

INTEGRATED MILP-BASED OPTIMIZATION AND FLEXIBILITY ASSESSMENT OF A DIGITALLY SUPPORTED POSTPONEMENT LOGISTICS SYSTEM

TAMÁS JÁRVÁS¹–BÉLA ILLÉS²–ÁGOTA BÁNYAI³

Abstract: Manufacturing companies are increasingly required to cope with a wider product mix, changing production volumes, and shorter response times. These challenges are particularly relevant in postponement environments, where final product differentiation takes place close to the actual customer demand and therefore places considerable pressure on internal logistics. This paper presents a mixed-integer linear programming (MILP) model developed for the operational optimization of a digitally supported postponement logistics system. The model links production-order assignment with the execution of the related intralogistics tasks and considers technological feasibility, milk-run operations, task sequencing, resource availability, timing requirements, service levels, and line-side storage limitations. The optimization is carried out in two stages. First, the maximum production volume that can be served with the available resources is determined. In the second stage, the logistics performance is improved while keeping the achievable volume close to this optimum. Besides supporting operational decisions, the model is also used to evaluate logistics flexibility. For this purpose, five indicators are defined, covering on-time task completion, logistics response time, mix flexibility, volume flexibility, and the stability of logistics performance under different operating conditions. The proposed approach therefore provides a common framework for analysing operational performance and flexibility and can support the comparison of alternative resource configurations and production scenarios in postponement logistics systems.

Keywords: postponement logistics; intralogistics; mixed-integer linear programming; MILP; logistics flexibility; milk-run; operational optimization; digital logistics.

1. INTRODUCTION

Manufacturing companies are increasingly expected to operate efficiently under conditions of high product variety, fluctuating order volumes, shorter product life cycles, and rapidly changing customer requirements. Cost efficiency and high service levels remain important, but they are no longer sufficient on their own. Production and logistics systems must also be able to adapt to changing operating conditions without excessive losses in time, cost, or performance. Flexibility has therefore become an important capability of modern manufacturing systems [1, 2].

Postponement is one of the supply chain strategies used to cope with demand uncertainty and product variety. Its basic principle is to delay selected activities related to final product differentiation (such as final assembly, configuration, labelling, or packaging) until more accurate information about actual customer demand becomes available [3-5]. This can reduce the inventory risk associated with finished-product variants and improve responsiveness to changing demand. At the same time, postponement does not remove operational complexity. Rather, part of that complexity is shifted towards the later stages of production and logistics,

¹ PhD student, University of Miskolc, Hungary
tamas.jarvas@hu.bosch.com

² professor, University of Miskolc, Hungary
bela.illes@uni-miskolc.hu

³ professor, University of Miskolc, Hungary
agota.banyaine@uni-miskolc.hu

where a wider product mix and changing order volumes must be handled within relatively short time windows.

This creates a particular challenge for intralogistics. In a high-mix, low-volume environment, production orders may be processed on several alternative production or packaging lines. The assignment of an order to a particular line therefore influences not only production capacity but also the location, timing, and volume of the corresponding material flows. Limited line-side storage further constrains material supply, since larger production orders may require several replenishment cycles instead of a single delivery. Deliveries that arrive too early occupy scarce line-side space, whereas late deliveries may interrupt production. Production assignment, material supply, routing, and resource allocation are consequently interdependent operational decisions.

Digital technologies provide the information infrastructure needed to manage these interdependencies more effectively. Digital identification, event-based data collection, integrated production and logistics data, and real-time or near-real-time status information make it possible to replace part of the static and experience-based coordination with data-driven decision support [6-8]. Digitalization, however, does not by itself determine how limited resources should be allocated or how competing logistics tasks should be prioritized. Once reliable operational data are available, the next question is how these data can be translated into coordinated decisions under technological, temporal, capacity, and service-level constraints.

The present research addresses this problem in the context of an operating, digitally supported postponement logistics system. Production planning, execution status, prepared materials, internal transport tasks, and milk-run operations are connected through a common information environment. The physical execution remains subject to finite production, material-handling, and line-side storage capacities. This combination of structured operational data and physical constraints provides the basis for a mathematical representation of the system.

The central research question is therefore formulated as follows:

How can production-order assignment and intralogistics decisions in a digitally supported postponement system be represented and optimized within a unified mathematical framework, and how can the logistics flexibility of the system be quantified based on the results of the model?

To address this question, a mixed-integer linear programming (MILP) model is developed. It integrates the assignment of production orders to technologically feasible lines with the scheduling, allocation, routing, and sequencing of the resulting intralogistics tasks. The model also considers material-handling resources, service-level requirements, production capacity, and line-side storage constraints.

The contribution of the study is threefold. First, production-order assignment and internal material supply are treated as interconnected decisions rather than independent optimization problems. Second, a two-stage optimization approach is introduced in which the achievable production volume is determined before the operational logistics performance is optimized without materially reducing that volume. Third, the model results are used to derive quantitative indicators for logistics flexibility, covering response time, delivery performance, mix flexibility, volume flexibility, and performance uniformity under different operating conditions.

The remainder of the paper first reviews the relevant literature on postponement, manufacturing and logistics flexibility, digital intralogistics, and integrated material-supply optimization. The mathematical formulation of the proposed model is then presented, followed by the definition of the flexibility indicators derived from the model.

2. LITERATURE REVIEW

This section reviews the literature relevant to the proposed research framework. Four main areas are considered: postponement and delayed product differentiation, manufacturing and logistics flexibility, digitalization and data-driven intralogistics, and milk-run-based material-supply optimization. The review provides the theoretical and methodological background for identifying the research gap and for developing the integrated MILP model presented in the following section.

2.1. Postponement and delayed product differentiation

The concept of postponement has a long history in logistics and operations management. Its underlying logic is to retain a product or process in a more generic state for as long as economically reasonable and delay differentiation until demand information becomes more reliable. Zinn and Bowersox examined postponement in physical distribution and distinguished several forms of delayed processing and distribution, including packaging, labelling, assembly, manufacturing, and shipment postponement. Their work demonstrated that the suitability of postponement depends on both product characteristics and demand conditions [3].

The concept became increasingly important with the growth of product variety and customer-specific configuration. Lee and Tang analysed delayed product differentiation from a product and process design perspective and showed how standardization, modularity, and process restructuring can move the differentiation point downstream [9]. The main advantage is that uncertainty can be pooled for a longer part of the production process, reducing the need to commit inventory to individual product variants before demand is sufficiently known.

Pagh and Cooper extended this perspective by positioning postponement and speculation as alternative supply chain strategies [10]. Their framework emphasized that postponement is not universally preferable; the appropriate configuration depends on factors such as demand uncertainty, product characteristics, logistics cost, and customer requirements. This is important from an operational perspective because postponement usually changes both the location and timing of logistics activities rather than simply eliminating them.

Van Hoek, in a broad review of the postponement literature, similarly argued for a more integrated supply chain perspective [4]. Postponement should not be understood solely as a product-design or inventory-management technique, since moving differentiation downstream changes the relationships between production, logistics, and customer-order information. Yang et al. further developed this argument by reviewing the enablers and constraints of postponement and proposing an integrated framework for its implementation [5].

