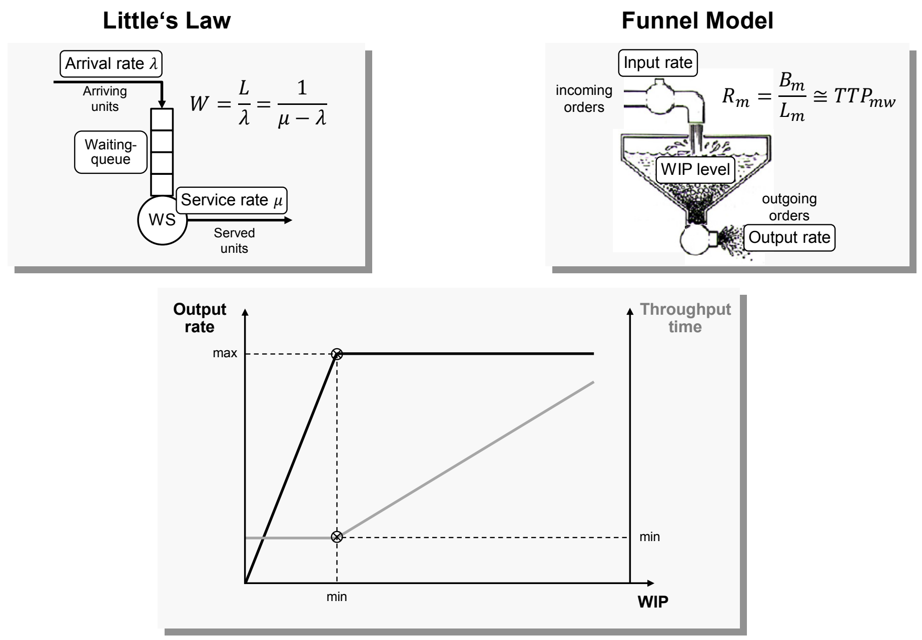

Fig. 1. Little's Law vs. Funnel Model.

Depending on the amount of already waiting units (jobs), Little's Law shows how long on average a newly arrived unit (job) will have to stay within the operating system until being processed (waiting time W). As long as the three parameters are measured in consistent units, Little's Law can be universally applied, with these findings coming chiefly from the observation of production lines and single workstations. If the WIP is then taken as an independent variable (WIP = the number of units L in the system), the interdependencies of Little's Law can be presented as an operating curve (sometimes called production-flow graph). [1]

There will be an ideal operating curve if no system variability is apparent. Once a minimally required WIP has been established, maximum system output rate will be provided accordingly. Underneath this critical WIP level, the system output rate will drop proportionally to the average amount of jobs [1, 2]. According to Little's Law, the curve of the throughput time is contrary (see **Fig. 1**).

2.2 Funnel Model

Extending Little's Law, [2] define what is called the Funnel Model, thereby referring to flow processes in chemical engineering. Here, production units (in general individual workstations) are regarded as funnels out of which the processed orders will flow out (cf. Figure 1 right). Depending on the maximum capacity available, the funnel opening, therefore, illustrates the performance of the observed production unit (incoming orders). The queuing orders together with the newly arriving orders within the funnel comprise the WIP of orders yet to be processed (outgoing orders).

Provided that there is no new order entering the system, the funnel model states how long it will take on average to completely finish processing the WIP with consistent mean performance [2]. This will result in so-called operating points which describe respective operating modes of a production system for specific reference periods (e.g. one week). When, in addition to that, output rate and throughput time are shown as a function of the WIP level, the system behaviour might be modelled accordingly ([2] refers to them as production or operating curves, cf. **Fig. 1**).

Such operating curves describing a system-specific behaviour that occurs individually for the production type depending on the formation conditions. It might even vary according to the determination coefficient of influencing factors. Therefore, a quantitative statement concerning the WIP makes only sense when having just a few influencing parameters, and will be used for the simulative design of production systems as well. It is crucial for such designs that the variance of the input parameters is deterministic in a way that these parameters are stipulated as design restrictions for the future management model. For existing systems that are subject to a certain change of input parameters it is highly useful to know – or determine for that matter – the progression of its operating curve.

2.3 Short Synopsis

A crucial characteristic of the Funnel Model is its time evaluation of WIP and output rate. Provided that the evaluation intervals are of sufficient length, the change of the WIP will, therefore, only be influenced by the output rate of the evaluated system when using the Funnel Model. The completion behavior of orders (resp. the sequencing rule) as well as the variance of its work contents actually have no direct impact for the Funnel Model.

The situation is different for Little's Law. As regards the modelling and calculation as a queuing formula, WIP and output rate are given as count variables. Especially when faced with a multitude of different jobs with many different work contents respectively, the measuring by means of count variables proves to be unsuitable [3]. On top of that, when it comes to applying queuing models, form and parameters of distribution for arrival and service times as well as their necessary stochastic independence have to be described in correlation with each other [2]. This is crucially disadvantageous when modelling complex production processes since both the arrival and service times can rarely be described in an exact manner with classical distribution functions.

