Advances in Industrial and Manufacturing Engineering 8 (2024) 100132

Contents lists available at ScienceDirect

Advances in Industrial and Manufacturing Engineering

journal homepage: www.sciencedirect.com/journal/advances-in-industrial-and-manufacturing-engineering

Lateness in production systems - In a nutshell: How to determine the causes of lateness at work systems?

Alexander Mütze *, Tobias Hiller, Peter Nyhuis

Leibniz University Hannover, Institute of Production Systems and Logistics, An der Universität 2, 30823, Garbsen, Germany

ARTICLE INFO

Keywords:
Lateness
Analytical model
Diagnostics
Logistical performance
Schedule reliability
Modelling

ABSTRACT

Adherence to customer due dates is the yardstick for the performance of manufacturing companies. In the era of same-day delivery, consumers expect reliable delivery of ordered goods and short delivery times. Also, in the field of business-to-business supply, it is evident that adherence to delivery dates is a fundamental logistical objective for companies. Contract manufacturers, in particular, are confronted with significant challenges: strong fluctuations in customer demand, shorter requested delivery times, and high competitive pressure require appropriate organisation, planning and control of production. However, companies often miss their schedule reliability targets and fail to identify the right causes for these failures. This raises the question of what factors influence the failure to meet schedule reliability targets, how to identify such factors, and what options are available to counteract them. This contribution addresses this issue and focuses on ways to analyse the emerging lateness at work systems in production areas as a deviation of the actual form the planned throughput time. We present existing approaches to analysing the lateness behaviour at work systems and extend the current theory of logistical modelling to determine the three drivers of the so-called relative lateness – planning influences, variance of work-in-process (WIP) and sequence deviations – at work systems systematically. Through this analysis, we enable the practical applicator to initiate target-oriented countermeasures to improve the schedule reliability of their work systems with acceptable analysis expenditure.

1. Introduction

Various studies, company surveys and performance measurement systems confirm the importance of high schedule reliability for manufacturing companies (Wiendahl and Tönshoff, 1988; Lödding and Kuyumcu, 2015; Schmidt et al., 2019). Along with throughput time as a performance indicator and the cost objectives of capacity utilisation and work-in-process (WIP), it is one of the four crucial production logistics objectives in the so-called matrix of production objectives (Nyhuis and Wiendahl, 2009). While the cost-oriented targets of WIP and capacity utilisation are only indirectly perceived by the customer and represent internal company targets, the objectives of logistical performance are directly noticeable by the customer in the form of delivery time and schedule reliability. This is especially true in the case of contract manufacturing. Therefore, companies must know their current logistical performance and derive targeted and appropriate countermeasures to reinstate the desired performance in case of deviations from the planned target level.

The first step in doing so is to identify the root causes of unsatisfactory logistical performance using appropriate analytical tools to monitor the processes taking place (Schmidt et al., 2019; Mütze et al., 2022). In the case of schedule reliability, a wide variety of influencing factors can be responsible for the failure to meet set targets. Examples of such factors include the processing sequence (dispatching) at the work system (Conway et al., 1967; Lödding, 2012), a backlog of production (Lödding, 2012; Yu, 2001) and insufficient scheduling quality in production planning (Lucht et al., 2021).

We present, systematise and refine analysis methods that partly have not yet been published for an international audience. Our contribution is based on existing axiomatic and empirical research (Nyhuis and Wiendahl, 2006), but aims to present a descriptive, analytical approach (cf. (Lödding, 2013)) in the form of a manageable diagnostic model that is generally applicable and can be used for the analysis of hypothetical production areas (e.g. in simulation studies) as well as actual production areas. This contribution thus aims to enhance the possibilities of schedule reliability analyses in production by considering the lateness behaviour of work systems as the smallest capacitive unit of a production area and the possible reasons for a deviation of the actual

* Corresponding author.
E-mail address: office@ifa.uni-hannover.de (A. Mütze).

https://doi.org/10.1016/j.aime.2023.100132
Received 23 April 2023; Received in revised form 23 September 2023; Accepted 26 October 2023
Available online 7 November 2023

throughput time and the planned throughput time causing relative lateness at them.

The paper is structured as follows. First, the current knowledge on the decomposition of lateness based on influencing variables and the analysis of lateness at work systems in general are presented. In this context, the theory of logistic operating curves (LOC) developed by Nyhuis and Wiendahl, 2006, 2009 is examined in detail. The concept of lateness decomposition at work systems, an extension of the LOC theory developed by Yu (2001), will be the modelling basis of this paper. The analysis is followed by the derivation of the need for research, leading to a refinement of Yu's model. A case study then is presented to demonstrate the practical applicability of the developed model. A simplification of the proposed model for diagnostic use in companies follows the case study. The contribution closes with a conclusion and a critical discussion of the presented model, elaborating on its value and limitations for research and practice.

2. State of the art in modelling lateness behaviour

As described in the introduction, schedule reliability resulting from the lateness behaviour of production is one of the key objectives of companies and thus for their configuration of production planning and control (PPC) as well as for the entire order fulfilment process (Lödding, 2013; Schmidt and Schäfers, 2017). This is demonstrated by the fact that schedule reliability is an integral part of various performance measurement systems for evaluating the logistical performance of production systems (cf. (Huan et al., 2004; Gunasekaran and Kobu, 2007)). It is also frequently used as a performance criterion in simulation studies, for example, in the research field of order release, evaluating a production system's configuration or the performance of a specific order release method (cf. (Wisner, 1995; Thürer et al., 2011)).

However, instead of a deeper analysis of lateness behaviour and the drivers of output lateness at a capacitive unit (i.e. the entire production system, a production area or a work system), the vast majority of investigations use lateness solely for performance evaluation. Depending on the object under consideration, these performance evaluations can focus on actual material flow, planned material flow (Conway et al., 1967) or the interaction between the two (Plossl and Wight, 1973).

However, some scholars further analyse lateness behaviour in the form of output lateness, dividing it into different parts to draw more nuanced conclusions about the prevailing cause-effect relationships in the production system. This also makes it possible to describe the impact of individual influencing factors quantitatively by calculating and comparing specific components of lateness (Lödding and Kuyumcu, 2015; Lödding et al., 2014; Bertsch et al., 2014). With the intention of creating a uniform understanding, the different lateness measures will be discussed in the next section. This will be followed by a discussion of the existing research approaches, focusing on the lateness behaviour of work systems.

2.1. Lateness measures at work systems

The lateness of an order at the output of a work system (output lateness) is calculated as the time difference between the planned end of order processing (the planned output date) and the actual end of order processing (the actual output date) (Lödding et al., 2014; Baker, 1974):

$$L_{out} = EDOP_{act} - EDOP_{plan} \quad (1)$$

where L_{out} is the output lateness [SCD], $EDOP_{act}$ the actual end of order processing [SCD] and $EDOP_{plan}$ the planned end of order processing [SCD].

As Equation (1) shows, this article specifies dates, such as the end of order processing, and time differences, such as throughput time and lateness, in shop calendar days (SCD), while, for example, the work content of orders (which is used later) is specified in hours (hrs). Output

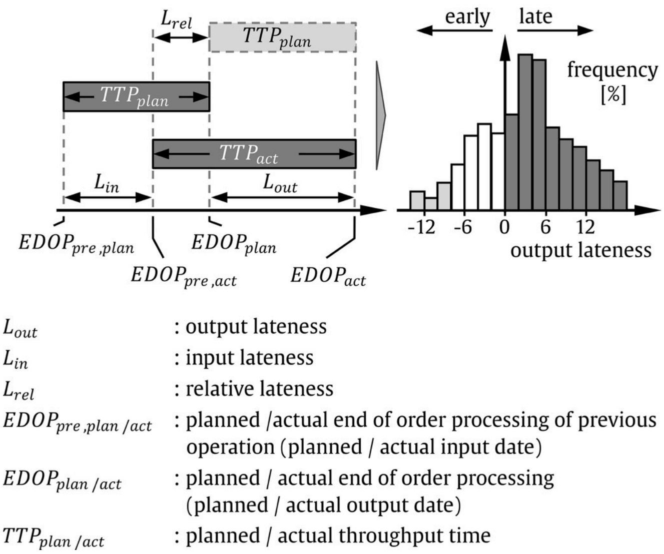

Fig. 1. Decomposition of lateness (Lödding, 2013; Bertsch et al., 2014; Dombrowski, 1988).

lateness calculated using Equation (1) is positive if the corresponding order is completed after its planned date (tardy) and negative if it is completed prior (early).

According to Dombrowski, two additional components of lateness can be defined to differentiate more precisely between the drivers of output lateness: input lateness and relative lateness (Lödding, 2013; Dombrowski, 1988).

The input lateness of an order at a work system indicates how late or early the order arrived at the work system. Equation (1) can be adapted to calculate input lateness using the corresponding actual and planned input dates. These, in turn, result from the actual and planned dates of the order's entry into production (order release) if the release is directly upstream in the routing of the order under consideration. Alternatively, these dates can be determined using the logic of the throughput element (cf. (Wiendahl and Tönshoff, 1988; Wiendahl, 1995)) from the output dates of the directly upstream processing step or work system.

Relative lateness indicates an order's relative acceleration or deceleration during processing and equals the difference between the planned and the actual throughput time at the work system under consideration. This measure is therefore also referred to as throughput time deviation (Lödding and Piontek, 2018). Negative relative lateness expresses that the actual throughput time of an order was shorter than planned (the order was decelerated) and vice versa.

$$L_{rel} = TTP_{act} - TTP_{plan} \quad (2)$$

where L_{rel} is the relative lateness [SCD], TTP_{act} the actual throughput time [SCD] and TTP_{plan} the planned throughput time [SCD].