Taken together, these studies establish postponement primarily as a response to demand uncertainty and product variety. They also point to an important operational consequence: delaying final differentiation tends to concentrate decision-making and material-handling requirements closer to the customer-order point. The later the differentiation takes place, the

more important it becomes that the supporting logistics system can react to changes in product mix, sequence, and volume without causing excessive inventory or service deterioration.

For the type of postponement system considered in this study, the operational implication is particularly relevant. Final packaging can involve a large number of product variants and substantially different order sizes. When products can be assigned to alternative packaging lines, postponement flexibility depends not only on the delayed differentiation itself but also on the ability of internal logistics to supply whichever line is selected at the required time. This shifts the problem from postponement as a strategic concept towards the operational coordination of flexible production and material supply.

2.2. Manufacturing and logistics flexibility

Flexibility has been studied extensively in operations management, but the literature has never converged on a single measure or dimension. Slack was among the early authors to identify flexibility as a manufacturing objective alongside more traditional performance criteria such as cost and productivity [1]. Sethi and Sethi later provided a broad survey of manufacturing flexibility and distinguished a range of flexibility types related to machines, products, routing, volume, expansion, operations, and production processes [11].

Upton proposed a more general view, defining flexibility through the ability to change or react with limited penalties in time, effort, cost, or performance [2]. This interpretation is useful because it shifts attention from the mere existence of alternatives towards the cost and speed of using them. Flexibility is therefore not simply a question of how many alternatives a system possesses, but also how efficiently the system can move between operating states.

Koste and Malhotra developed this idea into a more structured theoretical framework [12]. They identified four constituent elements of flexibility: range-number, range-heterogeneity, mobility, and uniformity. Range describes the number and diversity of operating states that can be accommodated. Mobility reflects the effort and transition penalties involved in moving from one state to another, while uniformity concerns the extent to which performance remains stable across the operating range. Their framework provided a basis for defining and comparing different forms of manufacturing flexibility in a consistent manner.

Later empirical work by Koste et al. developed measures for several manufacturing-flexibility dimensions, including machine, labour, material-handling, mix, new-product, and modification flexibility [13]. This work is especially relevant to the present study because it confirms that material handling and product mix are not peripheral aspects of manufacturing flexibility but distinct dimensions that can influence overall system adaptability.

The broader literature has continued to emphasize the multidimensional nature of flexibility. Beach et al. concluded from their review that flexibility has strategic importance but remains difficult to characterize using a single measure [14]. More recent research has expanded the concept beyond manufacturing itself. Jafari, in a systematic review of logistics flexibility, found substantial variation in both terminology and measurement and highlighted the need to distinguish logistics flexibility as a construct in its own right [15]. Manders et al. reached a similar conclusion for supply chain flexibility: research often focuses on individual dimensions rather than on a consistently defined overall construct [16].

The distinction is important for intralogistics. A manufacturing system may have high mix flexibility because several products can be produced on alternative lines, while its actual

ability to exploit this flexibility may still be limited by material supply. Similarly, additional production capacity does not automatically create volume flexibility if internal logistics cannot deliver the corresponding material quantities within the required time windows.

Chejarla et al. reinforce this point in their review of logistics flexibility [17]. They distinguish several logistics domains, including flexibility in in-house logistics, and show that logistics flexibility must be considered at different levels of the logistics system. This supports the view that the flexibility of internal material supply should be evaluated using indicators that reflect its own operating characteristics rather than inferred solely from production capacity.

For the mathematical model developed in this study, three elements of the flexibility literature are particularly useful: range, mobility, and uniformity. Range provides the conceptual basis for evaluating the product mix and production volume that can be handled. Mobility concerns the speed with which logistics responds to the requirements generated by production. Uniformity reflects whether an acceptable logistics performance can be maintained when operating conditions change. These dimensions provide the theoretical basis for the flexibility indicators derived later from the optimization model.

2.3. Digitalization and data-driven intralogistics

The emergence of Industry 4.0 has substantially changed the information environment in which production and logistics decisions are made. In traditional intralogistics, material movements and execution states are often only partially visible between individual transaction points. Industry 4.0 technologies can create a more continuous link between the physical flow of materials and its digital representation.

Winkelhaus and Grosse describe Logistics 4.0 as a developing logistics paradigm shaped by increasing digitalization, connectivity, decentralization, and real-time capabilities [6]. Their systematic review shows that the transformation is not based on one individual technology, but on the interaction of data acquisition, communication, processing, and decision-making capabilities.

Efthymiou and Ponis reached a similar conclusion in their systematic review of Industry 4.0 technologies in logistics [7]. Technologies such as IoT, data analytics, cyber-physical systems, cloud platforms, and automated identification can improve visibility and support more responsive decision-making. Their relevance to intralogistics lies particularly in the ability to capture execution data closer to the physical process.

More recent literature has moved beyond technology lists and has examined what digital technologies actually enable within logistics processes. Albrecht et al., based on a systematic literature review and expert interviews, identified a set of digital technology affordances specifically for intralogistics [8]. Their findings emphasize that digital technologies support capabilities such as process visibility, information integration, automated data processing, and more adaptive planning and execution. This process-oriented view is particularly useful because it distinguishes the existence of digital technology from the operational capabilities created through its use.

For the present research, this distinction is central. Digitalization provides the information needed for decision-making, but it does not automatically determine the decision itself. Knowing the location and status of a material unit, the expected start of a production order, or the availability of a milk-run resource creates decision-making potential. Determining

which task should be performed first, which production line should process an order, or how available transport capacity should be allocated still requires an explicit decision logic.

This creates a natural link between digital intralogistics and operations research. Once the physical process is represented through structured and sufficiently timely data, optimization models can be repeatedly updated as the operating state changes. A rolling-horizon formulation is therefore particularly suitable for digitally supported environments because the optimization problem does not need to represent an entire shift as a single static planning problem. Instead, the current system state can be incorporated whenever new planning or execution information becomes available.

2.4. Milk-run systems and intralogistics optimization

Milk-run systems are widely used to supply manufacturing areas because they allow several material-delivery and collection tasks to be combined within common transport cycles. Their effectiveness, however, depends on the relationship between transport capacity, delivery frequency, route structure, production demand, and line-side inventory.

Traditional milk-run concepts often rely on relatively stable routes and fixed delivery frequencies. Such an approach is well suited to repetitive production with predictable consumption but becomes more difficult to maintain when order sequences, product mix, or consumption rates change frequently. In these environments, the material-handling system must respond to time-dependent production requirements rather than simply repeat a predetermined route.

Industry 4.0 technologies have enabled a transition towards more dynamic milk-run concepts. Facchini et al. formulated the in-plant milk-run problem as a Vehicle Routing Problem with Time Windows (VRPTW) [18]. Their model dynamically assigns material deliveries to tugger-train routes while respecting the required delivery time windows and minimizing travel distance. This demonstrates the value of introducing explicit time requirements into intralogistics routing rather than optimizing distance alone.

Akkad and Bányai similarly examined Industry 4.0-supported milk-run systems and linked digital technologies with real-time scheduling, routing, and resource planning [19]. Their mathematical model considers the effects of Industry 4.0 capabilities on the performance and flexibility of in-plant supply. This is particularly relevant because it moves beyond route efficiency and recognises flexibility as an important characteristic of cyber-physical milk-run systems.

Facchini et al. took the integration one step further by combining an IoT-supported milk-run system with the time-dependent material requirements of manufacturing lines [20]. Their model plans dynamic tugger-train routes while ensuring that line-side material requirements are satisfied within the corresponding time slots. The study illustrates how real-time or near-real-time production information can be translated into dynamically changing logistics decisions.

