IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 3, 2020, accepted July 15, 2020, date of publication July 17, 2020, date of current version August 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3010050

Approaches for the Prediction of Lead Times in an Engineer to Order Environment–A Systematic Review

PETER BURGGRÄF, JOHANNES WAGNER, BENJAMIN KOKE, AND FABIAN STEINBERG
Chair of International Production Engineering and Management, University of Siegen, 57076 Siegen, Germany
Corresponding author: Fabian Steinberg (fabian.steinberg@uni-siegen.de)

ABSTRACT The interest of manufacturing companies in a sufficient prediction of lead times is continuously increasing - especially in engineer to order environments with typically a large number of individual parts and complex production processes. A multitude of approaches have been proposed in the literature for predicting lead times considering different data and methods or algorithms from operations research (OR) and machine learning (ML). In order to provide guidance at setting up prediction models and developing new approaches, a systematic review of the available approaches for predicting lead times is presented in this paper. Forty-two publications were analyzed and synthetized: Based on a developed framework considering the used data class (e.g. product data or system status), the data origin (master data or real data) and the used method and algorithm from OR and ML, the publications are classified. Based on the classification, a descriptive analysis is performed to identify common approaches in the existing literature as well as implications for further research. One result is, that mostly order data and the status of the production system are used for predicting lead times whereas material data are used seldom. Additionally, ML approaches primarily use artificial neural networks and regression models for predicting lead times, while OR approaches use mainly combinatorial optimization or heuristics. Furthermore, with increasing model complexity the use of real data decreased. Thus, we identified as an implication for further research to set up a complex data model considering material data, which uses real data as data origin.

INDEX TERMS Lead time reduction, machine learning, operations research, prediction methods.

I. INTRODUCTION

Production companies are in a constant state of change. They are challenged to assert themselves in international markets. Growing demands for individualized products with increasing quality and decreasing prices bring logistics performance, such as high adherence to deadlines or short delivery and lead times, to the fore as a competitive factor [1], [2, p. 2]. As a result, lead time is one of the key factors for meeting customer requirements [3]. By means of a valid prediction of the lead times, delivery dates can be determined at an early stage and deviations from schedule can be identified [4]. In contrast, an imprecise prediction of lead times can lead to delivery dates not being met, resulting in loss of customer confidence and consequential costs for late deliveries [5, p. 1]. Particularly relevant is the prediction of lead times for mechanical and plant engineering, a typical example of an engineer to order process. In addition to production and assembly, here the lead time includes all upstream processes such as design, order planning or the purchasing process for raw materials and finished parts [6, pp. 139-140]. Furthermore, the products of a machine and plant manufacturer often consist of a large number of components that are designed individually to achieve a tailor-made solution for the respective customer [7], [8]. Consequently, the product characteristics defined in the design process represent a unique selling point for the companies.

A primary cause of not meeting due dates and extended lead times are the negative effects of disruptions [9], [10]. The occurring disruptions are manifold and include, for example, machine breakdowns, missing material, lack of personnel or insufficient employee qualification [11], [12]. However, a recent study found that the majority of disruptions in

The associate editor coordinating the review of this manuscript and approving it for publication was Mouloud Denai.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

the assembly process occur repeatedly and are theoretically predictable [13]. If the occurrence of a disruption is known or predictable, this information should also be considered to predict lead times. Consequently, data containing that information about disruption and thus the causes of delays should be used for the prediction of lead times. The number of potential data classes, however, varies due to the large number of possible disturbances. In addition to the considered data, the methods and algorithms used for the prediction are relevant for the quality of the prediction [14], [15, p. 3]. For the prediction of lead times, methods and algorithms from the field of operations research (OR) such as heuristics or combinatorics and from the field of machine learning (ML) such as neural networks or random forest can be applied [16]–[19]. Consequently, the question arises which data should be considered in the context of the forecast and which method or algorithm can be utilized. Due to the multitude of possibilities, the choice is not easy. A systematic review can help to achieve an overview of the existing methods and thus facilitate the selection for the user.

In their often cited survey Cheng and Gupta [20] investigated relationships between due dates, dispatching rules and completion times in static and dynamic job shops. Öztürk *et al.* [21] comprehensively summarized the development of prediction models with a focus on dispatching rules and scheduling. Lingitz *et al.* [22] focused on approaches with regression models to predict lead times. Karaoglan and Karademir [23] provided a comprehensive overview of the mathematical approaches used in the field of machine learning as well as the data classes considered. In all publications, however, only parts of the current state of the art are considered. In addition, it is not always possible to identify whether a systematic procedure was used to review the literature. Even after a comprehensive search, no review was found that systematically summarized both the state of the art of the methods and algorithms used and the data considered.

The aim of this paper is therefore to conduct a systematic literature review to answer the following research question: 'Which is the current state of the art in predicting lead times in engineer to order environments and which data and methods or algorithms are used?'. Additionally, we ask as second research question 'How does the existing literature contribute to future research on the prediction of lead times?' to identify implications for further research. In our study we follow the structure of Vom Brocke *et al.* [24] supplemented by dedicated review concepts from other authors like a procedure model of Moher *et al.* [25] and a clustering approach of Weißer *et al.* [26]. Since we assume that the authors use different classes of data and methods or algorithms, we will develop a framework for the classification of the publications. Based on the classification, we will perform a descriptive analysis, which will then be used to identify focus topics in the existing literature as well as implications for further research.

Our paper is structured as follows. Section II first introduces the terms lead time and prediction. Section III elaborates the systematic literature review and details the applied methodological approach. In section IV a framework is derived as a result of the systemic review and a detailed analysis of the current state of the art in the body of literature is conducted. Based on this, the implications for further research are derived in section V. Finally, a summary is given in the last section.

II. LEAD TIME AND LEAD TIME PREDICTION

According to the Business Dictionary [27] *lead time* is defined as the 'number of minutes, hours, or days that must be allowed for an operation or process, or must elapse before a desired action takes place'. A definition for the term lead time with focus on manufacturing processes is given by the Cambridge Business English Dictionary [28] and Gunasekaran *et al.* [29] with the time that elapses between receiving a customer's order and the delivery of the goods or service to the customer. A more detailed definition for the manufacturing lead time is given by the Business Dictionary with the 'total time required to manufacture an item, including order preparation time, queue time, setup time, run time, move time, inspection time, and put-away time. For make-to-order products, it is the time taken from release of an order to production and shipment' [27]. Wiendahl [30, pp. 41-47] and Nyhuis and Wiendahl [31, pp. 17-24] divide an order into individual operations and differentiate accordingly between order lead time and operation lead time: The order lead time elapses between the start of the first operation and the end of the last operation. Each operation lead time is further divided into the interoperation and operation time. The interoperation time consists of the three components wait time after processing of the previous operation, time for transportation between previous and current workstations and another waiting time before processing on the current workstation. The operation time is divided into the setup time and the actual processing time. As it is well known, waiting times have a higher share in the lead time than the processing times [30, p. 37], [32], [33].

In a production environment the job's lead times are determined by the production schedule considering the available production capacity, technical restrictions, due dates and the system status [5], [34], [35]. The job sequence is defined according to certain rules to calculate the start and end dates of the jobs at the work stations [36]. One of the fundamental rules is to determine the job's waiting time depending on the machine's utilization [37]. Here, performance curves play a key role [38]. The performance curves, also called operating curve [39] or characteristic curve [40], can be generally understood as a tool to model performance indicators of a workstation's productivity considering functional relationships between logistic parameters such as lead times, throughput and stock [37]. To determine the performance curves, several different methods are known, which are subdivided mainly into the two areas *approximation function* and *queuing theory* [37], [38]. Within the area of *approximation functions* the main representative is a description of elementary relationships of flow processes based on the so-called ''funnel model'' and the flow diagram [30], [31], [41]. The funnel-model focuses on the representation of the performance-stock ratio and determines the capacity of a workstation as the upper performance limit. Here, the performance curve is defined as a so called C_{Norm}-function [31]. The area of *queuing theory* condenses approaches which are mainly based on the so-called Kingman equation [42], as well as their extensions to multi-operator systems and adaptations for practical use (see [38] and [43], and the references herein for further details). One exemplary extension of the Kingman's equation is given by the authors in [44], who approximated the curve by using a constant factor to replace the variability term in the Kingman's equation. The authors in [45]–[48] used this extension to quantify the productivity improvement of a semiconductor fabrication plant. Furthermore, historical data can be used in the determination of performance curves. Wu and Mcginnis [49] for example used historical lead times in the determination of the performance curves and based on that calculated queueing times and subsequently lead times.

