

Huchzermeier, Arnd; Mönch, Tobias

Article — Published Version

Mixed-model assembly lines with variable takt and open stations

Production and Operations Management

Provided in Cooperation with:

John Wiley & Sons

Suggested Citation: Huchzermeier, Arnd; Mönch, Tobias (2022) : Mixed-model assembly lines with variable takt and open stations, Production and Operations Management, ISSN 1937-5956, Wiley, Hoboken, NJ, Vol. 32, Iss. 3, pp. 704-722, <https://doi.org/10.1111/poms.13893>

This Version is available at:

<https://hdl.handle.net/10419/287859>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by-nc-nd/4.0/>

ORIGINAL ARTICLE

Mixed-model assembly lines with variable takt and open stations

Arnd Huchzermeier | Tobias Mönch

WHU – Otto Beisheim School of Management,
Burgplatz 2, Vallendar, Germany

Correspondence

Arnd Huchzermeier, WHU – Otto Beisheim
School of Management, Burgplatz 2, 56179
Vallendar, Germany.
Email: Arnd.Huchzermeier@whu.edu

Handling Editor: Panos Kouvelis

Abstract

Industry trends such as product customization, radical innovation, and local production accelerate the adoption of mixed-model assembly lines (MMALs) that can cope with a widening gap between model processing times and true build to order capability. The existing high work content deviations on such assembly lines stress production planning, especially the assembly line sequencing. Most manufacturers set the launching rate for all assembly line products to a fixed launching rate resulting in rising utility work and idle time when system load increases. We present an “ideal” variable rate launching (VRL) case resulting in minimal computation and achieving 100% productivity (full elimination of idle time and utility work) for balanced assembly times and homogeneous station lengths. Managers should foster the ideal circumstances where operators need not wait for a preceding task to be completed and product sequence restrictions are eliminated, thus enabling unmatched production flexibility. Furthermore, we present a mixed-integer model to analyze both closed and open workstations on an MMAL for fixed rate launching and VRL. This model incorporates costs not only for labor inefficiencies but also for extending the line length. We present a heuristic solution method when process times and station lengths are heterogeneous and demonstrate that the variable takt dominates the fixed takt. In a numerical, industrial benchmark study, we illustrate that a VRL strategy with open stations has significantly lower labor costs as well as a substantially reduced total line length and thus lower throughput time.

KEYWORDS

line length restriction, mixed-model sequencing, open stations, variable rate launching

1 | INTRODUCTION

During our visits with finalists in the INSEAD–WHU Industrial Excellence Award (IEA) competition (see www.industrial-excellence-award.eu), numerous executives from the automotive, machine tool, and electrical engineering industries emphasized that their respective businesses are being disrupted by three major trends: (i) increased product customization with a strong focus on high-end products (e.g., sport utility vehicles); (ii) radical innovation, as when electrical vehicles must be assembled alongside conventional vehicles; and (iii) natural hedging—that is, increased local production of their broader product portfolios because of

tariff and tax issues, carbon emissions, pandemics, and/or the need to mitigate risk by matching foreign revenue with local costs (Cohen et al., 2018, 2022; Knowledge@Wharton, 2019).

This dynamic business environment exerts pressure on existing linear and inflexible assembly lines to adopt a more efficient layout and to handle a wider range of products at a single site (oftentimes “brownfield” plants with an existing assembly setup). Many executives voiced the concern that assembly lines must be expanded considerably if they are to handle such product diversity. Most of these manufacturers strive to fulfill unique customer requirements in a true build-to-order environment—one of the predominant variants of mass customization—where the next unit scheduled for assembly is linked to the last incoming customer order

Accepted by Panos Kouvelis, after two revision.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Production and Operations Management* published by Wiley Periodicals LLC on behalf of Production and Operations Management Society.

(Agrawal et al., 2001; Wang et al., 2017). The central question that arises in this context is whether existing assembly line systems can effectively cope with such rapid changes and with the ever rising demand for greater flexibility and volume (Bloomberg, 2019).

Leading production systems are today based on four crucial objectives: takt, flow, zero defects, and pull; all of these derive from the Toyota Production System. Takt time (or cycle time) is the clock speed of every assembly process, which is determined by the available operating hours and the output per period that is needed to satisfy demand; it is the key lever of productivity (Monden, 2011). So when a new model with significantly greater work content is scheduled for production, plant managers can either plan a separate assembly line or adjust the existing line's parameters—that is, increase the fixed takt time. Yet, in either case, productivity suffers. Furthermore, the greater the difference in the workload of units sequenced on the line, the more difficult it is to assemble in a continuous flow of workload where the operator's individual workload is balanced—one of the cornerstones of lean. When assembling highly variable, customer-specific products in a fixed takt time, such a steady flow cannot be achieved. Mönch et al. (2022) show how variable takt time groups enable a convergence toward a perfect assembly flow, independent of workload differences within the product portfolio. It follows that determining the optimal launching strategy is critical when introducing multiple models with unequal work content onto a given assembly line. The approach most commonly used to determine the launching interval (λ) is fixed rate launching (FRL), where the interval between two adjacent units on the line approximates the fixed takt time (including all inefficiencies) (Bard et al., 1992). Our paper presents the concept of variable rate launching (VRL), under which models are placed onto the conveyor depending on their processing time. When this procedure is followed, labor inefficiencies are substantially reduced.

In 2017, Fendt—the German subsidiary of AGCO and an innovation leader within the agricultural machinery market—implemented such a VRL system, the “VarioTakt.” Adopting this approach allowed Fendt to assemble all 10 of its product series, including an unlimited number of model configurations, on a single assembly line and with relatively few concerns about sequence restrictions or product complexity (Huchzermeier et al., 2020a). Note that in 2020, the World Economic Forum and McKinsey & Company selected Fendt's Marktoberdorf plant as one of their global digital lighthouse factories (Betti et al., 2020). The four cornerstones of our “ideal” case are based on Fendt's current operations (see Theorem 1 in Section 5). Without the VarioTakt, Fendt would not have been able to cope with the increased customization resulting from greater competition and its expansion into new markets (e.g., North America) (Bebersdorf, 2017). A recent process innovation enabled by this novel approach is that the “Fendt 1000 Vario”—the largest premium tractor in the market, with 738% more horsepower than Fendt's smallest tractor—is now assembled with all other models on the same assembly line (Bebersdorf,

2017). The complex tasks of sequencing products can be omitted as models are perfectly balanced over the entire set of stations and all stations have the same length. These cornerstones of the ideal case are discussed in Section 5. This is particularly of great interest for manufacturing executives during times when the product mix changes quickly, for example, due to recessions or pandemics. Huchzermeier et al. (2020b) provide an in-depth description of the Fendt VarioTakt and how Fendt achieves the underlying requirement of perfectly balancing station processing times. Also, Mönch et al. (2021) describe how variable launching rates simplify the assembly line balancing problem and achieve high operator utilization levels by almost perfectly balancing assembly times.

Most research that explores how best to cope with mixed models in assembly has focused on minimizing idle time and utility work (or work overload) independently (Bolat & Yano, 1992b; Dar-El & Cothier, 1975; Okamura & Yamashina, 1979; Thomopoulos, 1967; Tsai, 1995; Yano & Rachamadugu, 1991) or on minimizing overall line length (Bard et al., 1992; Dar-El & Cothier, 1975; Dar-El & Cucuy, 1977). Utility work arises when processing times on a unit exceed the operators capacity. Idle time occurs when the operator cannot immediately start work on the next unit. We provide a more detailed description of both concepts in Section 4. The paper by Dar-El (1978) is the first to address overall line length and throughput time while considering both FRL and VRL. The numerical studies of Bard et al. (1992), Sarker and Pan (2001), and Fattahi and Salehi (2009) show that VRL outperforms FRL in terms of the costs associated with idle time and utility work—but only for the case of closed workstations. There has been hardly any research that models the VRL strategy with open stations, and no empirical studies have discussed the requirements and practical benefits of implementing a VRL strategy on assembly lines with open stations.

Our paper's contribution is to formulate a mixed-integer model of the optimization problem for both closed and open workstations on a mixed-model assembly line (MMAL) with FRL and VRL. We refer to this formulation as the *general* case, and it accounts for circumstances that have not previously been considered together. In addition to incorporating the costs associated with operator idleness and utility work, our analysis also considers—given the space restrictions commonly encountered in manufacturing sites—the cost of extending the assembly line length. Unlike most previous research, this study also accounts for sequence-dependent setup times when switching between the production of two distinct product models. Following the real-world example of Fendt, we identify several prerequisites for the *ideal* case: implementation of a VRL strategy under which employees are always busy, yet there are no workload peaks. This setting features 100% operator utilization and station lengths that are reduced to a bare minimum; these goals can be achieved by balancing assembly tasks and allowing parallel work on products as long as workers do not interfere with each other. Most importantly, this approach does not require extensive

computations and is flexible enough to handle almost any change in product mix at nearly any time. We also present crucial insights on the practical application of VRL. In particular, we show how a firm can switch easily from a fixed takt system requiring high computational effort to a variable takt system with full flexibility, 100% productivity, and hardly any computational burden.

The rest of our paper proceeds as follows. In Section 2, we describe the characteristics of an MMAL, after which Section 3 reviews the literature on FRL and VRL with closed and open stations. Our unifying model formulation—which we refer to as the general case—is proposed in Section 4. Section 5 then discusses the modeling and implementation of the ideal case for a VRL system. In Section 6, we develop heuristic solution procedures and conduct a benchmarking study. We conclude in Section 7 with summary remarks and managerial insights.

2 | MIXED-MODEL ASSEMBLY LINE SYSTEMS

Assembly lines are classified into single-model assembly lines (SMALs) and mixed-model assembly lines (MMALs), depending on the number of models being manufactured on the same line (Al-Hawari et al., 2015; Becker & Scholl, 2006). On SMALs, a single model is assembled. Whereas, on MMALs, multiple different product models are assembled with setup times between different models (Becker & Scholl, 2006; Boysen, Flidner, & Scholl, 2008). This section briefly outlines the basic structure of the MMAL systems considered in our paper (for a more detailed description of MMALs, see Buxey et al., 1973). Such assembly lines are used to assemble different workpieces (or products, or units), which vary in terms of assembly tasks and work content. A product model (model, hereafter) is defined by the same underlying design and physical characteristics as well as bill of materials (e.g., the BMW 3 series or the BMW 7 series). Thus, two workpieces can be of the same model. In this definition, model configurations, for example, interior upgrades, are not included. However, our approach also works for model configurations, where each workpiece belongs to a model configuration and not a model. Generally speaking, MMALs are partitioned into J work stations that are connected by a conveyor belt moving—from upstream to downstream—at a constant velocity v . There exists no buffer inventory between stations, and workpieces directly move into the next station. When looking to manufacture multiple models on a single assembly line, a manufacturer faces two closely related problems: the assembly line balancing problem and the assembly line sequencing problem. Line balancing defines the assignment of total workload among stations as evenly as possible (Boysen et al., 2007; Kilbridge & Wester, 1962; Thomopoulos, 1967). When a manufacturer decides to assemble models with differing work content on a single assembly line, the medium-term line balancing problem is linked to the short-term assembly line

sequencing problem (Thomopoulos 1967). Since assembly times differ between models and stations, it follows that operators are either underutilized or overloaded in terms of the assembly line's average work content. Once the best possible line balance is achieved, line sequencing determines the assembly order of products on the line based on leveling operator workload across total demand. The optimal assembly sequence has the effect of leveling workload deviations among the models based on a given assembly line balance, that is, by alternating single products (or batches) that have high and low levels of utilization. In the following, the focus is solely on the assembly line sequencing problem and we assume that stations are “balanced” in the sense that they reflect the best possible assignment of assembly tasks to stations. Then, the number of work stations is given and we do not consider the possibility of rebalancing the assembly line. Mönch et al. (2021) show how variable takt times reduce the complexity of the assembly line balancing problem. Moreover, the workpiece sequence is predetermined before assembly starts. Workpieces are processed on a first-come, first-serve basis and in the same sequence as they are launched onto the line. Different product model types m are assembled in a fixed station order, and the workpieces in the sequence are denoted by i . One operator is assigned to each station j ; that operator moves downstream with the conveyor while performing an assembly task on one workpiece at a time. Material supply is located at the upstream station boundary. We consider not only the deterministic assembly time t_{mj} but also sequence-dependent setup times s_{rmj} . The former is also seen as deterministic in practice, where a factor for unexpected downtimes and errors is included in the overall time. The latter occur when, at station j , an operator switches from processing a model r to a model m . Sequence-independent setup times (i.e., times that depend exclusively on the job to be processed) are included in the assembly time. To simplify the analysis, we assume that setup is performed by an operator while traveling with the workpiece on the conveyor. This assumption does not limit the generalizability of our model. In most real-world assembly lines, for example, in automotive, operators travel with the unit on the constantly paced assembly line. The timing with which models are launched onto the conveyor reflects one of three methods: (i) taking the maximum takt time over all models and stations; (ii) using the (weighted) average takt time; or (iii) dynamically adjusting the takt time—that is, adopting variable rate launching. These strategies are illustrated in Figure 1. The first two approaches necessarily lead to inefficiencies caused by idle time (Cases i and ii) and utility work (Case ii only). We shall demonstrate that, in the ideal case, idle time and utility work can be completely eliminated by varying the launching rate between successive units of production (Case iii). Note, that the utilization of operators is linked to operator idleness, where operator utilization is zero during idle time.