These studies demonstrate considerable progress in the optimization of internal material supply. Nevertheless, most milk-run models take production demand, delivery requirements, or workstation assignments as inputs to the routing problem. In practice, however, these inputs may themselves depend on upstream operational decisions. When a production order can be processed on several alternative lines, its line assignment determines where the corresponding material must be delivered and consequently affects routing, resource utilization, and line-side inventory.

This interdependence connects the problem to a broader stream of integrated production and transportation research. Hosseini et al. reviewed more than 140 studies dealing with manufacturing scheduling problems that explicitly include transportation [21]. Their review shows a substantial body of research on the interaction between production scheduling and transport resources, particularly in flexible manufacturing systems and assembly environments. It also demonstrates that integrated production–transport problems can differ considerably depending on transporter characteristics, buffer structures, material flows, and the sequence relationships between production and transportation.

The present problem shares this need for integration but has a somewhat different focus. The transport tasks are not simply movements of workpieces between consecutive machines. They include time-window-constrained material supply and return flows generated by production orders in a postponement environment. The production decision determines the location and timing of the logistics demand, while the logistics system must operate within its own route, resource, service-level, and storage constraints.

2.5. Synthesis and research gap

The reviewed literature provides strong foundations for each individual component of the problem. Postponement research explains why delayed differentiation can improve responsiveness to uncertain and diverse customer demand and also why it shifts operational complexity towards later stages of the supply chain [4, 5, 9]. Manufacturing- and logistics-flexibility research provides multidimensional frameworks for understanding the range of operating conditions a system can accommodate, the effort required to respond to change, and the stability of performance across those conditions [2, 12, 15, 17].

Research on Logistics 4.0 explains how digital identification, integrated information flows, and event-based operational data create the information basis for more responsive intralogistics [6-8]. At the operational level, milk-run and integrated scheduling studies demonstrate the potential of mathematical optimization for coordinating material-handling resources with time-dependent manufacturing requirements [18-21].

The literature is therefore substantial, but the individual research streams address different parts of the decision problem. Existing postponement studies predominantly analyse supply chain design and delayed differentiation; flexibility studies primarily develop conceptual dimensions and measures; digital-logistics research focuses on technological capabilities and information integration; and milk-run optimization typically considers routing and scheduling once the material requirements and service locations are already known.

This leaves room for an operational modelling approach in which these elements are more closely connected. In a flexible postponement environment, the production-line assignment is itself a decision and determines the location of the resulting logistics demand. At the same time, the feasibility and quality of that assignment depend on the ability of internal logistics to execute the generated tasks within time, resource, routing, and line-side storage constraints. Moreover, evaluating such a system solely on travel distance or resource utilization would not capture its ability to accommodate changes in volume, product mix, or operating conditions.

The present study addresses this gap by developing an integrated MILP framework in which production-order assignment and the resulting intralogistics execution are optimized jointly, and by using the same mathematical framework to derive quantitative measures of logistics flexibility. Rather than treating flexibility as an external qualitative characteristic,

the proposed approach links it to observable model outcomes such as achievable volume, alternative production possibilities, logistics response time, on-time task completion, and performance deterioration under changing operating conditions.

This combination provides the conceptual bridge between postponement, digital intralogistics, operations research, and flexibility measurement on which the mathematical model presented in the following section is based.

3. MATHEMATICAL MODEL FOR THE OPERATIONAL OPTIMIZATION OF THE POSTPONEMENT LOGISTICS SYSTEM AND FLEXIBILITY INDICATORS

A mixed-integer linear programming (MILP) model was developed to provide a mathematical description of the operational intralogistics subprocesses related to production within the digitally supported postponement logistics system. The purpose of the model is to provide a unified mathematical formalization of the system's operational decisions and to optimize the production-related internal logistics processes while taking into account the available resources, technological constraints, and temporal requirements.

3.1. General Mathematical Model of the Postponement Logistics System under Study

The mathematical model represents those operational decisions of the digitally supported postponement logistics system that directly influence the efficiency of internal material flow, material supply, and production support. Accordingly, the model integrates the assignment of production orders to production and/or packaging lines, the scheduling of intralogistics tasks generated by the production plan, milk-run-based internal material supply, and the associated resource and capacity constraints. The planned starting times of production orders are input parameters of the model, while the mathematical model uses these values to optimize order-to-line assignment and the execution of the resulting intralogistics tasks.

The problem can be decomposed into two interconnected decision levels:

- Assignment of production orders to production and/or packaging lines, taking the planned starting times into account.
- Assignment and sequencing of the resulting material-delivery and return-transport tasks within milk-run cycles.

The two decision levels can be operated on a rolling horizon. After each new production plan, status change, or execution event, the input data can be updated and a new optimization problem can be solved for the currently open tasks. Depending on the specific application, the input data may be provided by an external production-planning or scheduling system. This approach is better suited to a continuously changing operating environment than a static optimization model executed only once for an entire shift. In the remainder of this section, production and packaging lines are jointly referred to as production lines because they perform the same function from the perspective of the mathematical model.

The sets, parameters, decision variables, objective functions, and constraints form a unified mathematical system that provides a consistent representation of the operational behavior of the digitally supported postponement logistics system. The model is presented below through the definition of its sets, parameters, decision variables, objective functions, and constraints (see Table I - Table III).

Table I.
Definition of sets

Symbol	Definition
R	set of production (packaging) orders
L	set of production lines
T	set of discrete planning time intervals, where $t \in T$ denotes the starting time of the corresponding interval
F	set of open and executable material-delivery and return-transport logistics tasks within the decision horizon under consideration
F^+	set of outbound material-delivery logistics tasks
F^-	set of return-transport logistics tasks
F_0	$F_0 = F \cup \{0\}$, the set of logistics tasks augmented by the depot, denoted by 0, which serves as both the starting and return point
K	set of milk-run cycles that can be initiated within the given operational decision window

Table II.
Main input parameters

Symbol	Definition
q_r	quantity of the r -th order
A_{rl}	1 if the r -th order can be processed on the l -th production line; 0 otherwise
p_{rl}	processing time of the r -th order on the l -th production line
d_r	due date of the r -th order
w_f	priority weight of the f -th logistics task
c_{fg}	time required to execute the f -th and g -th logistics tasks consecutively, $c_{fg} > 0, \forall f, g \in F, f \neq g$
c_{0f}	time required from the depot denoted by 0 to the completion of the f -th logistics task, $c_{0f} > 0, \forall f \in F$.
d_{fg}	distance between the service locations of the f -th and g -th logistics tasks, where $f, g \in F_0$, and $F_0 = F \cup \{0\}$
C_l	line-side storage capacity available at the l -th production line
H	maximum number of material-delivery resources simultaneously available within the given operational decision window, each of which can be used to execute one milk-run cycle
M	sufficiently large positive constant used to linearize logical conditions
ε_z	small positive time value used to distinguish tardiness, $\varepsilon_z > 0$.
$\eta_{\min}$	minimum required proportion of logistics tasks completed on time
$E_{\max}$	maximum allowable average response time of logistics tasks
θ_f	task-generation lead time associated with the f -th logistics task
δ_f	safety time buffer associated with the f -th logistics task
t_r^0	planned starting time of the r -th order
N^+	maximum number of outbound material-delivery tasks that can be executed within one milk-run cycle
N^-	maximum number of return-transport logistics tasks that can be executed within one milk-run cycle
D_{lt}	quantity of material delivered to the l -th production line during the t -th time interval.
U_{lt}	quantity of material consumed on production line l during time interval t
B_{l0}	initial line-side inventory level at production line l

Table III.
Derived time parameters

Symbol	Definition
e_f	release time of the f -th logistics task
b_f	latest completion time of the f -th logistics task

The time parameters of the logistics tasks can be calculated from the planned starting times of the associated orders using the following relationships:

$$e_f = t_{r(f)}^0 - \theta_f, b_f = t_{r(f)}^0 - \delta_f, \quad (1)$$

where $r(f)$ denotes the order associated with the f -th logistics task.