After determining the production schedule, of course, disruptions can occur that lead to a deviation from the schedule. In this case a rescheduling is performed to update the scheduled according to the new situation [35]. There are also approaches that consider potential disruptions during scheduling to get a more robust schedule [35]. Jorge Leon *et al.* [50] for example analyze the effect of single disruptions for delaying a job and use a genetic algorithm that minimizes expected delays and lead times to find a robust schedule. Tadayonirad *et al.* [51] take unplanned machine breakdowns into account. Summarized in both scheduling and rescheduling the expected lead time is calculated based on the determined job sequence and available capacities.

Besides calculating the lead time based on a previous sequencing the lead time can also be predicted direclty. In the past, a large number of approaches have been established for predicting lead times. Cheng and Gupta [20] performed an early literature review and investigated relationships between due dates, dispatching rules and lead times in static and dynamic job shops. Their focus was on a particular segment of scheduling research in which the due date assignment is of primary interest. They reviewed methods for calculating a job's due date based on a given job starting time and a predicted flow allowance, which is equal to a lead time. They differentiated between exogenous and endogenous methods [20]. In exogeneous methods, a job's lead time is set as a fixed and given attribute of a job before entering the production system. Examples are Constant (CON), where all jobs are given exactly the same lead time, and Random (RAN), where the lead time for a job is randomly assigned. In endogenous methods the job's lead time is predicted as the job is entering the production system considering job characteristics and shop status information. Examples for considering job characteristics are Total Work (TWK), where the lead time is predicted based on a jobs processing time and Number of Operations (NOP), where lead times are predicted based on the number of operations to be performed on the job. Examples for considering shop information are Jobs in Queue (JIQ), where the lead times are predicted based on the number auf jobs in a queue of the production system or Work in Queue (WIQ), which is similar to JIQ but utilizes the processing times instead of the number of jobs. Comparing the predicted lead times of exogenous and endogenous methods, the endogenous methods are generally superior [52]. Combining job and shop status has proven to be more effective [53], [54]. Further details on the methods and its performance are given by [53], [55], [56]. All approaches reviewed by Cheng and Gupta have in common that they use analytical techniques for the prediction of lead times that are typically found in in the field of OR. One of the most fundamental analytical approaches is Little's Law, which determines the average number of items in a queue of a stationary system based on the average arrival rate of items to that system and the average waiting time [57]. With the increasing development of ML, new data analytics methods for directly predicting lead times have emerged. In their study, Burggräf *et al.* [58] have highlighted that scheduling and the prediction of lead times was traditionally one of the key research topics for ML in production. Öztürk *et al.* [21] for example used a regression tree to predict lead times considering several attributes from shop status and job characteristics which outperforms the traditional TWK, Alenezi *et al.* [59] utilize a support vector machine and Wang and Jiang [60] develop a deep neural network.

Concluding, there are two possible approaches to determine lead times: Firstly, indirect based on scheduling and approximating waiting times considering performance curves and secondly, by performing a direct prediction of lead times based on specific rules or historical data. To the best of our knowledge, no review article analyses the current status of available approaches for the direct prediction of lead times coming from both areas ML and OR. In the recent works the relevant state of the art is summarized. However, no systematic procedure is apparent.

III. CONDUCTING THE REVIEW

A systematic review is a type of literature review based on systematic methods to reproducibly answer a specific research question by identifying all relevant studies and synthesizing findings qualitatively or quantitatively [61], [62]. It is designed to provide a complete, exhaustive, transparent and replicable summary of current stare of the art [63].

The methodology used in this review is following the procedure model of Vom Brocke *et al.* which consists of five steps: (I) definition of review scope, (II) conceptualization of topic, (III) literature search, (IV) literature analysis and synthesis as well as (V) deduction of research agenda [24].

It is widely accepted within review theory [64] and not least it grants freedom of action for domain and process specific examinations.

A. DEFINITON OF REVIEW SCOPE

The review scope was characterized according to the taxonomy of literature reviews by Cooper [65] (cf. Fig.1). The research focus is on research outcomes and applications with the goal of knowledge integration using a conceptual structure. From a neutral perspective the review addresses specialized scholars considering all the relevant sources, but describing only a sample. So, the coverage is classified as exhaustive and selective.

	Characteristic	Categories			
1	Focus	Research outcomes	Research methods	Theories	Applications
2	Goal	Integration	Criticism	Central issue	
3	Organization	Historical	Conceptual	Methodological	
4	Perspective	Neutral representation	Espousal of position		
5	Audience	Specialized scholars	General scholars	Practitioners / politicians	General public
6	Coverage	Exhaustive	Exhaustive and selective	Representative	Central / pivotal

FIGURE 1. Taxonomy of literature reviews following Cooper [65].

The organization of prior research identifies a relationship between the considered data, algorithms and predicted lead times and serves to highlight the high multitude of possibilities to predict lead times (cf. Section II). The aim of this systematic literature review is consequently first to aggregate the latest state of the art for the prediction of lead times including used data and algorithms and second to develop an integrative framework for the further analysis and synthesis of the relevant publication. Here, we want to focus on the direct prediction of lead times only and leave out approaches focusing on scheduling, queueing theory or performance curves since these approaches rely on the determination of waiting or interoperation times and do not fully consider potential disruptions occurring during production process itself leading to an extension of the processing time. A direct prediction of lead times can include these disruptions as it considers always the complete lead time consisting of waiting and processing time instead of only a part of it. Furthermore, a direct prediction of lead times based on historical data is gaining new potentials with the enormous improvements in data acquisition combined with the upcoming research area of ML providing new data analytics methods. Accordingly, this leads to the following research questions:

- RQ1: Which is the current state of the art in directly predicting lead times for manufacturing companies and which data and methods or algorithms are used?
- RQ2: How does the existing literature contribute to future research on direct lead time prediction?

B. CONCEPTUALIZATION OF THE TOPIC

Before conducting a review to synthesize knowledge from literature, according to the authors in [66] it is strongly recommended to acquire a priori knowledge about the topic, to identify potential areas where synthetized knowledge may be needed and to properly conduct the review. Based on the explanations and definitions provided in Section I and II and reviewing over 40 publications with an explorative approach we identified concepts most relevant to our field of observation and mapped them to the topic. So, it is ensured to use a wide range of key terms that are locatable within literature. As a result, we generated a concept map [67] for lead time prediction (cf. Fig. 2). The concept map lists all relevant synonyms for the further literature search.

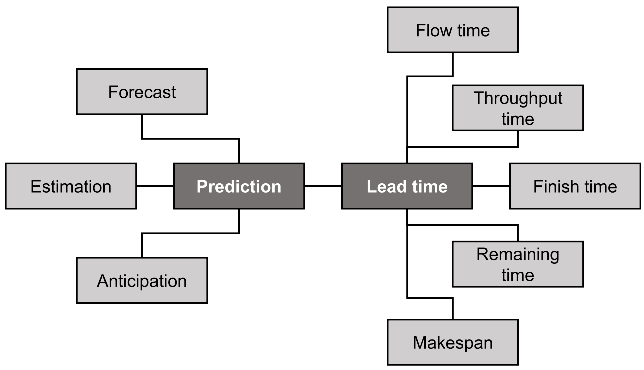

FIGURE 2. Conceptualization map for lead time prediction according to the procedure of Rowley and Slack [67].

C. LITERATURE SEARCH

Based on the concept map the search terms were transferred into the following search string including Boolean operators and wildcards: (“predict*” OR “forecast*” OR “estimat*” OR “anticipat*”) AND (“throughput time*” OR “flow time*” OR “remaining time*” OR “finish time*” OR “makespan*”). We used AND operators to exclude publications focusing on a single area of the search field only in order to increase the thematic relevance. The search strategy was enhanced by the elements of the STARLITE mnemonic framework [68]: We focus on journal articles and conference proceedings published in English between 1960 and 2019 in the electronic databases IEEE Xplore, Web of Science, EBSCO, ScienceDirect, and SpringerLink.