Depending on the nature of the task and on the assembly system's physical layout, station boundaries can be either open or closed. When stations are *closed*, the operator is not

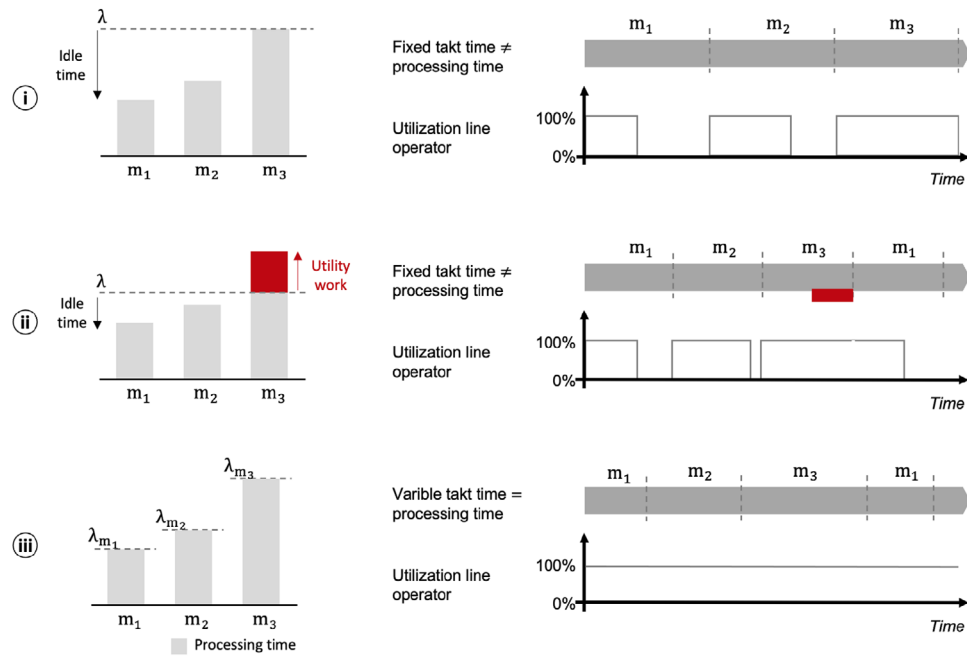


FIGURE 1 Launching rate strategies: Using the maximum process time (Case i), the average process time (Case ii), or variable process times (Case iii) [Color figure can be viewed at wileyonlinelibrary.com]

allowed to cross either its upstream or downstream boundary. This restriction might be needed, for instance, to preclude unproductive traveling or when technological or safety issues require that assigned tasks be performed within the station boundary—for example, in heating chambers or paint booths (Bard et al., 1992). Yet, when a workstation is *open*, the operator can cross its upstream boundary (to start work on a product outside that boundary) and/or its downstream boundary (to finish work on a product). Hence, the overlap or “drift” area between two adjacent stations extends an operator’s allowable work area: by L_j^{UO} and L_j^{DO} for, respectively, the upstream overlap and downstream overlap (recall, however that two adjacent operators cannot interfere with each other’s work). Even in open stations, the maximum overlap between subsequent stations can be restricted—as might be necessary when power tools, material handling equipment, or robotic devices have a limited range. Work overload can occur even in the presence of drift areas, but our model makes it easier to minimize overload. A hybrid MMAL is one that includes both closed and open work stations (Dar-El, 1978). In contrast to most researchers, we do not assume that assembly lines have homogeneous station characteristics. In case of work overload, specialized utility workers assist line operators facing utility work within the station. Another design criterion for MMALs is the strategy (FRL or VRL) for launching workpieces onto the paced conveyor. The chosen launching strategy depends on the takt time and the resulting distance between two successive units on the conveyor. If a constant-speed conveyor links all stations and if there are no buffers between those stations, then the launching interval (a.k.a. the feed-in rate) is equal to the takt time. In an assembly sequence, the distance between two consec-

utive workpieces i and $i + 1$ is determined by two factors: the launching interval λ_i and the conveyor velocity v (Bard et al., 1992). In the FRL scenario, workpieces are spaced equally on the assembly line irrespective of model type. The fixed launching interval is usually calculated as the weighted average of the total assembly time of all products over all stations (Dar-El, 1978; Kim et al., 1996; Sarker & Pan, 2001). If several models with assembly times that greatly exceed the average follow one another, then the operator is moved to the station’s downstream boundary—increasing the risk of work overload or even line stoppage. Since the *fixed* launching interval is the same for all pairs of successive workpieces, it follows that $\lambda = \lambda_i$ for all $i \in I$. In contrast, the *variable* launching interval λ_i allows for launching time variations in the placement of two adjacent products onto the conveyor regardless of the model type: $\lambda_i \geq \lambda_{i+1}$ for all $i \in I$ (Bard et al., 1992). This paper examines the VRL strategy. Our model’s assumptions are summarized in Table 1.

Note also that the product mix is predetermined and that sequences are built using the minimal part set (MPS). The notion of an MPS is widely used in the literature on MMAL sequencing, and it is defined as the smallest part set needed for the proportion of models that matches overall demand D_M (Bard et al., 1992).

Figure 2 presents an operator movement diagram for the case of open stations and VRL. This particular three-station scenario can be viewed as one component of a larger assembly line system. The horizontal dimension represents the operator’s movement, and the vertical position corresponds to unit i ’s location in the sequence. Here, we assume that the MPS is (1, 1, 1); the station length is L_j . Bold vertical lines in

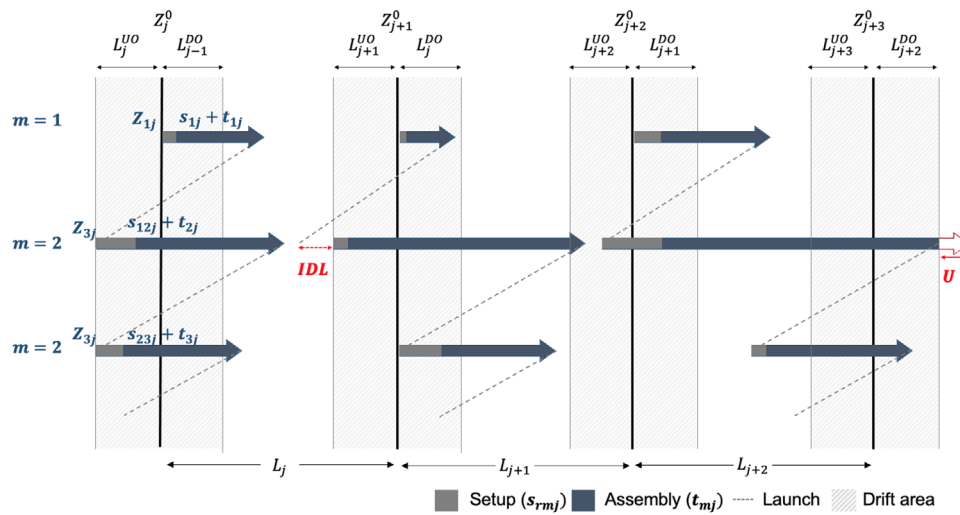


FIGURE 2 Operator movement diagram for the open-station case with variable launching rates [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Assumptions of the model

- The workstations of the assembly line can be open, closed, or hybrid
- No feeding lines or preassembly stations/lines are considered
- A moving conveyor with uniform speed carries the workpieces forward
- Each workstation is occupied by a single operator
- An operator can perform tasks on only one workpiece at a time
- No buffers exist within a workstation or between two adjacent workstations
- Extending the current assembly line is costly
- The workpiece order on the conveyor is fixed and does not vary between workstations
- Sequence-dependent setup times are not part of operator's assembly time
- Material supply is located near the upstream station boundary
- Utility workers assist operators facing utility work within the station

Figure 2 represent the station boundaries, where Z_j^0 represents the station upstream boundary (also called “zero reference point”). The thin vertical lines mark the maximum permitted upstream overlap (L_j^{UO}) and downstream overlap (L_j^{DO}) of station j ; thus stations $j - 1$ and $j + 1$ enclose the operator drift area. Finally, Z_{ij} defines the starting position of product i at station j .

In the figure, a horizontal arrow starting at location Z_{ij} shows the “length equivalent” of the processing time ($t_{mj} + s_{mj}$) needed to finish the work content of workpiece i at station j . The launching rate's length equivalent λ_{iv} between two consecutive workpieces is represented by a dashed line. Once the operator has either completed the workpiece task(s) or reached the downstream limit ($Z_j^0 + L_j + L_j^{DO}$), she returns upstream to continue work on the next item. We assume for the sake of simplicity that the operator moves instantaneously (i.e., in a negligible amount of time) from the completion of work on one product ZF to the starting point of work on the

next product Z (while picking up needed material near Z_j^0). Thus, the operator immediately starts working on the next product at its current location. If the operator is unable to finish work on a product within the station limits, then utility work U comes into play. Such utility work is the first source of inefficiency; the second source is operator idleness (Thomopoulos, 1967). Idle time (IDL) occurs when the operator either finishes work early on a unit or cannot immediately work on the next unit. When there is no idle time, the station operator remains continuously occupied during the entire sequence.