However, it is precisely the adjustment of the completion rate to the WIP which is the core task of production planning and control, in other words the essential control to influence the system performance, be it as a form of capacity adjustment or by means of load balancing. And this is exactly what queuing models will not enable with their modelling assumptions [2].

Consequently, with regard to practical questions concerning the configuration of production systems, queuing models do not comply with the requirements needed, which is why we will use the factual connections of the Funnel Model in the course of this paper.

3 Approach for the Aggregation of Operating Curves

3.1 Prerequisites

Our focus is set on the aggregated evaluation of complex production systems, especially versatile batch production scenarios in mechanical and plant engineering. Due to the production programmes and technological sequences typical for these scenarios, such systems are characterised by a complex organisational structure that has to deal with shifting bottleneck situations. Therefore, we will show in which way the relevant production parameters can be determined as an aggregation of specific individual parameters in order to describe the operating systems involved in the manufacturing process.

The aggregated evaluation of a production system with the help of the Funnel Model requires, above all, a determination of measuring points for evaluating the jobs that enter and leave the system. Basically, this determination can be performed starting with the smallest operating system which can not be reasonably broken down any further, and then continued for each higher aggregation level. Provided that the WIP can be determined with acceptable precision, it is irrelevant at this point how the individual workstations are linked with each other.

These measuring points (defined as input and output) are used as referential points for the aggregated evaluation of the work content per job in planned hours and will literally cover the scope of the evaluation, accordingly. Based on this, the present system state will then be determined by using the Funnel Model, and will be termed operating point. Here, ideal operating curves represent the suitable “reference coordinate system” for a present system state [4, 5].

Such state values are only helpful as a benchmark or reference values for decisions in production controlling if the statistical correlations of relevant system parameters are taken into account [6]. Consequently, a so-called black-box-approach is not feasible. It is rather necessary to aggregate the specific individual parameters from the workstations involved in the system processing according to their material flow-based interconnectedness. Using an intensity index, the material flow structure will be simplified to become a direct line of weighted predecessor-successor-relationships, when the aggregated parameters can ultimately be determined via the identified bottleneck [5, 7].

The description of the inner system structure is necessary and detailed, requiring a lot of effort and background knowledge. Since a long-term stable bottleneck is also required, such a white-box-approach is not feasible, either. The same goes for shifting bottlenecks, as is the case in our evaluation scenario. Therefore, this is our modified approach: We call it grey-box-approach for the aggregated evaluation of reference values for WIP (a), output rate (b) and throughput time (c).

3.2 Grey-Box-Approach

(a) The ideal minimum WIP is the minimally required WIP necessary for the full utilisation of all operational means within the evaluated system. When performing an aggregated I/O evaluation of a production system by means of the Funnel Model, the amount of available work (work load) within a production system can be regarded as its temporally evaluated total WIP level. Because of this system-wide perspective, it is irrelevant to determine which working step has been made for individual products according to their specific technological sequence on the production system's individual workstations. In fact, the cumulative process time of the individual jobs rather indicates their respective work load that they load the production system with, both in total and, consequently, each workstation according to their technological sequence. Therefore, the calculation of the so-called medium work content (WC_m) at first occurs as weighted mean by weighting the work content of each workstation i against the related quantity of given jobs m_p:

$$WC_{m,i} = \frac{\sum m_p \cdot WC_{p,i}}{\sum m_p} \tag{1}$$

It is necessary to include this aggregated work content because the minimum WIP is considered here for the entire production system and not for a single workstation. As a next step, the respective variation coefficients (WC_v) can be derived for each workstation:

$$WC_{v,i} = \frac{\sigma(WC_{p,i})}{\overline{WC_{p,i}}} \tag{2}$$

Then, by taking the system's composition (type i and number n of workstations) into account, the resource-based values can be accumulated to provide the minimum critical WIP_{min}:

$$WIP_{min} = \sum_i \left(n_i \cdot WC_{m,i} \cdot \left(1 + WC_{v,i}\right)^2\right) \tag{3}$$

Subsequently, this theoretically derived value will work as a system-wide parameter of the WIP, for which the production system will reach its maximum possible output rate; always in accordance with the planned production programme and performance losses inherent to the production system due to its complex interrelationship between various jobs.