The application of the search string to the metadata title, abstract and key words, considering the additional criteria from the STARLITE mnemonic, identified a total of 18,697 publications in all databases. Afterwards, we followed the procedure given in the PRISMA flow diagram according to Moher *et al.* [25] to consider relevant publications only. The procedure recommends to remove duplicates followed by a literature screening and detailed assessment of relevance based on the full text. The following quality criteria were defined for the screening and the detailed assessment:

- QC1: Addresses the domain of manufacturing.

- QC2: Publications are focusing on the prediction, estimation or forecast of lead times or parts of lead times.
- QC3: Publications focusing on algorithm development rather than methodological / domain specific applications are excluded.
- QC4: Publications focusing on job shop sequencing, queueing theory or performance curves rather than on a direct prediction of lead times are excluded.

The total number of publications included 3,786 duplicates. In the remaining 14,911 publications we identified various publications that do not comply with the applied search criteria. It turned out that some databases apply the search string to the full text in addition to title, abstract and key words. To comply with the search criteria, we additionally applied the search string to title, abstract and key words manually. After removing duplicates and the manual application of the search string a total number of 4,004 publications remain for the screening phase.

For screening the publications, we utilized a clustering approach by Weißer *et al.* [26] based on Natural Language Processing (NLP). Starting with a tokenization (word separation), the removal of stop words (stop words do not contain relevant information) and a TFIDF vectorization, a k-Means clustering is performed and the most relevant words (topwords) per cluster are identified. The topwords characterize each cluster and indicate its thematical relevance. We used title, abstract and key words without the search string as base for the clustering. Due to the resulting big text corpus we performed a dimensionality reduction by latent semantic analysis (LSA), as proposed by [69] and [70], to achieve better clustering results. Furthermore, to fully comply with the defined quality criteria, we did not solely rely on the topwords for excluding irrelevant clusters as proposed by Weißer *et al.* [26]. Based on the assumption of homogenous clusters, we have additionally taken a representative but random sample of publications of each cluster and read their full texts. Only if all of the publications in the sample do not match the quality criteria QC1-4, the whole cluster is assessed as irrelevant.

For the 4,004 remaining publications a clustering with ten clusters was performed and the topwords were extracted (cf. Table 1). The number of clusters was identified by applying the elbow method. Based on the analyzed samples and the topwords the clusters three, five and nine are assessed as relevant with a total number of 857 publications. Following the clustering, we analyzed the abstracts of all publications with respect to QC1-4. The remaining 367 publications were then further analyzed by reading the full text resulting in 39 relevant publications. With the relevant 39 publications we performed a forward and backward search, to identify models, theories and constructs that may not have been covered by the database search terms [71]. Thus, additional three relevant publications were identified, leading to the final data set of 42 publications for further analysis and synthesis in phase IV of the approach of Vom Brocke *et al.* [24].

TABLE 1. Clusters with topwords, cluster size and assessed relevance.

Cluster No.	Top Words	Cluster Size	Relevance
1	Model, data, based, system, using	1,349	Not relevant
2	Model, series, river, neural, network	463	Not relevant
3	Manufacturing, production, process, product, system	376	Relevant
4	Ensemble, precipitation, skill, model, weather	448	Not relevant
5	Abstract, copyright, may, users, abridged	193	Relevant
6	Flood, rainfall, model, river, warning	180	Not relevant
7	Traffic, series, network, model, term	68	Not relevant
8	Skill, enso, ocean, climate, sst	457	Not relevant
9	Inventory, demand, supply, chain, bullwhip	288	Relevant
10	Cancer, screening, patient, breast, survival	182	Not relevant

IV. RESULTS

The intention of this theoretical overview is to bring relevant concepts into a superordinate structure, to map the contribution of literature to our problem statements, and to provide starting points for future research [64]. Therefore, publications with different concepts are analyzed and synthesized considering how they contribute to our research questions (cf. section III A). Before performing the analysis and synthesis in section IV B we define a framework as a base in section IVA.

A. DEFINITION OF THE FRAMEWORK

Setting up a framework is a common approach to structure literature as recommended by [72] and [73]. Our framework is separated in the following three dimensions (cf. Fig. 3):

1) DATA CLASS

As a core differentiation we already mentioned the data class (cf. Section I and II). Edward Cronjäger [74] divides the recorded data of manufacturing companies into order data, machine data, employee data and material data. *Order data* define all specific dates, times and quantities of individual orders. In our framework we will further include operation specific dates, times and quantities in the order data since an operation is part of an order. *Machine data* define all characteristics of the machines that are used to process orders such as the machine ID, information about the tools or fault messages. *Employee data* contain information about the operators of the machines. This information is for example, the presence of employees or specific data such as the age or performance of an employee. *Material data* define all product characteristics of the product to be manufactured such as geometric specification, weights or the material itself. In addition, we identified publications that utilize information about the system status to directly

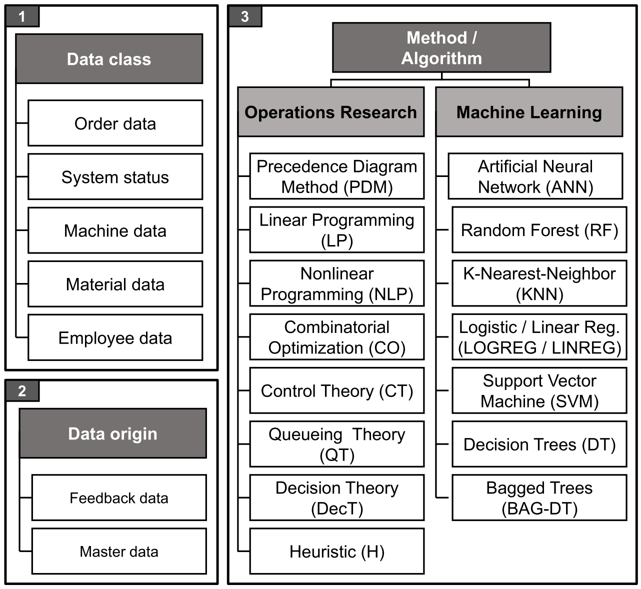

FIGURE 3. Dimensions of developed framework.

predict lead times such as the stock level in intermediate storage or the capacity utilization of the machines (compare [18], [75], [76]). We have therefore added the *system status* as a fifth data class.

2) DATA ORIGIN

The analysis of the relevant publications showed that data used to directly predict lead times have various origins such as a planning data, a simulation or feedback data from a real production. For example, Govind and Roeder [77] generate input data for a direct prediction of lead times from a simulation. Grabenstetter and Usher [78] consider historical data from a real production environment to directly predict lead times. Based on that we divided the second dimension of the framework *data origin* into the categories *feedback data* and *master data*. Feedback data describes data that was recorded in a real production environment during the production process. Master data are data used for planning without real feedback from a production environment. We included data that was generated from a simulation or whose origin is not further described within a publication in the category master data.

3) METHOD/ALGORITHM

Lead times can be predicted directly based on methods or algorithms from both research areas OR and ML (cf. section II). Since OR and ML are already established since many years, several overviews of these methods and algorithms are available in literature. For our framework we consider the basic works by Zimmermann and Stache [79] and Feichtinger and Hartl [80] to subdivide OR. They differentiate between Precedence Diagram Method (PDM), Linear Programming (LP), Nonlinear Programming (NLP), Combinatorial Optimization (CO), Control Theory (CT), Queuing Theory (QT), Decision Theory (DecT) and Heuristics (H). To subdivide ML we utilize the often-cited overview about supervised learning algorithms by Caruana and Niculescu-Mizil [81] to subdivide ML. They differentiate between Artificial Neural Networks (ANN), Logistic Regression (LOGREG), K-Nearest-Neighbor (KNN), Support Vector Machines (SVM), Random Forest (RF), Decision Trees (DT) and Bagged Trees (BAG-DT). In addition to that we extended the field of Logistic Regression by Linear Regression (LINREG).

B. ANALYSIS AND SYNTHESIS

Based on these defined dimensions we classified all publications accordingly and performed a descriptive analysis to identify the current state of the art in directly predicting lead times in manufacturing companies and in the used data classes and methods or algorithms (cf. RQ1). Additionally, we further deducted how the literature contributes to further research (cf. RQ2). A good overview of the development of a research area is given by the chronological development of the publications (cf. Fig. 4). Given the 42 identified publications, Fig. 4 shows an increasing number of publications focusing the direct prediction of lead times over time. Before the year 2000, we identified only three publications focusing the direct prediction of lead times, while the remaining 39 publications appeared after that date. Thus, a trend can be seen towards an increasing interest in the research area of directly predicting lead times.