3 | LITERATURE REVIEW

Over the past decades, a large amount of academic research has addressed the design criteria and configuration of MMAL systems. Thomopoulos (1967) distinguishes between two related barriers to an effective utilization of MMAL systems: the assembly line *balancing* problem and the assembly line *sequencing* problem. Line balancing involves the assignment of tasks (i.e., with and without precedence constraints) to a line's workstations so that the lowest takt time—which is usually the same for all products—is achieved. An overview of the line balancing problem can be found, for example, in Baybars (1986), Becker and Scholl (2006), Boysen et al. (2007), and Battaia and Dolgui (2013). Mönch et al. (2021) present a generalizable mixed-integer programming model that regards variable takt times and random customization, that is, random processing times, when solving the assembly line balancing problem. Moreover, Mönch et al. (2022) present a preliminary step to the line balancing and sequencing problem by defining variable takt times groups, where units with similar assembly times are assigned the same takt time. The results show that defining takt time groups can significantly reduce the effort connected to solving the assembly line

FRL	Case A <ul style="list-style-type: none"> • Dar-El and Cucuy (1977) • Okamura and Yamashina (1979) • Burns and Daganzo (1987) • Yano and Rachamadugu (1991) • Bolat and Yano (1992a,b) • Bard et al. (1994) • Bolat et al. (1994) • Xiaobo and Ohno (1994, 1997) • Tsai (1995) • Bolat (1997) • Goldschmidt et al. (1997) • Hyun et al. (1998) • Scholl et al. (1998) • Xiaobo and Ohno (2000) • Sarker and Pan (2001) • Bautista and Cano (2008) 	Case B <ul style="list-style-type: none"> • Thomopoulos (1967) • Dar-El and Cothier (1975) • Kim et al. (1996) • Sarker and Pan (1998) • Dar-El and Nadivi (1981)
	Case C <ul style="list-style-type: none"> • Fattahi and Salehi (2009) 	Case D <ul style="list-style-type: none"> • Dar-El (1978) • Bard et al. (1992) • Bock et al. (2006) • Tong et al. (2013)
VRL	CS	OS

FIGURE 3 Classification of MMAL research based on type of work station and type of launching rate

balancing and sequencing problem. Our paper focuses on the line sequencing problem (for an extensive overview of literature on MMAL sequencing, see Boysen et al., 2009).

In the area of MMAL product sequencing, scholars have highlighted the development of algorithms and (meta-)heuristics to ensure that parts are used at a constant rate and to minimize idle time, utility work, line length, throughput time, total setup costs, and the risk of stopping the conveyor. Because this paper focuses on the multiple objectives of minimizing idle time and utility work while accommodating line length restrictions, here we review only that literature devoted to optimizing one or more of these objectives. In this endeavor, we classify the research on MMAL sequencing in terms of the launching rate (i.e., fixed or variable) and the chosen station type: open station (OS) or closed station (CS) boundaries. For each of the resulting four cases, we discuss the related literature (as illustrated schematically in Figure 3) in turn.

Case A. Dar-El and Cucuy (1977) present, for the single-station case, an optimization algorithm for minimizing the overall assembly line length with zero idle time and utility work. Okamura and Yamashina (1979) introduce a heuristic that aims to minimize—via interchanging product pairs—the maximum downstream displacement that an operator faces during task completion. Burns and Daganzo (1987) use grouping and spacing rules to find an efficient sequence for minimizing setup and capacity costs. Yano and Rachamadugu (1991) develop a mathematical programming model that aims to minimize the work overload at a single station when operators are never idle and perform one of two distinct tasks. Bolat and Yano (1992a, 1992b) describe a “greedy” procedure based on spacing rules; this procedure identifies an optimal solution for minimizing total utility work at a single station that is responsible for two distinct tasks. The dual-criteria minimization model of Bard et al. (1994) uses weighted factors to analyze the trade-off between line length and parts usage. Bolat et al. (1994) formulate a mixed-integer mathe-

tical programming model that minimizes utility work and setup costs. Xiaobo and Ohno (1994, 1997) introduce a simulated annealing approach for large-scale problems, thereby minimizing the risk of conveyor stoppage. Tsai (1995) proves that the problems of minimizing conveyor stoppage risk and utility work are both NP-hard in the strong sense; for the (closed) single-station case with FRL, where all assembly times take one of two distinct values, he proposes an algorithm that reaches an optimal solution in $O(\log N)$ time. Bolat (1997) presents a computational study of a simulated annealing procedure under which units in the sequence are exchanged not randomly but rather based on specific knowledge about the problem, thus minimizing utility work. Goldschmidt et al. (1997) are the first to report complexity results for the sequencing problem of minimizing overall line length so that the operator can complete the required work on all products; the sequencing problem is strongly NP-complete not only for the single-station case but also for the case of multiple stations. Hyun et al. (1998) develop a genetic algorithm that aims to achieve the objectives of reducing utility work, achieving a constant rate of parts usage, and minimizing total setup costs; solutions generated by the algorithm are Pareto optimal (or nearly so). Scholl et al. (1998) develop a Tabu search algorithm that consists of a vocabulary-building strategy to minimize total work overload on the assembly line. Xiaobo and Ohno (2000) consider the operator’s upstream walking time in their heuristic for minimizing the risk of stopping the conveyor. The model of Sarker and Pan (2001) is based on mixed-integer programming and seeks to minimize total costs of idle time and utility work by determining the optimal line parameters: launching rate, station length, operator starting point of work, upstream walking distance, and locus of the operator’s movement. These authors use a sensitivity analysis to show that total costs for idle time and utility work decrease with increasing line length. Bautista and Cano (2008) extend the mathematical model of Okamura and Yamashina (1979) in proposing a heuristic algorithm

that yields a cyclical pattern of operator displacement while minimizing work overload and idleness.

Case B. Thomopoulos (1967) develops a simple approximate algorithm for determining a sequence that minimizes total costs of the inefficiencies—namely, those due to idleness, work deficiency, work congestion, and/or utility work—that result from sequencing an additional model onto the line. Dar-El and Cothier (1975) present a selection heuristic that minimizes the overall station length with no operator interference (i.e., zero operator idle time and utility work). Kim et al. (1996) present a genetic algorithm for minimizing the total line length in hybrid workstations while considering sequence-dependent setup times. Sarker and Pan (1998) formulate a mathematical model, for both the open and closed station approach, that minimizes idle time and utility work; they show that the total costs of idle time and utility work are lower for open stations than for closed stations. Dar-El and Nadivi (2007) detail a step-by-step application of the Dar-El–Cothier mixed-model sequencing algorithm to a telephone exchange–frame assembly line, thereby reducing the number of operators required from 160 to 117.

Case C. Fattahi and Salehi (2009) propose a hybrid metaheuristic, based on simulated annealing, to solve the sequencing problem. This approach incorporates their *launching interval between products* (LIBP) algorithm, which defines the launching intervals between two successive units in the sequence.

Case D. Dar-El (1978) classifies MMAL sequencing problems based on two objectives: minimizing (i) the overall line length or (ii) the throughput time. The optimal line length for zero idle time and utility work is found by defining the extreme operator movement locations as the station's upstream and downstream boundaries. Bard et al. (1992) introduce a mathematical framework to determine the optimal line length or throughput time for the mixed-model sequencing problem under the following design parameters: operator schedule, product mix, station boundaries, and launching strategy. Bard et al. establish that VRL is superior to FRL in terms of reducing line length and throughput time. Bock et al. (2006) extend the pure assembly line sequencing problem by incorporating specific real-time line balancing aspects and disturbances into an simulated annealing approach. Their model also allows assembly tasks to be moved from one workstation to a neighboring station during the execution process, which results in the movement of operators through several stations. Tong et al. (2013) present an advanced “scatter” search heuristic that yields a solution—for large-scale sequencing problems—that minimizes idle time and utility work.

The model that we develop minimizes restrictive modeling assumptions; in particular, it allows for (i) variable launching rates, (ii) open stations, (iii) unbalanced processing times, (iv) variable station lengths, (v) an unrestricted number of models, (vi) an unrestricted number of stations, and (vii) separation of setup and processing time. Most research considers setup time to be part of the processing time and hence independent of the sequence; however, Sun et al. (1999),

Vallada and Ruiz (2011), and Yoshida and Hitomi (1997) are among the scholars who disagree with that approach. Moreover, we exploit structural properties—leading to the ideal case—that enable optimal solutions for VRL with open stations. If these properties do not exist, that is, inhomogeneous assembly times, we introduce the line length algorithm for open stations (LLAOS). Space restrictions play a dominant role in the layout planning of today's assembly lines; it is for this reason that our objective function accounts for extending the line length. By fixing the launching rate and/or setting both the upstream and downstream workstation overlaps to zero, our model can be reduced to Case A, B, or C. This paper is the first to benchmark all four of the cases described above.

4 | VARIABLE RATE LAUNCHING MODEL FOR OPEN STATIONS (VRLMOS)

In many manufacturing settings, competitiveness depends not only on inefficiency costs associated with idle time and utility work but also on space limitations that affect production planning and the facility layout. Hence, our model simultaneously considers all three drivers: idle time, utility work, and line length. This model also accounts for setup times by addressing assembly environments in which products are fundamentally different. We shall employ the notation summarized in Table 2 when describing the mathematical model and its solution procedure.

For a station j , one unit of idle time IDL_{ij} costs C_j^{IDL} and one unit of utility work U_{ij} costs C_j^U . Therefore, idle and utility costs at station j are equal to $C_j^{IDL} \sum_{i=1}^I IDL_{ij} + C_j^U \sum_{i=1}^I U_{ij}$. Over all stations, these costs amount to $\sum_{j=1}^J (C_j^{IDL} \sum_{i=1}^I IDL_{ij} + C_j^U \sum_{i=1}^I U_{ij})$. The objective function also contains a linear penalty term C^L for the additional length ΔL that is added (without loss of generality) to the line length's lower bound. To simplify the computations, we assume a linear relation between the length of a production line and its costs. When we account for the MPS' frequency F , the total cost of satisfying overall demand D is formally expressed as Equation (1)—whose additional constraints are given in what follows. In many manufacturing settings, competitiveness depends not only on inefficiency costs associated with idle time and utility work but also on space limitations that affect production planning and the facility layout. Hence, our model simultaneously considers all three drivers: idle time, utility work, and line length. This model also accounts for setup times by addressing assembly environments in which products are fundamentally different. We shall employ the notation summarized in Table 2 when describing the mathematical model and its solution procedure.

For a station j , one unit of idle time IDL_{ij} costs C_j^{IDL} and one unit of utility work U_{ij} costs C_j^U . Therefore, idle and utility costs at station j are equal to $C_j^{IDL} \sum_{i=1}^I IDL_{ij} +$

TABLE 2 Notation

Indices	
i	Position of a unit in the sequence, $i = 1, \dots, I$
j	Workstation number, $j = 1, \dots, J$
k	Workstation index, $k = 1, \dots, J$
m	Model type, $m = 1, \dots, M$
n	Model type, $n = 1, \dots, M$
r	Model type, $r = 1, \dots, M$
Input parameters	
C_j^U	Costs of utility work at station j
C_j^{IDL}	Costs of idle time at station j
C^L	Costs of an incremental increase in the line length
D	Overall demand
D_m	Demand for model m 's minimal part set (MPS)
F	Number of times that MPS is repeated to meet overall demand
I	Number of workpieces in the MPS
J	Number of stations
L_{\max}	Upper bound of line length
M	Number of models within the MPS
s_{rmj}	Setup time required when changing from model r to model m at station j
t_{mj}	Assembly time of model m at station j
v	Conveyor velocity
β	Downstream shift area
ΔL	Line length increment
Computed parameters	
L_j^{UO}	Length of upstream overlap of station j
L_j^{DO}	Length of downstream overlap of station j
L_j^{\min}	Lower bound of station length
L_{\min}	Lower bound of line length
p_{mj}	Processing time of model m at station j
p_{\max}	Maximum processing time over all models and stations (incl. setup and assembly)
Z_j^0	Zero reference point for station j
Decision variables	
d_{ij}	Upstream walking distance of operator between workpiece i and workpiece $i + 1$
IDL_{ij}^1	Idle time of operator due to waiting for workpiece i to enter station j
IDL_{ij}^2	Idle time of operator because workpiece i is not finished
IDL_{ij}	Total idle time at station j for workpiece i
L_j	Length of station j
U_{ij}	Utility work at station j for workpiece i
X_{irm}	1 if workpieces $i - 1$ and i are models r and m , respectively; 0 otherwise
Z_{ij}	Starting position of work for operator on workpiece i at station j

(Continues)

TABLE 2 (Continued)

ZF_{ij}	Finishing position of work for operator on workpiece i at station j
λ_i	Launching interval between workpiece $i - 1$ and workpiece i

