Logistics Research (2020) 13:9
DOI_10.23773/2020_9

Discrete-Time Analysis of Levelled Order Release and Staffing in Order Picking Systems

Uta Mohring, Marion Baumann, Kai Furmans

Received: 07 December 2019 / Accepted: 05 August 2020 / Published online: 03 September 2020

ABSTRACT

Order picking systems are confronted with a volatile demand and short delivery time requirements. Manufacturing companies face the increasing variability requirements with Heijunka-levelling, one method of the Toyota Production System. The objectives of this publication are to develop a levelling concept for order picking systems, to analyse its performance based on a discrete-time analytical model and to develop a staffing algorithm determining the required workforce level in an order picking system with levelled order release. The levelling concept for order picking systems results from the existing models of Heijunka-levelling in the literature, which are adopted and expanded regarding the specific requirements of order picking systems. The order picking system with levelled order release is depicted as a discrete-time Markov chain. To analyse its performance, we derive several performance measures, such as service level, backlog duration and system utilisation, from the steady-state distribution of the Markov chain. The staffing algorithm is a binary search algorithm based on the Markov chain. The models developed in this publication enable a quantitative evaluation of the impact of several system parameters, such as variability of customer demand, workforce level and traffic intensity, on the performance measures of the order picking system. Furthermore, the staffing algorithm determines the workforce level which is required to guarantee a certain system performance, such as a service level of 99%, in an order picking system with levelled order release. By comparing levelled order release to FCFS-based order release strategies in a numerical example, we show the benefits of levelled order release.

KEYWORDS: discrete-time modelling · Markov chain · order picking · levelling · staffing

Uta Mohring, M.Sc.
eMail: uta.mohring@kit.edu

Dr.-Ing. Marion Baumann
eMail: marion.baumann@kit.edu

Prof. Dr.-Ing. Kai Furmans
eMail: kai.furmans@kit.edu

Institute for Material Handling and Logistics (IFL)
Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany

1. INTRODUCTION

Order picking systems are confronted with constantly increasing order volumes, volatile customer demand and permanent cost pressure. Since the order picking process directly depends on the customer orders, the variation of customer demand is assigned to the workload of order picking systems. Furthermore, customers require high flexibility and short delivery times (cf. [1]). Order picking systems often differentiate the customer orders regarding their lead time in express and standard orders. A typical example are the different delivery conditions of online retailers in B2C sectors. However, this differentiation is also common in B2B sectors: For example in the automotive aftermarket sector, express orders correspond to unexpected breakdowns of vehicles, whereas standard orders refer to planned and regular maintenance measures of vehicles.

Due to these flexibility requirements, many warehouses still prefer manual order picking systems with workers picking the required items from shelves or pallets while driving or walking through the warehouse instead of innovative, partly or fully automated order picking systems. Automated systems ensure higher picking rates than manual systems, but they require a high homogeneity of items. Furthermore, especially in contract logistics, automated order picking systems are not profitable, since the contract periods are too short and the customers' product ranges are too diverse. However, manual order picking is cost-intensive, especially in high-wage countries. According to [2] and [3], on average 45-55% of total warehousing costs can be assigned to the order picking process. Therefore, workforce efficiency is an important lever to reduce costs and to face the upcoming challenge of skills shortage.

Flexible workforce planning and levelled order release are two different appropriate solution approaches to face these partly opposite requirements in manual order picking systems (cf. Figure 1). The key idea of flexible workforce planning is to cover the volatile workload of the order picking system as precisely as possible with the available workforce capacities of the system by combining methods of flexible shift scheduling and flexible work time models. In this way, customer orders are processed in a timely manner and the order picking system can guarantee both short lead times with a high service level and high workforce efficiency. In contrast, the key idea of levelled order release is to convert the volatile workload of the order picking system into a smooth and regular workload per time interval. For this purpose, the deployed workforce capacities per time interval are constant and levelled order release takes advantage of the different lead times of the customer orders which allow a certain time flexibility to determine the time of processing of the orders. By considering the due dates of the customer orders, levelled order release compensates peaks in the workload of the order picking system resulting from peaks of the customer demand. Thus, the required workforce level is constant and expensive additional workforce capacities, such as overtime and temporary workers, are not necessary. It is also ensured that the customer orders are processed within the required lead times.

In this publication, we focus on the proposed approach of levelled order release. The approach of flexible workforce planning will not be considered further. The objectives of this publication are to develop a levelling concept for order picking systems, to analyse its performance based on a discrete-time analytical model and and to develop a staffing algorithm determining the required workforce level in an order picking system with levelled order release.

The remainder of this publication is structured as follows: Section 2 gives a short literature review on related research topics. In Section 3, we derive a levelling concept for order picking systems. To analyse the performance of this concept, we depict the order picking system with levelled order release as a discrete-time Markov chain and we derive several performance measures from its steady-state distribution in Section 4. In Section 5, we derive a staffing algorithm for order picking systems with levelled order release. The numerical studies in Section 6 give insights into system behaviour and show the benefits of levelled order release compared to FCFS-based order release strategies in a numerical example. Section 7 summarises the insights and provides directions for future research.

2. LITERATURE REVIEW

Review papers on warehouse operation (cf. [4], [5], [6]) and order picking systems (cf. [2]) state that past research focused on specific warehouse configurations or specific decision problems, e.g. routing, storage and batching policies. There is a lack of global models

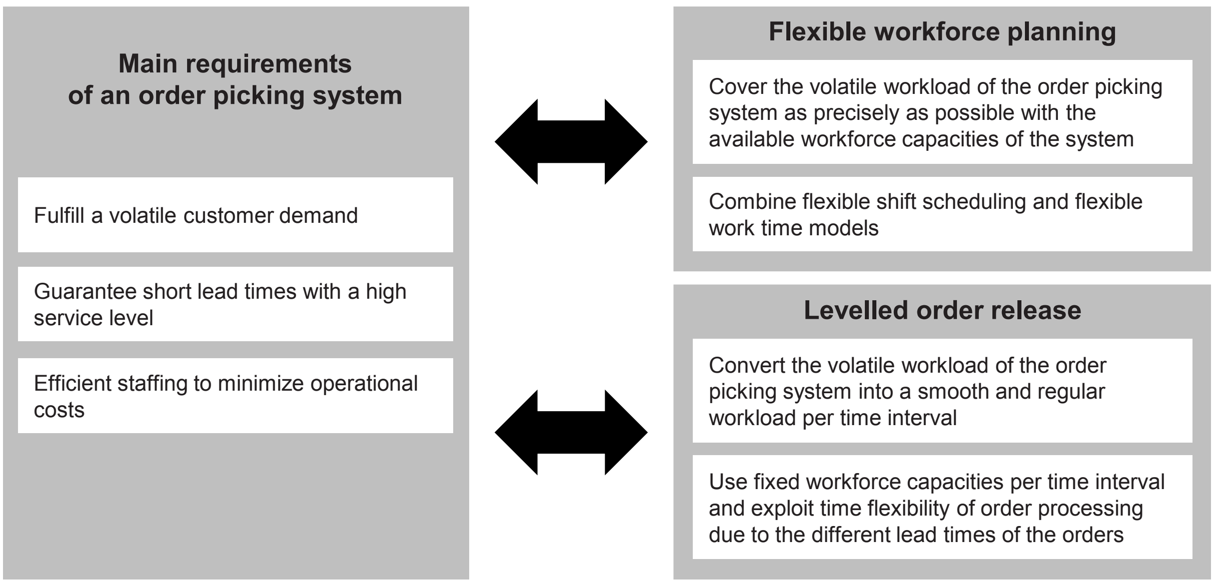

Fig. 1: Problem statement and possible solution approaches

and general procedures for order picking systems. Furthermore, past research predominantly focused on deterministic warehouse configurations assuming that all data is given in advance (cf. [2]). However, several questions in practice in warehouse operation include stochastics, e.g. stochastic customer demand and stochastic processing times, which is not explored in literature yet (cf. [7]).

An order picking policy describes the principle according to which orders are processed in the order picking system. Literature differentiates between strict order picking, batch picking, wave picking and zone picking (cf. [5], [8]). In contrast, an order release strategy describes the principle according to which orders are released for processing. Order release strategies in warehouses have hardly been investigated in literature so far. [9] differentiates between wave-based and waveless release policies: A wave-based release policy groups orders into batches by some criteria and these batches are released in a sequential manner. In case of waveless order release, individual orders are released continuously. [9] focuses on waveless order release for warehouses with an automated sorter and [10] deals with waved-based order release in order fulfillment systems with deadlines.

In contrast to warehouse literature, levelling customer demand and system workload in production systems is abundantly discussed in literature. The most known levelling concept in production systems is the Heijunka-levelling approach of the Toyota Production System. Past research on Heijunka-levelling can be classified into the following areas:

- Procedure models,
- Stochastic models and
- Models for level scheduling.

Procedure models of Heijunka-levelling, such as [11], [12] and [13], qualitatively describe the concept of Hejunka-levelling. Stochastic models, such as [14], [15], [16] and [17], focus on buffer sizing of a Heijunka-levelled Kanban system. [14]–[17] use discrete-time analytical models including stochastic parameters, such as production capacity and customer demand, to depict the Heijunka-levelled Kanban system and to compute several performance measures, such as service level and buffer size. The research area of level scheduling covers static optimization problems for production sequence planning in Heijunka-levelled production systems (cf. [18], [19]).

The philosophy and the methods of the Toyota Production System are transferred to other fields such as supply and distribution logistics. The so called “Lean Logistics” and “Lean Warehousing” are studied in academic literature to a limited extent: [20] describes the basic concepts of lean warehousing, [21] develops a lean assessment tool for warehouses and [22] investigates the impact of lean warehousing on the warehouse performance. None of these publications describes the different methods of lean warehousing, such as levelling, in detail.

[23], [24] and [25] provide comprehensive literature reviews on personnel scheduling categorizing the publications according to the solution method, the application area and several system characteristics. [25] focuses on staffing and scheduling approaches for systems with non-stationary demand. There are only few publications on staffing in warehouses, although researchers agree on the importance of workforce planning in warehouses (cf. [6], [26]). [27] develops a time series forecasting method to predict the workload in a zone order picking system. Based on the predicted workload and the productivity of one worker, the required workforce level is calculated.

To the best of our knowledge, this is the first publication dealing with levelled order release in order picking systems, analysing its performance based on an analytical, stochastic model and determining the workforce level of this system.

3. LEVELLING CONCEPT FOR ORDER PICKING SYSTEMS

In this section, we derive a levelling concept for order picking systems based on the key ideas of Heijunka-levelling in production systems. We initially describe the ideas and the procedure of Heijunka-levelling in production systems. Subsequently, we identify the differences between production systems and order picking systems regarding general conditions and decision problem. Finally, we present the levelling concept for order picking systems.

3.1. Heijunka-levelling in Production Systems

Heijunka-levelling is a simple and widespread concept for order release in production systems to manage the production of several different products on one common production line. The key idea of Heijunka-levelling is to convert the volatile customer demand into a regular, recurring and standardized production schedule to guarantee an even load of the given production capacity. Heijunka-levelling smooths both the volume and the product mix of the production system (cf. [11]).