(b) Viewed from the perspective of the planning manager, the scheduled production programme results in a medium workload that determines the maximum required output rate $OR_{req,\ prod}$ of the system in a certain period of time T:

$$OR_{req,\ prod} = \frac{\sum_p \sum_i WC_{p,i}}{T} \tag{4}$$

Since in the long run the average release rate is always less than the average capacity of the production programme [1, 2], the planning manager can determine the maximum achievable output rate $OR_{max,\ prod}$. Hence, for a fixed variety of jobs the following applies:

$$OR_{req,\ prod} \leq OR_{max,\ prod} \leq OR_{max,\ theo} \tag{5}$$

In this context, $OR_{max,\ theo}$ represents the theoretical optimum of the maximum possible output rate of a production system, resulting from the installed capacity. Basically, in practical circumstances, this maximum possible output rate $OR_{max,\ theo}$ is impossible to achieve, even without taking into account unforeseeable breakdowns.

(c) The fundamental basis, of course, for determining a reference value for the throughput time is again the production programme featuring the type and number of products m_p to be manufactured. Aggregated to a weighted mean, the lowest possible throughput time TTP_{min} can be determined according to their technological sequence, and will arise from the total of product-specific work content at each workstation i plus required transport times:

$$TTP_{min} = \frac{m_p \cdot TTP_{min,p}}{\sum m_p} \quad \text{with} \quad TTP_{min,p} = \sum WC_{p,i} \tag{6}$$

Therefore, this minimum value – which is mostly a theoretical value, especially in complex production systems – serves as an ideal reference value for the throughput time in the course of events.

4 Practical Verification and Findings

We verified the feasibility of the grey-box-approach by using a lab-sized production system. This multi-agent controlled system is composed of various Fischertechnik components simulating machining processes on work pieces. The use case was based on real-world production scenarios for which we gained the following aggregated reference values (see **Table 1**), using the given equations (1-6) of the grey-box-approach (note: the simulating processes were running with a reduced time scale in seconds).

Table 1. Data and aggregated reference values of the analysed production scenarios.

Production Scenario A						T_{av} = 250 d/a with 16 h/d	Production Scenario B							
A-10	A-20	A-30	A-40	A-50	A-60	Product	B-100	B-200	B-300	B-400	B-500	B-600	B-700	B-800
600	600	600	450	360	360	Order size	225	600	375	600	600	600	600	450
60	60	60	80	100	100	No. of orders	160	60	96	60	60	60	60	80
Production sequence and time per order [h]						Process (No. of machines)	Production sequence and time per order [h]							
			7,6	3,2	3,3	(1) Cutting				5,3		10,3		7,6
	40,3	50,4	45,4	30 / 31	75,3	(7) Turning			62,9	50,4	40,3	50,4	50,4	56,7
20,2				48,3	36,3	(3) Face-milling	22,8	20,2	25,3					
20,5				132,3	72,3	(7) Drilling	120,5	20,5						
	10,3	15,3				(1) Broaching					10,3		15,3	
	40,4	40,4		12,4		(2) Finish-milling					40,4	20,4	40,4	
10,1	10,1	10,1		12,1	6,1	(1) Deburring	7,6	10,1			10,1	10,1	10,1	
	20,3	15,3	19,2	36,3	24,3	(3) Grinding				40,3	20,3	20,3	15,3	19,2
	35,3	50,5				(2) Hob peeling					35,3		50,5	
dc = 1.29 **dl = 0.85**			**WIP_{min} = 6365 h** **$OR_{req,prod}$ = 326 h/d**				**WIP_{min} = 3973 h** **$OR_{req,prod}$ = 300 h/d**				**dc = 0.92** **dl = 1**			

The reference values have been verified using simulation data. In order to do so, the system behaviour was first determined as a "perfect world simulation". After that, the empirical system behaviour was simulated in a "real world scenario", which took place, firstly, as a result of an immediate job release ("uncontrolled system") and, secondly, by means of a WIP-based job release ("controlled system"). **Fig. 2** illustrates the results as output rate operating curve.

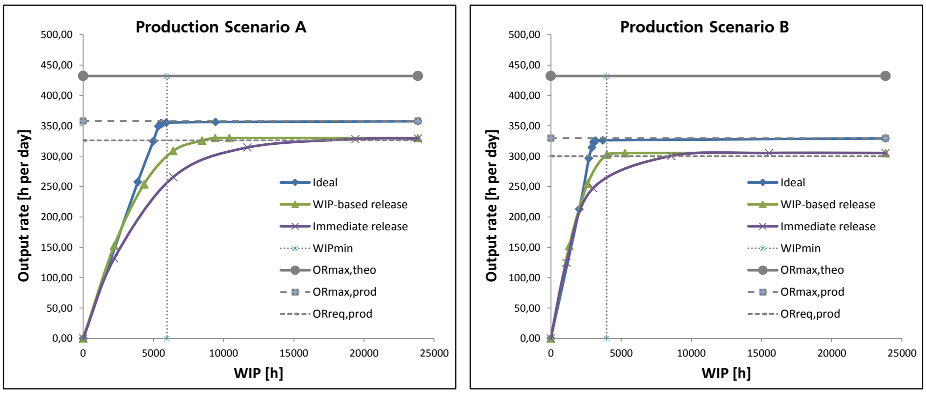

Fig. 2. Aggregated Operating Curves of the simulated production scenarios.