The output lateness of an order at a work system can be derived from the summation of input lateness and relative lateness. The relationship between the components of lateness is illustrated in Fig. 1.

Due to the mathematical relationship between the different lateness components, lateness can also be understood as a coupling variable between work systems. Thus, output lateness at a predecessor work system perpetuates itself in the input of successor work systems according to order routings. This influences the logistical performance of the successors and creates challenges in meeting the due dates of orders (Lödding et al., 2014).

Based on this primary subdivision of output lateness into input and relative lateness, several analyses can be carried out, for example, using correlation analyses or scatter matrices, or comparing the lateness of an order with its work content to determine whether orders are accelerated based on their work content. Outcomes of these analyses can include identifying the most significant drivers of lateness and gathering information on the dominant dispatching rule at work systems (Lödding and Kuyumcu, 2015; Nyhuis and Wiendahl, 2009).

2.2. The decomposition of lateness at work systems

In addition to the subdivision of output lateness into input and relative lateness, diagnostic approaches and models have been developed to divide lateness distributions into components to describe the causes of the lateness and to analyse lateness behaviour during the order fulfilment process. In section 2.2.1, the deduction of mean relative lateness at work systems using the theory of logistic operating curves (LOC), which is important for understanding this paper, will be presented. Later, the decomposition approach to output lateness using the calculation of rank deviations in order processing (section 2.2.2) and an embracing analytical transformation approach (section 2.2.3) will be introduced to complete the picture.

2.2.1. Components of mean relative lateness using the theory of logistic operating curves (LOC)

Yu was the first to present an approach for differentiating the mean relative lateness at a work system into a WIP-dependent and a sequence-dependent lateness component (Nyhuis and Wiendahl, 2009; Yu, 2001). The approach is based on the theory of LOC by Nyhuis and Wiendahl (2006), which itself is inspired by queueing theory (Hillier et al., 2005), Little's Law (Little, 1961) and the funnel model (Wiendahl and Tönshoff, 1988). Using deductive modelling as well as the tool of simulation (Zeigler et al., 2010; Law, 2015), the theory of LOC can be characterised as a deductive-experimental approach to modelling production logistics systems.

The theory of LOC focuses on modelling the relationship between the independent control variable of WIP and the dependent variables of *output rate*, *range* and *throughput time*, defining the following logistical operating curves:

- the Output Rate Operating Curve (OROC);
- the Range Operating Curve (ROC); and
- the Throughput Time Operating Curve (TTOC), in the case of First-in-First-out (FIFO) dispatching.

In this respect, the logistical operating curves are particularly suitable for estimating the operating behaviour of a work system in terms of, for example, output rate or range if influencing variables such as order batch size, the deviation of work contents or mean WIP are changed (cf. (Wiendahl and Kuprat, 1991)).

Based on the theory of LOC, Yu developed another logistic curve, that of relative lateness (see Fig. 2) (Nyhuis and Wiendahl, 2009; Yu, 2001). By using this as a specific work system characteristic curve, it is possible to express mean relative lateness as a function of the WIP of a work system and thus to determine its expected mean schedule reliability, taking into account the variation of the WIP and the customer's tolerance of schedule reliability.

Fig. 2 shows the four logistic operating curves with two exemplary operating points ($WIP_{m,OP1}$ & $WIP_{m,OP2}$) deviating from the planned operating point ($WIP_{m,plan}$), with the resulting mean relative lateness ($L_{m,rel}$). A detailed examination of the theory of LOC will not be undertaken in this paper, and the individual curves, their calculation equations and important influencing variables will be dealt briefly. For a more detailed discussion, see (Nyhuis and Wiendahl, 2009).

The three logistic operating curves of output rate (OROC), range (ROC) and throughput time (TTOC) in Fig. 2 can be calculated as follows:

- Equation (3) shows the calculation of the **OROC**.

$$WIP_m(ROUT_m) = WIPI_{min} * \frac{ROUT_m}{ROUT_{max}} + WIPI_{min} * \propto_1 * \left(1 - \sqrt[4]{1 - \frac{ROUT_m}{ROUT_{max}}}\right)^4 \quad (3)$$

$$with: WIPI_{min} = WC_m * \left(1 + {WC_v}^2\right)$$

where WIP_m is the mean WIP level [hrs], $ROUT_m$ the mean output rate [hrs/SCD], $WIPI_{min}$ the ideal minimum WIP level [hrs], $ROUT_{max}$ the maximum output rate [hrs/SCD], $\propto_1$ the stretch factor [-], WC_m the mean work content [hrs] and WC_v the coefficient of variation of the work content [-].

Significant influencing variables on the shape of the OROC are the ideal minimum WIP level of the work system, which results from the distribution (mean and variation coefficient) of the work contents of the orders, as well as the stretch factor $\propto_1$, which is derived from the load variation and the work system's capacity flexibility. This usually ranges from 5 (low load variation and high capacity flexibility) to 30 (high load variation and low capacity flexibility). A default value of $\propto_1 = 10$ has

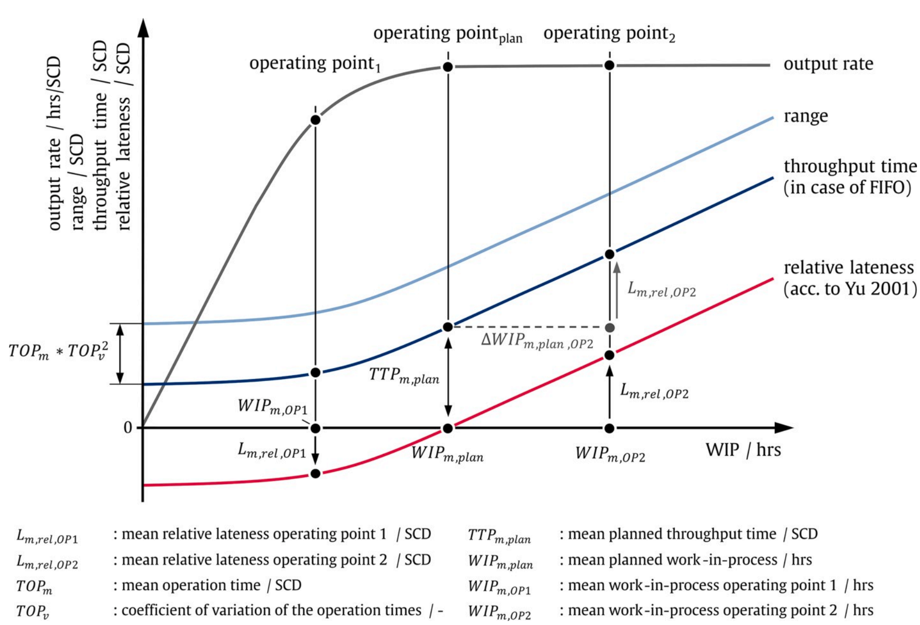

Fig. 2. The logistic operating curves of output rate, range, throughput time and relative lateness (Nyhuis and Wiendahl, 2009; Yu, 2001).

been confirmed in a large number of practical studies. The ideal minimum WIP level is defined as the mean WIP at which utilisation of 100% is achieved under ideal input and output conditions at the work system so that the mean output rate corresponds to the maximum possible output rate. In this case, the complete WIP is 'active' directly in the work system, and no order is waiting in the queue.

• The **ROC** can be calculated using the funnel formula (Wiendahl and Tönshoff, 1988; Wiendahl, 1995), which relates the WIP (x-axis) to the output rate and thus to the result of the calculation of the OROC (see Equation (4)).

$$R_m = \frac{WIP_m}{ROUT_m} \tag{4}$$

where R_m is the mean range [SCD], WIP_m the mean WIP level [hrs] and $ROUT_m$ the mean output rate [hrs/SCD].

The range defines how long a work system can stay productive with the corresponding mean output rate for a specified WIP level if no further order entries occur at the system. The range belongs to the class of time figures and is independent of the dispatching rule at the work system.

• The **TTOC** (in the case of FIFO dispatching) describes the mean throughput time as a function of the WIP level of the work system that is to be expected in the case of order dispatching according to FIFO or other dispatching rules that are not dependent on the work content. It is shifted compared to the range curve, where distance is calculated based on the distribution of operation times at the work system (see Equation (5)).

$$TTP_m = R_m - TOP_m * TOP_v^2 \tag{5}$$

$$with: TOP_m = \frac{WC_m}{ROUT_{max}} \Rightarrow TOP_v = WC_v$$

where TTP_m is the mean throughput time in the case of FIFO dispatching [SCD], t the running variable [-], R_m the mean range [SCD], TOP_m the mean operating time [SCD], TOP_v the variation coefficient of the operating time [-], WC_m the mean work content [hrs], $ROUT_{max}$ the maximum output rate of the work system [hrs/SCD] and WC_v the variation coefficient of the work content [-].

By shifting the TTOC, Yu derived the characteristic curve for the relative lateness (RLOC) of a work system. For this purpose, he defined the y-shift of the TTOC in the negative y-direction as the average planned throughput time ($TTP_{plan,m}$). This results in the relative lateness being 0 if the throughput time according to the TTOC is equal to the planned throughput time. This is the case when the work system is operated at the planned operating point, that is, the work system is operated at a mean WIP level corresponding to the planned throughput time according to the characteristic curve ($WIP_{plan,m}$), which can be determined using the TTOC. The RLOC is thus calculated using Equation (6).

$$L_{rel,m} = TTP_m - TTP_{plan,m} \tag{6}$$

where $L_{rel,m}$ is the mean relative lateness [SCD], TTP_m the mean throughput time in the case of FIFO dispatching [SCD] and $TTP_{plan,m}$ the mean planned throughput time [SCD].