The planning procedure of Heijunka-levelling refers to one planning period of the production system (e.g. one month), which is subdivided into smaller scheduling intervals (e.g. one week, one day, one shift). It consists of the following planning steps: System parametrisation and operational planning (cf. Figure 2). System parametrisation takes place at the beginning of each planning period and deals with smoothing both the production volume and the product mix. Volume smoothing determines the production capacity per product per scheduling interval which is reserved for the production of this product in each scheduling interval. For this purpose, the total customer demand of the planning period of each product is evenly distributed on the scheduling intervals. The reserved production capacity of a product corresponds to the average customer demand per scheduling interval of this product. Product mix smoothing determines the production sequence of the different products within a scheduling interval. Common objectives for production sequence planning are minimizing setup times or maximizing the regularity of the product mix. These decision problems are covered in detail by the research area of level scheduling. The reserved production capacities and the production sequence per scheduling interval are visualised in the levelling pattern on the Heijunka-board (cf. Figure 3). Based on the levelling pattern, the operational planning takes place at the beginning of each scheduling interval: The incoming customer orders of the current scheduling interval are fulfilled by taking the required products from the finished-goods-supermarket. The associated kanbans are returned to the Heijunka-board. These kanbans are assigned to the reserved production capacity of the corresponding product on the Heijunka-board according to First-Come-First-Served (FCFS). If the customer demand of a product exceeds its reserved production capacity in the current scheduling interval, the associated kanbans are kept in an overflow box. They are assigned to future scheduling intervals, when the customer demand of this product is below its reserved production capacity (cf. Figure 3, [16], [17]).

3.2. Delimitation from Heijunka-levelling in Production Systems

The concept of Heijunka-levelling in production systems cannot be directly applied in order picking systems because the general conditions and the decision problem of order picking systems differ from those of production systems to some extent: First, order picking systems and production systems differ in terms of lot size. Since orders are customer-specific concerning product type and product volume, the order lot size in order picking systems is usually one, whereas for reasons of setup times, lot sizes in production systems are often higher than one (cf. [22]). Second, setup times between different products are relevant in production systems, whereas setup times between the order picking processes of different customer orders are negligible small (cf. [22]). Third, Heijunka-levelling in production systems predominantly focuses on Make-To-Stock processes, which decouple workload and customer demand by a buffer, whereas the workload of order picking systems directly depends on the customer demand. Thus, order picking can be considered as a

Fig. 2: Planning procedure of Heijunka-levelling

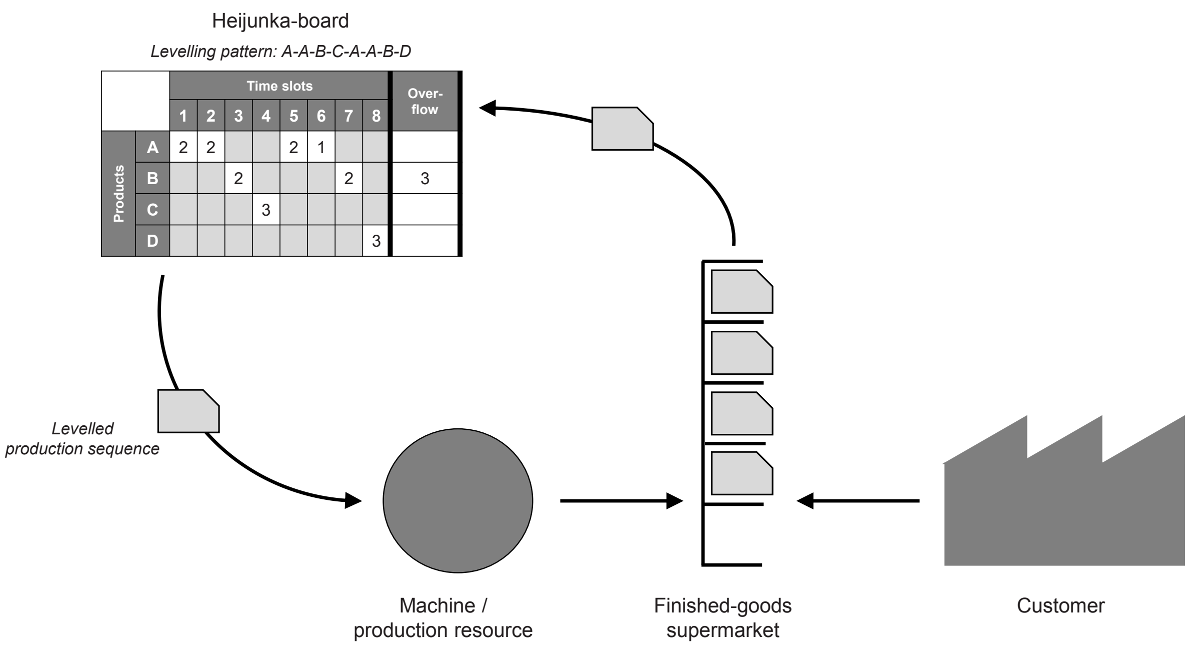

Fig. 3: Model of a Heijunka-levelled kanban system

Make-To-Order process. Finally, production capacity in production systems is fixed in the short term, due to a constant number of machines with a given performance. In manual order picking systems, capacity mainly depends on the number of assigned workers, which is rather flexible in short term. Thus, capacity can be easily adjusted to the current workload of the order picking system.

The decision problem of Heijunka-levelling in production systems focuses on choosing an appropriate buffer size to guarantee the required service level, whereby production capacity is constant. On the contrary, the decision problem of levelling in order picking systems determines the appropriate workforce capacity to guarantee the required service level.

Due to these differences between order picking systems and production systems, some adjustments and extensions are necessary to derive a levelling concept for order picking based on the key ideas of Heijunka-levelling. Furthermore, the above mentioned discrete-time stochastic models of Heijunka-levelled production systems focus on a different decision problem and cannot be used for the performance analysis of order picking systems with levelled order release.

3.3. Levelling Concept for Order Picking Systems

The levelling concept for order picking systems determines the release principles for picking orders. The picking orders result from the incoming customer orders depending on the order picking policy of the considered system. In the simplest case of strict order picking, each customer order corresponds to one picking order. For some order picking systems, it can be reasonable to classify the picking orders into different order types, e.g. regarding lead time (express vs. standard picking orders) or regarding used picking technology (picker-to-parts vs. parts-to-picker). Due to the Make-To-Order character of order picking, the levelling concept has to consider the individual due dates of the picking orders resulting from the due dates of the corresponding customer orders. Consequently, picking orders are differentiated regarding their due date into orders without failed due dates and those with failed due dates (cf. Figure 4). Picking orders with failed due dates are furthermore differentiated into backorders and lost sales: Backorders represent picking orders with failed due dates which still have to be fulfilled, whereas lost sales correspond to picking orders with failed due dates which are removed from the system without being processed, since their due date exceeds a certain maximum backlog duration.

Following the principles of Heijunka-levelling, the planning procedure of levelled order release in order picking systems consists of the planning steps system parametrisation and operational planning (cf. Figure 2). System parametrisation takes place at the beginning of each planning period and determines the picking capacity per order type per scheduling interval which is reserved for order processing of this order type in each scheduling interval (smoothing of volume) and the processing sequence of the different order types within a scheduling interval (smoothing of product mix). The procedure of both steps equals the one of Heijunka-levelling. The resulting levelling pattern is visualised on the Heijunka-board (cf. Figure 5), which is the starting point of the operational planning. Operational planning takes place at the beginning of each scheduling interval and allocates the picking orders of the different order types to the corresponding reserved picking capacities in the levelling pattern. The pool of assignable picking orders covers the incoming picking orders of the current scheduling interval and the remaining unprocessed picking orders of previous scheduling intervals stored in the overflow box. To determine the processing sequence of picking orders within one order type, their due dates are considered as follows:

- Picking orders are processed according to ascending due dates.
- Picking orders with identical due dates are processed in accordance of FCFS.
- Picking orders become lost sales, when their due date exceeds the maximum accepted backlog duration.

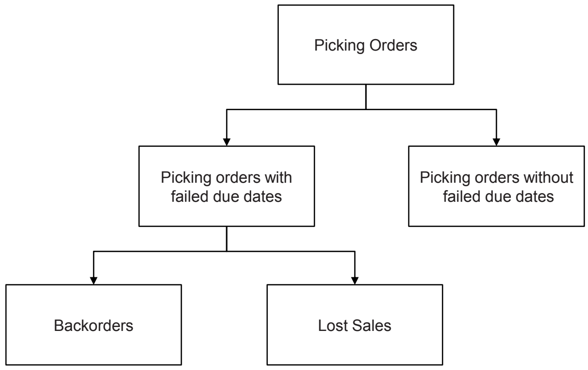

Fig. 4: Classification of picking orders regarding due date

If the number of assignable picking orders of an order type exceeds its reserved picking capacity in the current scheduling interval, the order backlog in the overflow box increases by the corresponding number of orders. Otherwise, the remaining capacity is used for training, maintenance and continuous improvement measures.

To sum up, the main characteristics of the levelling concept for order picking systems are the following:

- There is a fixed picking capacity per order type per scheduling interval which is reserved for order processing of picking orders of this order type in each scheduling interval.
- Size and sequence of the reserved picking capacities within one scheduling interval are visualised in the levelling pattern on the Heijunka-board.

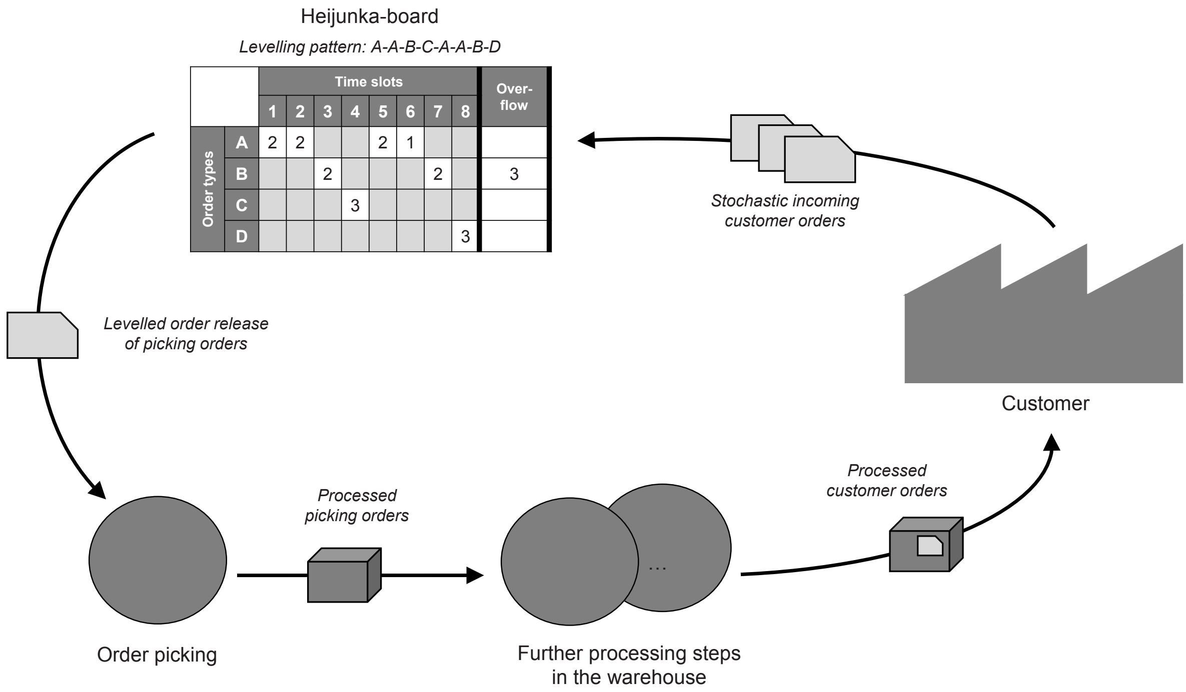

Fig. 5: Model of an order picking system with levelled order release

- During each scheduling interval, the reserved picking capacity per order type is used to process picking orders of this order type according to ascending due dates.

4. DISCRETE-TIME ANALYSIS OF LEVELLED ORDER RELEASE

In this section, we depict an order picking system with levelled order release as a discrete-time, analytical model to analyse the performance of the developed levelling concept. We firstly explain the reasons for choosing a discrete-time Markov chain. After describing the general conditions of the studied order picking systems with levelled order release, we introduce the corresponding Markov chain. Finally, several performance measures of interest are derived from the steady-state distribution of the Markov chain.