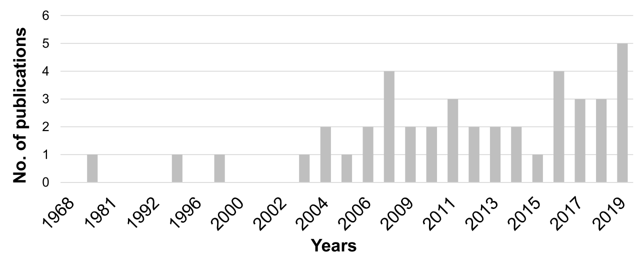

FIGURE 4. Chronological development of publications.

Next, we analyzed the dimensions of the framework (cf. Fig. 3) individually and subsequently combined two or more dimensions to identify common approaches and implications for further research. The following paragraphs are structured according to the considered dimensions.

1) DATA CLASS

Looking at the data classes, it was noticeable that with a share of 95% of all publications, almost every author takes order data into account to directly predict lead times (cf. Fig. 5). Jain and Raj *et al.* [82], Berlec *et al.* [83] or Gramdi [84] for example use order data such as start and end dates of orders or order-specific processing times for the prediction of lead times. Therefore, order data are relevant for the direct prediction of lead times. Furthermore, the system status with a share of 62 % of all publications is often used for direct

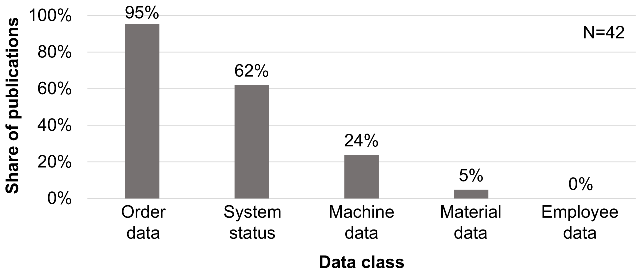

FIGURE 5. Overview of used data classes.

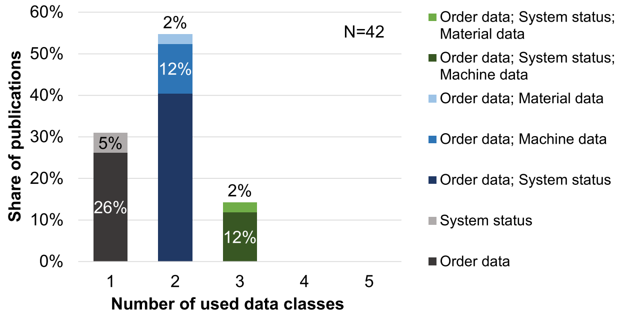

FIGURE 6. Overview of quantity of used data classes.

predicting lead times. In contrast, machine and material data with a share of 21 % and 5 % respectively are used relatively rarely and employee data with a share of 0 % have not been used for directly predicting lead times at all. One possible explanation for not using employee data could be, that due to data privacy restrictions employee data is not available for analysis. Furthermore, material data is commonly stored in the CAD-system, drawings or in the material master data in the ERP system, which might not be directly linked to the order data or system status. Gyulai *et al.* [85] and Karagolan and Karademir [23] are the only authors who use material data such as dimensions or specifications of the product for directly predicting lead times. Machine data are used by Weng and Fujimura [86], for example, in the form of the machine ID. Lingitz *et al.* [22] use so-called 'equipment data' containing information about machines and tools to predict lead times without describing these data in more detail. The small proportion of machine, material and employee data suggests that either there is no or only a small relation between lead times and these data classes, or the connection has a low research interest in previous research. Since products in an engineer to order environment are designed individually and therefore the materials differ greatly in their characteristics, we see a high potential for further research considering material data as an input for directly predicting lead times.

Analyzing the number of used data classes in more detail reveals that 86 % of all publications use two or less different data classes for directly predicting lead times (cf. Fig. 6). In case of using one data class only the majority of publications are considering order data like [87] or rarely system data like [75]. Machine and material data are not used solely. In case of using two or more data classes, order data is always included. With 40 % the majority combines order data and the system status like [21]. Only a minority of 14% of all publications is using three data classes for directly predicting lead times combining order and system status with either machine data like [88] or with material data like [85]. Furthermore, it can be seen that in none of the publications more than three data classes are used. Since different combinations of three data classes have already been successfully demonstrated, namely order data + system status + machine data and order data + system status + material data, it is also conceivable that a combination of all four data classes order data, system status, material data and machine data can provide good results in directly predicting lead times. Therefore, we see a high potential for further research in using three and more data classes for the direct prediction of lead times. Future researchers could, for example, develop a model using ML or OR in which, in addition to the system status and order data, they also use the material data to directly predict lead times.

2) METHOD/ALGORITHM

Over time, the number of publications with ML increases continuously, whereas the number of publications with OR remains almost constant. In the case of ML 18 of the 23 publications were published after 2010. Therefore, the emerging trend of ML can also be seen in the research field of directly predicting lead times. In total the comparison of the research areas ML and OR shows with 55% only a slight majority in the area of ML compared to OR with 45 % (cf. Fig. 7a). Looking at the ML methods and algorithms used in detail reveals that ANN (43% of all ML-publications), LINREG/LOGREG (30 %), DT (26 %) and RF (22 %) were primarily used (cf. Fig. 7b). Furthermore, we identified authors using more than one approach within a publication to directly predict lead times. For example, Asadzadeh *et al.* [19] combine two approaches (ANN and LINREG) in one model, the authors in [89], [90] compare two approaches (ANN and DT) and the authors in [91] use a linear regressor (LINREG) to predict lead times. Schuh *et al.* [33] present a three-step procedure with a DT regressor for predicting order-specific interoperation times. Gyulai *et al.* [85] compare OR (e.g. Little's Law) and ML approaches and conclude that ML provides more precise results than OR. In their proposed model, a random forest approach is finally chosen because of a higher model accuracy for the available input data. Furthermore, a digital twin of the production environment is created to provide the ML model with quasi real production data for predicting lead times. Looking on the used OR methods and algorithms in detail reveals that Combinatorial Optimization (26 % of all OR publications), Heuristics and Queuing Theory (both 21 %) were primarily used (cf. Fig. 7c). For example, Berlec and Starbek [17] use Combinatorial Optimization by setting up the lead times per operation of different orders in one

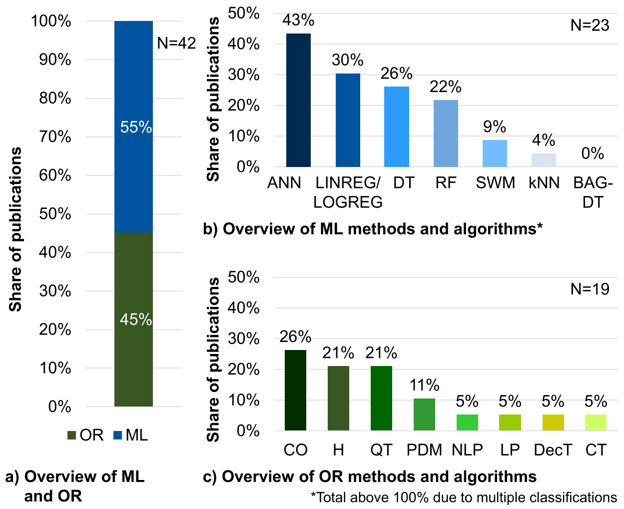

FIGURE 7. Overview of methods and algorithms used in ML and OR.

vector per workstation and then randomly select and combine individual elements of the vectors to determine the total lead time of the order following a given processing sequence. In conclusion, in both research areas ML and OR specific methods and algorithms are used more frequently for directly predicting lead times while others like SVM or Control Theory are used rarely.