The production line associated with the f -th logistics task is denoted by the mapping $l(f)$. Accordingly, the service location of each logistics task is linked to a potential production line of the associated order, so that the distance d_{fg} and execution time c_{fg} between logistics tasks can be predetermined.

An important condition is that

$$\theta_f \geq \delta_f \geq 0, \quad (2)$$

because it follows that

$$e_f \leq b_f \leq t_{r(f)}^0. \quad (3)$$

This ensures that the release time of the task is not later than its latest allowable completion time. The decision variables are shown in Table IV.

Table IV.
Decision variables

Symbol	Definition
y_{rlt}	1 if the r -th order is assigned to the l -th production line and its processing starts in the t -th time interval corresponding to the planned starting time; 0 otherwise
v_r	1 if the r -th order is selected for execution within the rolling horizon under consideration; 0 otherwise.
h_f	1 if the f -th logistics task is activated based on the actual production-line assignment of the associated order; 0 otherwise
g_{fk}	1 if the f -th logistics task is executed by the k -th milk-run cycle; 0 otherwise
x_{fgk}	1 if, within the k -th milk-run cycle, the service location associated with logistics task f is immediately followed by the service location associated with logistics task g ; 0 otherwise, where $g \in F_0$, $f \neq g$, and the 0 index denotes the depot serving as the starting and return point.
a_k	1 if the k -th milk-run cycle is activated; 0 otherwise
τ_f	completion time of the f -th logistics task as determined by the model (nonnegative continuous variable)
E_f	response time between the release and completion of the f -th logistics task (nonnegative continuous variable)
L_f	tardiness of the f -th logistics task (nonnegative continuous variable)
z_f	1 if the f -th logistics task is completed after its latest completion time; 0 otherwise
$U_{\max}$	maximum capacity-utilization ratio among the production lines
B_{lt}	inventory level available at the l -th production line at the end of the t -th time interval

The domains of the decision variables are:

$$\begin{aligned} y_{rlt}, v_r, h_f, g_{fk}, x_{fgk}, a_k, z_f \in \{0,1\}, \\ \tau_f, E_f, L_f, B_{lt} \geq 0. \end{aligned} \quad (4)$$

3.1.1. Objective Functions

The model uses a two-stage, ε -constraint optimization approach. In the first stage, the serviceable order volume is maximized subject to the available production and logistics resources and the required service level. In the second stage, operational logistics performance is optimized while keeping the optimal volume determined in the first stage nearly unchanged. This approach ensures that reductions in tardiness, response time, transport distance, and resource utilization are not achieved through an unjustified reduction in the serviceable volume.

First objective function: maximization of the serviceable volume

$$Q(H) = \sum_{r \in R} q_r v_r \rightarrow \max. \quad (5)$$

The first objective function maximizes the serviceable order volume subject to the production-line, line-side storage, milk-run, and material-delivery resource constraints and the required service level.

Let the result of the first optimization stage be: $Q_{\max}(H)$. The value $Q_{\max}(H)$ obtained in the first optimization stage represents the maximum order volume that can be served at the given resource level H and also provides the basis for determining the volume-flexibility indicator.

Second objective function: optimization of operational logistics performance

In the second stage:

$$\sum_{r \in R} q_r v_r \geq (1 - \varepsilon) Q_{\max}(H), \quad (6)$$

where $0 \leq \varepsilon < 1$. Under this condition, the operational objective function can be minimized:

$$\begin{aligned} F_{op} = & \alpha_1 \frac{\sum_{f \in F} w_f L_f}{L_{\text{ref}}} + \alpha_2 \frac{\sum_{f \in F} E_f}{E_{\text{ref}}} + \alpha_3 \frac{\sum_{k \in K} \sum_{f \in F_0} \sum_{g \in F_0} d_{fg} x_{fgk}}{D_{\text{ref}}} + \\ & + \alpha_4 \frac{\sum_{k \in K} a_k}{|K|} + \alpha_5 U_{\max} \rightarrow \min. \end{aligned} \quad (7)$$

The first term of the operational objective function (F_{op}) minimizes the sum of priority-weighted tardiness values, the second term minimizes the response times of the logistics tasks, the third term minimizes the total route length of the milk-run cycles, the fourth term minimizes the number of activated milk-run cycles, and the fifth term minimizes the maximum capacity utilization of the production lines.

The weights $\alpha_1, \dots, \alpha_5$ express the relative importance of the objective components and satisfy the following conditions:

$$\sum_{m=1}^5 \alpha_m = 1, \quad \alpha_1 > 0, \quad \alpha_m \geq 0, \quad m = 2, \dots, 5. \quad (8)$$

The weights may be determined on the basis of managerial preferences, expert assessment, or sensitivity analysis.

Table V.
Notation used in the objective function

Symbol	Definition
ε	allowable relative deviation from the maximum volume determined in the first stage
L_{ref}	positive reference value used to normalize tardiness, $L_{\text{ref}} > 0$
E_{ref}	positive reference value used to normalize response time, $E_{\text{ref}} > 0$
D_{ref}	positive reference value used to normalize total route length, $D_{\text{ref}} > 0$
α_m	weight of the m -th operational objective component, $m = 1, \dots, 5$

On-time material supply is a primary operational requirement. This is ensured both by the service-level constraints applied in the first optimization stage and by the priority-weighted tardiness component of the operational objective function.

3.1.2. Model Constraints

Assignment of Orders to Production Lines

Each order selected for completion within the rolling horizon under consideration must be assigned to exactly one technologically feasible production line. Processing of the order starts at the planned starting time specified as an input parameter.

$$\forall r \in R: \sum_{l \in L} \sum_{t \in T} y_{rlt} = v_r. \quad (9)$$

Because the planned starting time is an input parameter, the order cannot be started in another time interval:

$$\forall r \in R, l \in L, t \in T: y_{rlt} = 0, t \neq t_r^0. \quad (10)$$

An order may only be assigned to a technologically feasible line:

$$\begin{aligned} \forall r \in R, l \in L, t \in T: y_{rlt} &\leq A_{rl}, \\ \forall f \in F: h_f &= \sum_{t \in T} y_{r(f), l(f), t}. \end{aligned} \quad (11)$$

This constraint ensures that a logistics task is activated if and only if the associated order is assigned to the production line corresponding to that task.