4.1. Model Choice

We choose a discrete-time Markov chain to analyse the performance of the levelling concept for order picking systems due to the following aspects: On the contrary to static models, Markov chains are able to depict the stochastic character of several parameters. Furthermore, performance measures derived from the steady-state distribution of the Markov chain are exact in contrast to the approximate results of simulation models. A discrete-time model is preferred to continuous-time models, since the relevant parameters have discrete-time character and discrete-time models allow a more detailed analysis of the system: The performance analysis is not limited to expected values, but complete probability distributions are computed (cf. [28]).

4.2. System Description

The analytical model depicts order processing of one order type. An isolated consideration of each order type is possible, since the analytical model focuses on operational planning. When operational planning takes place, the levelling pattern of the order picking system has already been determined. Thus, the reserved picking capacity per order type is fixed and order processing of the different order types is independent of each other. We assume that there is one order income of picking orders per scheduling interval which is already known at the beginning of the scheduling interval.

The general conditions of the order picking system are described by the following parameters (cf. Table 1): Customer demand is specified by its volume and its lead time: Random variable A describes the number of incoming picking orders per scheduling interval and random variable E specifies the lead time associated to one picking order when arriving at the order picking system. We specify the order picking process by the parameters L and c: The individual picking performance L depicts the number of picking orders one worker is able to completely fulfill within one scheduling interval, whereas c corresponds to the number of workers assigned to the order picking system. Furthermore, parameter N specifies the maximum backlog duration.

Parameter	Variable	Value range
Parameters specifying the customer demand		
Number of incoming picking orders per scheduling interval	A	$\mathcal{A} = \{a_{min}, \ldots, a_{max}\}$
Lead time of a picking order	E	$\mathcal{E} = \{e_{min}, \ldots, e_{max}\}$
Incoming picking orders per scheduling interval	$\mathbf{G} = \begin{pmatrix} G_{-N} & \ldots & G_{e_{max}} \end{pmatrix}$	$\mathcal{G}$
Parameters specifying the order picking process		
Individual picking performance per scheduling interval	L	$\mathcal{L} = \{l_{min}, \ldots, l_{max}\}$
Number of workers	c	$\mathbb{N}$
Total picking performance per scheduling interval	B	$\mathcal{B} = \{(c \cdot l_{min}), \ldots, (c \cdot l_{max})\}$
Parameters specifying the levelling concept		
Maximum backlog duration	N	$\mathbb{N}$

Tab. 1: Overview of parameters of the discrete-time Markov chain

To aggregate all relevant information regarding the customer demand in a single parameter, we introduce random vector $\mathbf{G} = (G_{-N} \ldots G_{e_{max}})$. It characterises the incoming picking orders per scheduling interval, whereby random variable G_k corresponds to the number of incoming picking orders with a lead time of k scheduling intervals, $k \in \{-N, \ldots, e_{max}\}$. The value range $\mathcal{G}$ of $\mathbf{G}$ is defined based on the following conditions:

- No incoming picking order has a negative lead time (first component in equation (1)).
- The value range of each vector component g_k, $k \in \{0, \ldots, e_{max}\}$, is defined by zero and the maximum number of incoming picking orders a_{max} per scheduling interval (second component in equation (1)).
- The total number of incoming picking orders $\left(\sum_{k=-N}^{e_{max}} g_k\right)$ corresponds to a realisation of the number of incoming picking orders A per scheduling interval (third component in equation (1)).

$$\begin{aligned} \mathcal{G} = \Big\{ &\begin{pmatrix} g_{-N} & \ldots & g_{e_{max}} \end{pmatrix} \mid \\ & g_k = 0 \ \forall k \in \{-N, \ldots, -1\} \\ & \wedge \ g_k \in \{0, \ldots, a_{max}\} \ \forall k \in \{0, \ldots, e_{max}\} \\ & \wedge \ \sum_{k=-N}^{e_{max}} g_k \in \mathcal{A} \Big\}. \end{aligned} \tag{1}$$

The probability $P(\mathbf{G} = \mathbf{g})$ of realisation $\mathbf{g} = (g_{-N} \ldots g_{e_{max}})$ depends on

- the probability of having $\left(\sum_{k=-N}^{e_{max}} g_k\right)$ incoming picking orders per scheduling interval (first component in equation (2)),
- the number of possibilities of having g_k picking orders with a lead time of k scheduling intervals for each lead time $k \in \mathcal{E}$ (second component in equation (2)) and
- the probability of having g_k picking orders with a lead time of k scheduling intervals for each lead time $k \in \mathcal{E}$ (third component in equation (2)):

$$\begin{aligned} P(\mathbf{G} = \mathbf{g}) = P&\left(A = \sum_{k=-N}^{e_{max}} g_k\right) \\ &\cdot \left[\prod_{k=0}^{e_{max}} \frac{(\sum_{m=k}^{e_{max}} g_m)!}{g_k! \cdot ((\sum_{m=k}^{e_{max}} g_m) - g_k)!}\right] \\ &\cdot \left[\prod_{k=0}^{e_{max}} P(E = k)^{g_k}\right] \qquad \forall \mathbf{g} \in \mathcal{G}. \end{aligned} \tag{2}$$

To aggregate all relevant information regarding the order picking process in a single parameter, we introduce random variable B. This variable specifies the total picking performance per scheduling interval and depends on the number of workers c and their individual picking performance L. Assuming an identical and independent distribution of the individual performance of all workers, the probability distribution of B is computed as c-fold convolution of the probability distribution of L.

We assume that the considered order picking system is stable. A system is stable, if its traffic intensity U is smaller than one:

$$U = \frac{E(A)}{E(B)} < 1. \tag{3}$$

4.3. Discrete-time Markov Chain

The discrete-time Markov chain depicts the temporal development of the order backlog in an order picking system with levelled order release. We observe the number of picking orders at discrete-time points in time $t \in \mathbb{N}_0$, which correspond to the starting points of the scheduling intervals of the levelling concept.

System state $\mathbf{X}$ of the Markov chain depicts the number of unprocessed picking orders of the order picking system:

$$\begin{aligned}&\mathbf{X} = \left(X_{-N} \quad \dots \quad X_{e_{max}}\right), \\ &X_k \in \mathbb{N}_0 \; \forall k \in \{-N, \dots, e_{max}\},\end{aligned} \tag{4}$$

whereby X_k corresponds to the number of unprocessed picking orders with a due date of k scheduling intervals.

The state transition from an arbitrary state $\mathbf{X}^t = \mathbf{x}$ at the beginning of scheduling interval t to a state $\mathbf{X}^{t+1} = \mathbf{y}$ at the beginning of scheduling interval $(t + 1)$ depends on

- the total picking performance b of scheduling interval t,
- the incoming picking orders $\mathbf{g}$ at the beginning of scheduling interval $(t + 1)$ and
- the principles of the levelling concept.

Assuming independence of incoming picking orders $\mathbf{G}$ and total picking performance B per scheduling interval, the transition probability is computed as follows

$$\begin{aligned}&P(\mathbf{X}^{t+1} = \mathbf{y} \mid \mathbf{X}^t = \mathbf{x}) \\ &\qquad = \sum_{(\mathbf{g},b)\in\mathcal{I}(\mathbf{x},\mathbf{y})} P(\mathbf{G} = \mathbf{g}) \cdot P(B = b)\end{aligned} \tag{5}$$

with

$$\begin{aligned}\mathcal{I}(\mathbf{x},\mathbf{y}) = \Bigg\{ &(\mathbf{g}, b) \in \mathcal{G} \times \mathcal{B} \mid \\ &g_k + \max\left\{0; x_{k+1} - \max\left\{0; b - \sum_{l=-N}^{k} x_l\right\}\right\} = y_k \\ &\qquad \forall k \in \{-N, \dots, (e_{max} - 1)\} \\ &\wedge\; g_{e_{max}} = y_{e_{max}} \Bigg\}, \quad \mathbf{x}, \mathbf{y} \in \mathcal{X}.\end{aligned} \tag{6}$$

The number of unprocessed picking orders y_k with a due date of k scheduling intervals at the beginning of scheduling interval $(t + 1)$ is the sum of the number of incoming picking orders g_k with a lead time of k scheduling intervals at the beginning of scheduling interval $(t + 1)$ and number of unprocessed picking orders

$$\max\left\{0; x_{k+1} - \max\left\{0; b - \sum_{l=-N}^{k} x_l\right\}\right\} \tag{7}$$

with a due date of $(k + 1)$ scheduling intervals at the end of scheduling interval t. The number of unprocessed picking orders with a due date of $(k + 1)$ scheduling intervals at the end of scheduling interval t is either zero or it corresponds to the difference of the number of unprocessed picking orders x_{k+1} with a due date of $(k + 1)$ scheduling intervals at the beginning of scheduling interval t and the residual picking performance

$$\max\left\{0; b - \sum_{l=-N}^{k} x_l\right\} \tag{8}$$

of scheduling interval t remaining after all picking orders with a due date of $l < (k + 1)$ scheduling intervals have already been processed. The number of unprocessed picking orders $y_{e_{max}}$ with a due date of e_{max} scheduling intervals at the beginning of scheduling interval $(t + 1)$ equals the number of incoming picking orders $g_{e_{max}}$ with a lead time of e_{max} scheduling intervals at the beginning of scheduling interval $(t + 1)$.

Using the specific structure of the state transition, we derive an upper bound for the state space $\mathcal{X}$: The number of unprocessed picking orders of a particular due date is at its maximum,

- if the number of incoming picking orders per scheduling interval of this due date is at its maximum,
- if the residual picking performance per scheduling interval assigned to this due date is at its minimum and
- if the number of remaining unprocessed picking orders of the previous scheduling interval assigned to this due date is at its maximum.

The upper bound $\mathbf{O}$ of system state $\mathbf{X}$ is defined by

$$\begin{aligned}&\mathbf{O} = \left(O_{-N} \quad \dots \quad O_{e_{max}}\right) \\ &O_k = (e_{max} + 1) \cdot a_{max} \qquad \forall k \in \{-N, \dots, 0\}, \\ &O_k = (e_{max} - k + 1) \cdot a_{max} \quad \forall k \in \{1, \dots, e_{max}\}.\end{aligned} \tag{9}$$

Consequently, the Markov chain is finite and its state space $\mathcal{X}$ is defined by

$$\mathcal{X} = \{0, 1, \dots, O_{-N}\} \times \dots \times \{0, 1, \dots, O_{e_{max}}\}. \tag{10}$$

Several performance measures of interest can be derived from the steady-state distribution π with $\pi_{\mathbf{x}} = P\ (\mathbf{X} = \mathbf{x})$ of the Markov chain. For aperiodic, finite, irreducible and therefore ergodic discrete-time Markov chains, the steady-state distribution is obtained by solving a set of linear equations (cf. [29]).

The Markov chain modelled in this publication is aperiodic for all considered practical applications. According to equation (10), its state space is finite. Furthermore, it is possible to reach every state of the Markov chain from every other state either by a direct state transition or by an indirect transition via a finite number of other states. In case of unreachable states, these states are excluded from the computations. Consequently, it is possible to find an irreducible subset of the state space which is used as starting point for subsequent computations. The steady-state distribution π is therefore computed by solving the following set of linear equations

$$\begin{aligned}&\pi_{\mathbf{y}} = \sum_{\mathbf{x}\in\mathcal{X}} P\left(\mathbf{X}^{t+1} = \mathbf{y} \mid \mathbf{X}^t = \mathbf{x}\right) \cdot \pi_{\mathbf{x}} \qquad \forall \mathbf{y} \in \mathcal{X}, \\ &\sum_{\mathbf{x}\in\mathcal{X}} \pi_{\mathbf{x}} = 1.\end{aligned} \tag{11}$$

Since this set of linear equations is overdetermined by one equation, one equation of (11) is omitted when solving the system of equations. To obtain the exact steady-state distribution, this set of linear equations is solved by using the Gaussian Elimination (cf. [29]).

