International Journal of Production Research

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tprs20

Process time distribution simulation in robotic assembly line balancing

Dawid Stade, Jan Michael Spoor, Martin Manns & Jivka Ovtcharova

To cite this article: Dawid Stade, Jan Michael Spoor, Martin Manns & Jivka Ovtcharova (21 Oct 2024): Process time distribution simulation in robotic assembly line balancing, International Journal of Production Research, DOI: 10.1080/00207543.2024.2416570

To link to this article: https://doi.org/10.1080/00207543.2024.2416570

Published online: 21 Oct 2024.

INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH
https://doi.org/10.1080/00207543.2024.2416570

Process time distribution simulation in robotic assembly line balancing

Dawid Stade[a], Jan Michael Spoor[b], Martin Manns[a] and Jivka Ovtcharova[b]

[a]Mechanical Engineering Department, PROTECH-Institute of Production Technology, University of Siegen, Siegen, Germany; [b]Institut für Informationsmanagement im Ingenieurwesen, Karlsruhe Institute of Technology, Karlsruhe, Germany

ABSTRACT
Feedback control systems utilised in car body construction cause process time variance when compensating for external disturbances. By considering these in robotic assembly line balancing, the risk of cycle time violations can be controlled. This requires knowledge of the underlying process time distributions, which are not known in advance. Therefore, a simulation method is proposed to assess the impact of varying process time distributions on the balancing of robotic assembly lines. The initial step involves acquiring the process times of existing production processes. In the subsequent simulation, these are randomly and repeatedly selected as substitutes for the process times in the balancing of a new robotic assembly line. The impact of process time distribution variations on the result is investigated, and a single solution can be selected. The proposed method is evaluated based on the balancing of a robotic assembly line for a body-in-white rear compartment. Results are compared to normally distributed process times, which is a common assumption for modelling uncertain process times. Both approaches are evaluated utilising actual process time distributions. It is demonstrated that the proposed method yields fewer and less severe underestimations of cycle times, thereby reducing the number of uncontrolled violations of cycle times.

ARTICLE HISTORY
Received 7 May 2024
Accepted 2 October 2024

KEYWORDS
Robotic assembly line balancing; process time estimation; uncertainty; simulation; car body construction

SUSTAINABLE DEVELOPMENT GOALS
SDG 9: Industry; innovation andinfrastructure

1. Introduction

In car body construction, the vehicle's skeletal structure is formed by joining various components. This process is essential for ensuring the vehicle's structural integrity, safety, and durability.

The production schema in car body construction is largely uniform across the automotive industry (Spieckermann et al. 2000). It is segmented into several main assemblies, e.g. the front and rear compartments as well as side frames. The main assemblies are manufactured in distinct blocks, frequently decoupled by buffers. Each block is divided into numerous assembly lines.

Figure 1 shows a representative assembly line in car body construction, which consists of two sequential workstations. Each workstation is standardised and comprises two handling robots (HR1 to HR3), highlighted in green, which load and unload parts. The components are joined by four distinct robots per workstation, namely JR1 to JR8. At the beginning of a cycle, loading robots HR1 and HR2 place parts on the workstations. Subsequently, the respective joining robots establish a connection between the components by performing a series of joints.

Joining robots work simultaneously. Therefore, unloading can only be performed once all robots have finished their tasks. Ultimately, HR2 and HR3 unload both stations. Operations must be completed before a predetermined cycle time is reached, which is calculated based on the expected output of the line (Spieckermann et al. 2000).

During the planning of car body construction, several decisions must be made. One of these problems involves the combinatorial optimisation of assigning a set of product assembly operations, such as joining processes, to workstations and robots. Operation allocation is subject to technical restrictions, such as precedence constraints, that result from the assembly sequence, and their process times. This problem is known in the literature as the Robotic Assembly Line Balancing Problem (RALBP), which was first introduced by Rubinovitz and Bukchin (1991). It is an extension to the more general Assembly Line Balancing Problem by also selecting a tool required to perform the joining operation (Kim and Park 1995).

In car body construction, various joining techniques are utilised. These are, among other factors, selected

CONTACT Dawid Stade dawid.stade@uni-siegen.de Mechanical Engineering Department, PROTECH-Institute of Production Technology, University of Siegen, Siegen 57076, Germany

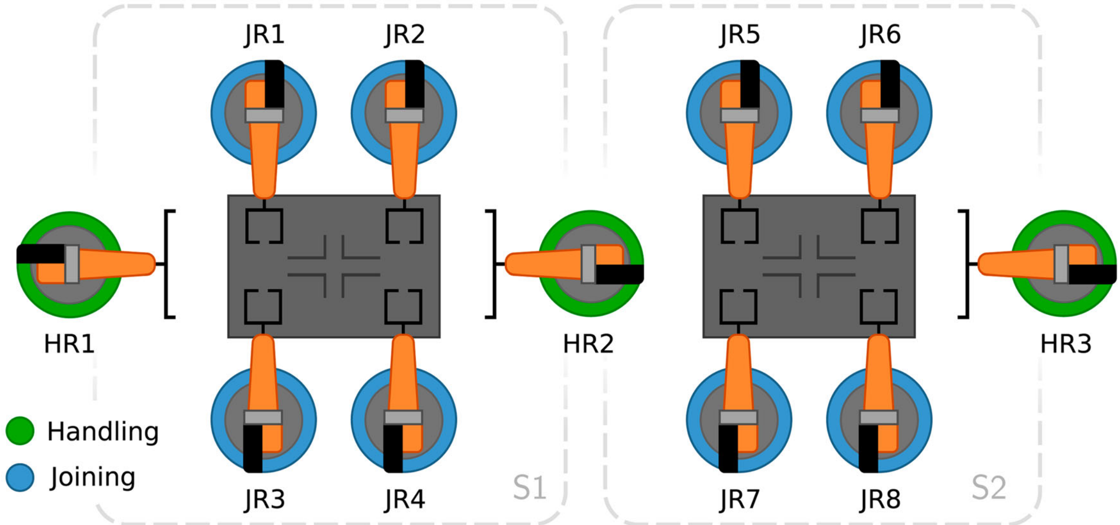


Figure 1. A typical car body construction assembly line consists of several workstations, here arranged sequentally, each equipped with tools such as handling and joining robots, which pick, place or join parts.

based on the materials of the joining partners. Similar materials can be welded utilising techniques like resistance spot welding (RSW) or laser welding. Adhesive bonding and mechanical fastening, on the other hand, are especially used to join diverse materials. To allow for fixing interior parts to the body in the final assembly, standard elements such as bolts and nuts have to be attached in car body construction. Arc stud welding (ASW) is frequently employed for this purpose.

All joining techniques in car body construction require highly specialised equipment, which often employs feedback control systems (FCS) to ensure process stability in the event of external disturbances. The FCS keeps process variables close to their intended values. Although they respond quickly to changes, adaptation takes time. An assembly line in car body construction of Mercedes-Benz is examined by Stade and Manns (2023a). Weld control systems for RSW and ASW contain several process steps that are driven by FCS and result in varying process times. In the case of RSW, the process step of closing the weld gun and welding, vary in time. This is especially cumulative in workstations, where multiple robots process a series of joints per cycle. The analysis of 194 production cycles of an RSW workstation used in car body construction shows that process times affect cycle time, which, in turn, fluctuates within a 5- to 95-percentile range of 0.94 s. As the welding equipment is utilised globally across all manufacturing plants of the company, and RSW is the most extensively used joining technique (see Stavropoulos et al. 2022), the impact of varying process times is substantial. This demonstrates the importance of considering varying processing times in the planning of car body construction. However, knowledge of the underlying process time distributions is required.

As 80% of the tools and equipment utilised in automotive production is standardised (Al-Zaher and ElMaraghy 2014), it is likely that joining equipment that is meant to be used in a new assembly line has already been utilised in ongoing production. Therefore, historical data can be employed to predict process time distributions during planning of car body construction. A distribution model may be parametrised with the help of historical planning data, e.g. material pairings and sheet thicknesses, as well as the respective process time distributions. In the RSW process under investigation, the shift and range of process time distributions can be predicted utilising planning data; however, it was not feasible for the shape of the distributions.

In practice, various suppliers are responsible for the welding setup. Although the company provides standard welding parameters as a starting point, the further course of action is unstandardised. Therefore, each supplier or even weld engineer may act differently. Furthermore, during production setup, requirements frequently change, which requires some weld controls to be repeatedly adapted. This may account for the challenges encountered in predicting process time distributions based on planning data.

Apparently, standardisation of the setup process could help address this issue, but is challenging and long-lasting to impose. Alternatively, the engineer in charge and the setup procedure could be documented and included in the distribution model. However, this has to be known already at the planning stage, which is not possible in practice.

Thus, the main contribution of this work is a simulation method that allows for assessing the effects of various process time distributions gathered from ongoing production on the balancing of robotic assembly lines. Based on the simulation results, it is possible to select a suitable balancing solution with a certain degree of confidence. The method enables the decision maker to determine an appropriate compromise between the risk of cycle time violations and cost reduction. It comprises three steps:

(1) For data acquisition and preprocessing, process times of a joining technique are collected from ongoing production. An outlier elimination, resampling, normalisation, and standardisation is conducted.
(2) In the simulation, samples are randomly selected from collected data. These are utilised as substitutes for process time distributions and used to solve the RALBP. The process is repeated if the balancing results alter beyond a predetermined threshold.
(3) The balancing results obtained from the simulation are used to investigate the impact of varying process time distributions on the RALBP. With a certain level of confidence, this allows for the selection of one of the outcomes and finding a compromise between elevated risks of cycle time violations and cost reduction.

The proposed method is evaluated utilising process data obtained from a car body construction plant of Mercedes-Benz during the summer of 2023. It contains process times and planning data for approx. 4000 RSW joints that are spread over 25 assembly lines, 115 workstations and 255 robots.

The remainder of this work is organised as follows: Section 2 presents an overview of existing solutions in research. The proposed method is described in Section 3 and then evaluated based on an example from car body construction. Therefore, the experimental setup and application-specific method implementation are given in Section 4. Results are presented in Section 5, which are discussed in Section 6. Finally, Section 7 concludes the findings of this study.

2. State of the art

2.1. Robotic assembly line balancing for the automotive Industry

In the production planning of robotic assembly lines, balancing is an important part. As described above, this can be formulated as an optimisation problem (Rubinovitz and Bukchin 1991), specifically RALBP, which extends the well-known Assembly Line Balancing Problem with the additional assignment of robots and tools. Due to its complexity, the exact solution of the RALBP is usually found only for small-scale problems (Chutima 2022). Due to the size of car body construction assembly lines, other solution techniques are also considered in the literature.

One approach is to segregate the two problems of line balancing and equipment selection and address them independently. An interactive three-step method is proposed by Michalos et al. (2015). First, it offers a guide for manual balancing of assembly lines. Based on this initial configuration, either an exhaustive search or a heuristic are employed to select equipment. A set of design alternatives are generated from this step. Evaluation metrics are determined using discrete event simulations and analytic formulas, such as the required space and line availability. Finally, the best design is chosen in a multi-criteria evaluation. The method aims to support the decision maker by exploring design alternatives and therefore improving assembly line configurations. A case study indicates that a car body construction assembly line can be improved by 15% to 60%.