3) DATA CLASS AND METHOD/ALGORITHM

Combining the data class with the used method and algorithms reveals that order data is used in combination with all methods and algorithms (cf. Fig. 8). This deducts a general relevance of order data for directly predicting lead times, regardless of the method or algorithm used. The system status is used in 12 of 13 methods and algorithms for directly predicting lead times and can therefore be classified as generally relevant as well. Only decision trees are not used in combination with the system status. Looking at the method of decision tree, we do not see any methodological reason for not using decision trees in combination with the system status. Considering machine data, it is noticeable that in more than 50 % of cases combinatorial optimization (e.g. [92]) and ANN (e.g. [89]) are used. One possible explanation for this could be, that the information about several machines within the machine data need to be combined according to the corresponding processing sequence which is a typical application for combinatorial optimization and ANN. When using product data, it is noticeable again that only ANN in [23] and Random Forest in [85] are used to predict lead times. This either indicates that material data are not analyzable with other methods and algorithms, material data do not correlate with the directly predicted lead times or that material data has received less attention in prior research. Since there are already approaches with good results using material data for directly predicting lead times, we consider the second option, that material data do not correlate with lead times, as negligible.

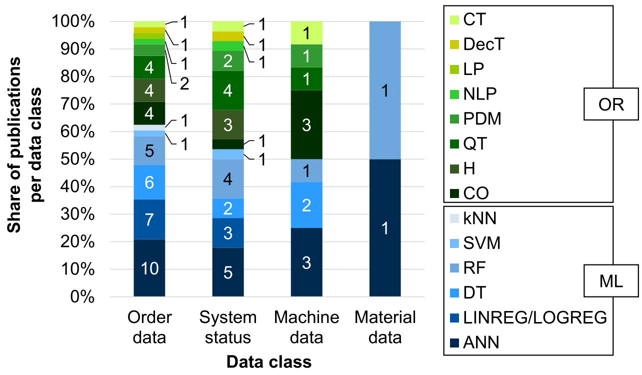

FIGURE 8. Overview of data classes combined with used methods and algorithms.

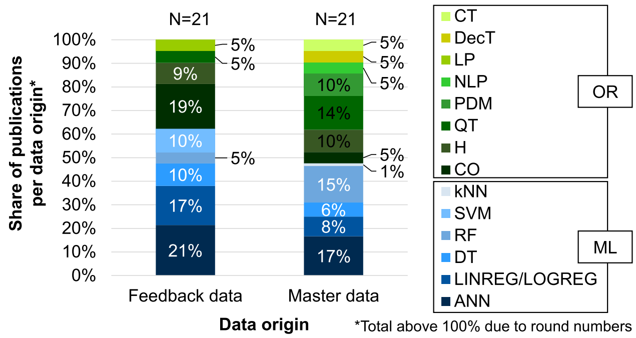

FIGURE 9. Overview of data origin combined with used methods and algorithms.

4) DATA ORIGIN AND METHOD/ALGORITHM

Looking at the data origin only, we recognized an equal distribution of publications between feedback data and master data (cf. Fig. 9). Combining the used methods and algorithms with the data origin enables a more detailed view: Publications considering feedback data as base for directly predicting lead times utilize ML approaches with a share of 63% more frequently than OR. Here, most authors use ANN or LINREG/LOGREG. On the other side, OR approaches based on feedback data are dominated by CO. This leads to the insight that, from the field of ML, ANN and LINREG/LOGREG and, from the field of OR, CO are solid approaches for directly predicting lead times based on feedback data. Karagolan and Karademir [23] for example perform a prediction of lead times using ANN and reach an accuracy up to 98.54 % comparing the predicted lead times with the real lead times. In publications considering master data instead of feedback data with a share of 55 % OR is used more frequently than ML. In detail ANN, RF, and QT are utilized almost equally. In conclusion, ML dominates the direct prediction of lead times based on feedback data whereas OR dominates the direct prediction of lead times based on master data. One possible explanation for this could be, that feedback data contain a larger amount of data sets which are predestined for ML, whereas the creation of master data is a manual and thus, expensive process which is suitable for OR.

5) DATA CLASS AND DATA ORIGIN

Analyzing the combination of data class and data origin reveals a trend in the considered data origin depending on the used number of data classes (cf. Fig. 10). If only one data class is used for the direct prediction of lead times, almost 70 % of the corresponding publications consider feedback data. If three data groups are used, the proportion of publications considering feedback data reduces to only 33 %. This shows that the proportion of publications using feedback data decreases as the number of considered data groups increases. Since the number of data classes is an indicator for the model complexity, the identified trend implicates a decreasing use of feedback data for a direct prediction of the lead times with an increasing model complexity. Therefore, we see a high potential for further research focusing on higher model complexity with a larger number of data classes combined with feedback data.

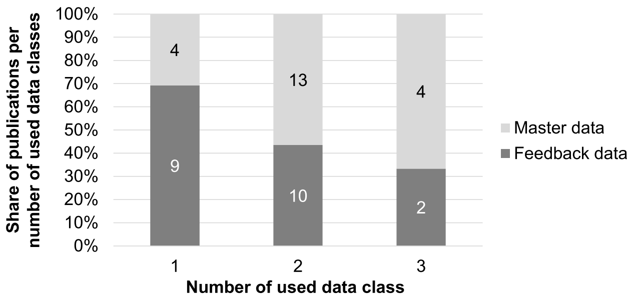

FIGURE 10. Overview of data origin combined with the quantity of used data classes.

The performed analysis and synthesis of the existing publications differentiated by the dimensions of our framework provided an extensive and detailed answer on RQ1. We identified data classes, data origins as well as methods and algorithms that are mainly used in the body of literature. We also identified implications for further research which we will summarize in the following section in detail.

V. IMPLICATIONS FOR FURTHER RESEARCH

As already stated, all of the publications found in literature focusing the direct prediction of lead times could be classified with our developed framework (cf. section IV). By performing a descriptive analysis, we were able to identify common approaches that were used by the majority of researchers. Furthermore, we identified white spots and noticeable trends that indicate the need for further research (RQ2). Looking at the considered data classes we identified material data as an almost complete white spot in the research area of directly predicting lead times. Only few researchers present results in directly predicting lead times considering material data. With our review focus on the engineer to order production, where products often consist of a large number of components that are designed individually to achieve a tailor-made solution for the respective customer [7], [8], we see a high potential for further research considering material data in the direct prediction of lead times. Furthermore, we identified only few publications considering three or more data classes. Since disruptions in production systems are widely spread over various root causes [13], each of the different data classes might contain relevant information that correlate with the lead time. Additionally, we identified a decreasing number of publications using feedback data, if the number of used data classes increases. Feedback data contain the real information about the production system. Consequently, we see a high potential for further research considering three or more data classes for directly predicting lead times based on feedback data from a real production environment. Those few researchers focusing material data as input for directly predicting lead times only used ANN and RF so far. Thus, analyzing the performance of other methods and algorithms for directly predicting lead times based on material data is another research potential.

VI. CONCLUSION

In this article an SLR was conducted to determine the state of the art of directly predicting lead times with focus on engineer to order production. The lead time is one of the key factors for meeting customer requirements and predicting lead times can help to identify potential deviations from agreed delivery dates at an early production stage. Based on the identified deviations, the responsible person for production can then set counter measures to meet the due dates. The aim of this study was therefore to identify relevant data classes as well as methods and algorithms from the field of OR and ML used for directly predicting lead times within the body of literature. We conducted our research according to the SLR procedure model according to Vom Brocke *et al.* [24] and integrated dedicated SLR concepts from other authors. Within the phase of literature search we identified a total of 18,697 publications, of which 42 publications were further considered in the core of our analysis. For the purpose of the selection of publications we utilized a clustering approach by Weißer *et al.* [26] to allow a more efficient and target oriented scanning and filtering. In the subsequent analysis phase a framework was developed to structure the considered publications followed by a descriptive analysis as the base to identify common approaches within the body of literature and to derive implications for further research.

A direct lead time prediction based on ML is a research field with increasing relevance. Concerning the considered data classes for the direct prediction, two data classes, namely order data and system status, are mainly used. Noticeable was the low usage of material data and feedback data in more complex models. From the field of ML, ANN and Regression models show high potential for further research in complex models considering material data and feedback data. With the performed detailed analysis all research questions stated in Section III A were eventually answered.

We believe this study has both theoretical and practical implications. It provides academics with an overview of the state of the art of approaches for the direct prediction of lead times and indicates potential for further research.