4.4. Performance Measures

Table 2 gives an overview of the performance measures of the order picking system with levelled order release derived from the steady-state distribution in the following subsections.

4.4.1. *Performance measures based on unprocessed orders:*

Random variable Q specifies the total number of unprocessed picking orders in the order picking system at the beginning of a scheduling interval. The total number of unprocessed picking orders q at the beginning of a scheduling interval is the sum of the number of unprocessed picking orders x_k with a due date of k scheduling intervals, $k \in \{-N, ..., e_{max}\}$. The probability distribution is derived from the steady-state distribution as follows

$$P(Q = q) = \sum_{\mathbf{x} \in \mathcal{X} \mid \sum_{k=-N}^{e_{max}} x_k = q} \pi_{\mathbf{x}} \qquad \forall q \in \mathcal{Q} \tag{12}$$

and the corresponding expected value is defined as

$$E(Q) = \sum_{\mathbf{x} \in \mathcal{X}} \left(\sum_{k=-N}^{e_{max}} x_k \right) \cdot \pi_{\mathbf{x}}. \tag{13}$$

Random variable M specifies the number of unprocessed backorders in the order picking system at the beginning of a scheduling interval. The number of unprocessed backorders m at the beginning of a scheduling interval is the sum of the number of unprocessed picking orders x_k with a negative due date of k scheduling intervals, $k \in \{-N, ..., -1\}$. The probability distribution is derived from the steady-state distribution as follows

$$P(M = m) = \sum_{\mathbf{x} \in \mathcal{X} \mid \sum_{k=-N}^{-1} x_k = m} \pi_{\mathbf{x}} \qquad \forall m \in \mathcal{M} \tag{14}$$

and the corresponding expected value is defined as

$$E(M) = \sum_{\mathbf{x} \in \mathcal{X}} \left(\sum_{k=-N}^{-1} x_k \right) \cdot \pi_{\mathbf{x}}. \tag{15}$$

Random variable S specifies the number of unprocessed backorders which become lost sales per scheduling interval, since their due date exceeds the maximum backlog duration N. Lost sales of s picking orders occur at the beginning of a scheduling interval, if the number of unprocessed picking orders x_{-N} with a due date of $(-N)$ scheduling intervals at the beginning of the previous scheduling interval exceeds the total picking performance b of the previous scheduling interval by s picking orders. The corresponding probability is computed as negative convolution of the steady-state distribution and the probability distribution of B as follows

$$P(S = s) = \sum_{j=s}^{O_{-N}} \sum_{\mathbf{x} \in \mathcal{X} \mid x_{-N} = j} \pi_{\mathbf{x}} \cdot P(B = j - s) \qquad \forall s \in \mathcal{S} \setminus \{0\}, \tag{16}$$

$$P(S = 0) = 1 - \sum_{j=1}^{O_{-N}} P(S = j).$$

System utilisation $\tilde{U}$ specifies the proportion of the available picking performance per scheduling interval which is used to process picking orders. The utilisation of a specific scheduling interval equals the ratio of the total number of unprocessed picking orders

Performance measure	Variable	Value range
Performance measures based on unprocessed orders		
Number of unprocessed picking orders	Q	$\mathcal{Q} = \{0, \ldots, (\sum_{k=-N}^{e_{max}} O_k)\}$
Number of unprocessed backorders	M	$\mathcal{M} = \left\{0, \ldots, \left(\sum_{k=-N}^{-1} O_k\right)\right\}$
Number of lost sales	S	$\mathcal{S} = \{0, \ldots, O_{-N}\}$
System utilisation	$\tilde{U}$	$[0, 1]$
Performance measures based on processed orders		
Processed picking orders per scheduling interval	$\mathbf{H}$	$\mathcal{H}$
Number of processed picking orders per scheduling interval	F	$\mathcal{F} = \{(c \cdot l_{min}), \ldots, (c \cdot l_{max})\}$
Number of processed backorders per scheduling interval	$F^{backlog}$	$\mathcal{F} = \{(c \cdot l_{min}), \ldots, (c \cdot l_{max})\}$
Number of processed picking orders without failed due date per scheduling interval	F^{buffer}	$\mathcal{F} = \{(c \cdot l_{min}), \ldots, (c \cdot l_{max})\}$
Time difference to order deadline of a processed picking order	D	$\mathcal{D} = \{-N, \ldots, e_{max}\}$
Backlog duration of a processed backorder	$D^{backlog}$	$\mathcal{D}^{backlog} = \{1, \ldots, N\}$
Time buffer of a processed picking order	D^{buffer}	$\mathcal{D}^{buffer} = \{0, \ldots, e_{max}\}$
Service levels		
β-service level	SL_β	$[0, 1]$
γ-service level	SL_γ	$[0, 1]$

Tab. 2: Overview of performance measures of the discrete-time Markov chain

$\left(\sum_{k=-N}^{e_{max}} x_k\right)$ at the beginning of this scheduling interval to the total picking performance b of this scheduling interval. To compute the system utilisation $\tilde{U}$, we sum this ratio for all tupels $(\mathbf{x}, b) \in \mathcal{X} \times \mathcal{B}$, whereby each summand is weighted by the product of the corresponding probabilities $\pi_{\mathbf{x}}$ and $P(B = b)$:

$$\tilde{U} = \sum_{(\mathbf{x},b)\in\mathcal{X}\times\mathcal{B}} \min\left\{1; \frac{\sum_{k=-N}^{e_{max}} x_k}{b}\right\} \cdot \pi_{\mathbf{x}} \cdot P(B = b). \quad (17)$$

4.4.2. *Performance measures based on processed orders:*

Random vector $\mathbf{H} = (H_{-N} \ \ldots \ H_{e_{max}})$ specifies the number of processed picking orders per scheduling interval considering their due dates at the time of processing, whereby random variable H_k corresponds to the number of picking orders having a due date of k scheduling intervals, $k \in \{-N, \ldots, e_{max}\}$, at their time of processing. The value range $\mathcal{H}$ of $\mathbf{H}$ is defined based on the following conditions:

- The value range of each vector component h_k, $k \in \{-N, \ldots, e_{max}\}$, is defined by zero and the maximum total picking performance b_{max} per scheduling interval (first component in equation (18)).
- The total number of processed picking orders $\left(\sum_{k=-N}^{e_{max}} h_k\right)$ corresponds to a realisation of the total picking performance B per scheduling interval (second component in equation (18)).

$$\mathcal{H} = \Big\{ (h_{-N} \ \ldots \ h_{e_{max}}) \mid h_k \in \{0, \ldots, b_{max}\} \ \forall k \in \{-N, \ldots, e_{max}\} \wedge \sum_{k=-N}^{e_{max}} h_k \in \mathcal{B} \Big\}. \quad (18)$$

The number of processed picking orders h_k with a due date of k scheduling intervals at a specific scheduling interval equals either the number of unprocessed picking orders x_k with a due date of k scheduling intervals at the beginning of this scheduling interval or the residual picking performance

$$\max\left\{0; b - \sum_{j=-N}^{k-1} x_j\right\} \quad (19)$$

of this scheduling interval remaining after all picking orders with due date of $j < k$ scheduling intervals have already been processed. The probability $P(\mathbf{H} = \mathbf{h})$ of the realisation $\mathbf{h} = (h_{-N} \ \ldots \ h_{e_{max}})$ is defined as

$$P(\mathbf{H} = \mathbf{h}) = \sum_{(\mathbf{x},b)\in\mathcal{I}(\mathbf{h})} \pi_{\mathbf{x}} \cdot P(B = b) \qquad \forall \mathbf{h} \in \mathcal{H} \quad (20)$$

with

$$\mathcal{I}(\mathbf{h}) = \Big\{ (\mathbf{x}, b) \in \mathcal{X} \times \mathcal{B} \mid \min\left\{x_k; \max\left\{0; b - \sum_{j=-N}^{k-1} x_j\right\}\right\} = h_k \quad \forall k \in \{-N, \ldots, e_{max}\} \Big\}, \quad \mathbf{h} \in \mathcal{H}. \quad (21)$$

Random variable F specifies the total number of processed picking orders per scheduling interval. The total number of processed picking orders f per scheduling interval is the sum of the number of processed picking orders h_k with a due date of k scheduling intervals, $k \in \{-N, \ldots, e_{max}\}$. The probability distribution is derived from the probability distribution of $\mathbf{H}$ as follows

$$P(F = f) = \sum_{\mathbf{h}\in\mathcal{H} \mid \sum_{k=-N}^{e_{max}} h_k = f} P(\mathbf{H} = \mathbf{h}) \qquad \forall f \in \mathcal{F} \quad (22)$$

and the expected value is defined as

$$E(F) = \sum_{\mathbf{h}\in\mathcal{H}} \left(\sum_{k=-N}^{e_{max}} h_k\right) \cdot P(\mathbf{H} = \mathbf{h}). \quad (23)$$

In the same manner, we determine the probability distribution of the number of processed backorders $F^{backlog}$ per scheduling interval and the probability distribution of the number of processed picking orders without failed due date F^{buffer} per scheduling interval:

$$P\left(F^{backlog} = f\right) = \sum_{\mathbf{h}\in\mathcal{H} \mid \sum_{k=-N}^{-1} h_k = f} P(\mathbf{H} = \mathbf{h}) \qquad \forall f \in \mathcal{F}, \quad (24)$$

$$P\left(F^{buffer} = f\right) = \sum_{\mathbf{h}\in\mathcal{H} \mid \sum_{k=0}^{e_{max}} h_k = f} P(\mathbf{H} = \mathbf{h}) \qquad \forall f \in \mathcal{F}. \quad (25)$$

The expected values of $F^{backlog}$ and F^{buffer} are defined as follows:

$$E\left(F^{backlog}\right) = \sum_{\mathbf{h}\in\mathcal{H}} \left(\sum_{k=-N}^{-1} h_k\right) \cdot P(\mathbf{H} = \mathbf{h}). \quad (26)$$

$$E\left(F^{buffer}\right) = \sum_{\mathbf{h}\in\mathcal{H}} \left(\sum_{k=0}^{e_{max}} h_k\right) \cdot P(\mathbf{H} = \mathbf{h}). \quad (27)$$

Random variable D quantifies the time difference to deadline of a processed picking order at its time of processing. The probability $P(D = d)$ that a processed picking order has a time difference of d scheduling intervals to its deadline at its time of processing is proportional to the sum of the number of processed picking orders h_d with a due date of d scheduling intervals at their time of processing for all realisations $\mathbf{h} \in \mathcal{H}$, whereby each summand is weighted by the corresponding probability $P(\mathbf{H} = \mathbf{h})$. By normalizing this weighted sum, the probability distribution of D is defined as follows

$$P(D=d)=\frac{\sum_{\mathbf{h}\in\mathcal{H}} h_d\cdot P(\mathbf{H}=\mathbf{h})}{\sum_{\mathbf{h}\in\mathcal{H}}\sum_{k\in\mathcal{D}} h_k\cdot P(\mathbf{H}=\mathbf{h})} \qquad \forall d\in\mathcal{D}. \quad (28)$$

The expected value of D is computed as

$$E(D)=\frac{\sum_{\mathbf{h}\in\mathcal{H}}\sum_{k\in\mathcal{D}} k\cdot h_k\cdot P(\mathbf{H}=\mathbf{h})}{E(F)}. \quad (29)$$

The backlog duration of a processed backorder is equal to the number of scheduling intervals which have already been passed since its deadline at its time of processing. The probability $P\left(D^{backlog}=d\right)$ that a processed picking order has a backlog duration of d scheduling intervals at its time of processing is proportional to the probability $P\ (D=-d)$ that a processed picking order has a time difference to deadline of $(-d)$ scheduling intervals. By normalizing the probabilities $P\ (D=d)$, $d\in\{-N,\ldots,-1\}$, the probability distribution of $D^{backlog}$ is computed as follows

$$P\left(D^{backlog}=d\right)=\frac{P(D=-d)}{\sum_{k=-N}^{-1} P(D=k)} \qquad \forall d\in\mathcal{D}^{backlog}. \quad (30)$$