$C_j^U \sum_{i=1}^I U_{ij}$. Over all stations, these costs amount to $\sum_{j=1}^J (C_j^{IDL} \sum_{i=1}^I IDL_{ij} + C_j^U \sum_{i=1}^I U_{ij})$. The objective function also contains a linear penalty term C^L for the additional length ΔL that is added (without loss of generality) to the line length's lower bound. To simplify the computations, we assume a linear relation between the length of a production line and its costs. When we account for the MPS' frequency F , the total cost of satisfying overall demand D is formally expressed as Equation (1)—whose additional constraints are given in what follows:

$$\min F \left(\sum_{j=1}^J \left(C_j^{IDL} \sum_{i=1}^I IDL_{ij} + C_j^U \sum_{i=1}^I U_{ij} \right) + C^L \Delta L \right) \quad (1)$$

subject to

$$L_j^{\min} = \arg \min_{t_{mj}, \forall m} t_{mj} \quad \forall j; \quad (2)$$

$$L_{\min} = \sum_{j=1}^J L_j^{\min}; \quad (3)$$

$$L_{\min} + \Delta L \leq L_{\max}; \quad (4)$$

$$\sum_{k=1}^{j-1} L_k = Z_j^0 \quad \text{for } j = 2, \dots, J, \quad Z_1^0 = 0; \quad (5)$$

$$L_{\min} + \Delta L = \sum_{j=1}^J L_j; \quad (6)$$

$$\sum_{i=1}^I \sum_{r=1}^M X_{irm} = D_m \quad \forall m; \quad (7)$$

$$\sum_{m=1}^M \sum_{r=1}^M X_{irm} = 1 \quad \forall i; \quad (8)$$

$$Z_{ij} = \max \left\{ Z_{(i-1)j} + v \sum_{m=1}^M \sum_{r=1}^M X_{(i-1)rm} (s_{rmj} + t_{mj} - U_{(i-1)j}) - d_{(i-1)j}, Z_j^0 - L_j^{UO} \right\} \quad \text{for } i = 2, \dots, I, \quad \forall j; \quad (9)$$

$$Z_{1j} = \max \left\{ Z_{1(j-1)} + v \sum_{m=1}^M \sum_{r=1}^M X_{1rm}(s_{rm(j-1)} + t_{m(j-1)} - U_{1(j-1)}), Z_j^0 \right\} \text{ for } j = 2, \dots, J; \quad (10)$$

$$d_{ij} = \min \left\{ Z_{ij} - Z_j^0 + L_j^{UO} + v \sum_{m=1}^M \sum_{r=1}^M X_{irm}(s_{rmj} + t_{mj} - U_{ij}) - (Z_{(i+1)(j-1)} + v \sum_{m=1}^M \sum_{r=1}^M X_{(i+1)rm}(s_{rm(j-1)} + t_{m(j-1)} - U_{(i+1)(j-1)})) \right\} \text{ for } i = 1, \dots, I-1, \quad j = 2, \dots, J; \quad (11)$$

$$d_{i1} = \min \left\{ Z_{i1} - Z_1^0 + L_1^{UO} + v \sum_{m=1}^M \sum_{r=1}^M X_{irm}(s_{rm1} + t_{m1} - U_{i1}), \lambda_i v \right\} \text{ for } i = 1, \dots, I-1; \quad (12)$$

$$d_{ij} = \min \left\{ Z_{ij} + v \sum_{m=1}^M \sum_{r=1}^M X_{irm}(s_{rmj} + t_{mj} - U_{ij}) - Z_j^0, \lambda_i v \right\} \forall j; \quad (13)$$

$$IDL_{ij}^1 = \max \left\{ 0, \left(Z_j^0 - L_j^{UO} - \left(Z_{(i-1)j} + v \sum_{m=1}^M \sum_{r=1}^M X_{(i-1)rm}(s_{rmj} + t_{mj} - U_{(i-1)j}) - \lambda_i v \right) / v \right) \right\} \text{ for } i = 2, \dots, I-1, \forall j; \quad (14)$$

$$IDL_{ij}^1 = \max \left\{ 0, Z_j^0 + \lambda_i v - \left(Z_{ij} + v \sum_{m=1}^M \sum_{r=1}^M X_{irm}(s_{rmj} + t_{mj} - U_{ij}) \right) \right\} + IDL_{(I-1)j} \quad \forall j; \quad (15)$$

$$IDL_{ij}^2 = \max \left\{ 0, Z_{i(j-1)} + v \sum_{m=1}^M \sum_{r=1}^M X_{irm}(s_{rm(j-1)} + t_{m(j-1)} - U_{i(j-1)}) - \left(Z_{(i-1)j} + v \sum_{m=1}^M \sum_{r=1}^M X_{(i-1)rm}(s_{rmj} + t_{mj} - U_{(i-1)j}) - \lambda_i v \right) - IDL_{ij}^1 \right\} \forall i, \forall j; \quad (16)$$

$$IDL_{ij} = IDL_{ij}^1 + IDL_{ij}^2 \quad \forall i, \quad \forall j; \quad (17)$$

$$U_{ij} = \max \left\{ 0, \left(Z_{ij} + v \sum_{m=1}^M \sum_{r=1}^M X_{irm}(s_{rmj} + t_{mj}) - (Z_j^0 + L_j + L_j^{DO}) \right) / v \right\} \text{ for } i = 1, \dots, I-1, \quad \forall j; \quad (18)$$

$$U_{ij} = \max \left\{ 0, \min \left\{ (s_{rmj} + t_{mj}), \left(Z_{ij} + v \sum_{m=1}^M \sum_{r=1}^M X_{irm}(s_{rmj} + t_{mj}) - Z_j^0 - \min\{\lambda_i v, L_j - L_j^{DO}\} \right) / v \right\} \right\} \quad \forall j; \quad (19)$$

$$d_{ij} \geq 0, IDL_{ij}^1 \geq 0, IDL_{ij}^2 \geq 0, IDL_{ij} \geq 0, U_{ij} \geq 0, Z_{ij} \geq 0 \quad \forall i, j;$$

$$\lambda_i \geq 0 \quad \forall i; X_{irm} \in \{0, 1\} \quad \forall i, r, m; \Delta L \geq 0.$$

4.1 | Line length and zero reference point

Most assembly line systems have a lower bound on the station length, L_j^{\min} , that reflects space requirements due to machines, robots, and so forth (Falkenauer, 2005). We define this lower bound on the station length as the minimum

assembly time over all models for that station (Equation 2). Hence, the lower bound on the overall line length, L_{\min} , is the sum of the lower bounds of all the station lengths (Equation 3). The upper bound on the available assembly line length, L_{\max} , is restricted by the available facility space (or possible extensions thereof). Since utility work decreases with increasing line length, it follows that the optimal line length for efficient operations lies between the lower and upper bounds on total line length. The existing total line length L is calculated as the line length increase ΔL plus the lower line length bound L_{\min} (Equation 4). We define the zero reference point Z_j^0 as the upstream boundary of station j (Equation 5). For the first station, the zero reference point is set to zero: $Z_1^0 = 0$. Also, the line length is equal to the sum of the station lengths (Equation 6).

4.2 | MPS and model assignment

The MPS required to meet demand D_m for model m is satisfied by the sequence (Equation 7), and only one model m is assigned to a position i in the sequence (Equation 8). The term X_{irm} is set equal to 1 if model r is sequenced at position $i - 1$ and model m is sequenced at position i ; otherwise, $X_{irm} = 0$.

4.3 | Operator starting position

An operator starts her work within the allowable upstream station overlap area $Z_j^0 - L_j^{\text{UO}}$, and the processing time that an operator works on the unit is given by $\sum_{m=1}^M \sum_{r=1}^M X_{irm}(s_{rmj} + t_{mj} - U_{ij})$. After finishing that task, the operator walks back the upstream distance d_{ij} to start work on the next item. Thus, the starting position Z_{ij} is either (a) the position of the operator after finishing her task on the previous product and returning upstream or (b) the allowable upstream station boundary (Equation 9). For $i = 1$, the operator's tasks start at the zero reference point (Z^0) if the previous task on the workpiece has already been completed; otherwise, the operator's starting position is the finishing position of the previous station's task (Equation 10). The zero reference point for the first station is set to 0 (i.e., $Z_{11} = 0$), a condition that ensures—provided all initial conditions remain the same—that every sequence is regenerative (Bolat, 1997).

4.4 | Operator's upstream walking distance

The operator's upstream walking distance d_{ij} is not always the length between two adjacent workpieces on the conveyor. When determining d_{ij} , three cases must be considered (Equation 11). In the first case, the next unit has entered the station and the previous operator has finished her task on it. In the second case, the workpiece has not yet arrived at the station boundary and so the operator must wait for it to enter the

station. In the third case, the workpiece has entered the station even though the previous operator has not yet finished her task on the unit; hence the operator must wait until the prior task is finished. For $j = 1$, there can be no unfinished work from a prior station and so the upstream walking distance simplifies to Equation (12). For $i = I$, the upstream walking distance is either the launching interval or the length between the finishing position and the zero reference point (i.e., since the operator finishes every sequence at the station's zero reference point). In the latter case, the operator leaves unfinished work content for a utility worker (Equation 13).

4.5 | Idle time

Idle time (IDL_{ij}) arises in one of two scenarios. First, the operator may have to wait at the upstream station boundary for the workpiece to enter that station (IDL_{ij}^1); this case is captured by Equation (14). Second, the operator may be idle because the task in the prior station has not been finished (IDL_{ij}^2) or the operator of the prior station is currently working on the unit within the focal operator's station limits (Equation 16). In either case, the worker is idle while waiting to start assembly work on the next unit. For the first workpiece in the sequence, idle time is necessarily zero: $IDL_{1j}^1 = 0$. For $i = I$, any subsequent idle time IDL_{Ij}^1 is determined by the unit's own launching rate and that of the prior unit (Equation 15). Total idle time IDL_{ij} is the sum of IDL_{ij}^1 and IDL_{ij}^2 (Equation 17). If total idle time is zero, then (a) the operator's upstream walking distance is equal to the space between two successive units on the assembly line and (b) the operator utilization rate is 100%.

4.6 | Utility work

Utility work U_{ij} arises when the station operator must move outside the allowable downstream drift area $Z_j^0 + L_j + L_j^{\text{DO}}$ to finish work content on the current unit (Equation 18). Such work can occur if the operator's finishing position, $Z_{ij} + v \sum_{m=1}^M \sum_{r=1}^M X_{irm}(s_{rmj} + t_{mj})$, exceeds the allowable downstream drift area. When an operator is about to reach her workstation's allowable downstream overlap, the work overload is assigned to a "utility operator" who assists the operator at that station (Hyun et al., 1998; Yano & Rachamadugu, 1991). Thus, the task is completed before the station downstream drift area ends, and the resulting work overload has no effect on the next station. For $i = I$, the operator ends the sequence at her station's zero reference point (Hyun et al., 1998). Because operators start at the zero reference point, this constraint guarantees that sequences are both comparable and replicable. If the distance between the workpiece's finishing position and the zero reference point exceeds the distance dictated by the launching rate, then additional utility work is incurred (Equation 19).

5 | IDEAL VRL APPROACH

In this section, we identify conditions that simplify the complex VRL model. These underlying assembly line requirements offer guidance for manufacturing companies—which face a wide variation in model process times—on where best to direct their investment and planning capacities. We show that the ideal VRL approach eliminates all labor inefficiencies (Theorem 1), almost entirely resolves the assembly line sequencing problem (Theorem 2), and minimizes the line length (Theorem 3). The conditions of Theorem 1 directly imply Theorem 2; yet given the latter's importance, we formulate it separately. For all three theorems, we set the upstream and downstream station overlaps to zero: $L_j^{\text{UO}} = L_j^{\text{DO}} = 0$ for all j . We shall use p_{ij} to abbreviate the station processing time, $\sum_{m=1}^M \sum_{r=1}^M X_{irm}(s_{rmj} + t_{mj})$; and $\hat{p} = \arg \max_{p_{ij}, \forall i,j} p_{ij}$ denotes the maximum processing time (including setup times) over all units.