As shown in Fig. 2, the mean relative lateness value at an operating point is equivalent to the difference in the TTOC values between the observed and planned operating points (see also the grey lines in Fig. 2).

2.2.2. *Decomposition of lateness using rank-deviations in order sequence*

Another approach is presented by Lödding et al. who differentiate the output lateness of a work system into the components of sequence-dependent and backlog-dependent output lateness (Lödding et al., 2014). This differentiation aligns with the influencing variables of schedule reliability, backlog and sequence deviation identified by Lödding in his model of manufacturing control (Lödding, 2012). These variables are significantly influenced by the manufacturing control tasks of capacity control and dispatching. The approach is distinctive as sequence-dependent lateness is not measured as the difference between two dates in order processing or as the difference between operating curves, but by calculating the so-called sequence deviation. The sequence deviation represents the difference between the priority ranks of the order in the planned and actual output, resulting from sorting the order data according to the planned respectively actual output date. It can be calculated both based on the unweighted rank and the rank weighted in accordance with the work contents of the orders. The weighted rank is preferred due to its better overall interpretability.

The weighted rank of an order is calculated using the following equation:

$$Rank_{w,i} = \sum_{j=1}^{Rank_i} WC_j \tag{7}$$

where $Rank_{w,i}$ is the weighted rank of order i [hrs], $Rank_i$ the unweighted rank of order i and WC_j the work content of the order with rank j [hrs].

By calculating the difference between the weighted actual and planned output rank of an order and dividing it by the planned output rate of the work system, the sequence-dependent output lateness of the order can be calculated, and by determining the mean sequence deviation the mean sequence-dependent output lateness of the work system, respectively.

$$L_{out,sd,i} = \frac{SD_{w,i}}{ROUT_{plan}} = \frac{Rank_{w,act,i} - Rank_{w,plan,i}}{ROUT_{plan}} \tag{8}$$

where $L_{out,sd,i}$ is the sequence-dependent output lateness of order i [SCD], $SD_{w,i}$ the weighted sequence deviation of order i [hrs], $ROUT_{plan}$ the planned output rate of the work system [hrs/SCD], $Rank_{w,act,i}$ the weighted, actual rank of order i and $Rank_{w,plan,i}$ the weighted, planned rank of order i.

The backlog-dependent output lateness of an order can be calculated in two ways. First, Equation (8) can be adapted by measuring the backlog of the work system at the actual end of processing of the order under consideration and dividing it by the planned output (Equation (9)).

$$L_{out,bl,i} = \frac{BL(EDOP_{act})}{ROUT_{plan}} \tag{9}$$

where $L_{out,bl,i}$ is the backlog-dependent output lateness of order i [SCD], $BL(EDOP_{act})$ the backlog at the actual end of operation processing [hrs] and $ROUT_{plan}$ the planned output rate of the work system [SCD/hrs].

Second, using to the decomposition assumption, the backlog-dependent lateness can also be calculated as follows:

$$\begin{aligned} L_{out,i} &= L_{out,sd,i} + L_{out,bl,i} \\ \Rightarrow L_{out,bl,i} &= L_{out,i} - L_{out,sd,i} \end{aligned} \tag{10}$$

where $L_{out,i}$ is the output lateness of order i [SCD], $L_{out,sd,i}$ the sequence-dependent output lateness of order i [SCD] and $L_{out,bl,i}$ the backlog-dependent output lateness of order i [SCD].

By calculating the mean values, it is possible to examine the mean backlog-dependent and mean sequence-dependent output lateness. In addition, calculating output lateness based on the weighted ranks enables a concrete estimate of the influence of dispatching rules on output lateness and its variation (Lödding and Piontek, 2017).

In addition, the calculation approach for determining the sequence-dependent output lateness at work systems using rank-deviations can be basically transferred onto the relative lateness. Thus, the relative acceleration or deceleration at a work system, as experienced by an order due to a specific dispatching rule, can be calculated by transforming

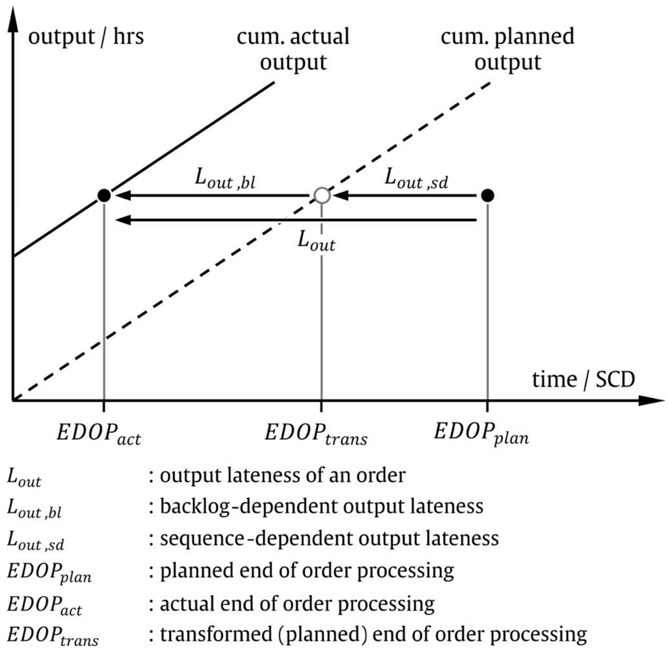

Fig. 3. Analytical Decomposition of Lateness (here: output lateness) (cf. (Lödding and Kuyumcu, 2015; Bertsch et al., 2016)).

Equation (8). In this way, any effects in the input of the work system can be isolated during the examination, enabling focusing on the individual (relative) contribution of the work system to the sequence-dependent lateness behavior of the work system in the output. Lödding and Piontek are showing this by calculating the difference of the weighted output and weighted input rank of an order at a work system and thus calculating the sequence-dependent relative lateness or in other words the sequence-dependent throughput time of an order at a work system (Lödding and Piontek, 2018).

2.2.3. Analytical decomposition of lateness using transformation

A third approach is presented by Kuyumcu and Bertsch (cf. (Lödding and Kuyumcu, 2015; Bertsch et al., 2016)). The basis of this approach is to determine a transformed planned end date of order processing for each order in addition to the existing planned end of order processing. This approach differs from the preceding approach in that no calculation of weighted rank deviations is necessary. The backlog-dependent output lateness can easily be determined (as demonstrated below), and the sequence-dependent output lateness is calculated as the remaining part of the lateness.

Fig. 3 illustrates this analytical decomposition in an idealised throughput diagram for a work system showing the cumulative actual and planned output as well as the mentioned transformed planned end dates of order processing.

As the figure shows, the transformed end of order processing represents the point in time when the cumulative planned output reaches the same value as the cumulative actual output at the time of the actual end of order processing. Although the resulting value for the backlog-dependent output lateness is calculated slightly differently, as the difference between the transformed and the actual end date of order processing, the result is the same as Equation (9), as long as the planned output rate is constant.

Assuming the system is in a steady state, the following equations are used to calculate the backlog- and sequence-dependent relative lateness of an order:

$$L_{rel,bl,i} = L_{out,bl,i} - L_{in,bl,i} = R_{act}\left(EDOP_{pre,i}\right) - R_{plan}\left(EDOP_{trans,i}\right) \tag{11}$$

where $L_{rel,bl,i}$ is the backlog-dependent relative lateness of order i [SCD], $L_{out,bl,i}$ the backlog-dependent output lateness of order i [SCD], $L_{in,bl,i}$ the backlog-dependent input lateness of order i [SCD], $R_{act}(EDOP_{pre,i})$ the actual range at the end of operation at the previous station of order i [SCD] and $R_{plan}(EDOP_{trans,i})$ the planned range at the transformed end of operation of order i [SCD].

$$L_{rel,sd,i} = \left(TTP_{act,i} - R_{act}\left(EDOP_{pre,i}\right)\right) - \left(TTP_{plan,i} - R_{plan}\left(EDOP_{trans,i}\right)\right) \tag{12}$$

where $L_{rel,sd,i}$ is the sequence-dependent relative lateness of order i [SCD], $TTP_{act,i}$ the actual throughput time of order i [SCD], $R_{act}(EDOP_{pre,i})$ the actual range at the end of operation at the previous station of order i [SCD], $TTP_{plan,i}$ the planned throughput time of order i [SCD], and $R_{plan}(EDOP_{trans,i})$ the planned range at the transformed end of operation of order i [SCD].

It is notable that the relative backlog-dependent lateness, according to Bertsch, is expressed as a difference of ranges, which is similar to the mean relative WIP-dependent lateness component used by Yu and corresponds to it on average.

Furthermore, Equation (12) indicates that relative sequence-dependent lateness results from the actual acceleration or deceleration of the order minus the planned acceleration or deceleration. The planned acceleration or deceleration thus depends on how orders are planned and can therefore have a major influence, in particular, if the planned throughput times of the orders fluctuate significantly, are not appropriate (e.g. through planning errors), or do not take into account the dynamics of a high variation of work contents.