The corresponding expected value is defined as

$$E\left(D^{backlog}\right)=\frac{\sum_{\mathbf{h}\in\mathcal{H}}\sum_{k=-N}^{-1}|k|\cdot h_k\cdot P(\mathbf{H}=\mathbf{h})}{E\left(F^{backlog}\right)}. \quad (31)$$

The time buffer of a processed picking order corresponds to the number of scheduling intervals which still remain at its time of processing until reaching its deadline. The probability $P\left(D^{buffer}=d\right)$ that a processed picking order has a time buffer of d scheduling intervals at its time of processing is proportional to the probability $P\,(D=d)$ that a processed picking order has a time difference to deadline of d scheduling intervals. By normalizing the probabilities $P\,(D=d)$, $d\in\{0,\ldots,e_{max}\}$, the probability distribution of D^{buffer} is computed as follows

$$P\left(D^{buffer}=d\right)=\frac{P(D=d)}{\sum_{k=0}^{e_{max}} P(D=k)} \qquad \forall d\in\mathcal{D}^{buffer}. \quad (32)$$

The corresponding expected value is defined as

$$E\left(D^{buffer}\right)=\frac{\sum_{\mathbf{h}\in\mathcal{H}}\sum_{k=0}^{e_{max}} k\cdot h_k\cdot P(\mathbf{H}=\mathbf{h})}{E\left(F^{buffer}\right)}. \quad (33)$$

4.4.3 Service Levels:

The service level of an order picking system specifies the proportion of on time processed picking orders of the total number of picking orders. Depending on the considered level of detail, we differentiate two different types of service level: The β-service level considers the number of on time processed orders, whereas the γ-service level additionally takes into account the backlog duration of processed backorders.

We define the β-service level as quotient of the expected value of the number of processed picking orders without failed due date $E\left(F^{buffer}\right)$ and the sum of the expected value of the number of processed picking orders $E(F)$ and the expected value of the number of lost sales $E(S)$:

$$SL_\beta=\frac{E\left(F^{buffer}\right)}{E(F)+E(S)}. \quad (34)$$

The γ-service level is computed as the complement of the following quotient: The numerator of the quotient covers the sum of the total number of processed backorders $\left(\sum_{k=-N}^{-1}|k|\cdot h_k\right)$, weighted by their backlog duration, for all realisations $\mathbf{h}\in\mathcal{H}$, whereby each summand is weighted by the corresponding probability $P\ (\mathbf{H}=\mathbf{h})$, and the expected value of the number of lost sales $E(S)$ weighted by a backlog duration of $(N+1)$ scheduling intervals. The denominator of the quotient considers the expected value of the number of processed picking orders $E(F)$ weighted by the maximum backlog duration of N scheduling intervals and the expected value of the number of lost sales $E(S)$ weighted by a backlog duration of (N + 1) scheduling intervals:

$$SL_\gamma=1-\left(\frac{\sum_{\mathbf{h}\in\mathcal{H}} P(\mathbf{H}=\mathbf{h})\cdot\left(\sum_{k=-N}^{-1}|k|\cdot h_k\right)}{N\cdot E(F)+(N+1)\cdot E(S)}+\frac{(N+1)\cdot E(S)}{N\cdot E(F)+(N+1)\cdot E(S)}\right). \quad (35)$$

5. STAFFING IN ORDER PICKING SYSTEMS WITH LEVELLED ORDER RELEASE

In this section, we develop a staffing algorithm for order picking systems with levelled order release determining the required workforce level to guarantee a certain system performance. First, we choose an appropriate model to implement the staffing algorithm based on the existing Markov chain. Second, we introduce the staffing algorithm for order picking systems with levelled order release. Finally, we analyse the impact of the workforce level on the performance measures of the order picking system.

5.1. Model Choice

The workforce level of a system corresponds to the number of workers assigned to this system. The workforce level of an order picking system with levelled order release is already captured in parameter c of the discrete-time Markov chain (cf. Table 1). Thus, the Markov chain determines the performance of an order picking system with levelled order release for a given workforce level. On the contrary, the question of staffing deals with determining the minimum workforce level to guarantee the required performance of the order picking system (cf. Figure 6). Therefore, we supplement the Markov chain by an appropriate search algorithm.

Search algorithms identify the element within a given search range whose value of a predefined key parameter equals the target value. We choose the binary search algorithm as appropriate algorithm for staffing, due to its applicability to our decision problem, its intuitive procedure and its computing time performance. The binary search algorithm halves the search range in each iteration. [30] provides a detailed description of its procedure.

5.2. Staffing Algorithm

The search range of the staffing algorithm corresponds to the set of possible workforce levels $\mathcal{C}$. It is derived from the value ranges of the number of incoming picking orders A per scheduling interval and the individual picking performance L per scheduling interval as follows

$$\mathcal{C} = \left\{ c \in \mathbb{N} \mid \left\lfloor \frac{a_{min}}{l_{max}} \right\rfloor \leq c \leq \left\lceil \frac{a_{max}}{l_{min}} \right\rceil \right\}. \tag{36}$$

There is an individual data set assigned to each workforce level $c \in \mathcal{C}$ which contains the values of the performance measures of the order picking system with levelled order release, if c workers are assigned to the system. It results from the computation of the associated Markov chain. Each performance measure can be chosen as key parameter of the staffing algorithm. The corresponding target value specifies the required performance of the order picking system. It is an input parameter of the staffing algorithm.

The decision of the staffing algorithm whether the lower or the upper half of the search range is chosen to continue the search depends on the relationship between the workforce level and the performance measure chosen as key parameter: The relationship has to be either positive monotonic or negative monotonic to correctly implement the procedure of the binary search in the staffing algorithm. In subsection 5.3, we show that this requirement is held. To ensure minimality of the identified workforce level, the staffing algorithm continues searching in the lower half of the search range until it fails.

If multiple performance measures determine the required performance of the order picking system, the staffing algorithm is implemented separately for each of these performance measures. The minimum required workforce level of the order picking system equals the maximum value of the different workforce levels identified in the separate implementations of the staffing algorithm.

5.3. Impact of the workforce level on the performance measures of the order picking system

Since the staffing algorithm requires a monotonic relationship between the workforce level and each performance measure of the order picking system, we analyse the relationship between the workforce level and the performance measures. The analysis exclusively considers the impact of an increasing workforce level on the performance measures. The impact of a decreasing workforce level is exactly opposite to the effects described in the following.

An increase of the workforce level c leads to an increase of the expected value of the total picking performance $E(B)$ per scheduling interval, since the total picking performance B per scheduling interval is computed as c-fold convolution of the individual picking performance L per scheduling interval. Several performance measures of the order picking system

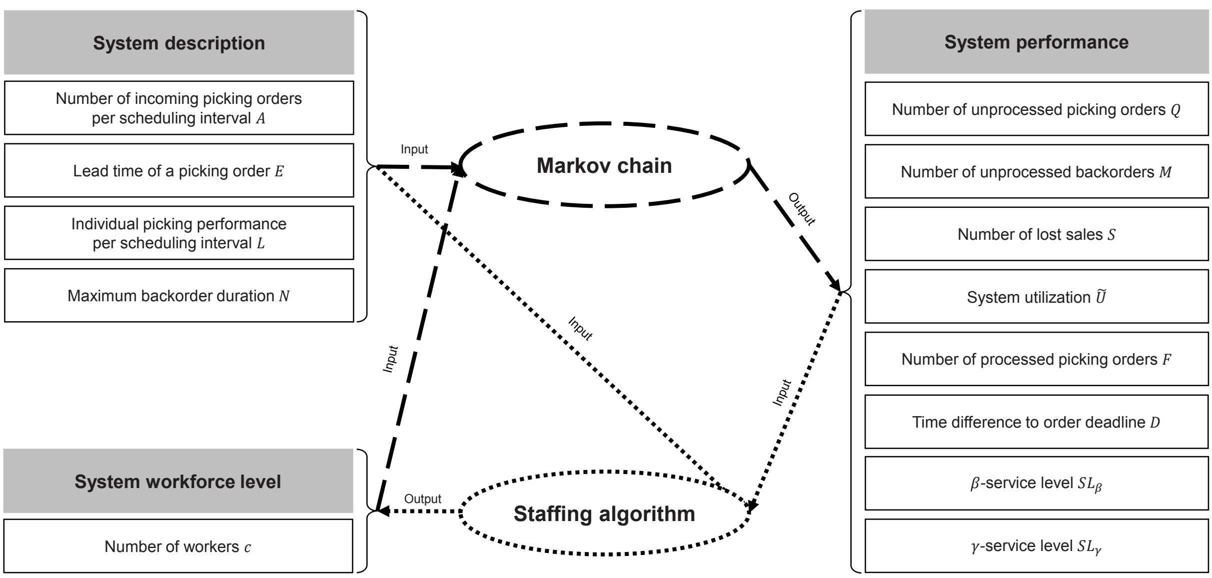

Fig. 6: Decision problem of Markov chain and staffing algorithm

directly depend on the total picking performance per scheduling interval. Thus, the analysis focuses on the impact of an increasing expected value of the total picking performance $E(B)$ per scheduling interval.

The expected value of the number of unprocessed picking orders $E(X_k)$ with a due date of k scheduling intervals, $k \in \{-N, ..., e_{max}\}$, either decreases or remains constant with increasing $E(B)$, since the expected value of the residual picking performance $E\left(B_k^{residual}\right)$ per scheduling interval for picking orders with a due date of k scheduling intervals,

$$E\left(B_k^{residual}\right) = \sum_{b\in\mathcal{B}}\sum_{\mathbf{x}\in\mathcal{X}} \max\left\{0; b - \sum_{l=-N}^{k} x_l\right\} \cdot P(B=b)\cdot\pi_{\mathbf{x}}, \tag{37}$$

either increases or remains constant (cf. equation (6)). Consequently, both the expected value of the number of unprocessed picking orders $E(Q)$ and the expected value of the number of unprocessed backorders $E(M)$ either decrease or remain constant with increasing $E(B)$ (cf. equations (13), (14)). System utilisation $\tilde{U}$ considers the ratio of the total number of unprocessed picking orders $\left(\sum_{k=-N}^{e_{max}} x_k\right)$ to the total picking performance B (cf. equation (17)). Due to the increase of the expected value of the total picking performance $E(B)$ per scheduling interval and the decrease of the expected value of the total number of unprocessed picking orders $E(Q)$, system utilisation $\tilde{U}$ either decreases or remains constant. The expected value of the number of lost sales $E(S)$ either decreases or remains constant with increasing $E(B)$, since the expected value of the number of unprocessed picking orders $E(X_{-N})$ with a due date of $(-N)$ scheduling intervals either decreases or remains constant and the expected value of the total picking performance $E(B)$ per scheduling interval increases (cf. equation (16)).

For short due dates k, $k \in \{-N, ..., e_{max}\}$, the number of processed picking orders h_k with a due date of k scheduling intervals corresponds to the number of unprocessed picking orders x_k with a due date of k scheduling intervals. Whereas for long due dates k, $k \in \{-N, ..., e_{max}\}$, the number of processed picking orders h_k with a due date of k scheduling intervals is determined by the residual picking performance (cf. equation (21)). We define the state-dependent parameter $k_{\mathbf{x}}^*$ as shortest due date k for which the residual picking performance determines the number of processed picking orders:

$$k_{\mathbf{x}}^* = \min\left\{k \in\{-N,\dots,e_{max}\} \mid h_k = \max\left\{0; b - \sum_{j=-N}^{k-1} x_j\right\}\right\} \tag{38}$$

For due dates $k < k^*$, the expected value of the number of processed picking orders $E(H_k)$ with a due date of k scheduling intervals either decreases or remains constant with increasing $E(B)$, since the expected value of the number of unprocessed picking orders $E(X_k)$ of this due date either decreases or remains constant. For due dates $k \geq k^*$, the expected value of the number of processed picking orders $E(H_k)$ with a due date of k scheduling intervals either increases or remains constant with increasing $E(B)$, since the expected value of the residual picking performance $E\left(B_k^{residual}\right)$ per scheduling interval either increases or remains constant. Furthermore, $k_{\mathbf{x}}^*$ either increases or remains constant with increasing $E(B)$. Consequently, there is a positive time shift in the time of processing of a picking order: There are fewer picking orders having a short due date at their time of processing, whereas more picking orders have a long due date at their time of processing.