Here it becomes clear again how important a detailed evaluation of the output rate as crucial parameter is (cf. Eq. 5). Based on a fixed mix of type and number of product variants (different jobs to be done), $OR_{max,\ prod}$ is determined as the maximum achievable output rate of the evaluated production system, resulting from the bottleneck in the production system. Correlating with the product variants to be produced, shifting bottlenecks often complicate this determination (cf. Ch. 3.1).

Apart from bottleneck-based analyses, we are using in our grey-box approach the maximum required output rate $OR_{req,\ prod}$ as a reference value; that is the output rate required for completing the given production programme. Regardless of other impact factors, $OR_{req,\ prod}$ serves as the saturation limit. According to economical considerations, it is by definition unprofitable to exceed $OR_{req,\ prod}$ significantly, except for additional work beyond the regular production programme.

A crucial parameter in this respect is the critical WIP determined by means of aggregation. Our investigations have shown that the interdependency between the output rate and WIP are specific to the system, which will individually appear according to the formation conditions of the production structure. It might even vary in accordance with the determination coefficient of those formation conditions. In order to assess the system behaviour imminent in both production scenarios, we can use two significant indicators: the degree of complexity dc and the degree of linearity dl (for further calculation details see [8]). Then the following can be stated.

- If dc is high and dl is low (Production Scenario A): trending towards a complex system structure = critical WIP works as lower threshold; i.e. a required output max will only be achieved by a multiple of the critical WIP (maximum 100%)
- If dc is low and dl is high (Production Scenario B): trending towards a simple system structure = critical WIP acts as upper threshold; i.e. a required output max will materialize within a narrow tolerance zone around the critical WIP level (± 10%)

On top of that, our investigation illustrates very well the sensitivity of the critical WIP as a control variable. As for the controlled system, even minor modifications in the critical WIP will lead to considerable changes in the system's performance. Contrarily, for the uncontrolled system, there is no immediate cancellation of jobs, which will run-up the WIP and, therefore, will affect the throughput time in a negative way.

5 Conclusions

As a matter of fact, our investigation has confirmed the basic principles of curve theory, extended by an aggregated evaluation and management of necessary parameters. It goes to show that the combination of $OR_{req,\ prod}$ as a saturation limit and critical WIP is a suitable indicator for setting a feasible operating point. This is much in line with experience-based routines of operation for the present application and also validates the methodological framework of the grey-box-approach introduced here.

Compared to the immediate job release, the advantages of the WIP-based job release are clearly visible. The latter benefits from the design parameters of the approach presented here.

Especially when viewed against the background of the WIP-oriented approach that is frequently used in industrial practice as production control, the introduced critical WIP proves its controlling effect as reference value. Furthermore, by using the present data base of a real production programme, it is relatively easy to determine the highest necessary output rate (saturation limit). The correlation of WIP and output rate limit quickly results in an operating point that seems ideally suited to get closer to the theoretical optimum in a WIP-controlled production system.

References

1. Hopp, W. J., Spearman, M. L.: Factory physics: foundations of manufacturing management. 3rd edn. Irwin McGraw-Hill, New York (2008).
2. Nyhuis, P., Wiendahl, H.P.: Fundamentals of Production Logistics Theory, Tools and Applications. Springer, Berlin (2014)
3. Pound, E. S., Bell, J. H., Spearman, M. L.: Factory physics for managers: How leaders improve performance in a post-lean six sigma world. McGraw-Hill Education, New York (2014).
4. Bergmann, U., Heinicke, M.: Resilience of productions systems by adapting temporal or spatial organization, Procedia CIRP 57, 183–188 (2016)
5. Nyhuis, P.: Beiträge zu einer Theorie der Logistik. Springer-Verlag, Berlin Heidelberg (2008)
6. Kemmner, A.: Ein heimlicher Leitfaden für ständig kritisierte Fertigungssteuerer. Abels & Kemmner Homepage, https://www.ak-online.de/de/2008/06/ps2008-14/, last accessed 2021/04/09.
7. Weigert, G.: Operating Curves of Manufacturing Systems: A Theoretical Discourse. In: International FAIM Conference: Advances in sustainable and competitive manufacturing systems: 23rd International Conference on Flexible Automation and Intelligent Manufacturing. pp. 887-897. Springer, Cham (2013).
8. Bergmann, U., Heinicke, M.: Approach for the Evaluation of Production Structures. In: Advances in Production Management Systems: The Path to Intelligent, Collaborative and Sustainable Manufacturing, Part II, 176–183. Springer, Cham (2017)