2.2.4. Lateness evolution and the influence of production planning

In addition to the three approaches to the causes and components of lateness at work systems presented above, there are also studies concerning the evolution of lateness during order processing, as well as investigations on the influence of production planning on lateness behaviour. The existing literature can be divided into three categories:

- Practice studies on the correlation of components of lateness in production areas (Nyhuis and Wiendahl, 2009; Wiendahl et al., 1998) and the exploitation of the emerging potential of data analytics and data mining for analysing lateness behaviour (Windt and Hütt, 2011).
- The development of analytical frameworks supporting companies in identifying the core causes of unsatisfactory schedule reliability (Schmidt et al., 2019; Soepenberg et al., 2012).
- The investigation of the time-dynamic development of lateness behaviour to determine the sources of lateness in the production area and analyse possible planning-induced lateness (Lucht et al., 2021; Soepenberg et al., 2008).

It is apparent that analysis of the causes of lateness is essential for companies. However, detecting the core causes of lateness is often complex, which is why analytical frameworks and diagnostic models are required for practical usage. As the studies by Lucht et al. (2021) have also shown, the analysis is made more difficult in the modern context because planning systems often undermine schedules and planned dates adapt dynamically to circumstances. This makes the detection of causes of lateness much more difficult. In addition, such planning systems tend to provoke planning errors if, for example, feedback data is not detected in time or is detected incorrectly. This can result in planning runs being carried out on the basis of 'wrong' data (such as faulty work plans). This implies that planning errors can have an important influence on the measured lateness of a work system and should therefore be included in the analysis.

2.3. Interim conclusion

The preceding presentation of the current state of knowledge shows that various approaches and models exist dealing with lateness at work systems and lateness in the order fulfilment process. It is notable that an exact differentiation for each order's lateness components is mathematically demanding and requires appropriate data, which impedes a fast and straightforward diagnosis of the causes of lateness at work systems. In addition, studies show that new planning systems, in particular, can hamper analysis and provoke planning inconsistencies. Planning inconsistencies have not or only partially been taken into account in the existing approaches to differentiating the causes of lateness at work systems.

Although the analytical calculation approach, according to Bertsch and Kuyumcu, is the most exact, it is only conditionally suitable for practical application due to its complexity. At the same time, it is evident that lateness caused by planning inconsistencies (e.g. not consistent throughput time targets) have not yet been fully considered in the analytical approaches and is therefore often included in the sequence-dependent lateness component. In addition to planning inconsistencies, other planning effects can also influence the lateness respectively the throughput time behavior of work systems. For example, if the orders in the production plan are already sequenced according to their work content, this may result in a detectable relative lateness respectively throughput time component at the work system, caused by production planning.

When comparing the three approaches (the theory of LOC, the rank-deviation approach and the analytical approach), a key difference becomes apparent. In Yu's approach, based on the theory of LOC, the range difference is described as WIP-dependent lateness, while the other two approaches use the term backlog-dependent lateness. However, backlog as a variable always establishes a relationship between an actual and a planned value and is, therefore, more challenging to determine than the directly detectable WIP level. For this reason, we will use the term WIP-dependent lateness in the following chapter.

3. Research gap and design

As explained in the previous chapter, Yu's extension of the theory of LOC allows the consideration of two influences on mean relative lateness: sequence-dependent and WIP-dependent lateness (see Fig. 4). A mean sequence-dependent lateness (Case 1) occurs if dispatching (partially) correlates with the order's work content (e.g. by using the shortest processing time (SPT) rule [cf. 6]). This lateness component can be obtained as the vertical distance of the actual throughput time to the TTOC at the mean actual WIP level.

A mean WIP-dependent lateness occurs if the mean actual WIP at a work system deviates noticeably from the mean planned WIP level (Case 2). This component can be obtained as the vertical distance of the planned throughput time to the throughput time according to the TTOC at the mean planned WIP level (planned throughput time according to LOC theory). In summary, no mean relative lateness occurs if the mean actual WIP equals the planned WIP and no work content-dependent dispatching takes place.

The WIP, generally represented by the difference between the cumulated output and input at a given time, corresponds to the work content of the orders to be processed (passive WIP) and the orders in processing (active WIP) at a work system. The mean planned WIP in a given time period [t_0, t_1] can be calculated accordingly by considering the planned inputs and outputs instead of the actual values (see Equation (13)).

$$WIP_{m,plan} = \frac{\int_{T=t_0}^{T=t_1} IN_{plan}(T)dT - \int_{T=t_0}^{T=t_1} OUT_{plan}(T)dT}{t_1 - t_0} \tag{13}$$

where $WIP_{m,plan}$ is planned, mean WIP [hrs], IN_{plan} the planned input (cumulative work content of the planned, incoming operations) [hrs], OUT_{plan} the planned output (cumulative work content of the planned outgoing operations [hrs], t_0 the beginning of the reference period, and t_1 the end of the reference period.

Without planning influences (inconsistencies or work content based order sequencing), the empirically determined planned throughput times ($TTP_{m,plan}$) would correspond to the throughput time according to the TTOC ($TTP^{loc}_{m,plan}$). However, if planning influences are present, $TTP_{m,plan}$ and $TTP^{loc}_{m,plan}$ deviate from each other, and the difference can be

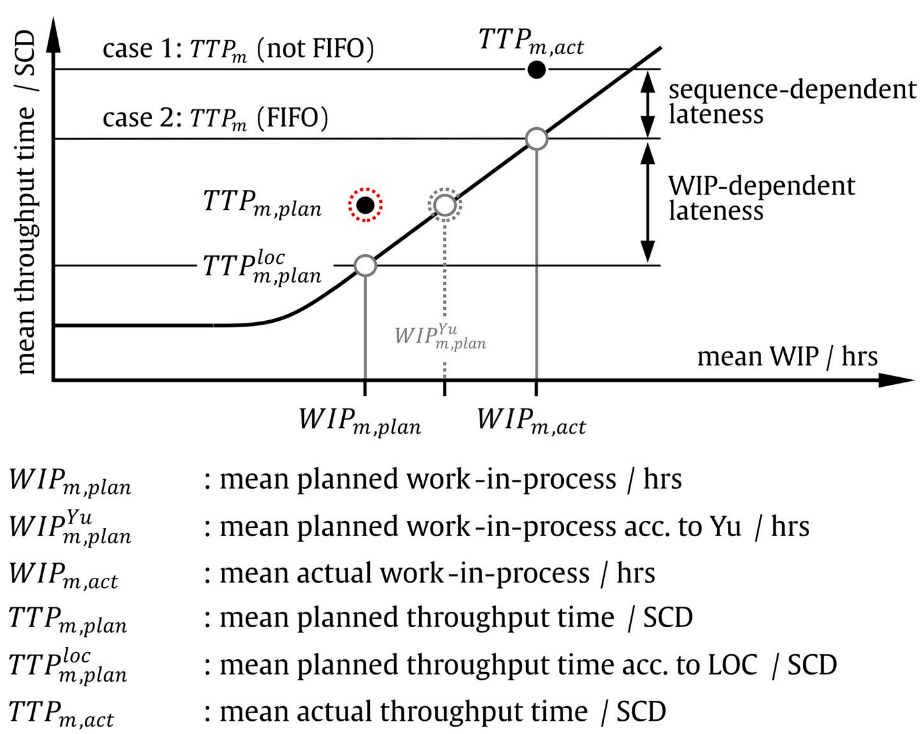

Fig. 4. Visualisation of a possible planning-dependent deviation in Yu's model.

described as a planning influence. Yu has not yet considered such a planning influence (neither inconsistencies nor planned order sequencing). As shown in Fig. 4, the mean planned throughput ($TTP_{m,plan}$) exceeds the mean planned throughput time ($TTP^{loc}_{m,plan}$) according to the LOC theory.

According to Yu, however, the planned WIP can be determined based on the mean planned throughput time (using TTOC). If this was carried out in the above example, the assumed mean planned WIP ($WIP^{Yu}_{m,plan}$) would no longer correspond to the calculated mean planned WIP (see Equation (13)), and an incorrect planned operating point would be used for the analysis. This results in two effects. First, the planning influence would be completely concealed, as there would be no difference between the TTOC and the mean planned throughput time at the 'wrongly' assumed planned operating point ($WIP^{Yu}_{m,plan}$). Second, the WIP-dependent lateness component would also not be calculated correctly. As the example in Fig. 4 shows, the WIP-dependent lateness is lower if the operating point according to Yu is assumed than the WIP-dependent lateness resulting from the exact determination of the planned WIP using Equation (13). In addition, it is clear that the resulting difference between the two assumptions is equal to the planning influence hidden by Yu's assumption. This, in turn, would lead to incorrect interpretation and the implementation of inappropriate countermeasures, which would not have the desired effect.

In the following chapter, therefore, an extension of Yu's model based on the theory of LOC is postulated that considers planning-dependent influences and thus enables a more detailed and accurate analysis of the components impacting the mean relative lateness of a work system. The derivation of this enhanced model is done by a formal-deductive model synthesis. A description of the entire model, an exemplary application and the derivation of a simple analysis approach for practical use will be presented.

4. Model synthesis | refinement of Yu's model

As shown, existing models for the analysis of the mean relative lateness of work systems have shortcomings. This leads to an incomplete analysis of the drivers of a deviation of the mean actual throughput time from the mean planned throughput time and to inadequate analysis of the root causes for such variations. Furthermore, some approaches require comparatively high mathematical input, for example, by calculating priority ranks, to decompose the resulting lateness. At the same time, for analysing work systems in operational practice, the analysis of mean values is usually wholly sufficient.