Furthermore, it can offer practical guidance to practitioners in selecting data classes as well as methods and algorithms to implement an approach for directly predicting lead times in their production environment.

REFERENCES

[1] C. Reuter and F. Brambring, "Improving data consistency in production control," *Procedia CIRP*, vol. 41, pp. 51–56, Jan. 2016, doi: 10.1016/j.procir.2015.12.116.

[2] G. Schuh and V. Stich, *Logistikmanagement Handbuch Produktion und Management 6*, 2nd ed. Berlin, Germany: Springer, 2013.

[3] S. Lee, Y. J. Kim, T. Cheong, and S. H. Yoo, "Effects of yield and lead-time uncertainty on retailer-managed and vendor-managed inventory management," *IEEE Access*, vol. 7, pp. 176051–176064, 2019, doi: 10.1109/ACCESS.2019.2957595.

[4] W. H. M. Raaymakers and A. J. M. M. Weijters, "Makespan estimation in batch process industries: A comparison between regression analysis and neural networks," *Eur. J. Oper. Res.*, vol. 145, no. 1, pp. 14–30, 2003, doi: 10.1016/S0377-2217(02)00173-X.

[5] H. Lödding, *Verfahren der Fertigungssteuerung: Grundlagen, Beschreibung, Konfiguration*, 2nd ed. Berlin, Germany: Springer-Verlag, 2008.

[6] G. Schuh and V. Stich, Eds., *Produktionsplanung und-Steuerung*, 4th ed. Berlin, Germany: Springer Vieweg, 2012.

[7] P. Fredriksson and L.-E. Gadde, "Flexibility and rigidity in customization and build-to-order production," *Ind. Marketing Manage.*, vol. 34, no. 7, pp. 695–705, Oct. 2005, doi: 10.1016/j.indmarman.2005.05.010.

[8] A. E. Coronado, A. C. Lyons, D. F. Kehoe, and J. Coleman, "Enabling mass customization: Extending build-to-order concepts to supply chains," *Prod. Planning Control*, vol. 15, no. 4, pp. 398–411, Jun. 2004, doi: 10.1080/0953728042000238809.

[9] A. Kampker, J. Wagner, P. Burggräf, and Y. Bäumers, "Criticality-focused, pre-emptive disruption management in low-volume assembly," in *Proc. Abstract 23rd Int. Conf. Prod. Res. (ICPR) Oper. Excellence Towards Sustain. Develop. Goals (SDG) Through Ind.*, Manila, Philippines, vol. 4, 2015, pp. 2–5.

[10] G. Schuh, T. Potente, and T. Jasinski, "Decentralized, Market-Driven coordination mechanism based on the monetary value of in time deliveries," in *Proc. Global Bus. Res.*, Kathmandu, India, 2013, pp. 1–13.

[11] R. J. Abumaizar and J. A. Svestka, "Rescheduling job shops under random disruptions," *Int. J. Prod. Res.*, vol. 35, no. 7, pp. 2065–2082, Jul. 1997, doi: 10.1080/002075497195074.

[12] P. Burggräf, J. Wagner, K. Lück, and T. Adlon, "Cost-benefit analysis for disruption prevention in low-volume assembly," *Prod. Eng.*, vol. 11, no. 3, pp. 331–342, Jun. 2017, doi: 10.1007/s11740-017-0735-6.

[13] J. Wagner, P. Burggräf, M. Dannapfel, and C. Fölling, "Assembly disruptions–empirical evidence in the manufacturing industry of Germany, Austria and Switzerland," *Int. Refereed J. Eng. Sci.*, vol. 6, no. 3, pp. 15–25, 2017.

[14] N. Levin and J. Zahavi, "Predictive modeling using segmentation," *J. Interact. Marketing*, vol. 15, no. 2, pp. 2–22, Jan. 2001, doi: 10.1002/dir.1007.

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[16] A. Raddon and B. Grigsby, "Throughput time forecasting model," in *Proc. IEEE/SEMI Adv. Semiconductor Manuf. Conf. Workshop (ASMC)*, Cambridge, MA, USA, Oct. 1997, pp. 430–433.

[17] T. Berlec and M. Starbek, "Forecasting of production order lead time in Sme's," in *Products and Services; from R&D to Final Solutions*, I. Fuerstner, Ed. Rijeka, Croatia: IntechOpen, 2010.

[18] A. Pfeiffer, D. Gyulai, B. Kádár, and L. Monostori, "Manufacturing lead time estimation with the combination of simulation and statistical learning methods," *Procedia CIRP*, vol. 41, pp. 75–80, 2016, doi: 10.1016/j.procir.2015.12.018.

[19] S. M. Asadzadeh, A. Azadeh, and A. Ziaeifar, "A neuro-fuzzy-regression algorithm for improved prediction of manufacturing lead time with machine breakdowns," *Concurrent Eng.*, vol. 19, no. 4, pp. 269–281, Dec. 2011, doi: 10.1177/1063293X11424512.

[20] T. C. E. Cheng and M. C. Gupta, "Survey of scheduling research involving due date determination decisions," *Eur. J. Oper. Res.*, vol. 38, no. 2, pp. 156–166, 1989, doi: 10.1016/0377-2217(89)90100-8.

[21] A. Öztürk, S. Kayalıgil, and N. E. Özdemirel, "Manufacturing lead time estimation using data mining," *Eur. J. Oper. Res.*, vol. 173, no. 2, pp. 683–700, Sep. 2006, doi: 10.1016/j.ejor.2005.03.015.

[22] L. Lingitz, V. Gallina, F. Ansari, D. Gyulai, A. Pfeiffer, W. Sihn, and L. Monostori, "Lead time prediction using machine learning algorithms: A case study by a semiconductor manufacturer," *Procedia CIRP*, vol. 72, pp. 1051–1056, 2018, doi: 10.1016/j.procir.2018.03.148.

[23] A. D. Karaoglan and O. Karademir, "Flow time and product cost estimation by using an artificial neural network (ANN): A case study for transformer orders," *Eng. Economist*, vol. 62, no. 3, pp. 272–292, Jul. 2017, doi: 10.1080/0013791X.2016.1185808.

[24] J. Vom Brocke, A. Simons, B. Niehaves, B. Niehaves, and K. Reimer, "Reconstructing the giant: On the importance of rigour in documenting the literature search process," in *Proc. ECIS*, 2009. [Online]. Available: https://aisel.aisnet.org/ecis2009/161

[25] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *Int. J. Surgery*, vol. 8, no. 5, pp. 336–341, 2010, doi: 10.1016/j.ijsu.2010.02.007.

[26] T. Weißer, T. Saßmannshausen, D. Ohrndorf, P. Burggräf, and J. Wagner, "A clustering approach for topic filtering within systematic literature reviews," *MethodsX*, vol. 7, 2020, Art. no. 100831, doi: 10.1016/j.mex.2020.100831.

[27] *Business Dictionary*. Accessed: May 15, 2020. [Online]. Available: http://www.businessdictionary.com/

[28] S. O'Shea, Ed., *Cambridge Business English Dictionary*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[29] A. Gunasekaran, C. Patel, and E. Tirtiroglu, "Performance measures and metrics in a supply chain environment," *Int. J. Operations Prod. Manage.*, vol. 21, nos. 1–2, pp. 71–87, Jan. 2001, doi: 10.1108/01443570110358468.

[30] H.-P. Wiendahl, *Load-Oriented Manufacturing Control*. Berlin, Germany: Springer, 1995.

[31] P. Nyhuis and H.-P. Wiendahl, *Logistische Kennlinien: Grundlagen, Werkzeuge und Anwendungen*, 3rd ed. Berlin, Germany: Springer, 2012.

[32] S. M. Meerkov and C.-B. Yan, "Production lead time in serial lines: Evaluation, analysis, and control," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 2, pp. 663–675, Apr. 2016, doi: 10.1109/TASE.2014.2365108.

[33] G. Schuh, J.-P. Prote, F. Sauermann, and B. Franzkoch, "Databased prediction of order-specific transition times," *CIRP Ann.*, vol. 68, no. 1, pp. 467–470, 2019, doi: 10.1016/j.cirp.2019.03.008.

[34] K. R. Baker and D. Trietsch, *Principles of Sequencing and Scheduling*. Hoboken, NJ, USA: Wiley, 2009.