5.1 | Productivity

Theorem 1. *In the ideal VRL approach, idle time and utility work are completely eliminated and operator utilization is 100% under the following conditions: (i) Station lengths are set to the maximum processing time (over all models) multiplied by the conveyor speed, $vp_{ij} \leq L_j$ for all i, j ; (ii) operators always start at the zero reference point, $Z_{ij} = Z_j^0$ for all i, j ; (iii) the launching rate is equal to the prior processing time of unit $i - 1$, $p_{(i-1)j} = \lambda_i$ for all i, j ; and (iv) a model has an uniform processing time over all stations, $p_{ij} = p_{ik}$ for all i, j, k .*

Proof. Utility work. Using condition (ii), we can write $Z_{ij} = Z_j^0$ for all i, j and thereby reduce Equation (17) to $U_{ij} = \max\{0, p_{ij} - L_j/v\}$ for all i, j . Condition (i) states that no processing time can exceed the station length, $vp_{ij} \leq L_j$ for all i, j ; hence $U_{ij} = 0$ for all i, j . It follows that no additional staffing is needed, since operators fulfill their tasks within the station boundary and are never overloaded. \square

Idle time. Conditions (iii) and (iv) reduce Equation (14) to $IDL_{ij}^1 = \max\{0, U_{(i-1)j}\}$ for all i, j . Since $U_{ij} = 0$ for all i, j , we have $IDL_{ij}^1 = 0$ for all i, j . Now, by conditions (iii) and (iv) and given that $U_{ij} = 0$ for all i, j , Equation (15) yields $IDL_{ij}^2 = 0$ for all i, j . Therefore, operator idleness is eliminated because workers can immediately start their task on the next unit at the upstream boundary without having to wait for that unit to enter the station or for the prior operator to finish her task.

Single-station case. From conditions (ii) and (iv) it follows that all stations have the same length: $L_j = v\hat{p}$ for all j . Moreover, condition (ii) ensures that all operators start at the upstream station boundary; hence all operators perform

all of their work at their respective stations (condition (i)). Thus, the complex general model is reduced to a single-station problem—that is, since all stations and operators have the same characteristics and are independent of each other. Figure 4 illustrates the operator movement diagram for the single-station case. Additional stations can be added before or after this station because the station's operators then do not interfere with one another. Line productivity is 100% because there are no labor inefficiencies, all stations have the same length, and stations are connected without buffers.

Key to the applicability of the ideal case is the assumption that work can be perfectly balanced across workstations. This assumption sounds rather restrictive, but in practice, companies like Fendt achieve an almost perfect balance of work for their models of over 94%; where 100% is a perfect line balance (Huchzermeier et al., 2020b). Note, however, there is a practical need for a balance in the mid-nineties range as this grants operators some slack in case of contingencies.

5.2 | Model-mix flexibility

Theorem 2. *The ideal VRL approach guarantees the highest efficiency for any product sequence, so models can be launched onto the assembly line—at any time and in any order—without causing idle time or utility work.*

Proof. We use the four conditions of Theorem 1 to prove Theorem 2. As shown in our model, the operator starting position Z_{ij} depends on the launching rate, the processing time, idle time, and the previous unit's utility work. However, condition (ii) of Theorem 1 sets $Z_j^0 = Z_{ij}$ for all i, j . So if we separate the sequencing problem into I independent parts—where I is the number of products in a sequence—then Z_{ij} can be decoupled from the previous workpiece. In this case, each unit i can be considered as an independent part of the sequence. Note also that the conditions of Theorem 1 eliminate both idle time ($IDL_{ij} = 0$ for all i, j) and utility work ($U_{ij} = 0$ for all i, j). Since all parts of the sequence are independent and since there is no factor that can induce labor inefficiencies, it follows that resequencing in any (even random) order itself precludes the absence of operator idleness and work overload.

For assembly systems with and without setup times, satisfying the conditions of Theorem 1 leads to a 100% workforce utilization—that is, because \hat{p} already includes the maximum setup time for a model. In other words, no maximum (resp., unbalanced) loading of the assembly line will overburden the assembly line (resp., its operators). If setup times are relevant, then a simple optimization will batch similar products and reduce setup effort. Because such batching proceeds in a straightforward manner, this myopic optimization is omitted from our model formulation. \square

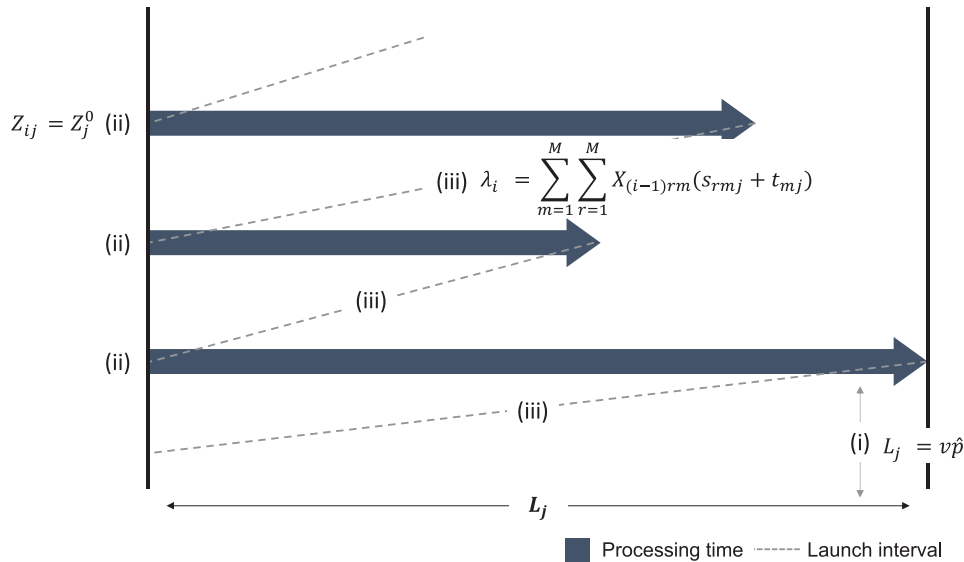


FIGURE 4 Operator movement under the ideal VRL approach for the single-station case [Color figure can be viewed at wileyonlinelibrary.com]

5.3 | Line length

Theorem 3. When multiple operators are allowed to work simultaneously on a product, the optimal line length L is determined by one of the following cases: (a) $L = v\hat{p}J$ for $\beta = 0$, (b) $L = v\hat{p}J - \beta(J - 1)$ for $\beta \in (0, v\hat{p})$, or (c) $L = v\hat{p}$ for $\beta = v\hat{p}$, when the four conditions of Theorem 1 hold also for Theorem 3. In the case of Fendt, this area β is negotiated with the works council and lies between 10% and 15%.

Proof. The extent to which a workstation's task may be shifted to the previous station is determined by the factor $\beta \in [0, v\hat{p}]$, where β 's value is the same for all stations. β determines the area in which operators are allowed to work simultaneously on the product. \square

Case (a). If $\beta = 0$, then stations cannot be shifted into one another and operators are not allowed to work simultaneously on products. In this case, the line length is $L = v\hat{p}J$; in other words, it is the sum of all station lengths $L_j = v\hat{p}$ for all j (condition (i) of Theorem 1).

Case (c). For $\beta = v\hat{p}$, stations are completely shifted into one another and operators are allowed to work simultaneously on a product at any time. Since each operator starts her work at the upstream station boundary $Z_j^0 = Z_{ij}$ for all i, j (condition (ii) of Theorem 1), it follows that the line collapses to the shared operating area $L = v\hat{p}$. The locus of each operator is identical to that of the others because their respective processing times are uniform for a given product model (condition (iv) of Theorem 1). It should be clear that any further reduction of the line length would result in utility work (condition (i) of Theorem 1).

Case (b). For $\beta \in (0, v\hat{p})$, all stations are shifted into one another by the amount β . Thus, the overall line length L is somewhere between the extremes represented by Case (a) and Case (c).

6 | SOLUTION METHOD FOR THE GENERAL CASE

In this section, we discuss our heuristic solution method for those manufacturing environments in which it is not possible for the assembly process to be aligned with the ideal case's four conditions (Theorem 1). For the general case, the planning unit has to solve the NP-hard assembly line sequencing problem in order to reduce labor inefficiencies and line length. In our simulation, we show that—in terms of costs and line length—VRL with closed stations (Case C) dominates FRL with closed stations (Case A) in all of the scenarios considered. The same holds true for open stations.

6.1 | Line length algorithm for open stations (LLAOS)

We have established that idle time and utility work are substantially reduced when production switches from fixed to variable takt times. In general, VRL (with launching rate λ_i for unit i in a product sequence) at minimum matches the costs of FRL (with a launching rate λ) as VRL offers the freedom to change the launching rate λ_i between products. If $\lambda_i = \lambda_{i+1}$ for all i , then VRL is reduced to FRL; thus the latter is a subset of the former. Nevertheless, the assembly line is still not entirely efficient because of deviations in station lengths and drift areas and in model processing times across stations. Our LLAOS (Algorithm 1a) helps to identify, for the case of variable launching rates, a product sequence that minimizes the costs associated with idle time and utility work while accounting for space requirements.

The LLAOS. The general idea behind the LLAOS is to increase (gradually) the length of only one station, j^* , in every iteration. In this procedure, we start with the line

ALGORITHM 1a Line length algorithm for open stations (LLAOS)

```

0   $C_{\min}^{\text{IDL},U,L} = \alpha$ ;  $C_l^{\text{IDL},U,L} = 0$ ;  $C_{\min}^{\text{IDL},U} = 0$ ;  $C^{\text{IDL},U} = 0$ ;  $C^L = 1$ ;
    $C_j^{\text{IDL}} = 0.2$ ,  $C_j^U = 0.5 \forall j$ 
1  For  $l = L_{\min}$  to  $L_{\max}$ 
2    For  $i = 1$  to  $I$ 
3       $ZF_{ij} =$ 
         $\min\{\sum_{k=1}^j L_k + L_j^{\text{DO}}, Z_{ij} + v \sum_{m=1}^M \sum_{r=1}^M X_{irm}(s_{rmj} + t_{mj})\} \forall j$ 
4       $p_i = \min\{(ZF_{ij} - \sum_{k=1}^{j-1} L_k - \max\{0, L_{j-1}^{\text{DO}}\})/v, (s_{rmj} + t_{mj})\} \forall j$ 
5       $\lambda_i = p_i$ 
6       $C_{\min}^{\text{IDL},U} = \sum_{j=1}^J \text{IDL}_{(i+1)j} C_j^{\text{IDL}} + \sum_{j=1}^J U_{(i+1)j} C_j^U$ 
7      While  $(C_{\min}^{\text{IDL},U} == C^{\text{IDL},U} \text{ OR } (i = I \text{ AND } \lambda_i \leq \min_{\forall j}\{L_j\}))$  Do
8         $\lambda_i = \lambda_i + 1$ 
9         $C^{\text{IDL},U} = \sum_{j=1}^J \text{IDL}_{(i+1)j} C_j^{\text{IDL}} + \sum_{j=1}^J U_{(i+1)j} C_j^U$ 
10       If  $(i \leq I \text{ AND } C^{\text{IDL},U} \leq C_{\min}^{\text{IDL},U})$  Then
11          $p_i = p_i + 1$ 
12          $C_{\min}^{\text{IDL},U} = C^{\text{IDL},U}$ 
13       ElseIf  $(i = I \text{ AND } C^{\text{IDL},U} \geq C_{\min}^{\text{IDL},U} \text{ AND } \lambda_i \leq \min_{\forall j}\{L_j\})$ 
         Then
14          $p_i = p_i + 1$ 
15       End If
16     End While
17      $\lambda_i = p_i$ 
18   End For
19    $C^{\text{IDL},U,L} = \sum_{j=1}^J \sum_{i=1}^I U_{ij} C_j^U + \sum_{j=1}^J \sum_{i=1}^I \text{IDL}_{ij} C_j^{\text{IDL}} + C^L \Delta L$ 
20   If  $(C_l^{\text{IDL},U,L} \leq C_{\min}^{\text{IDL},U,L})$  Then
21      $C_{\min}^{\text{IDL},U,L} = C_l^{\text{IDL},U,L}$ 
22   End If
23   If  $(l + 1 \leq L_{\max})$  Then
24      $L_{j^*} = L_{j^*} + 1$ 
25   End If
26 End For
```