For this reason, we have adopted Yu's modelling approach (Nyhuis and Wiendahl, 2009; Yu, 2001), which has not yet been published internationally, and the theory of LOC as a starting point and refined them to close the identified gap. The aim of this enhancement is twofold. First, to enable a better identification and analysis of the possible occurring lateness components by means of an easily applicable overall model. Second, to provide the practitioner with a tool for the simple and visual examination of the mean lateness situation at work systems.

In this chapter, we present the three different components determined as possible causes for a deviation of the mean actual throughput time from the mean planned throughput time:

- sequence-dependent component;
- WIP-dependent component; and
- planning deviation-dependent component.

In order to simplify the modelling and improve transparency, the components of relative lateness leading to deviations between the mean actual and the mean planned throughput time (throughput time deviation) are referred to as throughput time components. For example, the mean relative sequence-dependent lateness corresponds to the sequence-dependent throughput time of a work system.

4.1. Component 1: the sequence-dependent throughput time component

The sequence-dependent throughput time TTP^{seq}_m describes the influence of dispatching at the work system on the mean actual throughput time. It results from the difference between the mean actual throughput time and the FIFO throughput time corresponding to the TTOC at the operating point resulting from the mean actual WIP level (see Equation (14) and Fig. 5).

$$TTP^{seq}_m = TTP_{m,act} - TTP^{FIFO}_m(WIP_{m,act}) = TTP_{m,act} - \left(R_m(WIP_{m,act}) - TOP_m * TOP_v{}^2\right) \tag{14}$$

where TTP^{seq}_m is the sequence-dependent throughput time [SCD], $TTP_{m,act}$ the mean actual throughput time, $TTP^{FIFO}_m(WIP_{m,act})$ the mean throughput time in the case of FIFO dispatching according to the theory of LOC, $R_m(WIP_{m,act})$ the range [SCD], TOP_m the mean operation time [hrs] and TOP_v the coefficient of variation of the operation times at the work system [-].

Thus, it corresponds to the portion of the throughput time specified by Yu as sequence-dependent throughput time.

Three intervals can be defined for the sequence-dependent throughput time component, allowing a direct interpretation of the dominant type or rule of dispatching at the work system. A negative sequence-dependent throughput time (where the mean actual throughput time is smaller than the corresponding mean FIFO throughput time) indicates that on average orders with smaller work contents have been accelerated compared to orders with larger work contents. This is due to the fact that, by prioritising several small orders, correspondingly fewer proportionally large orders will be decelerated. This reduces the unweighted mean throughput time. The SPT rule is a widely known priority rule leading to such a deviation. On the other hand, a positive sequence-dependent throughput time indicates that jobs with large work contents are more likely to be accelerated, delaying several small jobs and increasing the mean unweighted throughput time., This may be due, for example, to the application of the LPT (longest-processing time) rule.

The intensity of the impact of the SPT or LPT effect is dependent on various factors, such as the work content variation of the orders and constraints in the applied dispatching rule concerning the maximum shift of orders. As such, this impact cannot yet be described analytically through an exact formula. However, it can be shown both argumentatively and simulation-based that the effects of SPT and LPT increase with increasing work content variation of orders as well as with increasing queue length at the work system (WIP level). This increases the potential for changes in order sequence.

An approximation formula for calculating the throughput time in the case of SPT or LPT dispatching is provided by Nyhuis and Wiendahl (2009). However, it should be noted that the effect of SPT and LPT dispatching on the sequence-dependent throughput time is often hard to determine. Particularly if the utilisation of the observed work system is high (98% and above), the system behaviour is very unstable leading to observations that are difficult to reproduce (e.g. using simulation) and thus also hard to approximate.

If the sequence-dependent throughput time is close to 0, this indicates that the dispatching rule systematically prefers neither small nor large orders at the work system, and thus the mean throughput time is not significantly affected. Examples of such 'neutral' rules are the FIFO rule and due-date-oriented rules if the production plan does not induce correlations of order work contents and order priorities.

4.2. Component 2: the WIP-dependent throughput time component

Due to a deviation of the mean actual WIP level from the mean planned WIP level, for example, as a result of the override of the production plan by order release or capacity control, the operating point of the work system, the output rate and the range of the work system

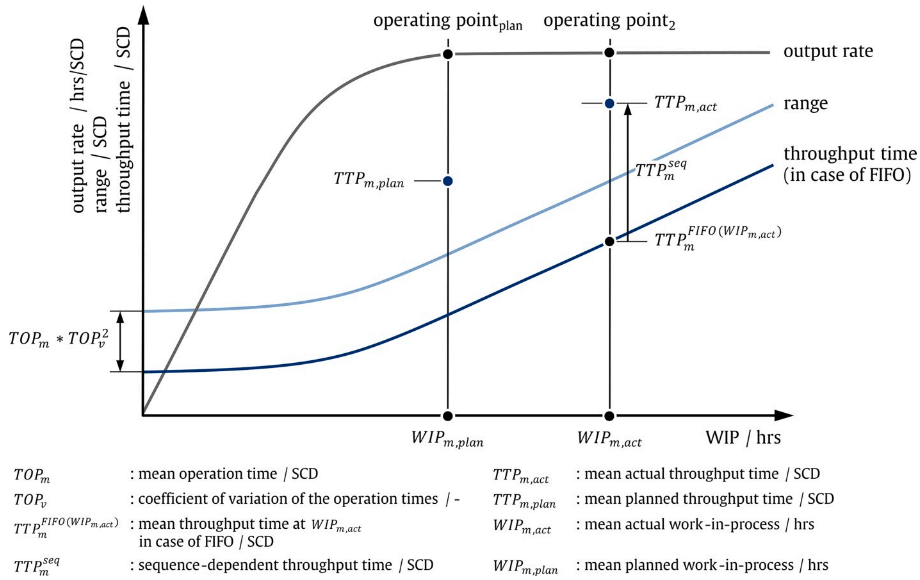

Fig. 5. The sequence-dependent throughput time component (TTP_m^{seq}).

changes according to the funnel formula (see Equation (4)). This change in the range is accompanied by a change in throughput time by the same proportion due to the parallel curve of the TTOC (see Fig. 6).

$$
\begin{aligned}
TTP_m^{\Delta WIP_m} &= R_m(WIP_{m,plan}) - R_m(WIP_{m,act}) \\
&= TTP_m^{FIFO}(WIP_{m,plan}) - TTP_m^{FIFO}(WIP_{m,act})
\end{aligned}
\tag{15}
$$

where $TTP_m^{\Delta WIP_m}$ is the WIP-dependent throughput time [SCD], $R_m(WIP_m)$ the range [SCD] and $TTP_m^{FIFO}(WIP_m)$ the mean throughput time in the case of FIFO dispatching according to the theory of LOC.

The WIP-dependent throughput time component calculation in the modified model differs from the calculation according to Yu. The planned WIP level is determined by the planning data or the production plan (see Equation (13)) and is not based on the transformation of the planned throughput times into a planned WIP using the theory of LOC. This method avoids conflating possible planning-dependent influences with the WIP-dependent throughput time. The outcome of the production plan is thus precisely represented in the model.

4.3. Component 3: the planning deviation-dependent throughput time component

As previously mentioned, existing approaches do not consider that the set mean planned throughput time potentially does not correspond to the load situation resulting from the production plan and, thus, to the resulting mean WIP or a work content based order sequence was established in production planning. As a result, the 'real' planned throughput time deviates from the mean planned throughput time (according to TTOC), corresponding to the mean planned WIP level, and a temporal difference occurs (see Fig. 7). This is referred to as the planning deviation-dependent throughput time component. It can be calculated as follows:

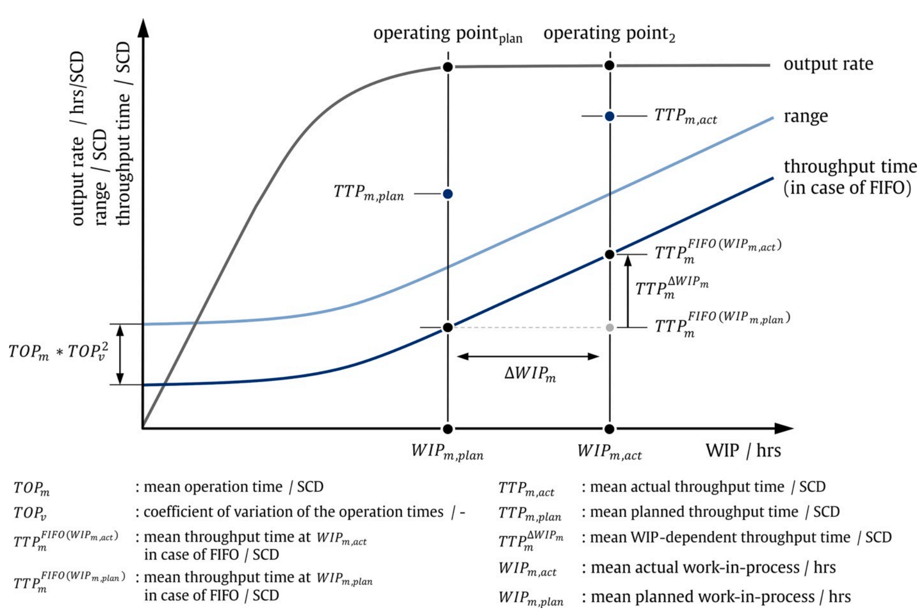

Fig. 6. The WIP-dependent throughput time component.