Production-Line Capacity

A production line can process at most one active order in any time interval:

$$\forall l \in L, t \in T: \sum_{r \in R} \sum_{\substack{\tau \in T: \\ \tau \leq t < \tau + p_{rl}}} y_{rl\tau} \leq 1. \quad (12)$$

The planned completion time of an order selected for execution must not exceed its due date:

$$\forall r \in R: \sum_{l \in L} \sum_{t \in T} (t + p_{rl}) y_{rlt} \leq d_r + M(1 - v_r). \quad (13)$$

Capacity utilization of the production lines:

$$\forall l \in L: \frac{\sum_{r \in R} \sum_{t \in T} p_{rl} y_{rlt}}{|T|} \leq U_{\max}, \quad (14)$$

$$0 \leq U_{\max} \leq 1.$$

Determination of the Time Parameters of Logistics Tasks

The intralogistics tasks associated with production orders are determined from the planned production starting times. In the model, the planned starting time of each order is an input parameter; therefore, the release time and latest completion time of the associated logistics tasks are not decision variables but time parameters derived from the starting time of the related order.

The order associated with the f -th logistics task is denoted by the mapping $r(f) \in R$. The release time of the task is:

$$\forall f \in F: e_f = t_{r(f)}^0 - \theta_f, \quad (15)$$

where θ_f is the task-generation lead time associated with the f -th logistics task. This relationship ensures that the task is released before the planned starting time of the associated order, taking into account the lead time required for its execution.

The latest completion time of the f -th logistics task is:

$$\forall f \in F: b_f = t_{r(f)}^0 - \delta_f, \quad (16)$$

where δ_f is the safety time buffer associated with the f -th task. The parameter b_f defines the latest time by which the task must be completed to ensure that the required material is available before the planned start of the order.

The following condition applies to the time parameters:

$$\forall f \in F: \theta_f \geq \delta_f \geq 0. \quad (17)$$

It follows that:

$$\forall f \in F: e_f \leq b_f \leq t_{r(f)}^0. \quad (18)$$

The condition ensures that the release time of the logistics task is not later than the latest allowable completion time and that the task completion deadline precedes, or at most coincides with, the planned starting time of the associated order.

The parameters θ_f and δ_f may take different values by task or task type. This allows logistics tasks with different preparation, picking, loading, transport, and line-side placement requirements to be handled within a common model. The assignment of tasks to individual milk-run cycles and their execution sequence are determined by the following constraints.

Assignment of Logistics Tasks to Milk-Run Cycles

The logistics tasks are executed by milk-run-based internal material-supply cycles. Each active logistics task must be assigned to exactly one milk-run cycle, while the capacity of each cycle and the available material-delivery resources must be taken into account.

$$\forall f \in F: \sum_{k \in K} g_{fk} = h_f \quad (19)$$

$$\forall k \in K: \sum_{f \in F^+} g_{fk} \leq N^+ a_k \quad (20)$$

$$\forall k \in K: \sum_{f \in F^-} g_{fk} \leq N^- a_k \quad (21)$$

$$\sum_{k \in K} a_k \leq H \quad (22)$$

These constraints ensure that every active logistics task is assigned to exactly one milk-run cycle, while the task load of each cycle and the number of available material-delivery resources do not exceed the permitted capacity.

Determination of the Sequence and Route of Logistics Tasks

The execution sequence of the logistics tasks assigned to milk-run cycles must be determined such that every active cycle starts from the depot denoted by 0, visits each assigned logistics task exactly once, and then returns to the depot.

An active milk-run cycle leaves the depot exactly once:

$$\forall k \in K: \sum_{f \in F} x_{0fk} = a_k. \quad (23)$$

An active milk-run cycle returns to the depot exactly once:

$$\forall k \in K: \sum_{f \in F} x_{f0k} = a_k. \quad (24)$$

Each logistics task assigned to the k -th milk-run cycle has exactly one outgoing route arc:

$$\forall f \in F, k \in K: \sum_{\substack{g \in F \cup \{0\} \\ g \neq f}} x_{fgk} = g_{fk}. \quad (25)$$

Each logistics task assigned to the k -th milk-run cycle has exactly one incoming route arc:

$$\forall f \in F, k \in K: \sum_{\substack{g \in F \cup \{0\} \\ g \neq f}} x_{gfk} = g_{fk}. \quad (26)$$

These constraints ensure that each active milk-run cycle starts from the depot, executes its assigned logistics tasks along a connected route, and then returns to the depot.

Temporal Constraints

The temporal constraints ensure that the execution of logistics tasks is consistent with the task sequence, release times, and planned starting times of the associated production orders.

$$\forall f, g \in F, f \neq g, k \in K: \tau_g \geq \tau_f + c_{fg} - M(1 - x_{fgk}), \quad (27)$$

$$\forall f \in F, k \in K: \tau_f \geq e_f + c_{0f} - M(1 - x_{0fk}).$$

This constraint ensures that, if the f -th logistics task is the first task of a milk-run cycle, its completion time accounts for the time required from the depot to the completion of that task.

$$\forall f \in F: 0 \leq \tau_f \leq Mh_f \quad (28)$$

$$\forall f \in F: E_f \geq \tau_f - e_f - M(1 - h_f) \quad (29)$$

$$\forall f \in F: E_f \leq \tau_f - e_f + M(1 - h_f) \quad (30)$$

$$\forall f \in F: 0 \leq E_f \leq Mh_f \quad (31)$$

$$\forall f \in F: L_f \geq \tau_f - b_f - M(1 - h_f) \quad (32)$$

$$\forall f \in F: 0 \leq L_f \leq Mh_f \quad (33)$$

$$\forall f \in F: \tau_f - b_f \leq Mz_f + M(1 - h_f) \quad (34)$$

$$\forall f \in F: \tau_f - b_f \geq \varepsilon_z - M(1 - z_f) - M(1 - h_f) \quad (35)$$

$$\forall f \in F: z_f \leq h_f \quad (36)$$

$$\sum_{f \in F} E_f \leq E_{\max} \sum_{f \in F} h_f \quad (37)$$

$$\sum_{f \in F} z_f \leq (1 - \eta_{\min}) \sum_{f \in F} h_f \quad (38)$$

These constraints ensure the temporal feasibility of logistics-task execution, determine response time and tardiness, and guarantee compliance with the required service level. If $h_f = 0$, the corresponding logistics task is inactive and τ_f , E_f , L_f , and z_f are equal to zero.

Line-Side Storage Capacity

The line-side inventory level is determined by the material deliveries and production material consumption specified for the given rolling decision horizon. The model ensures that the line-side inventory level does not exceed the available storage capacity throughout the planning horizon.

The evolution of the line-side inventory level over time is described by the following inventory-balance equation:

$$\forall l \in L, t \in T: B_{lt} = B_{l,t-1} + D_{lt} - U_{lt}, \quad (39)$$

where B_{lt} is the line-side inventory level at the l -th production line during the t -th time interval, D_{lt} is the quantity of material delivered to the l -th production line during the t -th time interval.

The parameters D_{lt} and U_{lt} are input data for the given rolling decision horizon and are updated before each optimization run based on the current production plan and logistics status. The inventory constraints in this model serve as a feasibility check of the current state within the rolling decision horizon. The values of D_{lt} and U_{lt} are exogenous input data; therefore, the evolution of line-side inventory is not directly determined by the milk-run routing and sequencing decisions.

The line-side inventory level cannot be negative and must not exceed the available storage capacity:

$$\forall l \in L, t \in T: 0 \leq B_{lt} \leq C_l. \quad (40)$$

These constraints ensure that the line-side inventory level remains physically feasible throughout the entire planning horizon while maintaining the balance between logistics deliveries and production material consumption.