Colledani et al. (2016) make use of heuristics to find a solution for the RALBP. In a design and management methodology comprising four individual modules, a heuristic is employed to render numerous assembly line configurations in a first step. Additional modules are available to (i) detail the configurations and model the dynamic behaviour, (ii) optimise the lot sizes and confirm the production plan, as well as (iii) reconfiguring the assembly line for different production scenarios, e.g. changes in production volume.

More recently, models have been introduced for the RALBP with a scope on car body construction, for which exact solutions can be found in a reasonable amount of time. It appears that this is facilitated by the availability of superior computational capabilities. However, both approaches are designed to reduce complexity. Michels et al. (2018), for instance, group operations based on their joining techniques and treat them as copies, thereby reducing the size of the solution space. In addition, cubic constraints are linearised. This enables the attainment of optimal solutions in three practical case studies in less than one hour of computation. However, computation time may be vulnerable to the grouping of joining operations. Therefore, various joining techniques may increase the computational effort and hinder the practical applicability of the model.

According to Hagemann and Stark (2020), an automotive body-in-white and the corresponding car body construction are typically organised hierarchically. This implies that a joining step, comprising all operations to join a minimum of two parts into a subassembly, equates to a single assembly line. All of these are interconnected, so that the output of one assembly line is the input for another. Since the number of operations per joining step is small, the RALBP can be optimally solved for every joining step.

None of the aforementioned publications consider uncertain processing times. Instead, deterministic values are required, which have to be determined by the decision maker. In the planning of manual assembly lines, well-known methodologies such as Methods-Time Measurement (MTM) can be used for this purpose. In the industrial practice of early planning in car body construction, however, the planner utilises standard process times, which have been established based on domain knowledge and experiments (see Hagemann 2022). Standard process times are valid only for a specific context, e.g. the joining technique, a group of material pairings, and the structural importance of the joint. The context of the standard process times is established by domain experts and is constantly updated. In each context, standard process times are calculated based on average process times gathered from experiments or similar production systems. Additional safety margins are usually added to compensate for various unknown parameters. The process of updating or introducing new contexts and standard process times is time-consuming.

Therefore, Mucientes et al. (2008) propose a fuzzy rule-based process time estimation method that intentionally maintains domain knowledge. Information on manufacturing contexts, originally defined by domain experts, as well as other influencing variables can be included. These are selected for a weighted parametrization of a polynomial that represents the process times. A genetic algorithm is then used to find the best combination of weights so that the process time can be estimated using the input variables. Hence, the method can be utilised to refine context rules. However, process times from manufacturing may include contexts which are not yet known. This would result in inaccurate estimates. To identify new contexts in production, Ringsquandl, Lamparter, and Lepratti (2015) introduce a framework that automatically detects context changes in production. A regression model is constantly updated by production data. Once, the prediction error exceeds a given limit, a context change has been identified. The framework utilises data from product lifecycle management, manufacturing execution and enterprise resource planning systems to identify the causes for context changes.

An alternative approach to the industrial practice is proposed by Zwicker and Reinhart (2016) that adopts the concept of MTM by dividing assembly operations into several steps. Unlike MTM, it does not exclusively rely on empirical data to determine the process times of each step. Instead, the usage of analytic models and simulations is also suggested. Another method is introduced by Denkena, Dittrich, and Settnik (2022), which propose a data-driven similarity-based process time estimation method that neglects any predefined contexts. Instead, process and setup times of a workpiece to be planned are estimated based on feedback data from other workpieces. Feedback data includes information on the workpieces that are also available in planning, such as dimensions and materials. Furthermore, it encompasses actual process and setup times from the past. The attribute impact is calculated using a feature importance analysis. Based on these, the similarity between a workpiece to be planned, and all other workpieces is computed. A group of the most similar workpieces is selected from this. The mean process time of the neighbours is used as the estimated process time for planning.

2.2. Uncertain process times in assembly line balancing

In the more general field of assembly line balancing, uncertain process times are modelled by stochastic, fuzzy, and interval representations (Boysen, Schulze, and Scholl 2022). Each of the approaches requires different levels of information and results in corresponding levels of details. Joslyn and Fersont (2004) suggest modelling uncertainty by consuming the exact amount of information that is available. The selection of an adequate approach, therefore, heavily depends on its application.

In stochastic approaches, uncertain process times are modelled by probability distributions. As a result, the parametrised class of distribution has to be known in advance, which may be difficult in certain settings. Most contributions therefore assume normally distributed process times, defined by their respective means and standard deviations (see Ağpak and Gökçen 2007; Fathi et al. 2019; Özcan 2018; Şahin and Tural 2023; Tiacci 2017). This reduces the required level of information, but also makes an assumption that may not be adequate in practice. The genetic algorithm proposed by Stade and Manns (2023b) considers any probability distribution to represent process times in RALBP. It adds a decoding procedure to the original algorithm of Levitin, Rubinovitz, and Shnits (2006) that determines the risk of cycle time violations and allows controlling it using a predefined probability limit. This approach requires knowledge of the underlying probability distributions.

Fuzzy approaches do not require a distribution assumption but a given membership function, which, to a limited extent, allows adjusting the level of information. Triangular fuzzy numbers, for instance, are represented by three parameters, which are the lowest expected, most likely, and highest expected values (see Zacharia and Nearchou 2013; Zacharia, Xidias, and Nearchou 2024). Trapezoidal fuzzy numbers, as employed by Salehi, Maleki, and Niroomand (2018), utilise one more parameter.

In robust assembly line balancing, the level of information on process times is the lowest. Numerous contributions adhere to the definition of robustness suggested by Bertsimas and Sim (2003) (see Gurevsky et al. 2013; Meng et al. 2023; Pereira and Álvarez-Miranda 2018). Each operation has a minimum, nominal, and maximum process time. The level of uncertainty is determined by Γ, which is the number of operations per workstation that need their maximum processing time. A high value of Γ therefore translates to a more risk-averse balancing. In contrast, $\Gamma = 0$ does not account for any variability. However, according to Hazır and Dolgui (2013), an increase in the number of operations per workstation is associated with a higher probability of process time deviation. Hence, a proportion of uncertainty θ is proposed, which establishes the number of deviated process times in relation to the number of operations in a workstation.

Other robust approaches are also considered in the literature. Instead of searching for a robust balance given deviating process times and a degree of uncertainty, a stability analysis is conducted. A stability radius is employed, which is the maximum independent process time deviation that any task can witness before a given assembly line configuration becomes unstable, i.e. a given cycle time cannot be met (see Sotskov et al. 2015; Sotskov, Dolgui, and Portmann 2006). Rossi et al. (2016) introduce a model for maximising the stability radius in line balancing, which is a special case of this approach. A literature review on the field of robust assembly line balancing is conducted by Hazır and Dolgui (2019).

3. Method

Given a robotic assembly line balancing problem, it is necessary to allocate tasks of size T to workstations and robots in an assembly line in such a way that a minimum number of workstations and robots is required. Each task t represents the process of one joining element being completed, for example, a welding joint or adhesive seam. A precedence matrix $\boldsymbol{P}$ restricts the order in which tasks can be accomplished, with $p_{ij} = 1$ indicating that task i must precede task j.

Each task requires a varying process time, expressed by the random variable X_t, which is not yet known in the planning stage. The proposed method uses production data from existing assembly lines filled with the same kind of equipment to simulate their effect on the balancing of a new robotic assembly line. All relevant variables are listed in Table 1. Relevant parameters for the method are given by Table 2. The proposed method is presented in Figure 2. It consists of three steps:

Table 1. Variables of the proposed method.

Variable	Description
t	Task of the balancing problem, $t \in \{1,\dots,T\}$
j	Joining element, $j \in \{1,\dots,J\}$
$n^{(j)}$	Number of process time samples of joining element j, $n^{(j)} \in \mathbb{N}$
$Y^{(j)}$	Set of process times of joining element j, $Y^{(j)} = \{y_1^{(j)},\dots,y_{n^{(j)}}^{(j)}\}$
$\boldsymbol{D}_{J\times J}$	Distance matrix, d_{ij} maps the distance from joining element i to j, $d_{ij} \in \mathbb{R}$
$Q^{(j)}$	Number of bins for process time histogram of joining element j, $Q^{(j)} \in \mathbb{N}$
$q^{(j)}$	Bin of process time histogram of joining element j, $q^{(j)} \in \{1,\dots,Q^{(j)}\}$
$H^{(j)}$	Process time histogram of joining element j, $H^{(j)} = \{h_1^{(j)},\dots,h_{Q^{(j)}}^{(j)}\}$
U	Multiset of balanced joining elements j, $U = \{1^{u(1)},\dots,J^{u(J)}\}$
$\bar{y}^{(j)}$	Mean process time for joining element j, $\bar{y}^{(j)} \in \mathbb{R}^+$
$\hat{Y}^{(j)}$	Normalized and standardised set of process times of joining element j
z	Objective of the RALBP, $z \in \mathbb{R}^+$
s	Number of workstations in balancing result, $s \in \mathbb{N}$
Δs	Relative utilisation of last workstation, $0 \le \Delta s \le 1$
w	Workstation, $w \in \{1,\dots,W\}$
r	Robot, $r \in \{1,\dots,R\}$
x_{trw}	Binary variable, $x_{trw} = 1$ assigns task t to robot r and workstation w
$f_{rw}(u)$	Probability density function of cycle time u for robot r in station w
$F_{rw}(v)$	Cumulative distribution function of cycle time v on robot r in station w
$\tilde{F}_w(v)$	Cumulative distribution function of maximum cycle time v on station w
s_w	Variable, $s_w = 1$ means at least one task being assigned to station w
B	Sequence of balancing solutions
t'	Substitute for task t
Z	Sequence of objective values, $Z = \{z_b \mid b \in B\}$
i^*	Index of selected balancing solution, $i^* \in \mathbb{N}$
b^*	Selected balancing solution

Note: Table of symbols and corresponding descriptions of the proposed method.

Data preparation Process data from ongoing production, utilising the same kind of equipment as the tasks of the balancing issue, is collected. In the outlier elimination, process data in the area of substantial low density is removed from the dataset. Then, the dataset is balanced using rejection sampling, followed by normalisation and standardisation. The collected data represents the dataset for the method.