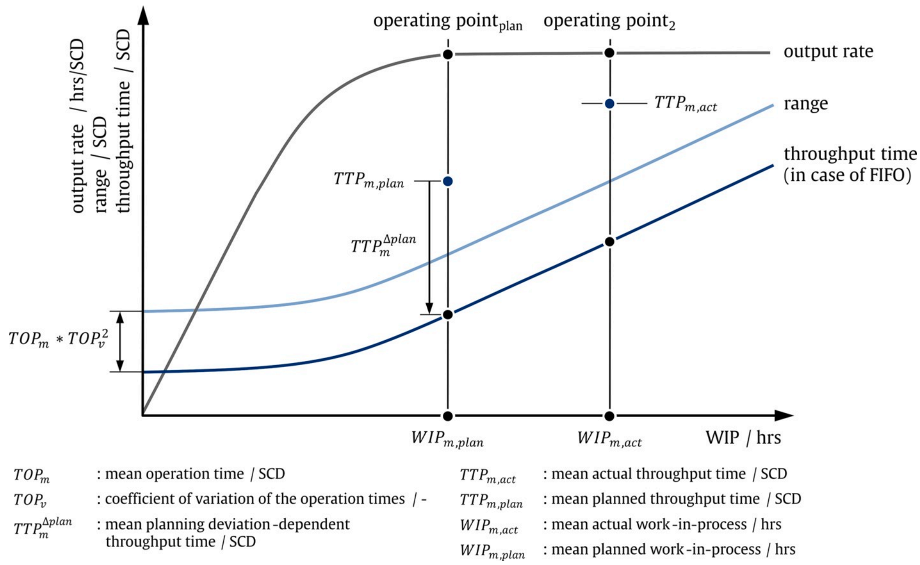

Fig. 7. The planning deviation-dependent throughput time component.

$$\begin{aligned} TTP_m^{\Delta plan} &= TTP_m^{FIFO}\left(WIP_{m,plan}\right) - TTP_{m,plan} \\ &= \left(R_m\left(WIP_{m,plan}\right) - TOP_m{*}TOP_v{}^2\right) - TTP_{m,Plan} \end{aligned} \tag{16}$$

where $TTP_m^{\Delta plan}$ is the planning deviation-dependent throughput time [SCD], $TTP_m^{FIFO}(WIP_{m,plan})$ the throughput time according to TTOC at the operating point of the planned WIP [SCD], $TTP_{m,act}$ the calculated mean throughput time according to LOC at the actual operating point [SCD], $R_m(WIP_{m,plan})$ the mean range according to the funnel formula [SCD], TOP_m the mean operation time [SCD] and TOP_v the coefficient of variation of the operation times at the work system [-].

By calculating this proportion, it is thus possible to explicitly determine which throughput time deviation beyond the control of the work system is induced by planning, affecting the work system's mean relative lateness. At the same time, calculating this dedicated component ensures that the resulting planning deviation-dependent throughput time is not added to a throughput time component as a residual component and is thus 'hidden' either in the sequence-dependent lateness component, WIP-dependent component, or both.

4.4. The resulting complete model at a glance

Fig. 8 shows the resulting overall model for detecting the different throughput time components leading to the mean relative lateness of a work system. Regarding the overall relationship between the various components, the following equation applies:

$$L_{m,rel} = TTP_m^{seq} + TTP_m^{\Delta WIP_m} + TTP_m^{\Delta plan} = TTP_{m,act} - TTP_{m,plan} \tag{17}$$

where $L_{m,rel}$ is the mean relative lateness [SCD], TTP_m^{seq} is the sequence-dependent throughput time [SCD], $TTP_m^{\Delta WIP_m}$ the WIP-dependent throughput time [SCD], $TTP_m^{\Delta plan}$ the planning deviation-dependent

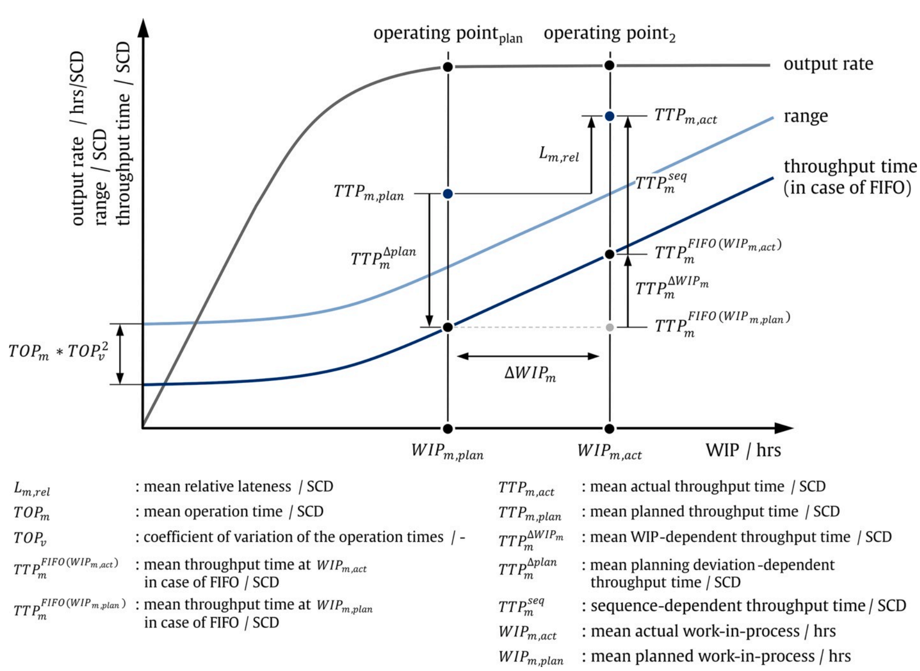

Fig. 8. The complete model for the detection of the throughput time components resulting in the mean relative lateness of a work system.

throughput time [SCD], $TTP_{m,act}$ the mean actual throughput time [SCD] and $TTP_{m,plan}$ the mean planned throughput time [SCD].

As the exemplary case shown in the figure illustrates, and as can be derived argumentatively, there might be a correlation between the planning deviation-dependent throughput time and the WIP-dependent throughput time. Thus, it is in general unlikely that in case of a planning deviation-dependent throughput time there will be no discrepancy between the planned and actual operating point according to LOC theory and thus no WIP-dependent throughput time. The strength of this correlation depends on various factors, such as the scheduling type and its parameterisation.

Although the effects, when added together, may (partially) neutralize each other with respect to the mean relative lateness, as in the example, this does not limit the application of the developed model. Instead, it provides greater insight into how the different unaccounted influencing factors affect the throughput time components and, ultimately, the mean relative lateness so that the practitioner gains deeper insight in the causes of relative lateness and can derive adequate measures for compensation.

5. Practical example

The presented model was tested in the context of a production logistics analysis in a manufacturing area of a German machinery and plant engineering company. The focus of the assessment was the question of why the schedule reliability of the production area was too low towards the downstream assembly area, which led to high delivery delays and poor performance for the customer. Within the scope of the analysis, the planned and actual input and output dates, as well as the work content of the orders at different single work systems of the manufacturing area, were recorded. Further, an analysis of the mean relative lateness was carried out for the work systems that generated the most significant amount of lateness using the presented model.

As an example, we demonstrate the outcome for a milling machine where the order flow was analysed over a period of 220 SCD. With this example the principal applicability of the developed model is demonstrated. The mean actual throughput time of the milling machine was 13.7 SCD, with a standard deviation of 16 SCD, and 180 orders were processed. The mean planned throughput time was 8 SCD. The $\propto_1$ value of the OROC (grey line) needed to finally calculate the TTOC (blue line) was examined using the proposed calculation method of Nyhuis and Wiendahl (2009) based on the Newtonian iteration method (also known as: Newton-Raphson method).

Fig. 9 shows the result of the analysis, which was calculated with a spreadsheet software, showing the application of the developed model.

As can be seen, the machine had a significant mean relative lateness of 5.7 SCD, which was primarily caused by two components. First, the actual operating point was significantly higher than the planned operating point, so the mean WIP-dependent throughput time was 10.45 SCD. However, it also became apparent that the mean planned throughput time did not match the mean planned WIP level, which is why a planning deviation-dependent throughput time of −4.1 SCD occurred, working against the WIP-dependent throughput time component. The impact of the sequence-dependent throughput time component, in comparison, was relatively low (−0.65 SCD) and showed slight tendencies toward dispatching according to SPT.

The components identified this way were subsequently subjected to further analysis, and, in particular, the planning system was critically reviewed. Among others, the analytical procedure of the bottleneck-oriented logistics analysis developed by Nyhuis and Wiendahl (2009) was carried out and correlation diagrams between planned throughput time and the work content of the orders were created. As a result of this, we concluded that the planning system did not adequately take into account the differing work contents of the individual orders. The correlation diagrams showed, for example, that the planned throughput times correlated only very weakly with the work content of the orders (the correlation coefficient was 0.08) and thus did not adequately reflect the strongly fluctuating work content (the coefficient of variation was close to 1) of the orders and the consequently also fluctuating time required to process them on the machine. A planned order sequence according to the work content of the orders was not identified.

In addition, there were substantial inconsistencies between the load calculation (feasible load per SCD) and the scheduling. Additional problems and inconsistencies could also be found in the planning logic and system itself, through which planned dates were frequently rescheduled, resulting in strong distortions between the planned throughput time and the planned WIP level. In order to investigate this impact, the so-called plan history diagram developed by Lucht et al. was used (Lucht et al., 2021).