3.1.3. Model Type and Solution Approach

The proposed model is a Mixed Integer Linear Programming (MILP) problem. The decision variables include both binary and continuous variables, while the objective functions and constraints are linear. The model applies a two-stage optimization procedure. In the first stage, it maximizes the serviceable order volume under the available resources; in the second stage, it optimizes the operational performance of the intralogistics processes while maintaining this volume.

3.2. Flexibility Indicators

The proposed mathematical model not only enables the optimization of operational decisions in the digitally supported postponement logistics system, but also provides a basis for the quantitative assessment of the system's logistics flexibility. Indicators can be derived from the model results that characterize the system's flexibility from different perspectives and allow the comparison of alternative operating configurations and resource levels.

In the scientific interpretation of logistics flexibility, it is useful to distinguish between the range of changes that the system can accommodate, the speed and resource requirements associated with implementing those changes, and the extent to which system performance remains uniform across different operating states. Based on the approaches of Upton and of Koste, Malhotra, and Sharma, these aspects correspond to the range, mobility, and uniformity dimensions of flexibility [2,13]. The range dimension expresses the breadth of the operating range within which the system can adapt to changing requirements; mobility characterizes the speed and effort required to implement changes; and uniformity indicates how stable the system's performance remains as operating conditions change.

Accordingly, evaluation of the mathematical-model results should not be limited solely to the serviceable order volume or resource utilization. A comprehensive characterization of logistics flexibility requires an indicator system that captures the different dimensions of flexibility. In the following, the logistics-flexibility indicators are denoted by $LF_i (i = 1, \dots, 5)$. For the task-based indicators LF_1 and LF_2 , only the actually executed elements of the set F are taken into account.

Based on the proposed mathematical model, the following five flexibility indicators can be defined:

- LF_1 - proportion of logistics tasks completed on time;
- LF_2 - average logistics response time;
- LF_3 - mix flexibility;
- LF_4 - volume flexibility;
- LF_5 - uniformity of logistics performance.

The indicators LF_1 and LF_2 characterize the mobility dimension of flexibility, LF_3 and LF_4 are associated with the range dimension, while LF_5 expresses the uniformity of system performance. This classification is consistent with approaches accepted in the international literature on logistics flexibility and with the volume, mix, temporal, transport, and routing flexibility dimensions used in the earlier questionnaire-based study.

3.2.1. Proportion of Logistics Tasks Completed on Time (LF_1)

One of the most important operational requirements of a digitally supported postponement logistics system is that logistics tasks be completed within the time window defined by the production process. Delayed material delivery or return transport may increase production waiting time, reduce resource utilization, and adversely affect overall system performance. Accordingly, the proportion of logistics tasks completed on time is one of the most important indicators of the system's temporal flexibility and operational responsiveness.

In the mathematical model, each logistics task is associated with a binary tardiness variable (z_f), which takes the value 1 if completion of the task exceeds the latest allowable completion time and 0 otherwise. On this basis, the proportion of actually executed logistics tasks completed on time is defined as follows:

$$LF_1 = \frac{\sum_{f \in F} (h_f - z_f)}{\sum_{f \in F} h_f}, \quad (41)$$

$$\sum_{f \in F} h_f > 0. \quad (42)$$

The distinction between on-time and tardy completion is made according to the time resolution represented by ε_z . The value of indicator LF_1 lies in the interval $[0,1]$. A value of $LF_1 = 1$ indicates that all actually executed logistics tasks in the period under consideration were completed within the allowable time window, while lower values indicate an increasing proportion of executed tasks completed late.

The indicator is directly linked to the service-level constraint of the mathematical model and is therefore suitable for comparing different operating configurations, resource capacities, and control strategies, as well as for evaluating the effect of digital decision support on logistics performance.

3.2.2. Average Logistics Response Time (LF_2)

One of the most important characteristics of logistics-system flexibility is the speed with which the system can respond to material-supply requirements generated by the production process. A shorter response time enables faster adaptation, reduces waiting time, and supports more efficient material supply. Average logistics response time is therefore a key indicator of the mobility dimension of flexibility.

In the mathematical model, each logistics task has an associated response-time variable (E_f) representing the time elapsed between the task's earliest feasible execution time and its actual completion time. The average response time of the actually executed logistics tasks is defined as follows:

$$LF_2 = \frac{\sum_{f \in F} E_f}{\sum_{f \in F} h_f}, \quad (43)$$

$$\sum_{f \in F} h_f > 0. \quad (44)$$

The indicator LF_2 represents the average response time of the actually executed logistics tasks. Lower values indicate faster logistics response and therefore more favourable operational flexibility, whereas higher values indicate slower response and less favourable operational flexibility.

Indicator LF_2 is derived directly from the decision variables of the mathematical model and is therefore suitable for comparing different resource capacities, control strategies, and operating configurations. The indicator characterizes the mobility dimension of logistics flexibility because it expresses how rapidly the system can adapt to changing production requirements and execute the necessary logistics tasks.

3.2.3. Mix Flexibility (LF_3)

One of the key flexibility characteristics of the logistics system is the extent to which the available production lines are capable of producing different product types. Greater technological flexibility increases the number of alternative production options, improves capacity utilization, and reduces operational risks arising from the failure or overload of

individual production lines. Mix flexibility is therefore one of the most important indicators of the range dimension of logistics flexibility.

Because mix flexibility is a structural indicator characterizing the technological capabilities of the system, it is determined independently of the current order portfolio. Accordingly, the set of product types P is used to determine this indicator. The set P is introduced solely for the assessment of structural mix flexibility and is not part of the decision variables of the operational optimization model.

The absolute mix-flexibility indicator expresses the average number of technologically feasible production lines available per product type:

$$LF_3 = \frac{\sum_{p \in P} \sum_{l \in L} A_{pl}}{|P|}, \quad (45)$$

where $|P|$ is the number of product types considered, A_{pl} a technological-feasibility parameter equal to 1 if the p -th product type can be produced on the l -th production line and 0 otherwise.

The value of indicator LF_3 shows the average number of different production lines on which a product type can be produced. A value of $LF_3 = 1$ means that every product type can be processed on only one production line, whereas higher values indicate an increasing number of alternative production options and therefore greater technological flexibility.

To enable objective comparison of systems of different sizes, it is also useful to introduce a normalized mix-flexibility indicator. The normalization takes into account that every product type must have at least one technologically feasible production line. Consequently, flexibility is represented not by the first feasible line but by the additional alternative processing options beyond it. The normalized mix-flexibility indicator is therefore defined as follows:

$$LF_3^{norm} = \frac{\sum_{p \in P} (\sum_{l \in L} A_{pl} - 1)}{|P| (|L| - 1)}. \quad (46)$$

The normalized indicator is defined for $|L| > 1$, and it is assumed that every product type under consideration can be produced on at least one production line. The normalized indicator takes values in the interval $[0,1]$. A value of $LF_3^{norm} = 0$ means that each product type can be processed on only one line and therefore no technological alternative exists, whereas a value of $LF_3^{norm} = 1$ indicates that all product types can be produced on all available production lines. The normalized indicator is particularly suitable for comparing manufacturing systems of different sizes or different technological configurations.