Simulation of process time distributions Based on the dataset from the previous step, a simulation to assess the impact of varying process time distributions on the RALBP is conducted. In each replication, process time distributions are randomly chosen to replace the tasks' process times of the balancing problem. The RALBP is then solved by using these

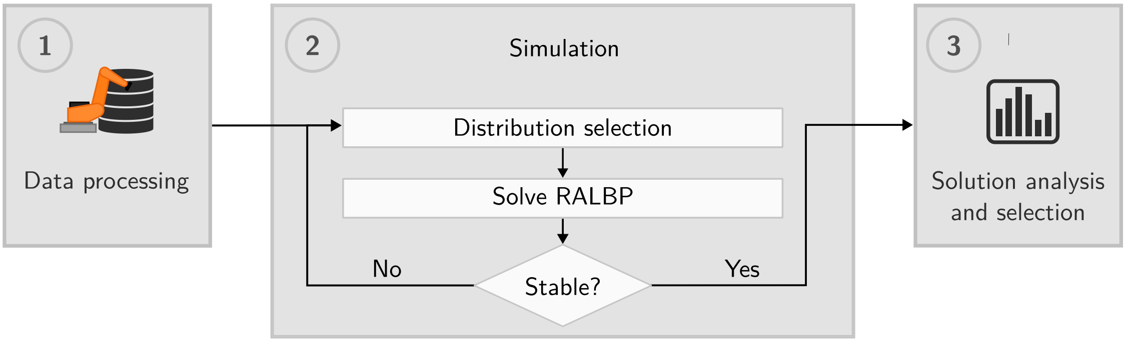


Figure 2. Illustration of the proposed simulation method.
Long Description. Based on collected and preprocessed data, the proposed simulation method repeatedly substitutes process time distributions with samples from the dataset. A stability analysis is conducted to terminate the simulation once its results remain stable. The method allows for assessing the impact of process time distribution variance on the RALBP and select an appropriate solution for further planning.

Table 2. Parameters of the proposed method.

Parameter	Description
T	Number of tasks, $T \in \mathbb{N}$
X_t	Random variable to represent process times of task t, $X_t \in \mathbb{R}^+$ [s]
$\boldsymbol{P}_{T\times T}$	Precedence matrix, $p_{ij} = 1$ means that task i precedes j, $p_{ij} \in \{0,1\}$
J	Number of joining elements in the dataset, $J \in \mathbb{N}$
$MinPts$	Minimum number of neighbours in a cluster, $MinPts \in \mathbb{N}$
Eps	Radius to search for neighbours, $Eps \in \mathbb{R}$
N	Size of the balanced dataset, $N \in \mathbb{N}$
$\hat{t}$	Deterministic standard time, $\hat{t} \in \mathbb{R}^+$ [s]
$\hat{r}$	Deterministic standard range, $\hat{r} \in \mathbb{R}^+$ [s]
R	Number of robots per workstation, $R \in \mathbb{N}$
α	Probability limit for cycle time violations, $0 \leq \alpha \leq 1$
C	Predetermined cycle time, $C \in \mathbb{R}^+$ [s]
W	Maximum number of workstations, $W \in \mathbb{N}$
r_{init}	Initial replications of simulation, $r_{init} \in \mathbb{N}$
r_{inc}	Increment in replications, $r_{inc} \in \mathbb{N}$
$Perc$	Percentiles for stability analysis, $Perc = \{perc \in \mathbb{R} \mid 0 \leq perc \leq 100\}$
r_{win}	Window size for stability analysis, $r_{win} \in \mathbb{N}$
$\Delta perc_{\min}$	Minimum relative change of percentile, $0 \leq \Delta perc_{min} \leq 1$
$r_{\max}$	Maximum number of replications in simulation, $r_{\max} \in \mathbb{N}$
π	Confidence level for solution selection, $0 \leq \pi \leq 100$

Note: Table of parameters and corresponding descriptions of the proposed method.

substitutes. A stability analysis on the simulation results decides whether additional replications are carried out or the simulation is terminated.

Solution analysis and selection The results of the simulation are examined. Given a degree of confidence, the decision maker can select one solution of the simulation.

3.1. Data preparation

In the first step of the method, process times for J joining elements are gathered from an existing assembly line for a predetermined time span. This requires the joining elements to employ the same kind of equipment as the tasks t in the balancing problem will. As Equation (1) shows, the collection yields a set of process times $Y^{(j)}$ of $n^{(j)}$ samples for each joining element $j \in \{1, \ldots, J\}$, which is considered the dataset for the method:

$$Y^{(j)} = \left\{ y_1^{(j)}, y_2^{(j)}, \ldots, y_{n^{(j)}}^{(j)} \right\}, \quad j \in \{1, \ldots, J\} \tag{1}$$

3.1.1. Outlier elimination

To eliminate outlier process data from the dataset, the approach introduced by Spoor (2022) is employed. The DBSCAN algorithm, which is based on density-based clustering, is employed to identify samples with a significantly lower density, referred to as outliers. DBSCAN defines a cluster by identifying the core points, which are points with at least $MinPts$ neighbours within a radius of Eps. Clusters are formed around these core points, including all directly connected points. If a point can be reached from a core point, it constitutes a component of the cluster. Points that are not connected to any particular cluster are marked as outliers. DBSCAN can detect clusters of varying shapes and densities using this method.

The principle of DBSCAN can be demonstrated through a numerical example that encompasses the points $1, 2, 2.5, 3$, and 10. Using DBSCAN with $Eps = 1.5$ and $MinPts = 2$, points 1, 2, 2.5, and 3 form a cluster. They are within distance Eps and meet the density requirement, i.e. a minimum of two points are located within a radius of 1.5. Point 10 is marked as an outlier because it has no neighbours within distance Eps.

A distance matrix $\boldsymbol{D}_{J\times J}$ is required, in which each element d_{ij} maps the distance between the joining elements i and j. To compute these distances, process times $Y^{(j)}$ are first compressed into histograms $H^{(j)}$, see Equation (2). Therefore, the data range of $Y^{(j)}$ is divided into $Q^{(j)}$ consecutive and non-overlapping bins, $q^{(j)} \in \{1, \ldots, Q^{(j)}\}$. Each histogram value, $h_q^{(j)}$, maps a probability density of all process times of a joining element, which fall into bin $q^{(j)}$. To determine the appropriate number of bins for all histograms, the median of individual bin widths is selected, of which each is determined by the estimator of Freedman and Diaconis (1981).

$$H^{(j)} = \left\{h_1^{(j)}, h_2^{(j)}, \ldots, h_{Q^{(j)}}^{(j)}\right\}, \quad j \in \{1, \ldots, J\} \tag{2}$$

Consequently, the Earth Movers Distance (EMD), as defined by Rubner, Tomasi, and Guibas (2000), is used to compute the distances between pairs of histograms. Information between divergent bins is considered by the EMD, unlike bin-by-bin measures such as Minkowski-Form distance. The EMD formalises the idea of calculating the minimal average cost to transform one probability distribution into another using a cost function $c(x, y)$, which specifies the cost to move a unit mass from x to y. It is, therefore, a solution to the popular transport problem. The concept can be illustrated by the example of two histograms. One can be regarded as a distribution of sand piles, whereas the other is a distribution of holes. The EMD represents the minimal amount of work required to fill in the gaps. The term *work* refers to the amount of sand being moved multiplied by a ground distance, which is calculated using the Manhattan metric.

3.1.2. Balancing

The distribution of process time distributions in the dataset may vary from that of actual process times. A dataset may have numerous bimodal distributions, for instance. In contrast, the actual process time distributions may mainly consist of unimodal distributions. To avoid underrepresentation of these unimodal distributions in simulation, a balancing of the dataset is proposed.

Therefore, a rejection sampling procedure is conducted, which resamples the dataset so that the means of process times are uniformly distributed. This approach aims to equally represent diverse types of process time distributions within the dataset and minimise the possibility of underrepresentations. Therefore, a multiset U is defined, in which each joining element j may appear multiple times, e.g. $u(1) = 3$. Likewise, it is possible that an element is not represented, e.g. $u(2) = 0$. The multiset U is populated using the following procedure.

(S1) Calculate the mean process times $\bar{y}^{(j)}$.
(S2) Split the mean process times into bins using the estimator of Freedman and Diaconis (1981).
(S3) Randomly select a bin. Find all joining elements whose mean process time falls within this bin. If there is no such element, repeat this step. Otherwise, randomly select one and add its index to the multiset U. If its size equals the desired size N, stop the procedure. Otherwise, repeat this step.

3.1.3. Normalization and standardisation

Standard process times are commonly adopted in planning of car body construction. These are typically determined by the average of historical process times, including additional safety margins to compensate for unknown parameters. Uncertain process times are part of these. Hence, the process times of the dataset are normalised in such a manner that the median process time of each joining element equals the predetermined standard process time $\hat{t}$; see Equation (3). Furthermore, the process times are standardised, so that the 5- to 95-percentile range equals a predetermined standard range $\hat{r}$, see Equation (4).

$$\dot{Y}^{(j)} = Y^{(j)} + \left(\hat{t} - \text{Median}\left(Y^{(j)}\right)\right) \tag{3}$$

$$\hat{Y}^{(j)} = \dot{Y}^{(j)} \frac{\hat{r}}{P_{95}\left(\dot{Y}^{(j)}\right) - P_5\left(\dot{Y}^{(j)}\right)} \tag{4}$$

3.2. Simulation of process time distributions

Based on the prepared dataset, a simulation algorithm is proposed to assess the effect of varying process time distributions on balancing robotic assembly lines. Therefore, in each replication, the tasks' process time distributions are replaced by a random set of distributions from the dataset and used to solve the RALBP. The algorithm terminates if the balancing results remain stable for a certain number of replications.

According to Rubinovitz and Bukchin (1991) proposal, the Robotic Assembly Line Balancing Problem is assuming that a singular robot will be assigned to each workstation. Task process times are assumed to be deterministic. Both assumptions are not valid in the practice of car body construction; therefore, the assumptions for RALBP are modified as follows:

- The process time of a task is represented by a random variable X_t. Tasks cannot be subdivided.
- Each workstation is equipped with R robots.
- The assignment of tasks to workstations and robots is limited by a probability limit α for violating a predetermined cycle time C and precedence constraints.
- The precedence relationships for all tasks are known, unchangeable, and represented by a precedence matrix $\boldsymbol{P}$.
- Material handling, setup, and tool changing times are neglected or included in the process times.
- The assembly line is balanced for a single product.

Table 3. Variables of the mixed-integer programming model.

Variable	Description
z	Objective function, $z \in \mathbb{R}^+$
s	Number of workstations, $s \in \mathbb{N}$
Δs	Utilization of the last workstation, $0 \leq \Delta s \leq 1$
w	Workstation, $w \in \{1, \ldots, W\}$
t	Task, $t \in \{1, \ldots, T\}$
r	Robot, $r \in \{1, \ldots, R\}$
x_{trw}	Binary variable, $x_{trw} = 1$ assigns task t to robot r in station w
$f_{rw}(u)$	Probability density function of cycle time u for robot r in station w
$F_{rw}(v)$	Cumulative distribution function of cycle time v on robot r in station w
$\tilde{F}_w(v)$	Cumulative distribution function of maximum cycle time v on station w
s_w	Binary variable, $s_w = 1$ means that station w has tasks assigned

Note: Table of symbols and corresponding descriptions of the mixed-integer programming model.

Table 4. Parameters of the mixed-integer programming model.