[35] G. E. Vieira, J. W. Herrmann, and E. Lin, "Rescheduling manufacturing systems: A framework of strategies, policies, and methods," *J. Scheduling*, vol. 6, no. 1, pp. 39–62, 2003, doi: 10.1023/A:1022235519958.

[36] O. Kocatepe, "An application framework for scheduling optimization problems," in *Proc. IEEE 8th Int. Conf. Appl. Inf. Commun. Technol. (AICT)*, Astana, Kazakhstan, Oct. 2014, pp. 1–4.

[37] A. Schömig, D. Eichhorn, and G. Obermaier, "Über verschiedene Ansätze zur Ermittlung von Betriebskennlinien—Eine Anwendungsstudie aus der Halbleiterindustrie," in *Operations Research Proceedings, v.2006, Operations Research Proceedings 2006: Selected Papers of the Annual International Conference of the German Operations Research Society (GOR), Jointly Organized with the Austrian Society of Operations Research (ÖGOR) and the Swiss Society of Operations Research (SVOR)*, 1st ed., K.-H. Waldmann and U. M. Stocker, Eds. Berlin, Germany: Springer-Verlag, 2007, pp. 467–472.

[38] C. Engelhardt, *Betriebskennlinien: Produktivität Steigern in der Fertigung*. München, Germany: Hanser, 2000.

[39] S. S. Aurand and P. J. Miller, "The operating curve: A method to measure and benchmark manufacturing line productivity," in *Proc. IEEE/SEMI Adv. Semiconductor Manuf. Conf. Workshop (ASMC)*, Cambridge, MA, USA, Oct. 1997, pp. 391–397.

[40] J. Fowler and J. Robinson, "Measurement and improvement of manufacturing capacities (MIMAC): Final report," SEMATECH, Austin, TX, USA, Tech. Rep. 95062861A-TR, 1995.

[41] W. Bechte, *Steuerung der Durchlaufzeit Durch Belastungsorientierte Auftragsfreigabe bei Werkstattfertigung:(Rckentitel: Belastungsorientierte Auftragsfreigabe)*. Düsseldorf, Germany: VDI-Verlag, 1984.

[42] J. F. C. Kingman, "The single server queue in heavy traffic," *Math. Proc. Cambridge Phil. Soc.*, vol. 57, no. 4, pp. 902–904, Oct. 1961, doi: 10.1017/S0305004100036094.

[43] W. J. Hopp and M. L. Spearman, *Factory Physics*, 3rd ed. Long Grove, IL, USA: Waveland Press, 2011.

[44] L. Sattler, "Using queueing curve approximations in a fab to determine productivity improvements," in *Proc. IEEE/SEMI Adv. Semiconductor Manuf. Conf. Workshop Theme-Innov. Approaches Growth Semiconductor Ind. (ASMC)*, Cambridge, MA, USA, Nov. 1996, pp. 140–145.

[45] F. G. Boebel and O. Ruelle, "Cycle time reduction program at ACL," in *Proc. IEEE/SEMI Adv. Semiconductor Manuf. Conf. Workshop. Theme-Innov. Approaches Growth Semiconductor Industry (ASMC)*, Cambridge, MA, USA, Nov. 1996, pp. 165–168.

[46] D. W. Collins, K. Williams, and F. C. Hoppensteadt, "Implementation of minimum inventory variability scheduling 1-Step ahead Policy(R) in a large semiconductor manufacturing facility," in *Proc. IEEE 6th Int. Conf. Emerg. Technol. Factory Autom. (EFTA)*, Los Angeles, CA, USA, Sep. 1997, pp. 497–504.

[47] O. Ruelle, "Continuous flow manufacturing: The ultimate theory of constraints," in *Proc. IEEE/SEMI Adv. Semiconductor Manuf. Conf. Workshop (ASMC)*, Cambridge, MA, USA, Oct. 1997, pp. 216–221.

[48] O. Rose, "The shortest processing time first (SPTF) dispatch rule and some variants in semiconductor manufacturing," in *Proc. Winter Simulation Conf. Crystal Gateway Marriott*, Arlington, VA, USA, Dec. 2001, pp. 1220–1224.

[49] K. Wu and L. McGinnis, "Performance evaluation for general queueing networks in manufacturing systems: Characterizing the trade-off between queue time and utilization," *Eur. J. Oper. Res.*, vol. 221, no. 2, pp. 328–339, Sep. 2012, doi: 10.1016/j.ejor.2012.03.019.

[50] V. J. Leon, S. D. Wu, and R. H. Storer, "Robustness measures and robust scheduling for job shops," *IIE Trans.*, vol. 26, no. 5, pp. 32–43, 1994, doi: 10.1080/07408179408966626.

[51] S. Tadayonirad, H. Seidgar, H. Fazlollahtabar, and R. Shafaei, "Robust scheduling in two-stage assembly flow shop problem with random machine breakdowns: Integrated meta-heuristic algorithms and simulation approach," *ASSEM Autom.*, vol. 39, no. 5, pp. 944–962, Nov. 2019, doi: 10.1108/AA-10-2018-0165.

[52] R. W. Conway, "Priority dispatching and job lateness in a job shop," *J. Ind. Eng.*, no. 16, pp. 228–237, 1965.

[53] J. K. Weeks, "A simulation study of predictable due-dates," *Manage. Sci.*, vol. 25, no. 4, pp. 363–373, Apr. 1979, doi: 10.1287/mnsc.25.4.363.

[54] J. W. M. Bertrand, "The use of workload information to control job lateness in controlled and uncontrolled release production systems," *J. Oper. Manage.*, vol. 3, no. 2, pp. 79–92, 1983, doi: 10.1016/0272-6963(83)90009-8.

[55] G. L. Ragatz and V. A. Mabert, "A simulation analysis of due date assignment rules," *J. Oper. Manage.*, vol. 5, no. 1, pp. 27–39, 1984, doi: 10.1016/0272-6963(84)90005-6.

[56] S. Eilon and I. G. Chowdhury, "Due dates in job shop scheduling," *Int. J. Prod. Res.*, vol. 14, no. 2, pp. 223–237, Mar. 1976, doi: 10.1080/00207547608956596.

[57] D. C. J. Little, "A proof for the queuing formula: $L = \lambda W$," *Oper. Res.*, vol. 9, no. 3, pp. 383–387, 1961.

[58] P. Burggraf, J. Wagner, and B. Koke, "Artificial intelligence in production management: A review of the current state of affairs and research trends in academia," in *Proc. Int. Conf. Inf. Manage. Process. (ICIMP)*, London, U.K., Jan. 2018, pp. 82–88.

[59] A. Alenezi, S. A. Moses, and T. B. Trafalis, "Real-time prediction of order flowtimes using support vector regression," *Comput. Oper. Res.*, vol. 35, no. 11, pp. 3489–3503, Nov. 2008, doi: 10.1016/j.cor.2007.01.026.

[60] C. Wang and P. Jiang, "Deep neural networks based order completion time prediction by using real-time job shop RFID data," *J. Intell. Manuf.*, vol. 30, no. 3, pp. 1303–1318, Mar. 2019, doi: 10.1007/s10845-017-1325-3.

[61] R. Armstrong, B. J. Hall, J. Doyle, and E. Waters, "'Scoping the scope' of a cochrane review," *J. Public Health*, vol. 33, no. 1, pp. 147–150, Mar. 2011, doi: 10.1093/pubmed/fdr015.

[62] J. Higgins, J. Thomas, and J. Chandler, *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester, U.K.: Wiley, 2019.

[63] A. P. Siddaway, A. M. Wood, and L. V. Hedges, "How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses," *Annu. Rev. Psychol.*, vol. 70, no. 1, pp. 747–770, Jan. 2019, doi: 10.1146/annurev-psych-010418-102803.

[64] G. Paré, M.-C. Trudel, M. Jaana, and S. Kitsiou, "Synthesizing information systems knowledge: A typology of literature reviews," *Inf. Manage.*, vol. 52, no. 2, pp. 183–199, Mar. 2015, doi: 10.1016/j.im.2014.08.008.

[65] H. M. Cooper, "Organizing knowledge syntheses: A taxonomy of literature reviews," *Knowl. Soc.*, vol. 1, no. 1, pp. 104–126, Mar. 1988, doi: 10.1007/BF03177550.