The indicators LF_3 and LF_3^{norm} characterize the range dimension of logistics flexibility because they express the breadth of the technological range within which the system can adapt to different production requirements. While LF_3 directly shows the average number of technological alternatives, LF_3^{norm} enables the flexibility of different systems to be compared on a common normalized scale.

3.2.4. Volume Flexibility (LF_4)

One of the most important flexibility characteristics of a logistics system is the extent to which it can adapt to changing production volume using the available resources. Volume

flexibility expresses the order volume that the material-supply system can serve under the operating conditions considered and how this performance changes when resources are increased or reduced. Volume flexibility is therefore a key indicator of the range dimension of logistics flexibility.

The first optimization objective of the proposed mathematical model is to maximize the order volume that can be served with the available resources. As a result of the model, the maximum achievable order volume can be determined for each resource configuration considered; this value is denoted by $Q_{\max}(H)$, where H represents the number of available logistics resources.

Volume flexibility can be defined as the relative change in volume compared with a reference state:

$$LF_4(H) = \frac{Q_{\max}(H) - Q_0}{Q_0}, \quad (47)$$

$$Q_0 = Q_{\max}(H_0) > 0.$$

where H_0 the reference resource level, Q_0 is the maximum serviceable order volume under the reference resource configuration.

Indicator $LF_4(H)$ expresses the extent to which the order volume that can be served by the system changes relative to the reference state. A positive value means that the operating configuration can serve a larger order volume, while a negative value indicates a reduction in system capacity. A value of $LF_4(H) = 0$ indicates that the serviceable order volume of the configuration under consideration is equal to that of the reference state.

The volume-flexibility indicator is determined from the results of repeated runs of the mathematical model under different resource configurations. It is therefore suitable for quantitatively evaluating the adaptability of the logistics system and objectively comparing different operating configurations. The indicator characterizes the range dimension of logistics flexibility because it expresses the extent to which the system can adapt to changing production volume as a function of the available logistics resources.

3.2.5. Performance Uniformity (LF_5)

Logistics flexibility means not only that the system is able to handle different operating states, but also that its performance remains stable as operating conditions change. Performance uniformity expresses the extent to which the system can maintain its service level when production volume, product mix, available resources, or the operating environment change. Indicator LF_5 therefore characterizes the uniformity dimension of logistics flexibility.

To determine indicator LF_5 , S denotes the set of operating states under consideration, $s \in S$ denotes the index of the individual states, and $s = 0$ denotes the reference state. The set S is introduced solely for the assessment of performance uniformity and is not part of the operational optimization model.

When different operating states are compared, performance deterioration relative to the reference state can be determined from several component indicators. To characterize the degree of volume service, the ratio of completed to required order volume ($FR = \text{completed volume}/\text{required volume}$) is used.

The aggregate performance deterioration of the s -th operating state is:

$$D_s = \lambda_1 d_{1s} + \lambda_2 d_{2s} + \lambda_3 d_{3s}, \quad (48)$$

where

$$d_{1s} = \max \left\{ 0, 1 - \frac{LF_{1,s}}{LF_{1,0}} \right\} \text{ and } LF_{1,0} > 0, \quad (49)$$

$$d_{2s} = \max \left\{ 0, \frac{LF_{2,s}}{LF_{2,0}} - 1 \right\} \text{ and } LF_{2,0} > 0, \quad (50)$$

$$d_{3s} = \max \left\{ 0, 1 - \frac{FR_s}{FR_0} \right\} \text{ and } FR_0 > 0. \quad (51)$$

Here, s denotes the index of the operating state under consideration, $LF_{1,s}$ is the proportion of logistics tasks completed on time in the s -th operating state, $LF_{2,s}$ is the average logistics response time in the s -th operating state, and FR_s is the ratio of completed to required order volume in the s -th operating state. The parameters λ_1 , λ_2 , and λ_3 are weights expressing the relative importance of the individual component indicators and satisfy the following conditions:

$$\lambda_1 + \lambda_2 + \lambda_3 = 1, \quad \lambda_j \geq 0. \quad (52)$$

The component indicators of performance deterioration take into account only changes in an unfavorable direction relative to the reference state. Accordingly, a decrease in the proportion of tasks completed on time or in the volume-service ratio, as well as an increase in average response time, increases the value of performance deterioration. If a performance characteristic improves relative to the reference state, the corresponding component indicator remains zero.

The performance-uniformity indicator is defined from the weighted performance deterioration across the operating states under consideration as follows:

$$LF_5 = 1 - \sum_{s \in S \setminus \{0\}} \omega_s D_s, \quad (53)$$

where S is the set of operating states under consideration; ω_s the weight expressing the importance of the s -th operating state.

The weights satisfy the following conditions:

$$\sum_{s \in S \setminus \{0\}} \omega_s = 1, \quad \omega_s \geq 0. \quad (54)$$

In this formulation, performance uniformity refers to the ability to maintain performance across different operating states; therefore, only deteriorations relative to the reference state are penalized, while improvements are not treated as a loss of uniformity.

A high value of indicator LF_5 means that system performance differs only slightly from the reference state across the operating states considered. Lower values indicate greater performance deterioration and therefore greater sensitivity of the system to changes in operating conditions. The value of indicator LF_5 may also become negative if the deterioration in performance relative to the reference state exceeds one.

Indicator LF_5 is particularly suitable for comparing the effects of different resource levels, volume peaks, product-mix changes, production-line failures, and route or node disruptions. It therefore provides a comprehensive measure of the extent to which the system can maintain its service level under different operating conditions.

The proposed mathematical model and the flexibility indicators derived from it provide a unified framework for the operational optimization and quantitative flexibility assessment of a digitally supported postponement logistics system. The model thus enables the systematic comparison of different operating configurations, resource levels, and production environments.

4. DISCUSSION AND CONCLUSIONS

This study developed an integrated MILP framework for the operational optimization and flexibility assessment of a digitally supported postponement logistics system. The model links production-order assignment with the execution of the resulting intralogistics tasks and considers technological feasibility, timing requirements, milk-run routing, resource limitations, service levels, and line-side storage constraints.

A key feature of the proposed approach is that production and intralogistics decisions are treated jointly. Assigning an order to a specific production or packaging line determines where and when the corresponding material demand occurs and therefore directly affects routing and resource utilization, while line-side inventory feasibility is evaluated based on the corresponding rolling-horizon delivery and consumption data. At the same time, limited logistics capacity may restrict which production configurations can be served within the required time windows. The integrated formulation makes these interdependencies explicit.

The model also reflects the dynamic nature of the operating environment. Logistics release and completion times are derived from planned production start times, and open tasks can be reoptimized on a rolling horizon when new planning or execution information becomes available. This provides a practical link between digital process transparency and mathematical decision support.

The two-stage optimization structure separates system capability from operational efficiency. In the first stage, the maximum production volume that can be served under the available resources and required service level is determined. In the second stage, logistics performance is optimized while keeping the achievable volume close to this optimum. This prevents improvements in travel distance, response time, or resource utilization from being achieved at the expense of a substantial reduction in serviceable production volume.

A further contribution of the model is that it provides a quantitative basis for evaluating logistics flexibility. Five indicators are used to represent different dimensions of flexibility: on-time task completion, average logistics response time, mix flexibility, volume flexibility, and performance uniformity. Together, these measures go beyond conventional efficiency indicators by describing not only how well the system performs under a given condition, but also how well it can adapt when production volume, product mix, resource availability, or other operating conditions change.