Parameter	Description
T	Number of tasks, $T \in \mathbb{N}$
X_t	Random variable to represent process times of task t, $X_t \in \mathbb{R}^+$ [s]
$\boldsymbol{P}_{T\times T}$	Precedence matrix, $p_{ij} = 1$ means that task i precedes task j, $p_{ij} \in \{0, 1\}$
R	Number of robots per workstation, $R \in \mathbb{N}$
α	Probability limit for cycle time violations, $0 \leq \alpha \leq 1$
C	Predetermined cycle time, $C \in \mathbb{R}^+$ [s]
W	Maximum number of workstations, $W \in \mathbb{N}$

Note: Table of parameters and descriptions of the mixed-integer programming model.

- The workstations and robots are not limited in their availability.

3.2.1. Mixed-Integer programming model for the RALBP

The variables utilised in the mathmatical problem definition are presented in Table 3. The parameters for the model are given by Table 4.

The planning of a car body construction assembly line often aims to minimise the number of workstations with a given cycle time, which is a RALBP type 1. This is represented by the objective function of the model given by Equation (5).

$$\min z = s + \Delta s \tag{5}$$

The first part represents the number of workstations s. The second part refers to the utilisation of the last workstation Δs, which is affected by the probability limit α. The alteration of this limit pertains to a modification in the utilisation of the assembly line. Therefore, for instance, increasing α allows for an increase in the utilisation of the assembly line while reducing the utilisation of the last workstation. Thus, the minimisation of Δs causes the probability limit to be exhausted without the need to save a whole workstation. The rest of the model is provided below:

$$\Delta s \geq \frac{\tilde{F}_s^{-1}(0.01)}{C} \tag{6}$$

$$f_{rw}(u) = \mathop{\circledast}_{t=1}^{T} f_{X_t}(u)^{x_{trw}} \tag{7}$$

$$x_{trw} \in \{0, 1\}; \quad \forall\, t, r, w \tag{8}$$

$$F_{rw}(v) = \int_{-\infty}^{v} f_{rw}(u)\,\mathrm{d}u \tag{9}$$

$$\tilde{F}_w(v) \geq F_{rw}(v); \quad \forall\, r \tag{10}$$

$$\tilde{F}_w(C) \leq s_w(1 - \alpha) \tag{11}$$

$$s_w \in \{0, 1\}; \quad \forall\, w \tag{12}$$

$$s \geq w^{s_w}; \quad \forall\, w \tag{13}$$

$$\sum_{w=1}^{W}\sum_{r=1}^{R} x_{trw} = 1; \quad \forall\, t \tag{14}$$

$$p_{ij}\left(\sum_{w=1}^{W}\sum_{r=1}^{R} w \cdot x_{irw} - \sum_{w=1}^{W}\sum_{r=1}^{R} w \cdot x_{jrw}\right) \leq 0; \quad \forall\, i, j \tag{15}$$

Constraint (6) defines the relative utilisation of the last workstation as the first percentile of the last workstation's process times divided by predefined cycle time C. The selection of the first percentile is supported by the normalisation of process time distributions. Normalization is implemented in such a way that the 95th percentile of process times for each task equals its standard time. Therefore, since the minimum relative utilisation is part of the objective function, tasks with large spread widths will be assigned to the last station. This leads to lower cycle time variations in the assembly line, which may be beneficial for its performance.

The utilisation of a workstation and robot is defined by Equation (7), which stipulates the convolution of the task times' probability density functions that are assigned to the respective workstations and robots by binary variable x, see Equation (8). In Equation (9), this is integrated to obtain the respective cumulative distribution function (CDF). The CDF is used to determine the maximum CDF for the process time of a workstation, see Equation (10). Constraint (11) imposes a maximum probability of process times per workstation remaining below the specified cycle time of $1 - \alpha$. It also sets the binary variable s_w, see Equation (12), which is used to determine the maximum number of workstations s, see Equation (13). Constraint (14) ensures that each task is allocated to exactly one workstation and robot. The precedence constraints defined by the precedence matrix $\boldsymbol{P}$ are enforced by Constraint (15).

Algorithm 1 Simulation-Algorithm

in: Precedence matrix $\boldsymbol{P}$; Cycle time C; Probability limit α; Multiset of the balanced dataset U; Number of tasks T, Sequence of percentiles $Perc$; Initial replications r_{init}; Max. replications r_{max}; Increment in replications r_{inc}; Window size r_{win}; Min. relative change of a percentile $\Delta perc_{min}$

out: Sequence of balancing solutions B

1: $B \leftarrow \langle\,\rangle\,; r \leftarrow 0$
2: **while** $r \leq r_{max}$ **do**
3:     **if** $r < r_{init}$ **then** $\Delta r \leftarrow r_{init}$ **else** $\Delta r \leftarrow r_{inc}$ **end if**
4:     **for** $r' \leftarrow 0 \ldots \Delta r$ **do**
5:         $t' \leftarrow \langle\, \textsc{random}(U)$ **for all** $i \in \{1, \ldots, T\}\,\rangle$
6:         $B \leftarrow B \,\|\, \langle \text{RALBP}(\boldsymbol{P}, t', C, \alpha)\rangle$ ▷ Solve RALBP using substitutes t'
7:     **end for**
8:     $r \leftarrow r + \Delta r$
9:     **if** $r > r_{win}$ **then**
10:         $Z \leftarrow \langle\, z_b$ **for all** $b \in B\,\rangle$
11:         **for** $i \leftarrow (|Z| - r_{win}) \ldots |Z|$ **do** ▷ Calculate change of percentiles in window
12:             **for all** $perc \in Perc$ **do**
13:                 $Z' \leftarrow \textsc{subseq}(Z, 0, i-1)$ ▷ Subsequence from 0 to $(i-1)$-th element
14:                 $Z'' \leftarrow \textsc{subseq}(Z, 0, i)$ ▷ Subsequence from 0 to i-th element
15:                 $perc' \leftarrow \textsc{percentile}(Z', perc)$
16:                 $perc'' \leftarrow \textsc{percentile}(Z'', perc)$
17:                 $\Delta perc \leftarrow |perc' - perc''|/perc'$ ▷ Relative change of percentile
18:                 **if** $\Delta perc > \Delta perc_{min}$ **then**
19:                     **goto** line 2
20:                 **end if**
21:             **end for**
22:         **end for**
23:     **end if**
24: **end while**
25: **return** B

3.2.2. Simulation algorithm

In the simulation, the mixed-integer programming model of the RALBP is repeatedly solved while random substitutes from the balanced dataset U replace the tasks' process times. A predetermined number of replications r_{init} is solved at first. Then, an additional number of replications r_{inc} is successively added. The process is repeated until a stability analysis on a given set of percentiles $Perc$ of the simulation results shows no more significant changes. A confidence interval (CI) can be established by providing a set of percentiles, such as the 5th and 95th. It should be noted that this CI pertains to the distribution of simulation results and is not a statistical estimator's CI. Once the results remain stable, the simulation terminates and the results B of the balancing replications can be examined. The pseudocode for the simulation algorithm is provided in Algorithm 1. The components of the algorithm are described below:

Solving the RALBP For each replication, the substitutes t' are created by randomly selecting process time distributions from the multiset U. In addition to the precedence matrix $\boldsymbol{P}$, the probability limit α and cycle time C are employed to solve the RALBP. The balancing solution is added to the solution sequence B.

Stability Analysis Once the RALBP has been solved, the stability analysis to determine if additional replications are required is conducted. Hence, the evaluation of the objective value of each solution z_b, for which the objective function of the problem definition is utilised, is performed. In the next step, relative changes of all percentiles given by $Perc$ are monitored during a window of r_{win} replications. The last r_{win} fitnesses are iterated, and the relative change of a percentile between the current and previous iteration is computed. The simulation is augmented with additional replications if any of the relative changes exceeds the minimum threshold $\Delta perc_{min}$. Otherwise, the simulation is terminated and the sequence of solutions B is returned.

3.3. Solution analysis and selection

As the stability of the simulation results is ensured by the stability analysis conducted during simulation, it is anticipated that the sequence of balancing solutions B will accurately represent the impact of varying process time distributions on the balancing of a robotic assembly line. Hence, the decision maker can select one of the solutions with a degree of confidence π, which is utilised to select the solution whose objective value is closest to the $(100-\pi)$-th percentile of all values. The confidence level of π can be interpreted as the percentage of solutions that are more conservative than the selected solution, i.e. have a higher objective value as defined by Equation (5). Therefore, Equation (16) first calculates the $(100-\pi)$-th percentile of all objective values. Then, the index i^* of the closest solution is determined using Equation (17). The selected solution b^* is given by Equation (18).

$$P_{100-\pi} = \text{Percentile}_{100-\pi}\left(\{z_b \mid b \in B\}\right) \tag{16}$$

$$i^* = \arg\min_{b \in B} |z_b - P_{100-\pi}| \tag{17}$$

$$b^* = B_{i^*} \tag{18}$$

4. Experimental setup

4.1. Implementation of the proposed method

To test the proposed method, it is implemented and used on a dataset that has been formed based on process data from ongoing production. Therefore, actual production data from car body construction was gathered over a three-month period in the summer of 2023. It contains process times of approx. 4000 RSW spots that are distributed on 25 assembly lines, 115 workstations, and a total of 255 robots. In the first step, a distance matrix has to be computed. For this purpose, the Earth Movers distance is calculated using the method proposed by Flamary et al. (2021). The DBSCAN algorithm, as implemented by Pedregosa et al. (2011), is then employed to identify outliers. It requires the setup of two parameters, for which the interactive approach by Ester et al. (1996) is applied. Balancing, normalisation and standardisation of the dataset are conducted according to Sections 3.1.2 and 3.1.3.

In the next step, the influence of varying process time distributions on the robotic assembly line balancing is investigated. Therefore, the RALBP as described in Section 3.2.1, has to be solved repeatedly as part of the simulation algorithm, see Section 3.2.2. The genetic algorithm proposed by Stade and Manns (2023b) is employed for this purpose. Once the simulation results remain stable and the algorithm terminates, the results are analysed, and a solution is selected according to the procedure suggested in Section 3.3.

4.2. Design of experiments

The proposed method is compared to normally distributed process times in the planning of a resistance spot welding assembly line for the production of a car body rear compartment. A normal distribution assumption (NDA) is made in various publications that consider uncertain process times (see Ağpak and Gökçen 2007; Fathi et al. 2019; Özcan 2018; Şahin and Tural 2023; Tiacci 2017). As the station cycle times can be determined analytically, in contrast to the convolution presented in the model formulation of this work, the NDA facilitates the solution computation for the RALBP. For comparison, process times of the normal distribution are normalised and standardised using the method suggested in Section 3.1.3. As RSW comprises two process steps that cause time variance, namely gun-closing and welding, the proposed method and NDA are applied to both.