Other work systems of the production area also showed similar behaviour, so the complete planning systematics, as well as the interface between actual data and planned data, were examined and analysed in more detail. As a result of the investigations, a fundamental revision of the planning system was subsequently proposed in order to reduce the planning deviation-dependent throughput time. The planned operating point resulting from the analyses was to be maintained; based on the

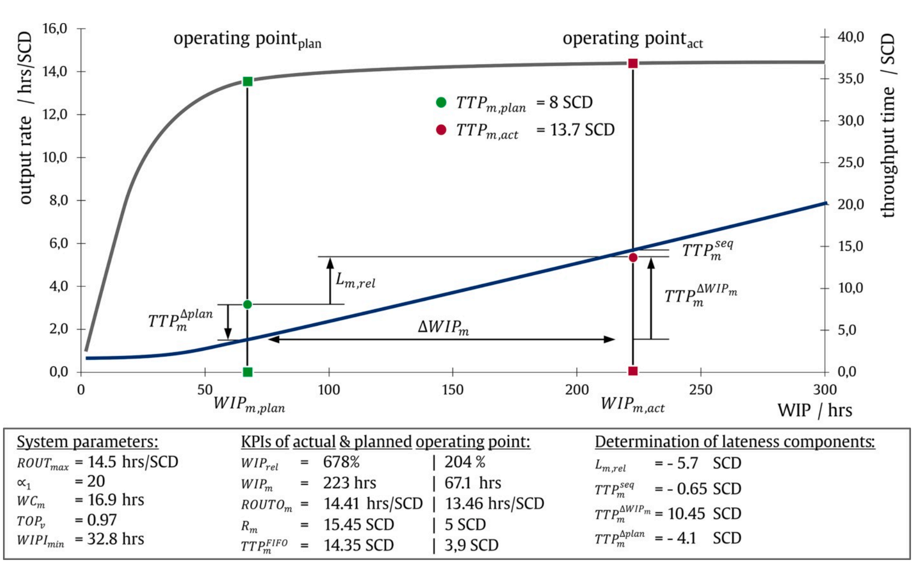

Fig. 9. Analysis of the mean relative lateness behaviour of a milling machine in a production area using the developed model.

TTOC, a correspondingly realistic planned throughput time standard was derived for each work system. The comparatively small influence of the order sequence on the throughput time was not countered at first and was proposed to be addressed in a second evaluation after the first two throughput time components had been rectified. By means of an improved workload control logic (order release), a measure for reducing the actual WIP was derived, so that the planned and actual operating points could be aligned. In combination, it is therefore to be expected that the specified production logistics positioning of the work systems will be achieved in future and that the mean relative lateness will be reduced.

6. A simplified model for practical application

In addition to the clear decomposition of the mean relative lateness into the three throughput time components of sequence-dependent, WIP-dependent and planning deviation-dependent, the presented model also offers the advantage that it can be applied in practice with limited available data to provide a starting point for an initial assessment of the lateness and its components. Further, this method can also be applied to evaluate the production logistics positioning of the entire work system.

For the application of the overall model, including the creation of all logistic characteristic curves, the following key figures are sufficient: maximum possible output rate of the work system (capacity), mean value and standard deviation (or the coefficient of variation) of the work content, estimation of the $\propto_1$ value, mean actual WIP level and mean planned WIP level, as well as mean planned and actual throughput time. This is because the curves of the LOC theory can be interpolated sufficiently based on a few support points. This means that a computer-aided calculation of the lateness components can be omitted, and analyses can be drawn up on paper (comparable with a hand-based value stream analysis).

Table 1 shows possible supporting points for drawing the OROC for different $\propto_1$ values. Although a drawing of the OROC is not necessary for drawing the required TTOC, it increases understanding of the operational behaviour of the work system. It thus supports the derivation of possible countermeasures against a possibly occurring mean relative lateness, an estimation of the achievable output rate at the operating point and the impact of any changes to the system parameters (e.g. the variation of the order's work content). In addition, the OROC can be used to estimate the extent to which the estimated $\propto_1$ value fits the operating behaviour of the work system by comparing the characteristic curve at the actual operating point with the measured actual output rate.

The values given in Table 1 for the supporting points are relative values. For application to a specific work system, the utilisation (U_m) must be multiplied by the maximum possible output rate ($Rout_{max}$) and the relative WIP (WIP_{rel}) by the ideal minimum WIP of the work system ($WIPI_{min}$). The components of the WIP (active and passive), which are also indicated, additionally show how the WIP is subdivided functionally into the WIP currently in processing (active) and the WIP actually in the queue of the work system (passive). While the active WIP corresponds to the mean utilisation as a percentage, the passive WIP depends on the empirical stretch factor ($\propto_1$) value. In addition to the $\propto_1$ values 10 and 20, the table contains the general calculation formula for the passive WIP and the overall relative WIP.

where U_m is mean utilisation [%], $WIP_{a,rel}$ the active WIP as % of the ideal WIP, $WIP_{p,rel}$ the passive WIP in the queue of the work system as % of the ideal WIP, WIP_{rel} the total WIP of the work system as % of the ideal WIP, and $\propto_1$ the empirical stretch factor [-].

In general, the OROC can be drawn with three support points with high accuracy. These are usually a support point at an average output rate of 49.4%, a support point close to the ideal minimum WIP (mean utilisation of 83.2%) and, depending on the $\propto_1$ value and the drawing, a support point close to full mean utilisation (usually 98.8%). To support the drawing, the maximum output rate, as well as the ideal relationship between mean WIP and output rate, can be drawn as a straight line between the coordinate origin and the point ($WIPI_{min}$ | $ROUT_{max}$), which represent the outer boundaries for the OROC and the ideal curve (corresponding to an $\propto_1$ value of 0), where there is only active WIP in the work system.

Similarly, two support points and a support line are to be selected for the range. The first support point should be chosen in the area of the ideal minimum WIP, that is, a mean utilisation of 83.2%. Applying the funnel formula, it becomes apparent that the range at this point corresponds to 120%, that is, a factor of 1.2 of the minimum range, which results from the division of $WIPI_{min}$ and $ROUT_{max}$. The second support point should be chosen at the very end of the diagram to be created and can be determined by dividing the WIP level by the maximum possible output rate of the work system. This is only applicable if the mean output rate according to the OROC is close to the maximum possible output rate. Otherwise, the value according to OROC must be used.

The required support line is the straight line connecting the second support point with the coordinate origin (the light blue dotted line), which, together with the R_{min} line, represents the limits for the ROC. The basic principle is that the ROC adapts to the created supporting straight line with increasing WIP level and an ever-closer approximation of the output rate to the maximum output rate.

Based on the ROC curve created this way, the TTOC can be created using the previously presented Equation (5), which can then be used for the previously presented analyses.

Fig. 10 shows the result of the application of the support point logic and the sketched characteristic curves, as well as the values of the components of the throughput time derived from the curves. As can be seen, the drawn model is very close to the analytically calculated values, confirming that the model is suitable for simple and quick practical application. However, the correct estimation of the $\propto_1$ value represents an important issue. Nyhuis and Wiendahl (2009) have extensively investigated the effects of an incorrect estimation of the $\propto_1$ value on the characteristic curves and the calculated values.

7. Conclusion/outlook

By enhancing Yu's model, it was possible to bring an easy-to-use model up to the current state of the art. This allows the mean lateness at a work system to be easily subdivided into three components and thus enables the reliable identification of the causes of lateness and the determination of suitable countermeasures. The methodology based on

Table 1
Selected supporting points for the creation of the Output Rate Operating Curve (OROC).

Basic variables		$\propto_1$ = variable		$\propto_1$ = 10		$\propto_1$ = 20	
U_m [%]	$WIP_{a,rel}$ [%]	$WIP_{p,rel}$ [%]	WIP_{rel} [%]	$WIP_{p,rel}$ [%]	WIP_{rel} [%]	$WIP_{p,rel}$ [%]	WIP_{rel} [%]
49.4	49.4	0.6 * $\propto_1$/10	**49.4 + 0.6 * $\propto_1$/10**	0.6	**50.0**	1.2	**50.6**
83.2	83.2	16.8 * $\propto_1$/10	**83.2 + 16.8 * $\propto_1$/10**	16.8	**100.0**	33.6	**116.8**
93.1	93.1	56.9 * $\propto_1$/10	**93.1 + 56.9 * $\propto_1$/10**	56.9	**150.0**	103.8	**196.9**
98.8	98.8	201.2 * $\propto_1$/10	**98.8 + 201.2 * $\propto_1$/10**	201.2	**300.0**	402.4	**501.2**
99.8	99.8	400.2 * $\propto_1$/10	**99.8 + 400.2 * $\propto_1$/10**	400.2	**500.0**	800.4	**900.2**

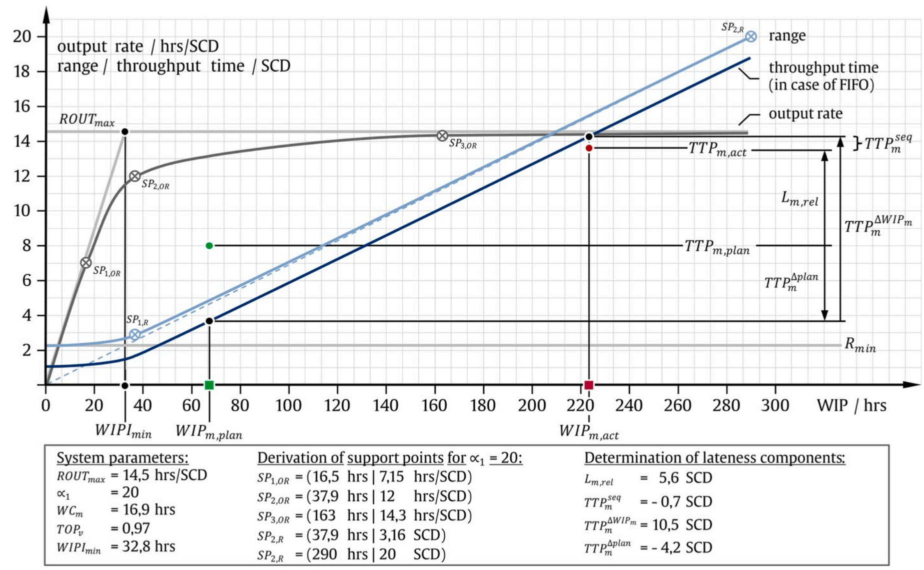

Fig. 10. Analysis of the considered work system using the support point logic.

the theory of LOC is distinguished from the alternative approaches presented, particularly by the significantly lower data requirements, but in return, it does not offer the possibility of showing the specific values of lateness components for each individual order.