Regarding the time difference to deadline D of a processed picking order, the probability $P(D = d)$ of having a time difference of $d < k^*$ scheduling intervals either decreases or remains constant with increasing $E(B)$, since the expected value of the number of processed picking orders $E(H_d)$ with a due date of d scheduling intervals either decreases or remains constant for due dates $d < k^*$. On the contrary, the probability $P(D = d)$ of having a time difference of $d \geq k^*$ scheduling intervals either increases or remains constant with increasing $E(B)$, since the expected value of the number of processed picking orders $E(H_d)$ with a due date of d scheduling intervals either increases or remains constant for due dates $d \geq k^*$ (cf. equation 28). Consequently, the expected value of the time difference to deadline $E(D)$ of a processed picking order either increases or remains constant (cf. equation (29)).

The range of possible due dates $k \in \{-N, ..., e_{max}\}$ of a picking order is independent of the total picking performance per scheduling interval. Due to the constant range of due dates and the positive time shift in the time of processing of a picking order, we conclude the following: The expected value of the number of processed backorders $E(F^{backlog})$ per scheduling interval either decreases or remains constant with increasing $E(B)$ and the expected value of the number of processed picking orders $E(F^{buffer})$ without failed due date per scheduling interval either increases or remains constant (cf. equations (26), (27)). For the same reasons, the probability $P(D^{backlog} = d)$ that a processed picking order has a negative time difference to deadline $d \in \{-N, ..., -1\}$ either decreases or remains constant with increasing $E(B)$, whereas the probability $P(D^{buffer} = d)$ that a processed order has a non-negative time difference to deadline $d \in \{0, .., e_{max}\}$ either increases or remains constant. Consequently, the expected value of the backlog duration $E(D^{backlog})$ of a processed backorder either decreases or remains constant and the expected value of the time buffer $E(D^{buffer})$ of a processed picking order increases or remains constant (cf. equations (31), (33)).

The expected value of the total number of processed picking orders $E(F)$ per scheduling interval either increases or remains constant with increasing $E(B)$.

Based on these findings, we are able to analyse the impact of the workforce level on the service levels SL_β and SL_γ: Since the total number of outgoing picking orders corresponds on average to the total number of incoming picking orders, the sum of the expected value of the number of processed picking orders $E(F)$ per scheduling interval and the expected value of the number of lost sales $E(S)$ per scheduling interval is equal to the expected value of the number of incoming picking orders $E(A)$ per scheduling interval. This sum is constant, since the expected value of the number of incoming picking orders $E(A)$ is independent of the workforce level. As a result, β-service level SL_β either increases or remains constant with increasing $E(B)$, since the expected value of the number of processed picking orders $E(F^{buffer})$ without failed due date increases or remains constant (cf. equation (34)).

Regarding the analysis of the γ-service level, we consider the numerator and the denominator of the quotient in equation (35) separately: The denominator can be rewritten as

$$N \cdot (E(F) + E(S)) + E(S). \tag{39}$$

The first summand remains constant with increasing $E(B)$, since the sum $(E(F)+ E(S))$ is independent of the total picking performance per scheduling interval. The second summand $E(S)$ either decreases or remains constant with increasing $E(B)$. Consequently, the denominator of the quotient in equation (35) either decreases or remains constant with increasing $E(B)$. Using equation (31), the first summand of the numerator can be rewritten as follows

$$\begin{aligned}&\sum_{\mathbf{h}\in\mathcal{H}} P(\mathbf{H}=\mathbf{h}) \cdot \left(\sum_{k=-N}^{-1} |k| \cdot h_k\right)\\ &= E\left(D^{backlog}\right) \cdot \sum_{\mathbf{h}\in\mathcal{H}} \sum_{k=-N}^{-1} h_k \cdot P(\mathbf{H}=\mathbf{h})\\ &= E\left(D^{backlog}\right) \cdot E\left(F^{backlog}\right).\end{aligned} \tag{40}$$

Since $E\ (D^{backlog})$ and $E\ (F^{backlog})$ both decrease or remain constant with increasing $E(B)$, the first summand of the numerator decreases or remains constant. The second summand of the numerator $((N+1) \cdot E(S))$ either decreases or remains constant with increasing $E(B)$, since the expected value of the number of lost sales $E(S)$ either decreases or remains constant. Consequently, the numerator of the quotient in equation (35) either decreases or remains constant with increasing $E(B)$. The decrease of the numerator is at least equal to the decrease of the denominator, since the first summand of the numerator decreases or remains constant and the decrease of the second summand of the numerator is greater than the decrease of the denominator due to its greater coefficient: In the numerator, $E(S)$ is weighted by $(N + 1)$, whereas $E(S)$ is unweighted in the denominator. Consequently, γ-service level SL_γ increases or remains constant with increasing $E(B)$ (cf. equation (35)).

In this analysis, we show that the relationship between the workforce level and each performance measure is either positive monotonic or negative monotonic:

- The performance measures Q, M, S, $\tilde{U}$, $F^{backlog}$ and $D^{backlog}$ are negatively correlated with the workforce level.
- The performance measures F, F^{buffer}, D, D^{buffer}, SL_β and SL_γ are positively correlated with the workforce level.

A	$E(A) \in \{1.0, 1.5, 2.0\}$	$c^2(A) \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$
E	$E(E) = 0.6$	$c^2(E) = 0.667$
L	$E(L) = 1.05$	$c^2(L) = 0.315$
c	2	
N	2	

Tab. 3: Parameter setting for numerical performance analysis for different variabilities of incoming picking orders

6. NUMERICAL STUDIES

In this section, we investigate the impact of several variations of system parameters on different performance measures of the order picking system and we compare levelled order release with FCFS-based order release strategies in a numerical example.

6.1. Numerical Performance Analysis for Different Variabilities of Incoming Picking Orders

To investigate the impact of the variability of incoming picking orders on the performance measures of the order picking system with levelled order release, we examine several variabilities of the number of incoming picking orders per scheduling interval in the range between $c^2(A) = 0$ and $c^2(A) = 1$ for different traffic intensities of the system $U \in \{0.4762, 0.7143, 0.9524\}$ (cf. Table 3). A typical example for an order picking system with $c^2(A) = 0$ is a just-in-time material supply of a production line, whereas $c^2(A) = 1$ refer to order picking systems with a high number of different customers.

Figure 7 shows that β-service level SL_β and γ-service level SL_γ decrease with increasing variability of incoming picking orders $c^2(A)$. The variability of the incoming picking orders reflects the volatility of the workload of the order picking system. Higher and more frequent peaks of the customer demand result in an increasing volatility of the workload of the order picking system which is generally compensated using additional picking performance. However, an order picking system with a given picking performance can compensate an increasing volatility of its workload only to some extent. Thus, an increasing volatility of

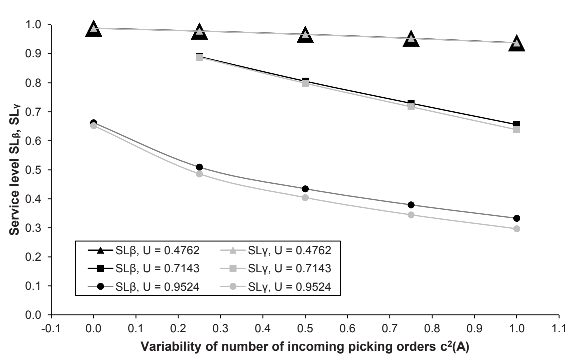

Fig.7: Impact of variability of incoming picking orders $c^2(A)$ and traffic intensity U on β-service level SL_β and γ-service level SL_γ (Note: There is no value of SL_β, SL_γ for scenario ($E(A) = 1.5$, $c^2(A) = 0.0$), since the value range of A is integer.)

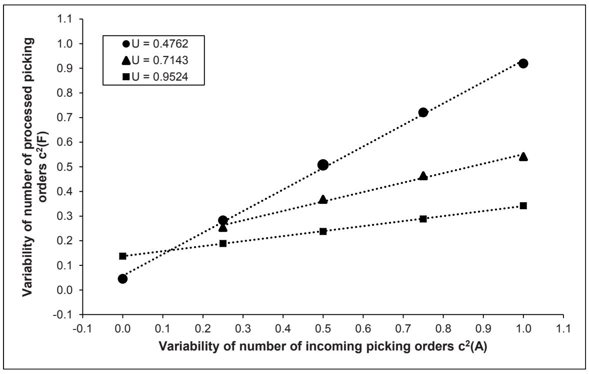

Fig.8: Impact of variability of incoming picking orders $c^2(A)$ and traffic intensity U on variability of processed picking orders $c^2(F)$ (Note: There is no value of $c^2(F)$ for scenario ($E(A) = 1.5$, $c^2(A) = 0.0$), since the value range of A is integer.)

A	$E(A) \in \{1.0, 1.5, 2.0\}$	$c^2(A) = 0.5$
E	$E(E) \in \{0.0, 1.0, 2.0, 3.0\}$	$c^2(E) = 0$
L	$E(L) = 1.05$	$c^2(L) = 0.315$
c	2	
N	2	

Tab. 4: Parameter setting for numerical performance analysis for different lead times

the workload has a negative impact on the performance of the order picking system (cf. Figure 7).

Furthermore, Figure 7 shows a disproportionate decrease in β-service level SL_β and γ-service level SL_γ with increasing traffic intensity *U*. In order picking systems with a low traffic intensity, the average available picking performance $E(B)$ exceeds the average needed picking performance $E(A)$ to a remarkable extent. This performance excess is used to compensate occurring peaks of the workload. On the contrary, in order picking systems with a high traffic intensity, the average available picking performance $E(B)$ is only slightly higher than the average needed picking performance $E(A)$, so that there is less picking performance to compensate occurring peaks of the workload. Consequently, the negative impact of an increasing volatility of the workload on the performance of the order picking system increases with increasing traffic intensity of the system (cf. Figure 7).

Figure 8 shows a positive correlation between the variability of processed picking orders $c^2(F)$ and the variability of incoming picking orders $c^2(A)$. For $c^2(A) > 0$, the value of the variability of processed picking orders is smaller than the corresponding value of the variability of incoming picking orders, especially for medium and high traffic intensities. Thus, the levelled order release succeeds in reducing the volatility of the workload of the order picking system. The smoothing effect of levelled order release increases with increasing traffic intensity.

6.2. Numerical Performance Analysis for Different Lead Times

To investigate the impact of the lead time of a picking order on the performance measures of the order picking system with levelled order release, we examine several expected values of the lead time of a picking order in the range between $E(E) = 0$ and $E(E) = 3$ for different traffic intensities of the system $U \in \{0.4762, 0.7143, 0.9524\}$ (cf. Table 4). Express orders have to be processed within the same day ($E(E) = 0$), whereas for replenishment orders, the lead time is typically greater than one day.

Figure 9 shows an increase in β-service level SL_β and γ-service level SL_γ with increasing expected value of the lead time $E(E)$ of a picking order. The lead time of a picking order affects the time flexibility of the order picking system to determine the time of processing of this picking order: The time flexibility to determine the time of processing of a picking order with a long lead time is higher than the one of a picking order with a short lead time. In an order picking system with a given picking performance and a given number of incoming picking orders, an increasing time flexibility increases the number of on time processed picking orders and thus results in an increase of the performance of the order picking system (cf. Figure 9).