The rear compartment has to be assembled in eight subsequent joining steps, totalling 279 weld spots, which are split into 65 primary and 214 secondary welds. Primary weld joints fix parts in position so that they always need to be executed before any secondary or primary welds of subsequent joining steps can be done. Secondary welds are used to secure, seal, or finish the assembly, so they can usually be processed later. Given the sequence of joining steps and rules for processing primary and secondary welds, a precedence constraint matrix $\boldsymbol{P}$ can be formed.

The predetermined cycle time C totals 60 s, which is decreased by 10 s per handling operation. Assuming that there are two parts to be loaded into and one assembly off each workstation, this reduces the cycle time by 30 s to $C = 30$ s. In the proposed definition of the RALBP, the risk of cycle time violations can be controlled by a predetermined probability limit α. Increasing the limit allows for higher utilisation of the assembly line resources, which therefore saves capacity and reduces equipment usage. Therefore, a good compromise between cycle time violation risk and cost reduction is desired. As a result, the experiments are repeated for probability limits of $\alpha =$ 0%, 5%, 10% and $\alpha = 50\%$.

The normalisation and standardisation step is executed in both methods and requires deterministic standard times $\hat{t}$ and ranges $\hat{r}$. In practice, these are valid only for a specific manufacturing context, such as material pairings and sheet thickness ranges. The utilised contexts and corresponding standard times are given in Table 5. While standard times $\hat{t}$ are practical values from the industry, standard ranges $\hat{r}$ are determined by the average

Table 5. Manufacturing contexts and corresponding standard times and ranges.

Context		Standard time $\hat{t}$		Standard range $\hat{r}$	
Sheet thickness	Priority	Welding	Gun-closing	Welding	Gun-closing
Small	Primary	3 s		78 ms	32 ms
Small	Secondary	2.3 s		83 ms	34 ms
Medium	Primary	3.2 s		110 ms	40 ms
Medium	Secondary	2.5 s		99 ms	32 ms
High	Primary	3.5 s		117 ms	36 ms
High	Secondary	2.8 s		111 ms	42 ms

Note: Table of standard times and ranges for manufacturing contexts partitioned by sheet thickness and joint priority for welding and gun-closing.

Table 6. Relevant input parameters of the experiments.

Parameter	Value
Population size	500
Maximum generations	1000
Crossover rate	0.2
Mutation rate	0.02
Unimproved generation limit	5

Note: Table of parameters for the genetic algorithm that is executed during the experiments.

5- to 95-percentile ranges of all weld joints in the dataset allocated to a specific context.

For the simulations, $r_{init} = 100$ initial replications, an increment of $r_{inc} = 10$ replications and a maximum of $r_{\max} = 1500$ replications are selected. A window of $r_{win} = 10$ replications is used to monitor the 5th and 95th percentiles of objective values in the stability analysis. The simulation will be terminated if all of the percentiles in the monitoring window change by less than $\Delta perc_{\min} = 0.1\%$. In practice, the decision maker aims to avert overly optimistic planning results and therefore is likely to make conservative assumptions. Therefore, the confidence level is set to $\pi = 10\%$. Additional parameters for the experiments are summarised in Table 6.

Application of the proposed method requires a dataset of known RSW process times, and the assembly line must be planned to use the same equipment as in the dataset. Both of these requirements are satisfied in these experiments. Additionally, the actual process time distributions of the car body rear compartment are known and serve as comparison data for the experiments.

5. Experimental results

5.1. Solution analysis and selection

Variable process time distributions of the welding and gun-closing process step of RSW can be assessed by the simulations. The simulations of all probability limits were completed within 100 to 130 replications with the given parameters. Figure 3 illustrates the objective values of each process step, probability limit, and replication in histograms. Moreover, the objective values of confidence level π are indicated by vertical lines. As all balancing solutions result in the same number of workstations, changes in objective values can be interpreted as changes in relative utilisation of the last workstation. A decrease in the utilisation of the last workstation indicates a greater overall utilisation of all previous workstations.

At a probability limit of $\alpha = 0\%$, the process time distribution variations have the highest impact. Considering the 5- to 95-percentile range of objective values, the impact of varying process time distributions amounts to 16.9% of relative utilisation in the scenario of welding. The varying shapes of the distributions are the sole reason for this effect, as they are all normalised and standardised. In all other instances, the impact is lower, but also varies from 5% to 2.7%.

In both scenarios, the highest capacity savings can be achieved by a small increase in probability limit from 0% to 5%. 16.6% and 9.5% are the respective amounts. The capacity gains achieved by elevating the probability limit to 10% or even 50% result in a maximum of 3.4% of savings. In the scenario of gun-closing, an increase in α from 10% to 50% does not result in any change of assembly line utilisation. This observation is due to positively skewed process time distributions that allow for large capacity savings by small elevations of the probability limits.

5.2. Benchmarking the proposed method

For comparison of the results obtained by the proposed method and the normal distribution assumption (NDA), four evaluation metrics are introduced and described below:

Assembly line utilisation (ALU) The assembly line utilisation metric considers the median utilisation of the last workstation. The value of a balancing result b obtained by one of the approaches is compared to the utilisation of the last workstation of the balancing result b_a, which is computed using the ground truth distributions of the rear compartment, see Equation (19).

$$\mathrm{ALU}\,(b, b_a, s) = \tilde{F}_{b,s}^{-1}\,(0.5) - \tilde{F}_{b_a,s}^{-1}\,(0.5) \tag{19}$$

In practice, these are not available at the time of planning. To minimise the risk of overly optimistic planning results, the ALU should be positive but close to zero, which indicates a modestly lower anticipated utilisation of the assembly line.

Capacity savings (CS) As with the ALU, the capacity savings metric also utilises the median utilisation of

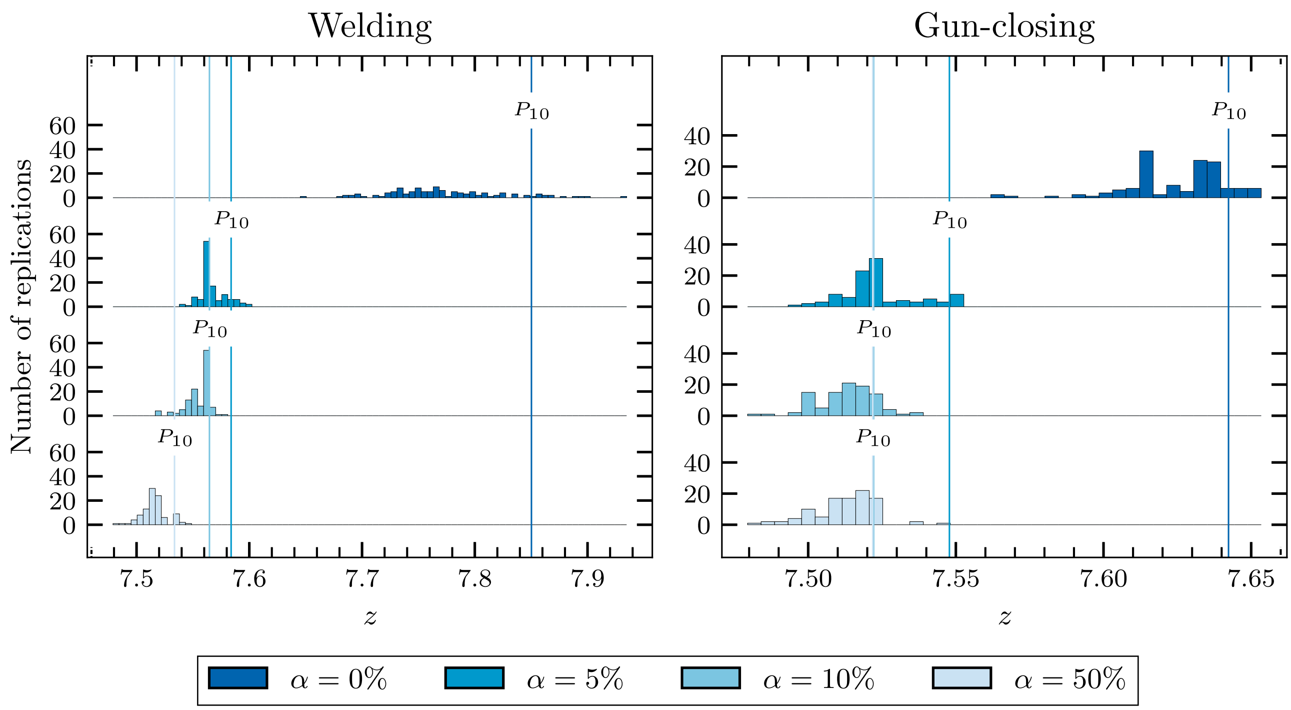


Figure 3. Analysis of the simulation results and solution selection.
Long Description The simulation of probability limits of $\alpha = 0\%$, 5%, 10% and 50% indicate an impact of varying process time distributions between 2.7% and up to 16.9%. A solution selection with a confidence level of $\pi = 10\%$ suggest the highest capacity savings by an elevation of the probability limit from 0% to 5%.

the last workstation and compares the result to those obtained by the ground truth distributions. In this case, however, the anticipated capacity savings due to an increase of α_1 to α_2 are compared to the actual savings, see Equation (20).

$$\mathrm{CS}\left(b, b_a, \alpha_2, \alpha_1, s\right) = \tilde{F}^{-1}_{b_a,\alpha_2,s}(0.5) - \tilde{F}^{-1}_{b_a,\alpha_1,s}(0.5) - \tilde{F}^{-1}_{b,\alpha_2,s}(0.5) + \tilde{F}^{-1}_{b,\alpha_1,s}(0.5) \tag{20}$$

As before, a positive CS is favourable but should be close to zero, which indicates lower anticipated capacity savings and therefore leads to more conservative results.

Probability limit violation (PLV) Once a balancing solution is implemented in practice, actual process time distributions apply. Therefore, the cycle times at the workstations may differ from what was anticipated. The specified probability limit α is not violated by the balancing solutions, but actual distributions may. The probability limit violation metric determines the probability of actual cycle times violating the given probability limit, see Equation (21).

$$\mathrm{PLV}\left(w, \alpha\right) = \tilde{F}'_w\left(C\right) - \alpha \tag{21}$$

The cumulative distribution function of the cycle times of a workstation $\tilde{F}'_w$ utilises the task assignments of a balancing solution but replaces the process time distributions by their actual counterparts. A positive PLV therefore indicates a violation of the specified probability limit.