[66] R. J. Torraco, "Writing integrative literature reviews: Guidelines and examples," *Human Resource Develop. Rev.*, vol. 4, no. 3, pp. 356–367, Sep. 2005, doi: 10.1177/1534484305278283.

[67] J. Rowley and F. Slack, "Conducting a literature review," *Manage. Res. News*, vol. 27, no. 6, pp. 31–39, Jun. 2004, doi: 10.1108/01409170410784185.

[68] A. Booth, "'Brimful of STARLITE': Toward standards for reporting literature searches," *J. Med. Library Assoc.*, vol. 94, no. 4, p. 421-e205, 2006.

[69] C. C. Aggarwal and C. Zhai, "A survey of text clustering algorithms," in *Mining Text Data*, 2012nd ed. C. C. Aggarwal and C. Zhai, Eds. Boston, MA, USA: Springer, 2012, pp. 77-128.

[70] S. Adinugroho, Y. A. Sari, M. A. Fauzi, and P. P. Adikara, "Optimizing K-means text document clustering using latent semantic indexing and pillar algorithm," in *Proc. 5th Int. Symp. Comput. Bus. Intell. (ISCBI)*, Dubai, United Arab Emirates, Aug. 2017, pp. 81–85.

[71] Y. Levy and T. J. Ellis, "A systems approach to conduct an effective literature review in support of information systems research," *Informing Sci. Int. J. Emerg. Transdiscipline*, vol. 9, no. 1, pp. 181–212, 2006, doi: 10.28945/479.

[72] P. Salipante, W. Notz, and J. Bigelow, "A matrix approach to literature reviews," *Res. Organizational Behav. Annu. Ser. Anal. Essays Crit. Rev.*, vol. 4, no. 1, pp. 321–348, 1982.

[73] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS Quart.*, vol. 26, no. 2, pp. 13–23, 2002.

[74] L. Cronjäger, Ed., *Bausteine für die Fabrik der Zukunft: Eine Einführung in die Rechnerintegrierte Produktion (CIM)*. Berlin, Germany: Springer, 1994.

[75] G. Ioannou and S. Dimitriou, "Lead time estimation in MRP/ERP for make-to-order manufacturing systems," *Int. J. Prod. Econ.*, vol. 139, no. 2, pp. 551–563, Oct. 2012, doi: 10.1016/j.ijpe.2012.05.029.

[76] H. Sabeti and F. Yang, "Flow-time estimation by synergistically modeling real and simulation data," in *Proc. Winter Simulation Conf. (WSC)*, Las Vegas, NV, USA, Dec. 2017, pp. 3230–3241.

[77] N. Govind and T. Roeder, "Estimating expected completion times with probabilistic job routing," in *Proc. Winter Simulation Conf.*, Monterey, CA, USA, Dec. 2016, pp. 1804–1810.

[78] D. H. Grabenstetter and J. M. Usher, "Determining job complexity in an engineer to order environment for due date estimation using a proposed framework," *Int. J. Prod. Res.*, vol. 51, no. 19, pp. 5728–5740, Oct. 2013, doi: 10.1080/00207543.2013.787169.

[79] W. Zimmermann and U. Stache, *Operations Research: Quantitative Methoden zur Entscheidungsvorbereitung*, 10th ed. Münich, Germany: Oldenbourg, 2001.

[80] G. Feichtinger and R. F. Hartl, *Optimale Kontrolle ökonomischer Prozesse: Anwendungen des Maximumprinzips in den Wirtschaftswissenschaften*. Berlin, Germany: W. de Gruyter, 1986.

[81] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, Pittsburgh, PA, USA, 2006, pp. 161–168.

[82] V. Jain and T. Raj, "An adaptive neuro-fuzzy inference system for makespan estimation of flexible manufacturing system assembly shop: A case study," *Int. J. Syst. Assurance Eng. Manage.*, vol. 9, no. 6, pp. 1302–1314, Dec. 2018, doi: 10.1007/s13198-018-0729-6.

[83] T. Berlec, E. Govekar, J. Grum, P. Poto?nik, and M. Starbek, "Predicting order lead times," *J. Mech. Eng.*, vol. 54, no. 5, pp. 308–321, 2008.

[84] J. Gramdi, "Elaborating actual lead time with both management and execution data," in *Proc. Int. Conf. Comput. Ind. Eng.*, Troyes, France, Jul. 2009, pp. 674–677.

[85] D. Gyulai, A. Pfeiffer, G. Nick, V. Gallina, W. Sihn, and L. Monostori, "Lead time prediction in a flow-shop environment with analytical and machine learning approaches," *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1029–1034, 2018, doi: 10.1016/j.ifacol.2018.08.472.

[86] W. Weng and S. Fujimura, "Estimating job flow times by using an agent-based approach," in *Proc. 5th IIAI Int. Congr. Adv. Appl. Informat. (IIAI-AAI)*, Kumamoto, Japan, Jul. 2016, pp. 975–979.

[87] S. Singh and U. Soni, "Predicting order lead time for just in time production system using various machine learning algorithms: A case study," in *Proc. 9th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence)*, Noida, India, Jan. 2019, pp. 422–425.

[88] W. Fang, Y. Guo, W. Liao, K. Ramani, and S. Huang, "Big data driven jobs remaining time prediction in discrete manufacturing system: A deep learning-based approach," *Int. J. Prod. Res.*, vol. 13, no. 1, pp. 1–16, 2019, doi: 10.1080/00207543.2019.1602744.
[89] I. Tirkel, "Forecasting flow time in semiconductor manufacturing using knowledge discovery in databases," *Int. J. Prod. Res.*, vol. 51, no. 18, pp. 5536–5548, Sep. 2013, doi: 10.1080/00207543.2013.787168.
[90] I. Tirkel, "Cycle time prediction in wafer fabrication line by applying data mining methods," in *Proc. IEEE/SEMI Adv. Semiconductor Manuf. Conf.*, Saratoga Springs, NY, USA, May 2011, pp. 1–5.
[91] A. Aburomman, M. Lama, and A. Bugarin, "A vector-based classification approach for remaining time prediction in business processes," *IEEE Access*, vol. 7, pp. 128198–128212, 2019, doi: 10.1109/ACCESS.2019.2939631.
[92] T. Berlec and M. Starbek, "Predicting order due date," *Arabian J. Sci. Eng.*, vol. 37, no. 6, pp. 1751–1766, Sep. 2012, doi: 10.1007/s13369-012-0279-1.

PETER BURGGRÄF was born in 1980. He received the Dr.Ing. and M.B.A. degrees. He studied mechanical engineering in Aachen and London. He wrote the Ph.D. thesis at the Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, in the field of factory planning. From 2011 until 2017, he was the Chief Engineer of the Chair of Production Engineering at the WZL. Since 2013, he has been the Managing Director of StreetScooter Research GmbH. Since 2017, he has been holding the Chair of International Production Engineering and Management (IPEM), University of Siegen, Germany.

JOHANNES WAGNER was born in 1986. He received the Dr.Ing. and M.B.A. degrees, and the M.Sc. degrees in industrial engineering from RWTH Aachen University and Tsinghua University, Beijing. He wrote the Ph.D. thesis at the Chair of Production Engineering at the Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, in the field of disruption management in low-volume assembly. Since 2017, he has been the Chief Engineer of the Chair of International Production Engineering and Management (IPEM), University of Siegen. He is also the Founder and the CEO of the Smart Demonstration Factory Siegen (SDFS).

BENJAMIN KOKE was born in 1988. He received the M.Sc. degree from the Chalmers University of Technology and the M.Sc. degree from Northumbria University. He is currently pursuing the Ph.D. degree with the Chair of International Production Engineering and Management (IPEM), University of Siegen. He is writing the Ph.D. thesis in the die field of automated factory planning based on artificial intelligence. Since 2017, he has been working at the Chair of IPEM, University of Siegen, and he has been managing the group factory planning for medium-sized businesses at the aforementioned chair, since 2019.

FABIAN STEINBERG was born in 1990. He received the M.Sc. degree in production engineering from RWTH Aachen University. He is currently pursuing the Ph.D. degree with the Chair of International Production Engineering and Management (IPEM), University of Siegen. He is writing the Ph.D. thesis in die field of lead time prediction in engineering to order environment based on artificial intelligence. Since 2018, he has been working at the Chair of IPEM, University of Siegen.