length's lower bound, L_{\min} , and end with its upper bound, L_{\max} (Step 1). The process requires that we determine the variable launching rates by using the *variable rate launching algorithm for open stations* (VRLAOS; Steps 2–18). For that purpose, we extend the LIBP algorithm (see Case C in Section 3) developed by Fattahi and Salehi (2009) to meet the requirements of our model in an open station setting. Fattahi and Salehi (2009) were the first to introduce a heuristic approach, the LIBP algorithm, for determining the variable launching intervals between successive units on MMALs with closed stations minimizing only labor costs; hence these authors do not consider variations in assembly line length and do not allow for operators drifting into other stations or costs for increasing line length. We extend this approach by allowing for open station boundaries while simultaneously minimizing labor costs and overall assembly line length. Also, we regard setup times enabling the adoption of our algorithm in a larger group of production

environments. The VRLAOS algorithm's goal is to shift the operator starting position as far upstream as possible to avoid the risk of utility work. After the launching intervals are determined, we compute the total costs $C_l^{\text{IDL},U,L}$ of the current iteration by summing up the costs for idle time, utility work, and longer line length (Step 19). These costs are then compared with the minimum costs over all iterations, $C_{\min}^{\text{IDL},U,L}$, and result in an update if a reduction is achieved (Steps 20–22). Finally, the length of station j^* is increased by an increment of one line length (Steps 23–25).

The procedure for selecting j^* is presented in the next paragraph. The LLAOS allows overall costs to increase over iterations in the short term because, in some iterations, the cost reduction due to less utility work does not fully compensate for idle time and line length costs. However, this increase in overall costs may well be absorbed in subsequent iterations. We therefore choose the line length upper bound as a natural termination criterion, thereby limiting the number of iterations performed; thus the total cost calculated during each iteration is compared with the lowest cost calculated previously.

Procedure for selecting j^* . Here, we focus on practical, iterative improvement procedures for determining which station j^* should be increased in length within an iteration of the LLAOS. The first selection procedure (Option 1) is based on a local search approach that, before increasing the line length, evaluates all neighboring solutions. Thus, the neighborhood solution space is constructed by increasing the length of each station, one by one, and comparing the total cost for idle time and utility work that results; the station j^* that yields the greatest reduction in cost when its length is increased is then selected (Equation 21). Option 2 instead considers the current costs of idle time and utility work without regard to information on neighboring solutions, thereby reducing the computational effort required (Equation 22). Options 3 and 4 concentrate exclusively on utility work while ignoring idle time; the goal is a horizontal smoothing of the workload over every station. These latter two options aim either to reduce and align utility work at all stations (Equation 23) or to reduce the time that a utility worker is bound to one station (Equation 24).

Option 1. Selecting the station for which the *future* total cost of idle time and utility work would be highest:

$$j^* = \arg \max_{\forall j} \left(\sum_{i=1}^I U_{ij} C_j^U + \sum_{i=1}^I \text{IDL}_{ij} C_j^{\text{IDL}} \right) \text{ for } L_j = L_j + 1. \quad (20)$$

Option 2. Selecting the station for which the *current* total cost of idle time and utility work is highest:

$$j^* = \arg \max_{\forall j} \left(\sum_{i=1}^I U_{ij} C_j^U + \sum_{i=1}^I \text{IDL}_{ij} C_j^{\text{IDL}} \right) \text{ for } L_j. \quad (21)$$

ALGORITHM 1b Extension of the LLAOS

```

27   For  $l$  in  $J$ 
28     For  $x$  in  $(-1, 1)$ 
29        $L_l = L_l + x$ 
30       VRLAOS
31        $C_l^{\text{IDL},U,L} = \sum_{i=j}^J \sum_{i=1}^I \text{IDL}_{ij} C_j^{\text{IDL}} + \sum_{i=j}^J \sum_{i=1}^I U_{ij} C_j^U + C^L \Delta L$ 
32       If  $(C_{\min}^{\text{IDL},U,L} \geq C_l^{\text{IDL},U,L})$  Then
33          $C_{\min}^{\text{IDL},U,L} = C_l^{\text{IDL},U,L}$ 
34       End If
35     End For
36   End For

```

Option 3. Selecting the station with the highest sum of utility work over *all units* in the sequence:

$$j^* = \arg \max_{vj} \sum_{i=1}^I U_{ij}. \quad (22)$$

Option 4. Selecting the station with the most utility work for *one unit* in the sequence:

$$j^* = \arg \max_{vj} (U_{1j}, \dots, U_{Ij}). \quad (23)$$

Extending the LLAOS. We now discuss a simple procedure for improving the LLAOS even further. Starting with its solution $C_{\min}^{\text{IDL},U,L}$, Algorithm 1b investigates the neighboring solution space by gradually increasing and decreasing—in one-length increments—the length of each station in turn (Step 28). For every entity in the solution space, we apply the VRLAOS (Step 30) and then calculate total costs (Step 32). Finally, the improvement yielding the greatest cost reduction is stored (Step 33). Our results when employing this simple extension are presented in the next section's benchmark analysis.

6.2 | Benchmark study and analysis

This section presents the findings derived from our large-scale analysis of 2070 product sequences. We benchmarked the LLAOS selection procedures for Cases A, B, C, and D with respect to total cost of inefficiencies and assembly line length.

6.2.1 | Data generation

The data set consists of 30 scenarios, each featuring a unique combination of the underlying MPS and model processing times. Each MPS is randomly generated under the condition that all models occur at least once in the sequence. Processing times are derived from a uniform distribution whose intervals are given in Table 3, and they include both assembly time and setup time. All scenarios incorporate the following

TABLE 3 Intervals for uniform distribution of processing times

	t_{m1}	t_{m2}	t_{m3}
Model 1	[8, 12]	[8, 14]	[8, 10]
Model 2	[4, 8]	[6, 9]	[4, 6]
Model 3	[5, 9]	[6, 11]	[5, 8]

line-related parameters: frequency, $F = 1$; number of products in the MPS, $I = 6$; number of stations, $J = 3$; number of models, $M = 3$; models, $m = (1, 2, 3)$; cost of idle time, $C_j^{\text{IDL}} = 0.2$; cost of utility work, $C_j^U = 0.5$; cost of increasing the line length, $C^L = 1$; and conveyor velocity, $v = 1$. In addition, the upstream and downstream overlaps are set to 10% of the lower station bound and then rounded up to the next higher integer value. These parameters anchor in our review of real-world assembly systems and thus can be seen as representative for a real-world production setup, such as Fendt's production in Marktoberdorf, Germany. Furthermore, the three-station layout is treated as a subsection of a larger assembly line. Depending on the MPS, each scenario has 30, 60, or 90 sequences (or permutations); hence our benchmark analysis involves 2070 compared sequences. Rather than focusing on stochastic variations of processing time, this computational setup allows to solve the more relevant problem of implementing a total mix of sequences to be assembled on the same line and thus on the same space requirement. In today's assembly lines, stochastic processing times are rather rare since during the last decades, a great amount of effort has been deployed to reduce such deviations. Peter Bebersdorf, Director Manufacturing at AGCO-Fendt, mentioned that deviations may exist when a product is newly introduced on the line, but never reaches a value of 10% or higher. To the contrary, at AGCO-Fendt the workers council demands, at the minimum, three measurements of process times before the start of production. These three measurements set the baseline for an approval of the processing times by the workers council. In general, large deviations of process times raise red flags with the workers council and are addressed by management directly.

6.2.2 | Benchmarking the LLAOS options

In order to identify the preferred LLAOS option, we compare the minimal costs of all 2070 sequences for each option and case. Table 4 reports the percentage of sequences for which the specified option yields the best result, where the value in parentheses is the percentage of sequences for which only this option achieves the single best outcome of all options. Option 1 is clearly the dominant choice for at least 84% of the sequences. In fact, for the most restrictive strategy (Case A), the first selection procedure always leads to the best sequence; and it leads to the single best solution in 57% of the cases. Another noteworthy observation is that Option 1 becomes less dominant as flexibility is increased—for

TABLE 4 Comparison of the LLAOS options by percentage of sequences yielding best (single best) result

	Case A	Case B	Case C	Case D
Option 1	100 (57)	98 (57)	92 (37)	84 (32)
Option 2	33 (0)	39 (1)	52 (7)	63 (12)
Option 3	27 (0)	23 (1)	26 (0)	30 (1)
Option 4	15 (0)	15 (0)	12 (0)	34 (0)

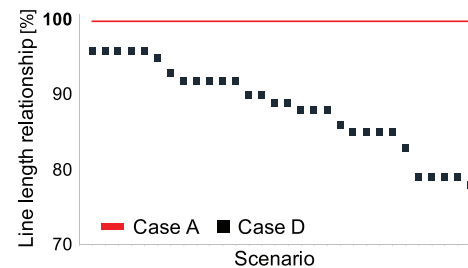
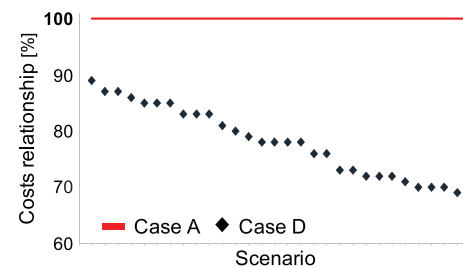
example, by opening up station boundaries and/or implementing VRL. Even so, Option 1 remains the best choice for flexible systems: For Case D, it delivers 84% of best solutions and 32% of the single best solutions.

These results can be improved if one undertakes the simple local search procedure described in Section 6.1. By examining neighboring solutions, our extended LLAOS for Option 1 (hereafter, LLAOS-1) reduces the costs associated with Cases B, C, and D in (respectively) 37%, 26%, and 52% of the sequences. We remark that LLAOS-1b does not yield any significant improvement for Case A; the reason is that it obtains the same station lengths (for most instances) as the LLAOS and thus results in the same inefficiency costs for each sequence.

With these parameters, a full enumeration over the line length was performed; starting with the minimal station lengths and iterating through all combinations until arriving at the maximum line length. The performed enumeration showed a runtime of roughly 2700 s. Moreover, the average optimality gap is 2.3% with one outlier of 8.5% (but still achieving better results than the fixed launching rate setting).

6.2.3 | Analysis of deviations in line length and total costs

Our purpose here is to show that shifting the assembly strategy from Case A to Case D reduces assembly line length and overall assembly costs. Figure 5 and Figure 6 plot the percentage of improvement in (respectively) line length and overall cost when the LLAOS-1 is implemented with variable takt times and open stations. As expected, Case A results in the highest average costs for all 30 scenarios. We can also see that Case A (resp., Case D) corresponds to the highest (resp., lowest) average cost and minimal cost for every scenario. These graphs show that changing from Case A to Case D reduces the line length of the best sequence by anywhere from approximately 4% to 22% (Figure 5) and also reduces total costs by approximately 11% to 37% (Figure 6). For any randomly chosen sequence, changing from Case A to Case D results in an average cost reduction of approximately 22%.