Cycle time estimation (CTE) The cycle time estimation metric specifies the estimation error of anticipated cycle times in a workstation. It is the difference between the anticipated and actual median utilisation of a workstation, see Equation (22). A positive CTE is desirable and indicates the overestimation of cycle times.

$$\mathrm{CTE}\left(w\right) = \tilde{F}^{-1}_w(0.5) - \tilde{F}'^{-1}_w(0.5) \tag{22}$$

5.2.1. Assembly line utilisation

Table 7 presents the ALU results of the proposed simulation method and the NDA. The proposed method is in six out of eight instances closer to zero and constantly positive. Therefore, the last workstations' utilisation is overestimated by a range of 0.04 s to 1.22 s. The NDA, on the other hand, leads to constant underestimations and therefore negative ALUs. At a probability limit of $\alpha = 0\%$, the underestimation reaches its maximum in both scenarios. In the scenario of welding, however, the worst ALU of −8.28 s occurs. In the other cases, the underestimation varies between −0.02 s to −1.89 s. These underestimations are also highlighted by Figure 4, which shows

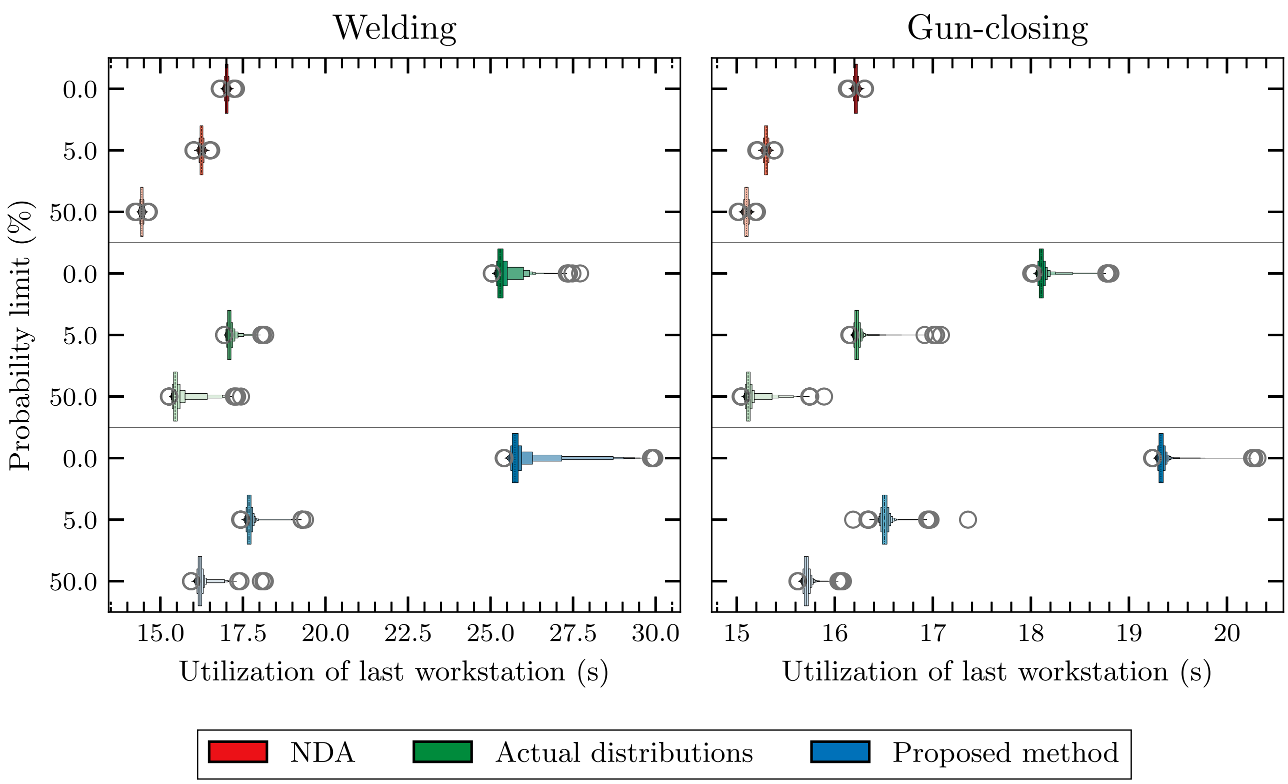


Figure 4. Last workstations' utilisation obtained by the NDA as well as proposed method and comparison to ground truth distributions. **Long Description** The utilisation of the last workstation at probability limits of 0%, 5% and 50% indicates that the proposed simulation method well represents the ground truth distributions by showing a similar effect at an increase of α = 0% to 5%. The NDA cannot replicate this.

Table 7. The ALU metric indicates a constant overestimation of the assembly line utilisation by the NDA. The proposed simulation method outperforms the NDA.

		ALU			
Scenario	Method	α = 0%	α = 5%	α = 10%	α = 50%
Welding	Simulation	0.45 s	0.61 s	0.04 s	0.75 s
	NDA	−8.28 s	−0.84 s	−0.82 s	−1.0 s
Closing	Simulation	1.22 s	0.29 s	0.26 s	0.59 s
	NDA	−1.89 s	−0.92 s	−0.13 s	−0.02 s

Note: Table of assembly line utilisation metric values for welding and gun-closing as well as the proposed method and normal distribution assumption. The normal distribution assumption is consistently outperformed by the proposed method.

the utilisation distributions of the last workstation of both scenarios and approaches for probability limits of α = 0%, α = 5% and α = 50%. The results obtained by the ground truth process time distributions are also presented. In comparison to the ground truth, the spread width of the utilisation distribution of the NDA is significantly smaller. The spread widths of the proposed method, in contrast, are closer to the ground truth.

5.2.2. Capacity savings

Table 8 presents CS results of both approaches and scenarios, as well as all probability limits. As indicated by the results regarding the ALU metric, the anticipated capacity savings of the NDA by increasing the probability limit from 0% to 5% is significantly underestimated. This can also be observed in Figure 4. Especially in the scenario of welding, the CS metric amounts to 7.45 s. The proposed simulation method yields a respective value of 0.15 s, which is substantially closer to the ground truth. As has been observed in the solution analysis, see Section 5.1, the substantial capacity savings from 0% to 5% in the ground truth and simulation results are due to positively skewed process time distributions and possible outliers. This cannot be reproduced by the NDA.

In all other cases, however, the NDA overestimates the capacity savings once by 0.19 s. The proposed simulation approach, on the other hand, results in negative CS values in three instances.

5.2.3. Probability limit violation

The probability limit violation metric is categorised in Table 9 into 10% wide intervals that range from 0% to 50% of probability limit violation. The simulation method leads in total to four probability limit violations, of which three are in a range between 0% and 10%. In one instance, the violation lies between 10% and 20%. All violations are in the scenario of welding.

Table 8. The NDA underestimates the capacity savings of the ground truth distributions at an increase of the probability limit from $\alpha = 0\%$ to 5%, but partly outperforms the proposed simulation method later.

		CS		
Scenario	Method	$\alpha = 0\% \rightarrow 5\%$	$\alpha = 5\% \rightarrow 10\%$	$\alpha = 10\% \rightarrow 50\%$
Welding	Simulation	0.15 s	−0.57 s	0.72 s
	NDA	7.45 s	0.02 s	−0.19 s
Closing	Simulation	−0.94 s	−0.02 s	0.33 s
	NDA	0.97 s	0.79 s	0.12 s

Note: Table for the comparison of the capacity savings metrics for the proposed simulation method and normal distribution assumption at welding and gun-closing scenarios. The normally distributed process times underestimate the large capacity savings between a probability limit of 0% and 5%, but later partly outperfom the proposed method.

Table 9. The proposed simulation method violates the given probability limit less often and less severe.

		PLV			
Scenario	Method	0%–10%	10%–20%	30%–40%	40%–50%
Welding	Simulation	3	1	0	0
	NDA	7	0	2	1
Closing	Simulation	0	0	0	0
	NDA	4	0	0	0

Note: Table for the comparison of probability limit violation metrics of the proposed simulation method and normal distribution assumption. The proposed method violates the specified limit less often and less severe.

In the scenario of welding, the NDA results in a higher number and more severe probability limit violations. Seven instances fall in between 0% and 10%. In two cases, however, the probability limit is exceeded by 30% to 40% and in one instance by 40% to 50%. In the scenario of gun-closing, four marginal violations varying between 0% and 10% can be observed. Hence, the proposed simulation method clearly outperforms the NDA in both scenarios. This is emphasised by Figure 5, which depicts instances of probability limit violations for both approaches in the welding scenario, with probability limits of 10% and 50%. The anticipated cycle times and their ground truth equivalents are presented for workstations two, four, and five of the balancing solutions. As previously noted in the ALU and CS metrics, the NDA significantly underestimates the spread widths and positive skewness of cycle times. This leads to numerous breaches of probability limits, particularly at lower probability limits. The suggested simulation approach better captures the distinctive features of the underlying ground truth distributions. Nevertheless, in the majority of instances, it also underestimates the spread width and positive skewness, although by a smaller margin.

5.2.4. Cycle time estimation

Figure 6 presents the negative values of the CTE metric in histograms for both scenarios and approaches. It can be observed that the NDA underestimates the median cycle times in all instances by up to 161 ms and 31 ms. Less severe and frequent underestimations of the proposed simulation approach are evident. The corresponding maximum underestimates are 71 ms and 20 ms. 13 and 7 instances are underestimated, respectively.

6. Discussion

As has been described in Section 1 on the example of RSW, the shape of process time distributions cannot be estimated utilising planning data. This is likely due to the unstandardised and individual process of weld control parametrization. The proposed simulation method allows for the evaluation of the effects of varying process time distributions on the balancing of robotic assembly lines.

The solution analysis of the experimental results on the welding and gun-closing process times of RSW indicates a substantial impact. A small increase in probability limits from 0% to 5% results in large capacity savings thanks to the positive skewness and potential outliers of the underlying process time distributions. Further increases of the probability limit result in significantly diminished savings.

This can be validated by the comparative analysis considering the NDA and ground truth distributions. The RALBP solved by utilising the actual process time distributions shows similar effects. This behaviour is not replicated by the NDA, and the capacity enhancements at low probability limit elevations are clearly underestimated. Capacity gains induced by further increases are better estimated, partly better than those predicted by the simulation method. However, the utilisation of the assembly line is consistently underestimated. This means, that the anticipated cycle times of workstations are lower than the actual cycle times. As a result, the prespecified probability limit is violated in numerous instances, three of which by a considerable margin. The experimental findings suggest that the NDA is inadequate for the process times of welding and gun-closing, thereby limiting its practical application.