The proposed framework may therefore support several practical decisions. Repeated model runs can be used to compare alternative resource configurations, evaluate the effect of production-volume changes, analyse product-mix scenarios, or examine the consequences of temporary resource or production-line losses. The model can thus support both operational planning and selected capacity-related decisions.

The proposed approach also has limitations. First, the current formulation is deterministic, although actual production and logistics systems are affected by uncertainty in processing times, transport durations, resource availability, and execution conditions. The rolling-horizon structure allows the model to react to updated information, but uncertainty is not explicitly represented.

Second, production start times are treated as input parameters. The model therefore optimizes line assignment and intralogistics execution within a given production plan rather than jointly optimizing the complete production sequence. Third, some flexibility indicators depend on reference configurations and weighting parameters, which should be defined consistently and, where appropriate, tested through sensitivity analysis. Mix flexibility should also be interpreted as a structural measure of technological alternatives rather than as a direct measure of realized operational performance.

Another important limitation is that the current study provides the mathematical formulation and flexibility measures, but does not yet present a full empirical validation of the model under multiple operating scenarios. Future research should therefore focus on computational testing with operational data and scenario-based evaluation under different production volumes, product mixes, resource levels, and disruption conditions.

Further extensions may include stochastic or robust formulations to account explicitly for uncertainty, as well as computational studies to assess the scalability of the integrated MILP model. Future work may also consider closer integration between production sequencing and intralogistics planning, together with additional constraints such as congestion, heterogeneous vehicle fleets, energy consumption, or temporary route restrictions.

In conclusion, the proposed framework combines operational optimization and logistics-flexibility assessment within a single mathematical model. Its main contribution lies in integrating production-order assignment, intralogistics execution, and multidimensional flexibility evaluation rather than treating these aspects separately. The approach provides a basis for assessing not only the efficiency of a postponement logistics system under a given operating condition, but also its ability to maintain performance when those conditions change.

REFERENCES

- [1] Slack, N. (1983). Flexibility as a manufacturing objective. *International Journal of Operations & Production Management*, **3**(3), 4–13, <https://doi.org/10.1108/eb054696>
- [2] Upton, D. M. (1994). The management of manufacturing flexibility. *California Management Review*, **36**(2), 72–89, <https://doi.org/10.2307/41165745>
- [3] Zinn, W. & Bowersox, D. J. (1988). Planning physical distribution with the principle of postponement. *Journal of Business Logistics*, **9**(2), 117–136.
- [4] van Hoek, R. I. (2001). The rediscovery of postponement: A literature review and directions for research. *Journal of Operations Management*, **19**(2), 161–184, [https://doi.org/10.1016/S0272-6963\(00\)00057-7](https://doi.org/10.1016/S0272-6963(00)00057-7)

- [5] Yang, B., Burns, N. D. & Backhouse, C. J. (2004). Postponement: A review and an integrated framework. *International Journal of Operations & Production Management*, **24**(5), 468–487, <https://doi.org/10.1108/01443570410532542>
- [6] Winkelhaus, S. & Grosse, E. H. (2020). Logistics 4.0: A systematic review towards a new logistics system. *International Journal of Production Research*, **58**(1), 18–43, <https://doi.org/10.1080/00207543.2019.1612964>
- [7] Efthymiou, O. K. & Ponis, S. T. (2021). Industry 4.0 technologies and their impact in contemporary logistics: A systematic literature review. *Sustainability*, **13**(21), 11643, <https://doi.org/10.3390/su132111643>
- [8] Albrecht, T., Baier, M.-S., Gimpel, H., Meierhöfer, S., Röglinger, M., Schlüchtermann, J. & Will, L. (2024). Leveraging digital technologies in Logistics 4.0: Insights on affordances from intralogistics processes. *Information Systems Frontiers*, **26**(2), 755–774, <https://doi.org/10.1007/s10796-023-10394-6>
- [9] Lee, H. L. & Tang, C. S. (1997). Modelling the costs and benefits of delayed product differentiation. *Management Science*, **43**(1), 40–53, <https://doi.org/10.1287/mnsc.43.1.40>
- [10] Pagh, J. D. & Cooper, M. C. (1998). Supply chain postponement and speculation strategies: How to choose the right strategy. *Journal of Business Logistics*, **19**(2), 13–33.
- [11] Sethi, A. K. & Sethi, S. P. (1990). Flexibility in manufacturing: A survey. *International Journal of Flexible Manufacturing Systems*, **2**(4), 289–328, <https://doi.org/10.1007/BF00186471>
- [12] Koste, L. L. & Malhotra, M. K. (1999). A theoretical framework for analyzing the dimensions of manufacturing flexibility. *Journal of Operations Management*, **18**(1), 75–93, [https://doi.org/10.1016/S0272-6963\(99\)00010-8](https://doi.org/10.1016/S0272-6963(99)00010-8)
- [13] Koste, L. L., Malhotra, M. K. & Sharma, S. (2004). Measuring dimensions of manufacturing flexibility. *Journal of Operations Management*, **22**(2), 171–196, <https://doi.org/10.1016/j.jom.2004.01.001>
- [14] Beach, R., Muhlemann, A. P., Price, D. H. R., Paterson, A. & Sharp, J. A. (2000). A review of manufacturing flexibility. *European Journal of Operational Research*, **122**(1), 41–57, [https://doi.org/10.1016/S0377-2217\(99\)00062-4](https://doi.org/10.1016/S0377-2217(99)00062-4)
- [15] Jafari, H. (2015). Logistics flexibility: A systematic review. *International Journal of Productivity and Performance Management*, **64**(7), 947–970, <https://doi.org/10.1108/IJPPM-05-2014-0069>
- [16] Manders, J. H. M., Caniëls, M. C. J. & Ghijsen, P. W. T. (2017). Supply chain flexibility: A systematic literature review and identification of directions for future research. *The International Journal of Logistics Management*, **28**(4), 964–1026, <https://doi.org/10.1108/IJLM-07-2016-0176>
- [17] Chejarla, K. C., Vaidya, O. S. & Kumar, S. (2023). Flexibility in logistics: A literature review. *International Journal of Services and Operations Management*, **46**(1), 60–110, <https://doi.org/10.1504/IJSOM.2023.133411>
- [18] Facchini, F., Mossa, G. & De Tullio, S. (2022). A milk-run routing and scheduling model for a smart manufacturing system. *IFAC-PapersOnLine*, **55**(10), 1122–1127, <https://doi.org/10.1016/j.ifacol.2022.09.540>
- [19] Akkad, M. Z. & Bányai, T. (2023). Energy consumption optimization of milk-run-based in-plant supply solutions: An Industry 4.0 approach. *Processes*, **11**(3), 799, <https://doi.org/10.3390/pr11030799>
- [20] Facchini, F., Mossa, G., Sassanelli, C. & Digiesi, S. (2024). IoT-based milk-run routing for manufacturing system: An application case in an automotive company. *International Journal of Production Research*, **62**(1–2), 536–555, <https://doi.org/10.1080/00207543.2023.2254408>
- [21] Hosseini, A., Otto, A. & Pesch, E. (2024). Scheduling in manufacturing with transportation: Classification and solution techniques. *European Journal of Operational Research*, **315**(3), 821–843, <https://doi.org/10.1016/j.ejor.2023.10.013>