7.1. Contribution to theory

The analysis and differentiation of the sources of lateness at work systems and in production areas is the subject of various scholars' contributions. The paper comprehensively presented previous approaches, identified gaps in previous works and further developed Yu's existing model. As described, the model is intended to support the diagnosis and analysis of occurring lateness at work systems and is thus practically oriented. It is integrated into the existing theory of LOC.

7.2. Contribution to practice

With the model, and especially the sketch model for the practical application, we have developed an easy-to-use analysis approach and tested it with empirical data. In various projects and discussions with practitioners, we have observed that companies often lack practical tools for analysis and, especially in the age of digitalization, often do not know what data is needed for analysis and what cause-effect relationships exist in a production system. In this article, we, therefore, provide a tool for the practical user to diagnose and derive targeted measures to counter occurring lateness at work systems and, at the same time, strengthen the system- and process-logistic understanding of cause-effect relationships. Through transparent visualisation and the coupling of the lateness analysis with the overall theory of LOC, it is also possible to consider the performance objectives of utilisation, output rate, throughput time and lateness in one model, to display them transparently, and to prognosticate the effects of any changes on the logistical performance of the work system with comparatively little data available.

7.3. Limitations and further research

The main limitation of the presented approach is that it is not designed to find an exact solution, and, in particular, as discussed, it depends on the correct determination of the $\propto_1$ value. This value must be explicitly defined for a work system and, currently, cannot be determined analytically. However, it can be verified empirically. Further research is necessary for the analytical determination.

In addition, the approach does not allow for the consideration of dispersion parameters, which, in addition to the mean value, are significant in influencing schedule reliability. For example, a random sequence formation is to be mentioned here, which leads to the fact that the sequence-dependent throughput time at the work system per order varies strongly, but the mean has a value around 0. The same applies to any variance of the WIP and scatter of the planning influence.

Funding

This project is funded by the German Federal Ministry of Education and Research, as part of the Aviation Research and Technology Program of the Lower Saxony Ministry of Economics, Labor, Transport and Digitalization (funding code ZW 1-80157862).

Availability of data and material

None.

Code availability

None.

Authors' contributions

Alexander Mütze: Conceptualisation, Methodology, Writing - Original Draft, Tobias Hiller: Conceptualisation, Methodology, Writing - Original Draft, Peter Nyhuis: Conceptualisation, Funding acquisition, Writing - Reviewing and Editing, Supervision.

Declaration of competing interest

None

Data availability

No data was used for the research described in the article.

Acknowledgements

None.

References

Baker, K.R., 1974. Introduction to Sequencing and Scheduling. Wiley, New York, p. 305.

Bertsch, S., Schmidt, M., Nyhuis, P., 2014. Modeling of lateness distributions depending on the sequencing method with respect to productivity effects. CIRP Annals 63 (1), 429–432.

Bertsch, S., Felix, C., Nyhuis, P., 2016. Modeling the sequence dependent lateness of a single workstation. AMR (Adv. Magn. Reson.) 1140, 443–448.

Conway, R.W., Maxwell, W.L., Miller, L.W., 1967. Theory of Scheduling, vol. X. Addison-Wesley, Reading, Mass, p. 294. S.

Dombrowski, U., 1988. Qualitätssicherung im Terminwesen der Werkstattfertigung: [Quality assurance in the scheduling system of a job shop production]. In: At the Same Time. Univ., Hannover. PhD, 1988, Düsseldorf.

Gunasekaran, A., Kobu, B., 2007. Performance measures and metrics in logistics and supply chain management: a review of recent literature (1995–2004) for research and applications. Int. J. Prod. Res. 45 (12), 2819–2840.

Hillier, F.S., Lieberman, G.J., 2005. In: internat (Ed.), Introduction to Operations Research, vol. 8. McGraw-Hill, New York, NY, p. 1061.

Huan, S.H., Sheoran, S.K., Wang, G., 2004. A review and analysis of supply chain operations reference (SCOR) model. Supply Chain Manag.: Int. J. 9 (1), 23–29.

Law, A.M., 2015. Simulation Modeling and Analysis, fifth ed. McGraw Hill Education, New York, NY, xviii. 776, 4 ungezählte Seiten.

Little, J.D.C., 1961. A proof for the queuing formula: $L = \lambda W$. Oper. Res. 9 (3), 383–387.

Lödding, H., 2012. A manufacturing control model. Int. J. Prod. Res. 50 (22), 6311–6328.

Lödding, H., 2013. Handbook of Manufacturing Control: Fundamentals, Description, Configuration. Springer, Berlin Heidelberg. Online-Ressource (XXV, 577 S.).

Lödding, H., Kuyumcu, A., 2015. Modelling schedule reliability. Int. J. Prod. Res. 53 (9), 2871–2884.

Lödding, H., Piontek, A., 2017. The surprising effectiveness of earliest operation due-date sequencing. Prod. Plann. Control 28 (5), 459–471.

Lödding, H., Piontek, A., 2018. Extending Little's Law to single order throughput times. Prod. Plann. Control 29 (1), 1–8.

Lödding, H., Nyhuis, P., Schmidt, M., Kuyumcu, A., 2014. Modelling lateness and schedule reliability: how companies can produce on time. Prod. Plann. Control 25 (1), 59–72.

Lucht, T., Mütze, A., Kämpfer, T., Nyhuis, P., 2021. Model-based approach for assessing planning quality in production logistics. IEEE Access 9, 115077–115089.

Mütze, A., Lucht, T., Nyhuis, P., 2022. Logistics-oriented production configuration using the example of MRO service providers. IEEE Access 10, 20328–20344.

Nyhuis, P., Wiendahl, H.-P., 2006. Logistic production operating curves – basic model of the theory of logistic operating curves. CIRP Annals 55 (1), 441–444.

Nyhuis, P., Wiendahl, H.-P., 2009. Fundamentals of Production Logistics: Theory, Tools and Applications. Springer Berlin Heidelberg, Berlin, Heidelberg, p. 320.

Plossl, G.W., Wight, O.W., 1973. Capacity planning and control. Prod. Inventory Manag. 14 (3), 31–67.

Schmidt, M., Schäfers, P., 2017. The Hanoverian Supply Chain Model: modelling the impact of production planning and control on a supply chain's logistic objectives. Prod. Eng. Res. Dev. 11 (4–5), 487–493.

Schmidt, M., Maier, J.T., Härtel, L., 2019. Data based root cause analysis for improving logistic key performance indicators of a company's internal supply chain. Procedia CIRP 86, 276–281.

Soepenberg, G.D., Land, M.J., Gaalman, G.J.C., 2008. The order progress diagram: a supportive tool for diagnosing delivery reliability performance in make-to-order companies. Int. J. Prod. Econ. 112 (1), 495–503.

Soepenberg, G.D., Land, M.J., Gaalman, G.J.C., 2012. A framework for diagnosing the delivery reliability performance of make-to-order companies. Int. J. Prod. Res. 50 (19), 5491–5507.

Thürer, M., Stevenson, M., Silva, C., 2011. Three decades of workload control research: a systematic review of the literature. Int. J. Prod. Res. 49 (23), 6905–6935.

Wiendahl, H.P., 1995. Load-Oriented Manufacturing Control. Springer Berlin Heidelberg, Berlin, Heidelberg, p. 368.

Wiendahl, H.-P., Kuprat, T., 1991. Logistic analysis of production processes by operating curves. CIRP Annals 40 (1), 475–478.

Wiendahl, H.-P., Tönshoff, K., 1988. The throughput diagram — an universal model for the illustration, control and supervision of logistic processes. CIRP Annals 37 (1), 465–468.

Wiendahl, H.-P., Nyhuis, P., Helms, K., 1998. Durchlaufzeit/2-Die Potentiale liegen auf der Hand!: [Throughput time/2-The potentials are obvious!]. Werkstattstechnik-Forschung und Entwicklung fur die Produktion 88 (4), 159–163.

Windt, K., Hütt, M.-T., 2011. Exploring due date reliability in production systems using data mining methods adapted from gene expression analysis. CIRP Annals 60 (1), 473–476.

Wisner, J.D., 1995. A review of the order release policy research. Int. J. Oper. Prod. Manag. 15 (6), 25–40.

Yu, K.-W., 2001. Terminkennlinie. Eine Beschreibungsmethodik für die Terminabweichung im Produktionsbereich: [Scheduling curve. A description methodology for lateness in the production area]. Düsseldorf At the same time: Hannover, Univ., PhD, 2001. VDI-Verl XIII, 122.

Zeigler, B.P., Prähofer, H., Kim, T.G., 2010. Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems. In: Transferred to Digital Print, 2. ed. Academic Press, Amsterdam, Heidelberg, p. 510.