The proposed simulation method, by contrast, underestimates cycle times less often and severely. Therefore, the probability limit is violated in fewer instances, none of which is as significant as the NDA. The method surpasses the NDA by utilising historical process time distributions of the respective joining techniques, which possess characteristics of the actual distributions. In practice, the approach can be employed to establish an appropriate probability limit, which serves as a compromise between capacity reduction and the potential for cycle time violations. However, the decision maker must determine a confidence level based on the simulation results and the

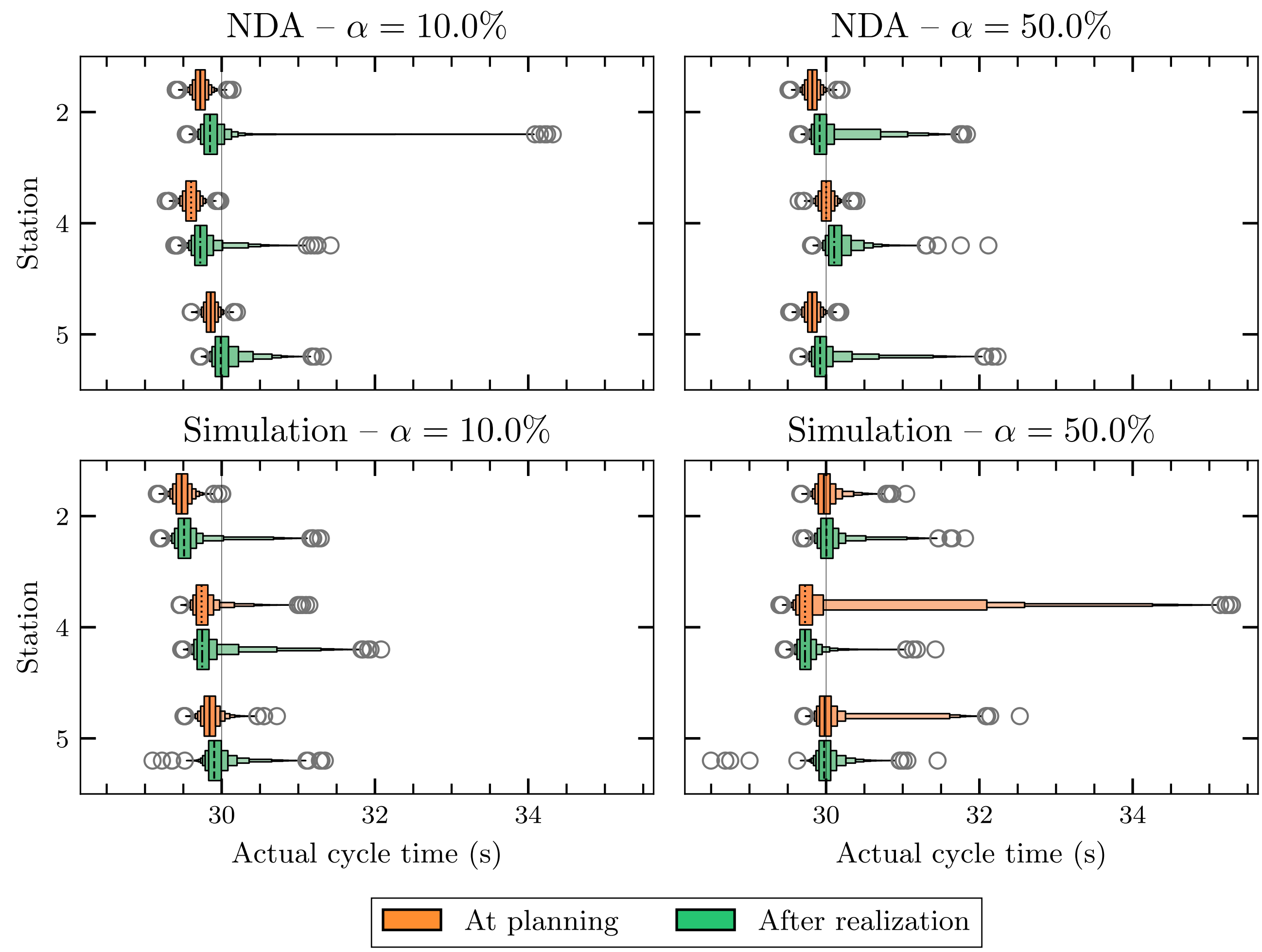


Figure 5. Comparison of anticipated (at planning) and actual cycle times (after realisation) for the normal distribution assumption and proposed simulation method.

Long Description The anticipated cycle times obtained by both methods are compared to actual cycle times that apply after realisation of the assembly line in workstations 2, 4 and 5 for probability limits 10% and 50%. It indicates that the NDA constantly underestimates the cycle times, which leads to violation of cycle times and probability limits.

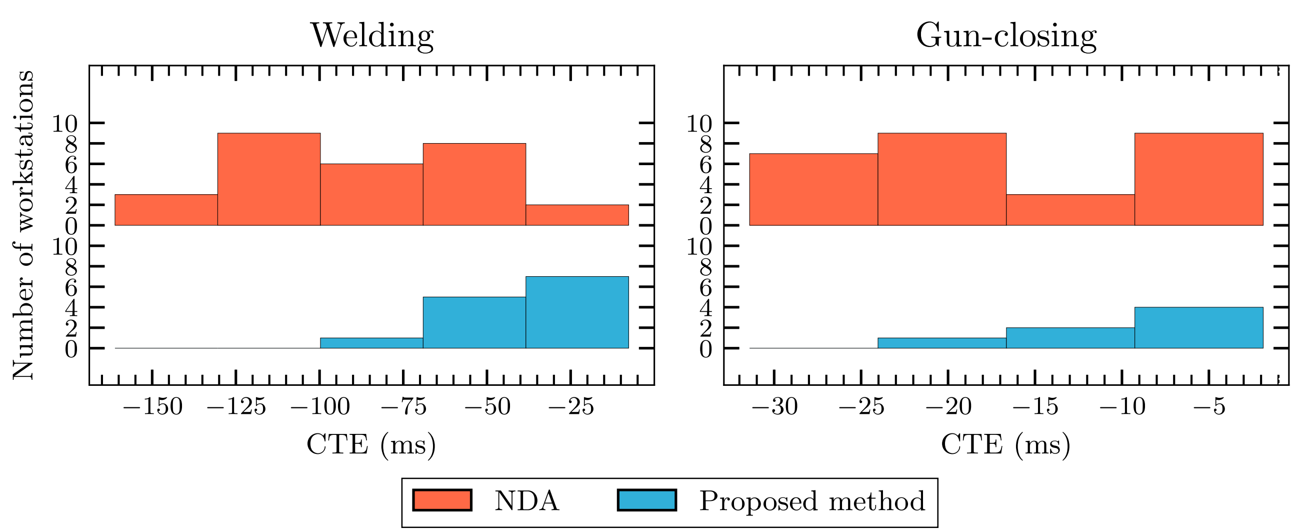


Figure 6. The CTE metric shows more and more severe underestimations of cycle times by the NDA compared to the proposed method.

use case. As a low confidence level translates to a more conservative balancing solution, the assembly line utilisation in general is lower. Given the effect of skewness and outliers observed in the experiments, this may increase the sensitivity of balancing for small increases in probability limit. Therefore, a low confidence level may lead to overly optimistic expectations for capacity savings.

The balancing of the dataset in the proposed method greatly contributes to its improved performance. The rejection sampling technique is employed to draw distributions from the dataset to evenly distribute the mean process times. This is important because it is not known how the process time distributions to be predicted relate to the dataset. The balancing mitigates the possibility of an unfavourable ratio, which may lead to an inadequate representation of the dataset. Although the shift and scale of the distributions correlate with the planning data, the shape does not. The utilisation of the mean process time for balancing of the distributions is therefore inadequate. Future research may therefore focus on the balancing of the dataset considering the distributions' shapes to further improve the robustness of the approach. This needs to be validated and quantified by extended experiments.

Although the proposed simulation method outperforms the NDA, the findings emphasise that the positive skewness of the ground truth distributions is a contributing factor. Hence, alternative distribution assumptions that consider skewness have the potential to yield superior outcomes. Furthermore, the computational effort may be greatly reduced. This, however, needs to be validated and may not be adequate for every joining technique and corresponding process time distributions. In direct comparison, the proposed data-driven simulation method, including its limitations, is robust to altering distribution shapes and provides good results. The simulations are based on historical process data, which allows for a straightforward adaptation to other joining techniques. Different balancing objectives or even assembly line types are also possible through the adaptation of the underlying problem formulation and the respective solution technique.

7. Conclusion and future research

This work introduces a simulation method to assess the impact of varying process time distributions on the balancing of robotic assembly lines. It focuses on the early planning phases of car body construction, where many requirements and premises, such as process times, are still uncertain. However, as car body construction widely employs standardised equipment, knowledge of existing assembly lines can be transferred to planning. Based on two practical examples, the application of the method was demonstrated. With the help of process time distributions of existing assembly lines, the impact of varying process time distributions on the balancing of a new assembly line is simulated. This allows for the consideration of process time variance in the early planning of car body construction before actual process times are available.

It becomes apparent that the proposed simulation method provides a contribution to the consideration of uncertain process times in the early planning of car body construction. The simulation method is compared to normally distributed process times, which is a common assumption for process times in the literature. As the experiments show, the balancing results of the simulation method are more stable and reliable than those obtained by the normal distribution assumption. As a result, costs may be reduced by improving the quality of early planning. However, in some cases of the experiments, the capacity savings by increasing the probability limit for cycle time violations, are overestimated. This is due to the selection of a simulation solution based on the objective function of the balancing results, which do not reflect the capacity savings. Further research should be conducted to investigate and mitigate these observations.

As car body construction is highly automated, many process steps cause process time variance. The experiments of the present work focus on resistance spot welding for one main assembly of a car body-in-white using production data of a single plant. Nonetheless, the approach is capable of being adapted to other joining techniques, assembly line types, and objective functions. Future research may expand to explore and evaluate these further research directions.

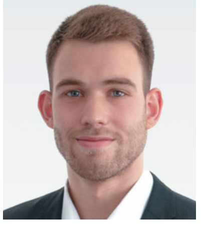
Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research is funded by the Mercedes-Benz Group AG.

Notes on contributors



Dawid Stade is a Ph.D. student at the Chair of Production Automation and Assembly at the University of Siegen (Germany) and a researcher at Mercedes-Benz AG. His field of research is production planning in the automotive industry, with a focus on the balancing of robotic assembly lines in the early design stages.

Jan Michael Spoor pursues his doctoral degree at the Institute for Information Management in Engineering at the Karlsruhe Institute of Technology in cooperation with Mercedes-Benz. His research topics are Anomaly Detection, Digital Twins, and AI support systems in production planning. Currently, he works as an IT consultant at BCG Platinion. Prior, he worked at Homburg & Partner as a consultant specialised in B2B market strategy, sales, and pricing.

Martin Manns is a Full Professor, holding the Chair of Production Automation and Assembly at the University of Siegen (Germany). He holds a doctoral degree from the University of Hannover (Germany). His research and projects focus on human activity recognition, modelling, and simulation of manual assembly processes, collaborative robots, and the additive manufacturing of soft robotic actuators.

Jivka Ovtcharova has been a Professor for Computer Application in Planning and Design and Head of the Institute for Information Management in Engineering (IMI) at the Karlsruhe Institute of Technology (KIT) since 2003. Since 2004, she has been Director of Process and Data Management in Engineering in the Intelligent Systems and Production Engineering division of the Research Center for Information Technology (FZI) in Karlsruhe. Prof. Ovtcharova is the author of more than 250 scientific and technical publications, has supervised 23 doctoral theses as first reviewer, and has received numerous awards for her teaching and research.

Data availability statement

The data that support the findings of this study are available from the corresponding author, Dawid Stade, upon reasonable request.

References

Ağpak, K., and H. Gökçen. 2007. "A Chance-Constrained Approach to Stochastic Line Balancing Problem." *European Journal of Operational Research* 180 (3): 1098–1115. https://doi.org/10.1016/j.ejor.2006.04.042.