**FIGURE 5** Line length improvement, from Case A to Case D, of best sequence [Color figure can be viewed at wileyonlinelibrary.com]**FIGURE 6** Cost reduction, from Case A to Case D, of best sequence [Color figure can be viewed at wileyonlinelibrary.com]

We next discuss whether manufacturing executives who seek to push, in a stepwise fashion, their Case A assembly process toward a more flexible approach should start by opening up station boundaries or rather by adopting VRL. From the *line length* perspective, opening up the station boundaries (Case B) has a significantly greater effect on the final length than does introducing variable takt times (Case C): In the former case, line length is reduced by approximately 12% on average. Of course, operators at open stations can start work earlier (than can those at closed stations), thanks to the adjacent station's drift area—thereby reducing the space required to fulfill their task. The combination of open stations and VRL has the greatest effect on line length: an average reduction of approximately 12%.

From the *cost* perspective, Case A and Case D can be seen as amounting to (respectively) an upper and lower bound on total inefficiency costs due to idle time, utility work, and line length—bounds that subsume the outcomes in Cases B and C. Unlike the results of our line length analysis, recommendations to open station boundaries or to introduce VRL are strongly dependent on the sequence.

Figure 7 plots the cost distribution of the sample scenario and its 90 sequences (the connecting lines are solely for the purpose of illustration). If the goal is to extend Case A, then the choice of opening stations versus establishing VRL clearly depends on the sequence—that is, since Case B yields a lower cost than does Case C in just over half (52%) of the sequences. This sequence dependency of production planning can be reduced significantly by the conditions presented in Section 5.

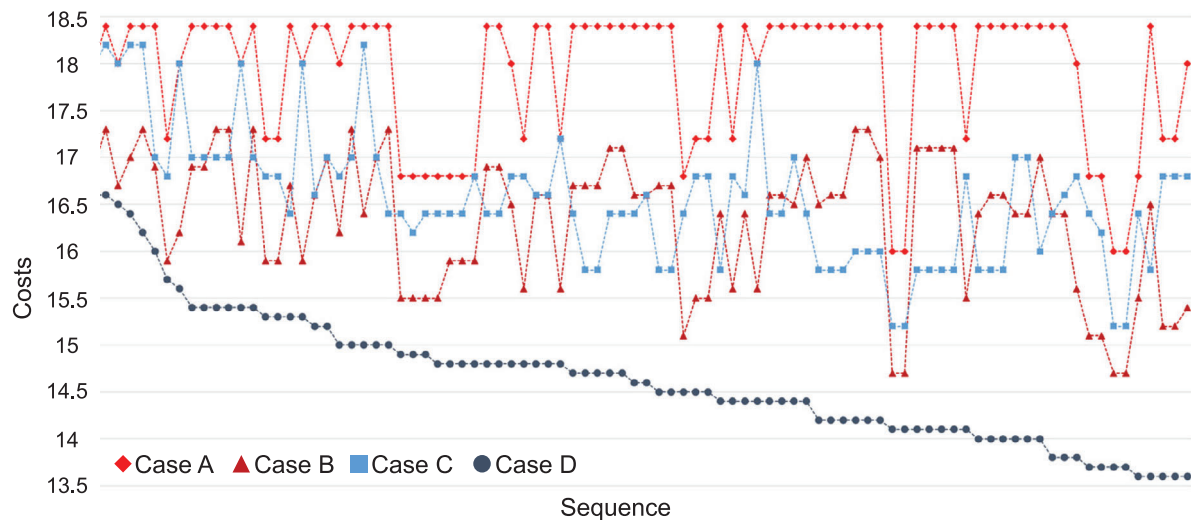


FIGURE 7 Total costs of every sequence in the sample scenario after implementing LLAOS-1 [Color figure can be viewed at wileyonlinelibrary.com]

7 | DISCUSSION AND PRACTICAL IMPLICATIONS

Our paper consists of three main parts. First, we introduce a mixed-integer programming model for minimizing labor inefficiencies and line length simultaneously on MMALs with variable takt times and open stations. Second, we discuss how companies can establish an ideal case, where labor inefficiencies are eliminated and operators are constantly utilized. Third, for manufacturing environments that are unable to align their operations with the ideal case, we present the LLAOS. In the rest of this section, we discuss the practical implications of all three elements and present managerial implications.

VRLMOS. Our general model extends the literature in that it is unrestricted with regard to launching rates and setup times as well as to variations in models, processing time, station length, and line length. This model applies to a large number of business cases, including those in which a leveled workload has proved to be elusive.

Ideal case. We present evidence that the general NP-hard problem of MMAL sequencing can be broken down into an “ideal” VRL approach that maximizes the benefits of a VRL strategy and overcomes most implementation barriers to employing variable takt times. Successful implementation requires that manufacturing executives focus on five key planning objectives. First, processing times for every model should be perfectly balanced over all stations so that a product model has the same processing time at each station. Second, station lengths should be set to the maximum processing time over all models. Third, operators should start work on each unit at the upstream station boundary. Fourth, the launching rate should be determined by the previous product’s processing time at the first station. Fifth, when parallel work is feasible, then workstations should overlap as much as possible.

If enough assembly space is available to set the station lengths equal to the maximum processing time, then fulfilling the first two conditions should not be difficult. Yet, achieving the third condition requires a highly flexible conveyor system, one that allows the spacing between units to vary; for that purpose, products can be placed on automated guided vehicles. The last condition requires considerable planning effort. For example, Fendt established a highly productive assembly design with an almost perfect line balance by assigning single tasks or entire tractor components to preassembly stations and utility workers (Huchzermeier et al., 2020b; Mönch et al., 2021).

Implementing variable takt time leads to five benefits. First, worker productivity rises significantly because operators can immediately start work on the next unit; hence the manufacturer’s workforce is constantly occupied. Second, a system with zero idle time and no utility work ensures the best possible throughput time. Third, eliminating sequence restrictions substantially reduces planning efforts. Fourth, the decrease in utility work naturally increases quality by lowering the stress level of operators; beneficial outcomes include lower error rates and less rework. Finally, dispensing with the restrictions due to sequence dependency greatly increases manufacturing flexibility to fulfill a broad range of different customer demands in real time. Our description and solution of the general case covers manufacturing environments that deviate from the ideal case in terms of one or more aspects.

Solution method for the general case. The general case applies for manufacturing companies at which station lengths vary and/or processing times cannot be perfectly balanced, circumstances that violate requirements of the ideal case. We are the first to formulate a mixed-model assembly optimization problem that accounts for both the cost of inefficiencies and line length. As discussed in the literature review, such combinatorial problems are solved efficiently by a variety of meta-heuristics. In our paper, we emphasize the ideal case

over the general case and thus do not focus on such standard solution methods. However, we suggest four novel and efficient search strategies—as defined above—for choosing the best station improvement for our MMAL problem. We tested these rules on a sample assembly line scenario, with three stations and three products. We develop a solution method for the general case—the LLAOS—that is an extension of Fattahi and Salehi's (2009) LIBP algorithm, which ignores open station boundaries and their effects on line length and labor costs. The LLAOS enables us to solve the unrestricted model by identifying the best sequence and its associated line length for open stations. We conduct a benchmark analysis to evaluate four station selection options for the LLAOS. This analysis involves comparing the minimum costs and line length of 2070 scenarios in terms of policies defined by the four cases that arise from the possible combinations of workstation type (closed or open) and launching rate (fixed or variable). To the best of our knowledge, this study is the first to compare the effects of all four cases on labor inefficiencies and line length. We find that the best outcome results from choosing, in every iteration of the LLAOS, the station predicted to yield the greatest reduction in both idle time and utility work. Another important finding is that, when the assembly strategy is changed from closed and fixed to open and variable, the line length and total costs are reduced by (respectively) 12% and 22% on average for the three-station scenario. Also, we performed a cost analysis for an assembly setup with 10 stations and compared the costs of Case A and Case D. (The processing times of the extension are based on the distribution presented in Table 3.) On average, switching from Case A to Case D leads to a cost reduction of approximately 9% over all scenarios. The range of cost savings over all scenarios lies between 5% and 16%. These results are in line with our observations made at Fendt: After switching to VRL, overall line length productivity increased by 9% and annual productivity savings rose to 6% (Bebersdorf, 2017). Moreover, we assess the optimality gap by performing a full enumeration over the possible line length, starting from the minimal station length. The results we obtain show that the optimality gap over 30 scenarios is a mere 2.3% with negligible run times. This states the effectiveness of the proposed algorithm. Our computational results indicate that companies should strive to implement variable launching rates. Clearly, the ideal case should be the “true north” for such assembly strategies. However, the more stations and models considered, the more difficult it will be to obtain optimal solutions if processing times are not balanced and space restrictions exist, certainly, there is no guarantee that idle time and utility work can be eliminated at all and most likely never will be. Even for small problem instances, for example, three stations, such an optimization proves difficult and does not come even close to the performance of the variable takt approach. The stark difference between an optimal result, that is, 100% utilization of frontline employees, with no need for any optimization—in the case of the variable takt—becomes really evident. This should be one of the key arguments for abandoning the fixed takt as it is unsuited for the optimization of an MMAL.

Finally, the LLAOS results in the ideal case when processing times are balanced over all stations for fixed launching rates (for variable launching rates, see the approach presented by Mönch et al. (2021)—the results show that VRL is superior to FRL for uniform processing times as well). This is the case in many standardized assembly lines, for example, automotive final assembly, where all stations have the same work content. If model processing times are uniform over all stations and space is not limited (no costs for increasing stations lengths), then the stations length would increase up to the maximum processing time (multiplied by the conveyor speed) and variable launching rates would equal the prior processing times, thus resembling the ideal case conditions from Theorem 1.

Managerial implications. Mass customization, highly innovative products, and localized production facilities have increased the pressure on static manufacturing designs, even those that previously operated at high levels of productivity. The prevailing management view wrongly supposes that all models should be handled using the same fixed takt time and that sequencing will, in itself, somehow accommodate variations in workload—beliefs that stem from a decades-long quest to eliminate the variability from manufacturing systems (Monden, 1983). Yet, the inflexibility of employing a single takt time renders sequence optimization, a computationally expensive procedure. We want to point out that in a real-world assembly context, additional factors play a role: Planners in multistage operations strive to achieve an average workload that matches the time it takes for a line carrier to pass through a workstation. If the workload is (much) larger, workers tend to drift off, that is, in the downstream direction. Moreover, if this is the case, workers drift in the upstream direction unless they pause at the start of their workstation (Mönch et al., 2022). Therefore, process times tend to “not vary that much,” at least they are bounded by a certain factor, which needs to be agreed by the workers council, worker representative, respectively (at Fendt, this factor is 15%). If process times are exceeding the time for passing of workstations, preassembly, suppliers, hybrid assembly (including matrix assembly), or flexible workers are scheduled. As one cannot ensure that workers use the entire space for assembly tasks, long process times have a high risk of significantly interfering with activities in the subsequent station and thus are often curtailed before the data are optimized. Conversely, operation tasks with little workload lead to idle time, which is also to be avoided (Bebersdorf & Huchzermeier, 2022). For those reasons, it is unreasonable, in our view, to simulate extreme realizations or variations in process times—as is typical for many operations research (OR) approaches. In our experiments, we took great care to assemble a meaningful data set for this specific context. Note that the variable takt solves the issue of accepting more variance in process times by leveling the work content of each station's operator even though the spread of that content widens. As each product is assigned its own takt, much larger variations can be accepted, and there is limited or no drifting off whatsoever. This can never be achieved, as outlined

above, with the fixed takt. It is not a matter of combinatorial optimization finesse; it is a matter of a fundamental change in approach.

For those who seek to move quickly from a fixed to a variable takt system, we suggest that product process times be rebalanced by distinguishing between tasks that must be performed directly on the production line and tasks that can be shifted to preassembly stations or lines located near the final assembly line (Huchzermeier et al., 2020b). In the medium term, it can be helpful to set up an interface between manufacturing and R&D; doing so will make it easier to integrate manufacturing objectives with the product development process. At Fendt, this is called Design for Takt (Bebersdorf & Huchzermeier, 2022).