Furthermore, Figure 9 shows a disproportionate increase in β-service level SL_β and γ-service level SL_γ with increasing traffic intensity *U*. In order picking systems with a low traffic intensity, the average available picking performance $E(B)$ exceeds the average needed picking performance $E(A)$ to a remarkable extent, so that there is enough picking performance to process almost every picking order on time independent of its lead time. In order picking systems with a high traffic intensity, the average available picking performance $E(B)$ is only slightly greater than the average needed picking performance $E(A)$, so that a higher time flexibility has a remarkable positive effect on the number of on time processed picking orders. Thus, the positive impact of an increasing lead time of the picking orders on the performance of the order picking system increases with increasing traffic intensity of the system (cf. Figure 9).

These interaction effects between the lead time of a picking order, the traffic intensity of the system and its performance could be relevant for practitioners when negotiating their supply agreements with customers: To guarantee a high service level, either the customers have to transmit their orders with a sufficiently long lead time, or the system has to be run at a sufficiently small traffic intensity. The models developed in this publication enable a quantitative evaluation of these interaction effects: For instance, for a given probability distribution of the lead time of a picking order, we are able to determine the maximum possible traffic intensity of the order picking system while guaranteeing the required service level of the customer at the same time.

6.3. Comparison of Order Release Strategies in a Numerical Example

We investigate a manual order picking system of a German company of the automotive aftermarket sector. The observation period lasts from 2 January 2015 to 31 December 2015. The time series of the order data within this period consists of 260 data sets (cf. Figure 10). Each data set represents the order data of one working day and includes the daily order volume as well as the lead times of the customer orders. The daily order volume fluctuates between its minimum of 1144 orders per day and its maximum of 14193 orders per day. The average order volume is 7842 orders per day. The customer orders are subdivided into the following order types: express orders with an average lead time of 0.84 days and standard orders with an average lead time of 4.04 days. Regarding the picking performance, it is known that one worker picks on average 14 orders per hour and has a daily working time of 8 hours. The required performance of this order picking system is measured by β-service level and its target value is 98%.

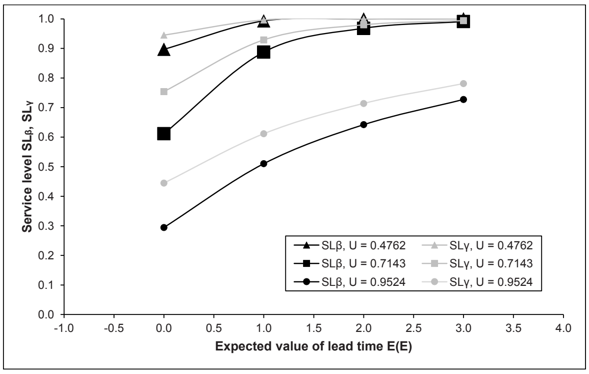

Fig. 9: Impact of expected value of lead time E(E) and traffic intensity U on β-service level SL_β and γ-service level SL_γ

For this order picking system, we aim to quantify the benefits of levelled order release compared to FCFS-based reference order release strategies. Therefore, we compare levelled order release with the order release strategies FCFS-DD and FCFS-RAND:

- FCFS-DD: Orders arriving at different points in time are released according to FCFS and orders arriving at the same point in time are released according to ascending due dates.
- FCFS-RAND: Orders arriving at different points in time are released according to FCFS, whereas the release sequence of orders arriving at the same point in time is determined randomly.

We initially have to derive the parameters of the Markov chain (cf. Table 1) from the given data. There are two possibilities to model the different order types of the numerical example: On the one hand, we can consider each order type separately modelling one Markov chain for each order type. On the other hand, the order types only differ regarding their lead time in this example. We can depict these differences in the probability distribution of the lead time and thus the differentiation of two order types is not necessary any further. For reasons of simplicity, we choose the latter

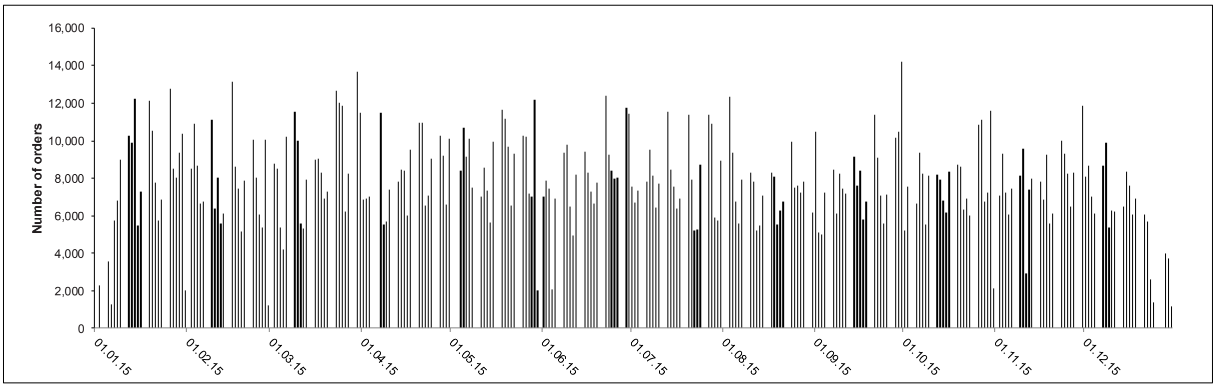

Fig. 10: Time series of order volume in 2015

and model the order picking system of the numerical example based on a single Markov chain. The scheduling interval of the levelling concept is one day, in particular one working day, due to the time interval of the order data. Furthermore, the data of the daily order volume is classified into 14 equally-sized classes in the range from 1000 to 15000 orders per day, whereby the mean of each class is chosen as class representative. The resulting probability distribution of the number of incoming picking orders per scheduling interval is shown in Figure 11. The probability distribution of the lead time of a picking order is derived from the absolute frequency distribution of the order lead time. The peaks of the probability distribution for a lead time of one day and a lead time of four days indicate the two different order types (cf. Figure 12). Regarding the individual picking performance per scheduling interval, the mean value is given by 112 picking orders per scheduling interval. Based on our experiences on common probability distributions of processing times in order picking systems, we assume a discrete log-normal distribution with a variability of $c^2(L) = 0.4$ as appropriate probability distribution for the individual picking performance per scheduling interval (cf. Figure 13). Since there is no information regarding the maximum backlog duration, we assume that it equals the maximum lead time of an order, which is 8 days.

Based on these parameters, the state space of the corresponding Markov chain consists of $8.65 \cdot 10^{83}$ states. For reasons of computing time and memory, it is not possible to analyse levelled order release in the order picking system of the numerical example by means of the Markov chain. Instead, we exploit a simulation model which has been validated based on a comparison with the Markov chain for numerous example scenarios. The results of a Chi-Square Goodness-of-Fit Test with a 5% level of significance show that the empirical steady-state distribution resulting from the simulation model deviates from the exact one to a negligible small extent. Both FCFS-based order release strategies FCFS-DD and FCFS-RAND are also each implemented in a simulation model. To obtain robust results, we perform ten replications each and compute the required workforce level and the performance measures as average values of these replicates.

Initially, we compare the different order release strategies regarding the workforce level which is required to guarantee a β-service level of 98% in the order picking system of the numerical example. When using levelled order release, at least 84 workers have to be assigned to the order picking system. In contrast, in case of FCFS-DD order release, at least 87 workers and in case of FCFS-RAND order release, at least 90 workers have to be assigned to the order picking system. The respective values of the performance measures are summarized in Table 5. Thus, in the order picking system of the numerical example, the required workforce level decreases by three workers (3.13%) when using levelled order release instead of FCFS-DD order release or by six workers (6.29%) when using levelled order release instead of FCFS-RAND order release respectively.

Furthermore, we compare the different order release strategies regarding β-service level for a given workforce level. For a workforce level of 84 workers, the β-service level of the order picking system is 99.09% in case of levelled order release, 96.30% in case of FCFS-DD order release and 93.87% in case of FCFS-RAND order release. This corresponds to an increase in β-service level by 2.79% when using levelled order

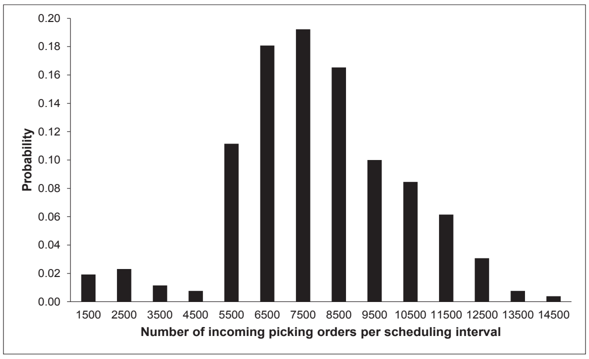

Fig. 11: Number of incoming picking orders per scheduling interval

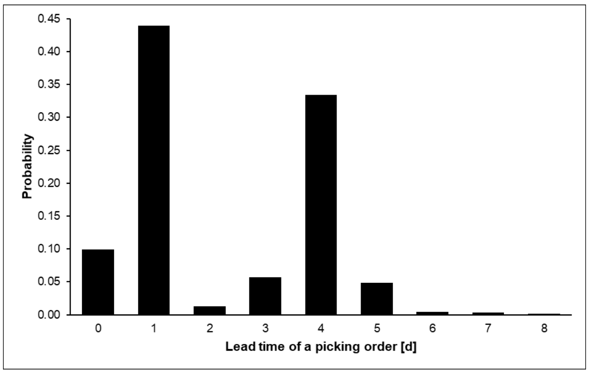

Fig. 12: Lead time of a picking order

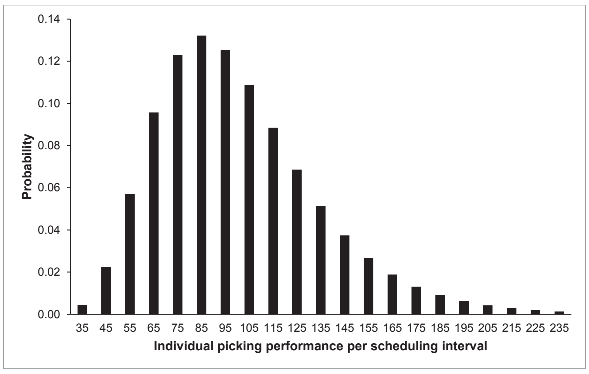

Fig. 13: Individual picking performance per scheduling interval

		LEVELLING	FCFS-DD	FCFS-RAND
Workforce level				
Number of workers	c	83.5000	86.2000	89.1000
Performance measures based on unprocessed orders				
Number of unprocessed picking orders	Q	12206.8671	10338.0943	9606.7460
Number of unprocessed backorders	M	127.5067	113.7588	121.2315
Number of lost sales	S	0	0	0
System utilisation	$\tilde{U}$	0.9426	0.9130	0.8845
Performance measures based on processed orders				
Number of processed picking orders	F	7879.7851	7892.9886	7893.6698
Number of processed backorders	$F^{backlog}$	102.9403	103.0287	113.7604
Number of processed picking orders without failed due date	F^{buffer}	7776.8447	7789.9599	7779.9094
Time difference to order deadline of a processed picking order	D	1.7239	1.9634	2.0560
Backlog duration of a processed backorder	$D^{backlog}$	1.2045	1.0989	1.0486
Time buffer of a processed picking order	D^{buffer}	2.0430	2.0039	2.1012
Service level				
β-service level	SL_β	98.77%	98.81%	98.70%
γ-service level	SL_γ	99.81%	99.84%	99.83%

Tab. 5: Workforce level and performance measures of the order picking system of the numerical example for different order release strategies: levelled order release strategy LEVELLING and FCFS-based reference order release strategies FCFS-DD and FCFS-RAND (average values of ten replications)

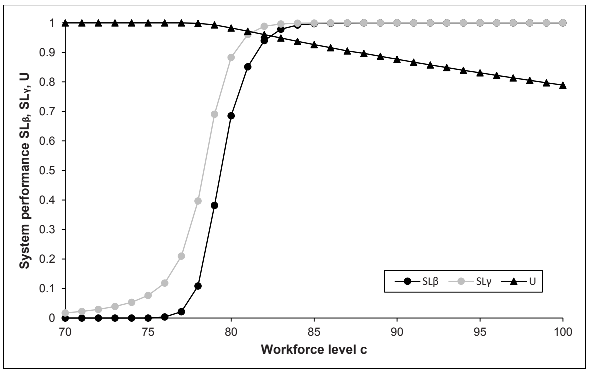

Fig. 14: Impact of workforce level c on several performance measures SL_β, SL_γ, $\tilde{U}$ of the order picking system with levelled order release in the numerical example

release instead of FCFS-DD order release or by 5.22% when using levelled order release instead of FCFS-RAND order release respectively.