Al-Zaher, A., and W. ElMaraghy. 2014. *Design Method of Under-Body Platform Automotive Framing Systems*. In *Variety Management in Manufacturing. Proceedings of the 47th CIRP Conference on Manufacturing Systems*, Windsor, Canada, Vol. 17. Elsevier B.V.

Bertsimas, D., and M. Sim. 2003. "Robust Discrete Optimization and Network Flows." *Mathematical Programming* 98 (1–3): 49–71. https://doi.org/10.1007/s10107-003-0396-4.

Boysen, N., P. Schulze, and A. Scholl. 2022. "Assembly Line Balancing: What Happened in the Last Fifteen Years." *European Journal of Operational Research* 301 (3): 797–814. https://doi.org/10.1016/j.ejor.2021.11.043.

Chutima, P. 2022. "A Comprehensive Review of Robotic Assembly Line Balancing Problem." *Journal of Intelligent Manufacturing* 33 (1): 1–34. https://doi.org/10.1007/s10845-020-01641-7.

Colledani, M., D. Gyulai, L. Monostori, M. Urgo, J. Unglert, and F. V. Houten. 2016. "Design and Management of Reconfigurable Assembly Lines in the Automotive Industry." *CIRP Annals* 65 (1): 441–446. https://doi.org/10.1016/j.cirp.2016.04.123.

Denkena, B., M.-A. Dittrich, and S. J. Settnik. 2022. *Similarity-Based Process and Set-Up Time Estimation*. In *Proceedings of the 8th Changeable, Agile, Reconfigurable and Virtual Production Conference (CARV2021) and the 10th World Mass Customization & Personalization Conference (MCPC2021)*, Aalborg, Denmark, October/November 2021. Springer Science and Business Media Deutschland GmbH.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu. 1996. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, USA, 226–231. AAAI Press.

Fathi, M., A. Nourmohammadi, A. H. C. Ng, and A. Syberfeldt. 2019. "An Optimization Model for Balancing Assembly Lines with Stochastic Task times and Zoning Constraints." *IEEE Access* 7:32537–32550. https://doi.org/10.1109/ACCESS.2019.2903738.

Flamary, R., N. Courty, A. Gramfort, M. Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, et al. 2021. "Pot: Python Optimal Transport." *Journal of Machine Learning Research* 22 (78): 1–8.

Freedman, D., and P. Diaconis. 1981. "On the Histogram as a Density Estimator: L_2 Theory." *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 57 (4): 453–476. https://doi.org/10.1007/BF01025868.

Gurevsky, E., Ö. Hazïr, O. Battaïa, and A. Dolgui. 2013. "Robust Balancing of Straight Assembly Lines with Interval Task times." *Journal of the Operational Research Society* 64 (11): 1607–1613. https://doi.org/10.1057/jors.2012.139.

Hagemann, S. 2022. "Algorithmische Konzeption von hochautomatisierten Fließmontagesystemen am Beispiel des automobilen Karosseriebaus." PhD diss., TU Berlin, Berlin.

Hagemann, S., and R. Stark. 2020. "An Optimal Algorithm for the Robotic Assembly System Design Problem: An Industrial Case Study." *CIRP Journal of Manufacturing Science and Technology*31:500–513. https://doi.org/10.1016/j.cirpj.2020.08.002.

Hazır, Ö., and A. Dolgui. 2013. "Assembly Line Balancing under Uncertainty: Robust Optimization Models and Exact Solution Method." *Computers & Industrial Engineering* 65 (2): 261–267. https://doi.org/10.1016/j.cie.2013.03.004.

Hazır, Ö., and A. Dolgui. 2019. "A Review on Robust Assembly Line Balancing Approaches." *IFAC-PapersOnLine* 52 (13): 987–991. https://doi.org/10.1016/j.ifacol.2019.11.323.

Joslyn, C., and S. Fersont. 2004. "Approximate Representations of Random Intervals for Hybrid Uncertainty Quantification in Engineering Modeling." In *SAMO 04 Conference, Sensitivity Analysis of Modal Output*, Santa Fe, New Mexico, USA.

Kim, H., and S. Park. 1995. "A Strong Cutting Plane Algorithm for the Robotic Assembly Line Balancing Problem." *International Journal of Production Research* 33 (8): 2311–2323. https://doi.org/10.1080/00207549508904817.

Levitin, G., J. Rubinovitz, and B. Shnits. 2006. "A Genetic Algorithm for Robotic Assembly Line Balancing." *European Journal of Operational Research* 168 (3): 811–825. https://doi.org/10.1016/j.ejor.2004.07.030.

Meng, K., Q. Tang, Z. Zhang, and Z. Li. 2023. "Robust Mixed-Model Assembly Line Balancing and Sequencing Problem considering Preventive Maintenance Scenarios with Interval Processing times." *Swarm and Evolutionary Computation* 77:101255. https://doi.org/10.1016/j.swevo.2023.101255.

Michalos, G., A. Fysikopoulos, S. Makris, D. Mourtzis, and G. Chryssolouris. 2015. "Multi Criteria Assembly Line Design and Configuration – An Automotive Case Study." *CIRP Journal of Manufacturing Science and Technology* 9:69–87. https://doi.org/10.1016/j.cirpj.2015.01.002.

Michels, A. S., T. C. Lopes, C. G. S. Sikora, and L. Magatão. 2018. "The Robotic Assembly Line Design (rald) Problem: Model and Case Studies with Practical Extensions." *Computers & Industrial Engineering* 120:320–333. https://doi.org/10.1016/j.cie.2018.04.010.

Mucientes, M., J. C. Vidal, A. Bugarin, and M. Lama. 2008. "Processing Times Estimation in A Manufacturing Industry Through Genetic Programming." In *2008 3rd International Workshop on Genetic and Evolving Systems*, Witten-Bommerholz, Germany, 95–100. IEEE.

Özcan, U. 2018. "Balancing Stochastic Parallel Assembly Lines." *Computers & Operations Research*99:109–122. https://doi.org/10.1016/j.cor.2018.05.006.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12:2825–2830.

Pereira, J., and E. Álvarez-Miranda. 2018. "An Exact Approach for the Robust Assembly Line Balancing Problem." *Omega* 78:85–98. https://doi.org/10.1016/j.omega.2017.08.020.

Ringsquandl, M., S. Lamparter, and R. Lepratti. 2015. "Estimating Processing times within Context-Aware Manufacturing Systems." *IFAC-PapersOnLine* 48 (3): 2009–2014. https://doi.org/10.1016/j.ifacol.2015.06.383.

Rossi, A., E. Gurevsky, O. Battaïa, and A. Dolgui. 2016. "Maximizing the Robustness for Simple Assembly Lines with Fixed Cycle Time and Limited Number of Workstations." *Discrete Applied Mathematics* 208:123–136. https://doi.org/10.1016/j.dam.2016.03.005.

Rubinovitz, J., and J. Bukchin. 1991. "Design and Balancing of Robotic Assembly Lines." In *Proceedings of the Fourth World Conference on Robotics Research*, Pittsburgh, Pennsylvania, USA. The Society.

Rubner, Y., C. Tomasi, and L. J. Guibas. 2000. "The Earth Mover's Distance as a Metric for Image Retrieval." *International Journal of Computer Vision* 40 (2): 99–121. https://doi.org/10.1023/A:1026543900054.

Şahin, M. C., and M. K. Tural. 2023. "Robotic Stochastic Assembly Line Balancing." *Flexible Services and Manufacturing Journal* 35 (4): 1076–1115. https://doi.org/10.1007/s10696-023-09494-x.

Salehi, M., H. R. Maleki, and S. Niroomand. 2018. "A Multi-Objective Assembly Line Balancing Problem with Worker's Skill and Qualification Considerations in Fuzzy Environment." *Applied Intelligence* 48 (8): 2137–2156. https://doi.org/10.1007/s10489-017-1065-2.

Sotskov, Y. N., A. Dolgui, T.-C. Lai, and A. Zatsiupa. 2015. "Enumerations and Stability Analysis of Feasible and Optimal Line Balances for Simple Assembly Lines." *Computers & Industrial Engineering*90:241–258. https://doi.org/10.1016/j.cie.2015.08.018.

Sotskov, Y. N., A. Dolgui, and M.-C. Portmann. 2006. "Stability Analysis of An Optimal Balance for An Assembly Line with Fixed Cycle Time." *European Journal of Operational Research* 168 (3): 783–797. https://doi.org/10.1016/j.ejor.2004.07.028.

Spieckermann, S., K. Gutenschwager, H. Heinzel, and S. Voß. 2000. "Simulation-Based Optimization in the Automotive Industry – A Case Study on Body Shop Design." *SIMULATION: Transactions of The Society for Modeling and Simulation International* 75: 276–286.

Spoor, J. M. 2022. "Improving Customer Segmentation Via Classification of Key Accounts as Outliers." *Journal of Marketing Analytics* 11:747–760. https://doi.org/10.1057/s41270-022-00185-4.

Stade, D., and M. Manns. 2023a. "Influence of Time-Variant Robotic Welding Processes in Car Body Construction." In *Production Processes and Product Evolution in the Age of Disruption*, Bologna, Italy, 581–588. Springer.

Stade, D., and M. Manns. 2023b. "Robotic Assembly Line Balancing With Multimodal Stochastic Processing Times." In *Advances in Automotive Production Technology – Towards Software-Defined Manufacturing and Resilient Supply Chains: Stuttgart Conference on Automotive Production (SCAP2022)*, Stuttgart, Germany, 78–84. Springer.

Stavropoulos, P., K. Sabatakakis, A. Papacharalampopoulos, and D. Mourtzis. 2022. "Infrared (ir) Quality Assessment of Robotized Resistance Spot Welding Based on Machine Learning." *The International Journal of Advanced Manufacturing Technology* 119 (3-4): 1785–1806. https://doi.org/10.1007/s00170-021-08320-8.

Tiacci, L. 2017. "Mixed-Model U-shaped Assembly Lines: Balancing and Comparing with Straight Lines with Buffers and Parallel Workstations." *Journal of Manufacturing Systems* 45:286–305. https://doi.org/10.1016/j.jmsy.2017.07.005.

Zacharia, P., and A. C. Nearchou. 2013. "A Meta-Heuristic Algorithm for the Fuzzy Assembly Line Balancing Type-e Problem." *Computers & Operations Research* 40 (12): 3033–3044. https://doi.org/10.1016/j.cor.2013.07.012.

Zacharia, P. T., E. K. Xidias, and A. C. Nearchou. 2024. "The Fuzzy Human-Robot Collaboration Assembly Line Balancing Problem." *Computers & Industrial Engineering* 187:109774. https://doi.org/10.1016/j.cie.2023.109774.

Zwicker, C., and G. Reinhart. 2016. "System for the Estimation of Robot Cycle times in Early Production Planning Phase." *Applied Mechanics and Materials* 840:99–106. https://doi.org/10.4028/www.scientific.net/AMM.840.