Such integration, by enabling a perfect balance of workload across the assembly line for each product, can ease efforts to improve shop-floor planning and can also reduce assembly costs substantially. The typically time-intensive planning of the next sequence can thus be eliminated in systems without setup times, since all sequences yield the same efficiency. As for setup costs, an optimal batching of products in an MPS is both straightforward and independent of the assembly line status and layout. When combined with open stations, VRL not only significantly reduces operator idleness and work overload but also shortens the line length considerably.

We conclude that, with regard to future developments in assembly-intensive industries, the introduction of variable launching rates into MMALs is an efficient way to meet the challenges arising from mass customization and the need for more flexibility/responsiveness during turbulent times. Our results show that VRL outperforms FRL even when the conditions of the ideal case are violated. A major change in the manufacturing landscape is already evident in the form of new dynamic line configurations and layouts. The VRL strategy proposed here is a practical way to cope with the ever increasing demand for assembly line flexibility while keeping investment costs under control. As labor inefficiencies inevitably rise, we are certain that the VRL adoption rate will likewise increase. For example, during several visits at premium car manufacturers, we observed first experiments with variable takt times in subassembly stations, for example, in the assembly of two- and four-door cars. Employing variable takt times eliminates the major limitations of a fixed takt time system and yet retains all of the conventional assembly line's advantages.

REFERENCES

- Agrawal, M., Mercer, G. A., & Kumares, T. V. (2001). The false promise of mass customization. *The McKinsey Quarterly*, 01(3), 62–71.
- Al-Hawari, T., Ali, M., Al-Araidah, O., & Mumani, A. (2015). Development of a genetic algorithm for multi-objective assembly line balancing using multiple assignment approach. *The International Journal of Advanced Manufacturing Technology*, 77(5), 1419–1432.
- Bard, J. F., Dar-El, E., & Shtub, A. (1992). An analytic framework for sequencing mixed model assembly lines. *International Journal of Production Research*, 30(1), 35–48.
- Bard, J. F., Shtub, A., & Joshi, S. B. (1994). Sequencing mixed-model assembly lines to level parts usage and minimize line length. *International Journal of Production Research*, 32(10), 2431–2454.
- Battaïa, O., & Dolgui, A. (2013). A taxonomy of line balancing problems and their solution approaches. *International Journal of Production Economics*, 142(2), 259–277.
- Bautista, J., & Cano, J. (2008). Minimizing work overload in mixed-model assembly lines. *International Journal of Production Economics*, 112(1), 177–191.
- Baybars, İ. (1986). A survey of exact algorithms for the simple assembly line balancing problem. *Management Science*, 32(8), 909–932.
- Bebersdorf, P. (2017). Strategy development at AGCO Fendt. <http://de.industrial-excellence-award.eu/ie-konferenz/industrial-excellence-conference-2017/>
- Bebersdorf, P., & Huchzermeier, A. (2022). *Variable takt principle: Mastering variance with limitless product individualization*. Springer, 256 p.
- Becker, C., & Scholl, A. (2006). A survey on problems and methods in generalized assembly line balancing. *European Journal of Operational Research*, 168(3), 694–715.
- Betti, F., de Boer, E., & Giraud, Y. (2020). *Industry's fast-mover advantage: Enterprise value from digital factories*. World Economic Forum and McKinsey & Company.
- Bloomberg. (2019). *BMW takes on VW and Tesla with electric mini cooper*. <https://www.industryweek.com/leadership/bmw-takes-vw-and-tesla-electric-mini-cooper>
- Bock, S., Rosenberg, O., & van Brackel, T. (2006). Controlling mixed-model assembly lines in real-time by using distributed systems. *European Journal of Operational Research*, 168(3), 880–904.
- Bolat, A. (1997). Stochastic procedures for scheduling minimum job sets on mixed model assembly lines. *Journal of the Operational Research Society*, 48(5), 490–501.
- Bolat, A., Savsar, M., & Al-Fawzan, M. A. (1994). Algorithms for real-time scheduling of jobs on mixed model assembly lines. *Computers & Operations Research*, 21(5), 487–498.
- Bolat, A., & Yano, C. A. (1992a). Scheduling algorithms to minimize utility work at a single station on a paced assembly line. *Production Planning & Control*, 3(4), 393–405.
- Bolat, A., & Yano, C. A. (1992b). A surrogate objective for utility work in paced assembly lines. *Production Planning & Control*, 3(4), 406–412.
- Boysen, N., Flidner, M., & Scholl, A. (2007). A classification of assembly line balancing problems. *European Journal of Operational Research*, 183(2), 674–693. <https://doi.org/10.1016/j.ejor.2006.10.010>
- Boysen, N., Flidner, M., & Scholl, A. (2008). Assembly line balancing: Which model to use when? *International Journal of Production Economics*, 111(2), 509–528.
- Boysen, N., Flidner, M., & Scholl, A. (2009). Production planning of mixed-model assembly lines: Overview and extensions. *Production Planning & Control*, 20(5), 455–471. <https://doi.org/10.1080/09537280903011626>
- Burns, L. D., & Daganzo, C. F. (1987). Assembly line job sequencing principles. *International Journal of Production Research*, 25(1), 71–99.
- Buxey, G. M., Slack, N. D., & Wild, R. (1973). Production flow line system design—A review. *AIIE Transactions*, 5(1), 37–48.
- Cohen, M. A., Cui, S., Doetsch, S., Ernst, R., Huchzermeier, A., Kouvelis, P., Lee, H. L., Matsuo, H., & Tsay, A. (2022). Putting supply chain resilience theory into practise. *Management and Business Review*. Forthcoming.
- Cohen, M. A., Cui, S., Ernst, R., Huchzermeier, A., Kouvelis, P., Lee, H. L., Matsuo, H., Steuber, M., & Tsay, A. A. (2018). Om forum—Benchmarking global production sourcing decisions: Where and why firms offshore and reshore. *Manufacturing & Service Operations Management*, 20(3), 389–402.
- Dar-El, E. M. (1978). Mixed-model assembly line sequencing problems. *Omega*, 6(4), 313–323.
- Dar-El, E. M., & Cothier, R. F. (1975). Assembly line sequencing for model mix. *International Journal of Production Research*, 13(5), 463–477.
- Dar-El, E. M., & Cucuy, S. (1977). Optimal mixed-model sequencing for balanced assembly lines. *Omega*, 5(3), 333–342.

- Dar-El, E. M., & Nadivi, A. (2007). A mixed-model sequencing application. *International Journal of Production Research*, 19(1), 69–84.
- Falkenauer, E. (2005). Line balancing in the real world. *Proceedings of the International Conference on Product Lifecycle Management PLM*, (5), 360–370.
- Fattahi, P., & Salehi, M. (2009). Sequencing the mixed-model assembly line to minimize the total utility and idle costs with variable launching interval. *The International Journal of Advanced Manufacturing Technology*, 45(9–10), 987–998.
- Goldschmidt, O., Bard, J. F., & Takvorian, A. (1997). Complexity results for mixed-model assembly lines with approximation algorithms for the single station case. *International Journal of Flexible Manufacturing Systems*, 9(3), 251–272.
- Huchzermeier, A., Mönch, T., & Bebersdorf, P. (2020a). The Fendt variotakt: Revolutionizing mixed-model assembly line production: Case article. *INFORMS Transactions on Education*, 20(3), 134–140.
- Huchzermeier, A., Mönch, T., & Bebersdorf, P. (2020b). The Fendt variotakt: Revolutionizing mixed-model assembly line production: Case study. *INFORMS Transactions on Education*, 20(3), 141–153.
- Hyun, C. J., Kim, Y., & Kim, Y. K. (1998). A genetic algorithm for multiple objective sequencing problems in mixed model assembly lines. *Computers & Operations Research*, 25(7–8), 675–690.
- Kilbridge, M., & Wester, L. (1962). A review of analytical systems of line balancing. *Operations Research*, 10(5), 626–638.
- Kim, Y. K., Hyun, C. J., & Kim, Y. (1996). Sequencing in mixed model assembly lines: A genetic algorithm approach. *Computers & Operations Research*, 23(12), 1131–1145.
- Knowledge@Wharton. (2019). *How the U.S. can regain its manufacturing edge*. <https://knowledge.wharton.upenn.edu/article/future-of-us-manufacturing/>
- Mönch, T., Huchzermeier, A., & Bebersdorf, P. (2021). Variable takt times in mixed-model assembly line balancing with random customization. *International Journal of Production Research*, 59(15), 4670–4689.
- Mönch, T., Huchzermeier, A., & Bebersdorf, P. (2022). Variable takt time groups and workload equilibrium. *International Journal of Production Research*, 60(5), 1535–1552.
- Monden, Y. (1983). *Toyota production system: A practical approach to production management*. Industrial Engineering and Management.
- Monden, Y. (2011). *Toyota production system: An integrated approach to just-in-time* (4th ed.). Productivity Press. <https://doi.org/10.1201/b11731>
- Okamura, K., & Yamashina, H. (1979). A heuristic algorithm for the assembly line model-mix sequencing problem to minimize the risk of stopping the conveyor. *International Journal of Production Research*, 17(3), 233–247.
- Sarker, B. R., & Pan, H. (1998). Designing a mixed-model assembly line to minimize the costs of idle and utility times. *Computers & Industrial Engineering*, 34(3), 609–628.
- Sarker, B. R., & Pan, H. (2001). Design configuration for a closed-station, mixed-model assembly line: A filing cabinet manufacturing system. *International Journal of Production Research*, 39(10), 2251–2270.
- Scholl, A., Klein, R., & Domschke, W. (1998). Pattern based vocabulary building for effectively sequencing mixed-model assembly lines. *Journal of Heuristics*, 4(4), 359–381.
- Sun, X., Noble, J. S., & Klein, C. M. (1999). Single-machine scheduling with sequence dependent setup to minimize total weighted squared tardiness. *IIE Transactions*, 31(2), 113–124.
- Thomopoulos, N. T. (1967). Line balancing-sequencing for mixed-model assembly. *Management Science*, 14(2), B–59–B–75.
- Tong, K.-n., Xu, K.-l., & Zheng, Y.-q. (2013). Sequencing mixed-model flexible assembly lines with variable launching intervals. *Journal of Shanghai Jiaotong University (Science)*, 18(4), 460–467.
- Tsai, L.-H. (1995). Mixed-model sequencing to minimize utility work and the risk of conveyor stoppage. *Management Science*, 41(3), 485–495.
- Vallada, E., & Ruiz, R. (2011). A genetic algorithm for the unrelated parallel machine scheduling problem with sequence dependent setup times. *European Journal of Operational Research*, 211(3), 612–622.
- Wang, Y., Ma, H.-S., Yang, J.-H., & Wang, K.-S. (2017). Industry 4.0: A way from mass customization to mass personalization production. *Advances in Manufacturing*, 5(4), 311–320.
- Xiaobo, Z., & Ohno, K. (1994). A sequencing problem for a mixed-model assembly line in a JIT production system. *Computers & Industrial Engineering*, 27(1–4), 71–74.
- Xiaobo, Z., & Ohno, K. (1997). Algorithms for sequencing mixed models on an assembly line in a JIT production system. *Computers & Industrial Engineering*, 32(1), 47–56.
- Xiaobo, Z., & Ohno, K. (2000). Properties of a sequencing problem for a mixed model assembly line with conveyor stoppages. *European Journal of Operational Research*, 124(3), 560–570.
- Yano, C. A., & Rachamadugu, R. (1991). Sequencing to minimize work overload in assembly lines with product options. *Management Science*, 37(5), 572–586.
- Yoshida, T., & Hitomi, K. (1997). Optimal two-stage production scheduling with setup times separated. *AIIE Transactions*, 11(3), 261–263.

How to cite this article: Huchzermeier, A., & Mönch, T. (2023). Mixed-model assembly lines with variable takt and open stations. *Production and Operations Management*, 32, 704–722. <https://doi.org/10.1111/poms.13893>