In conclusion, for the numerical example, we find that the order picking system benefits from levelled order release: Compared to FCFS-based order release strategies, levelled order release enables either a decrease in the workforce level required to guarantee a certain system performance or it enables an increase in system performance for a given workforce level.

For operational planning and control of order picking systems, further performance measures besides service levels are relevant. For instance, the system utilisation of the order picking system may not exceed 90%. As shown in Figure 14 for levelled order release, system utilisation decreases with increasing number of assigned workers. Thus, by assigning more than the required minimum number of workers to the order picking system, system utilisation of the order picking system decreases. However, including this additional requirement regarding a maximum possible system utilisation raises a conflict of interests: Since system utilisation is independent of the selected order release strategy, assigning additional workers to the order picking system reduces the benefits of levelled order release compared to FCFS-based order release strategies.

7. CONCLUSION AND FUTURE DIRECTIONS

In this publication, we investigate the approach of levelled order release in order picking systems. Based on the concept of Heijunka-levelling in production systems, we derive a levelling concept for order picking systems: There is a fixed picking capacity per order type per scheduling interval which is reserved for order processing of picking orders of this order type in each scheduling interval. Size and sequence of the reserved picking capacities within one scheduling interval are visualised in the levelling pattern. During each scheduling interval, the reserved picking capacity per order type is used to process picking orders of this order type according to ascending due dates. To analyse the performance of the levelling concept, we depict the order picking system with levelled order release as a discrete-time Markov chain and we derive several performance measures from its steady-state distribution, such as the number of unprocessed and processed picking orders, the number of lost sales, the time difference to order deadline, the system utilisation and the service levels. Subsequently, we develop a staffing algorithm for order picking systems with levelled order release as binary search based on the Markov chain. In contrast to the majority of staffing models in practice, our staffing algorithm is not only a rough estimation based on average data and empirical knowledge, but it considers the stochastic variations of customer demand and picking performance. Furthermore, our approach establishes a direct link between the performance of the order picking system and the required workforce level. The models developed in this publication can be used for two different purposes: On the one hand, the discretetime Markov chain exactly determines the system performance of a given order picking system with levelled order release. On the other hand, the staffing algorithm determines the workforce level which is required to guarantee a certain system performance in an order picking system with levelled order release.

Numerical studies show that the variability of incoming picking orders has a negative impact on the system performance, whereas the expected value of the lead time of a picking order has a positive impact on the system performance. These effects increase with increasing traffic intensity of the system. Furthermore, the numerical studies show the smoothing effect of the levelling concept: The variability of processed picking orders is smaller than the variability of incoming picking orders, especially for systems with a high traffic intensity. The structure of these effects is as expected. However, the models developed in this publication enable an exact and quantitative evaluation of these effects. Thus, we are able to quantify the impact of variations in the system parameters, such as variability of incoming picking orders, workforce level or traffic intensity, on the performance measures of the order picking system.

The comparison of levelled order release with FCFS-based order release strategies in a numerical example shows the benefits of levelled order release: Compared to FCFS-based order release strategies, levelled order release enables either a decrease in the workforce level required to guarantee a certain system performance or it enables an increase in system performance for a given workforce level.

Further research can be conducted on generalising the models described here. Different types of workers regarding performance and work time models can be included in the staffing algorithm, since the algorithm actually assumes an identical picking performance and full-time jobs for each worker. Additionally, the models abstract from seasonal fluctuations and auto correlation of the customer demand. Furthermore, to investigate the benefits of levelled order release, the concept of levelled order release should be compared to alternative order release strategies in a comprehensive numerical study. Other future directions concern the extension of the levelling concept to the whole warehouse. Levelling the global workload of a warehouse can be a meaningful approach, if workers can be flexibly assigned to different processes of the warehouse and switches between different processes within one shift are possible. We mentioned flexible workforce planning and levelled order release as two appropriate solution approaches to face the current requirements in manual order picking systems (cf. Figure 1) and only focused on levelled order release in this publication. The combination of these two approaches would be a meaningful further potential future field of research.

ACKNOWLEDGMENTS

The authors wish to thank two anonymous referees for their many helpful comments which led to a much improved form of the paper.

This research is supported by the research project "Smoothing and Levelling in Order Picking Systems" (original title: "Glätten und Nivellieren in der Kommissionierung") by the Bundesministerium für Wirtschaft und Energie (BMWi) (reference number 20509N).

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] N. Boysen, R. de Koster, and F. Weidinger. "Warehousing in the e-commerce era: A survey". In: *European Journal of Operational Research* (2018). DOI: 10.1016/ j.ejor.2018.08.023.

[2] R. de Koster, T. Le-Duc, and K. J. Roodbergen. "Design and control of warehouse order picking: A literature review". In: *European Journal of Operational Research* 182.2 (2007), pp. 481–501. DOI: 10.1016/j.ejor.2006.07.009.

[3] C. Huber. "Throughput Analysis of Manual Order Picking Systems with Congestion Consideration". PhD thesis. Karlsruhe: Karlsruhe Institute of Technology (KIT), 2011.

[4] B. Rouwenhorst et al. "Warehouse design and control: Framework and literature review". In: *European journal of operational research* 122.3 (2000), pp. 515–533. DOI: 10.1016/S0377-2217(99)00020-X.

[5] J. Gu, M. Goetschalckx, and L. F. McGinnis. "Research on warehouse operation: A comprehensive review". In: *European Journal of Operational Research* 177.1 (2007), pp. 1–21. DOI: 0.1016/j.ejor.2006.02.025.

[6] H. Davarzani and A. Norrman. “Toward a relevant agenda for warehousing research: literature review and practitioners’ input”. In: *Logistics Research* 8.1 (2015), p. 78. DOI: 10.1007/s12159-014-0120-1.

[7] Y. Gong and R. B. M. de Koster. “A review on stochastic models and analysis of warehouse operations”. In: *Logistics Research* 3.4 (2011), pp. 191–205. DOI: 10.1007/s12159-011-0057-6.

[8] C. G. Petersen. “An evaluation of order picking policies for mail order companies”. In: *Production and Operations Management* 9.4 (2000), pp. 319–335.

[9] J. Gallien and T. Weber. “To Wave or Not to Wave? Order Release Policies for Warehouses with an Automated Sorter”. In: *Manufacturing & Service Operations Management* 12.4 (2010), pp. 642–662. DOI: 10.1287/ msom.1100.0291.

[10] E. Çeven and K. R. Gue. “Optimal Wave Release Times for Order Fulfillment Systems with Deadlines”. In: *Transportation Science* 51.1 (2017), pp. 52–66. DOI: 10.1287/trsc.2015.0642.

[11] J. K. Liker. *The Toyota way: 14 management principles from the world’s greatest manufacturer*. New York, NY: McGraw-Hill, 2004.

[12] A. Smalley et al. *Creating level pull: A lean production-system improvement guide for production-control, operations, and engineering professionals ; a lean toolkit method and workbook*. Version 1.0, March 2009. Cambridge, MA: Lean Enterprise Institute, 2004.

[13] Y. Monden. *Toyota production system: An integrated approach to just-in-time*. 4. ed. Boca Raton, Fla. and Norcross, Ga.: CRC Press and IIE Inst. of Industrial Engineers, 2012.

[14] K. Furmans. “Models of heijunka-levelled kanbansystems”. In: *5th International Conference on Analysis of Manufacturing Systems-Production and Management*. 2005, pp. 243–248.

[15] C. R. Lippolt and K. Furmans. “Sizing of Heijunka-controlled Production Systems with Unreliable Production Processes”. In: *Lean Business Systems and Beyond*. Ed. by T. Koch. Boston, MA: International Federation for Information Processing, 2008, pp. 11–20.

[16] J. Matzka, M. Di Mascolo, and K. Furmans. “Buffer sizing of a Heijunka Kanban system”. In: *Journal of Intelligent Manufacturing* 23.1 (2012), pp. 49–60. DOI: 10.1007/s10845-009-0317-3.

[17] K. Furmans and M. Veit. “Models of Leveling for Lean Manufacturing Systems”. In: *Handbook of stochastic models and analysis of manufacturing system operations*. Ed. by J. M. Smith and B. Tan. New York, NY: Springer, 2013, pp. 115–138.

[18] T. N. Dhamala and W. Kubiak. “A brief survey of just-in-time sequencing for mixed-model systems”. In: *International Journal of Operations Research* 2.2 (2005), pp. 38–47.

[19] N. Boysen, M. Fliedner, and A. Scholl. “Level Scheduling for batched JIT supply”. In: *Flexible Services and Manufacturing Journal* 21.1-2 (2009), pp. 31–50. DOI: 10.1007/s10696-009-9058-z.

[20] D. T. Jones, P. Hines, and N. Rich. “Lean Logistics”. In: *International Journal of Physical Distribution & Logistics Management* 27.3/4 (1997), pp. 153–173. DOI: 10.1108/09600039710170557.

[21] E. B. Sobanski. “Assessing Lean Warehousing: Development and Validation of Alean Assessment Tool”. PhD thesis. Oklahoma: Oklahoma State University, 2009.

[22] P. Dehdari. “Measuring the Impact of Lean Techniques on Performance Indicators in Logistics Operations”. PhD thesis. Karlsruhe: Karlsruhe Institute of Technology (KIT), 2014.

[23] A. T. Ernst et al. “Staff scheduling and rostering: A review of applications, methods and models”. In: *European Journal of Operational Research* 153.1 (2004), pp. 3–27.

[24] J. Van den Bergh et al. “Personnel scheduling: A literature review”. In: *European Journal of Operational Research* 226.3 (2013), pp. 367–385.

[25] M. Defraeye and I. Van Nieuwenhuyse. “Staffing and scheduling under nonstationary demand for service: A literature review”. In: *Omega* 58 (2016), pp. 4–25.

[26] S. S. Chakravorty. “Improving distribution operations: Implementation of material handling systems”. In: *International journal of production economics* 122.1 (2009), pp. 89–106. DOI: 10.1016/j.ijpe.2008.12.026.

[27] T. van Gils et al. “The use of time series forecasting in zone order picking systems to predict order pickers’ workload”. In: *International Journal of Production Research* 55.21 (2017), pp. 6380–6393. DOI: 10.1080/00207543.2016.1216659.

[28] M. Schleyer and K. Furmans. “An analytical method for the calculation of the waiting time distribution of a discrete time G/G/1-queueing system with batch arrivals”. In: *OR Spectrum* 29.4 (2007), pp. 745–763.

[29] G. Bolch. *Queueing networks and Markov chains: Modeling and performance evaluation with computer science applications*. 2. ed. Hoboken, NJ: Wiley-Interscience, 2006.

[30] T. H. Cormen et al. *Introduction to algorithms*. 3. ed. Cambridge, Mass.: MIT Press, 2009.