ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.1129

Data-Driven CONWIP: Similarity-Aware Admission for Job-Shop Scheduling

Guido Vinci Carlavan[1]; Tadeo Vega[2]; Francisco Yuraszeck[3] ; Daniel Rossit[2,4]

[1]Aluar, Puerto Madryn, Argentina, *gvincicarlavan@gmail.com*
[2]Engineering Department, Universidad Nacional del Sur, Argentina, *tadeo076@gmail.com*
[3]Department of Engineering Sciences, Universidad Andres Bello, Chile, *francisco.yuraszeck@unab.cl*
[4]INMABB, CONICET, Argentina, *daniel.rossit@uns.edu.ar*

Abstract– This work investigates a data-driven enhancement of CONWIP (Constant Work In Process) for job shop environments characterized by high product variety and routing heterogeneity. Standard CONWIP stabilizes flow by limiting total WIP, but blind admission can induce abrupt workload swings when a freed slot is filled by a job with a radically different routing or workload. We propose and evaluate similarity-aware admission rules (Sim-A and Sim-A-EDD) that use real-time job attributes) to select the candidate that best matches the profile of the job that just exited, while preserving due-date sensitivity. Using a simulation model calibrated to OKP-type production, we compare these rules against classical dispatchers (EDD, FIFO, Slack) across three arrival-rates scenarios. Each configuration was tested with 50 independent replications and a 95% confidence analysis after a 2000-day warm-up and 2000-day measurement horizon. Results show that workload-aware, data-driven admission significantly reduces both the percentage of late jobs and the magnitude of lateness under moderate and heavy loads, while retaining CONWIP's operational simplicity under light load. The findings demonstrate a practical pathway to combine pull control with online data to deliver customization without sacrificing predictability or throughput.

Keywords: production planning; mass customization; CONWIP; scheduling; data-driven.

I. Introduction

Manufacturing is undergoing a rapid shift from standardized mass production toward highly customized, customer-driven models [1]. Customization increases product value and market differentiation but also multiplies the number of possible variants and undermines the assumptions that traditional planning and scheduling methods rely on. In high-mix, low-volume job shops, each order can require distinct operations, setups, processing times and routing sequences, producing large variability in instantaneous workload and resource demand. [2] This variability complicates capacity planning, forecasting and resource allocation, and it amplifies the risk of bottlenecks, excessive work-in-process (WIP) and missed due dates [3]. Conventional dispatching rules and fixed WIP limits (designed for repetitive environments) often perform poorly under these conditions: a heavy job followed by a light one can leave some machines idle while others become overloaded, generating oscillating loads and unstable throughput [4].

Pull systems such as CONWIP (Constant Work In Process) offer a promising control paradigm by limiting the total number of jobs in the system and thereby stabilizing flow [5,6]. However, in job shops the benefits of a simple card-based CONWIP can be eroded by routing heterogeneity: admitting the next waiting job whenever a card is freed may inadvertently replace a completed heavy job with a light one (or vice versa), producing abrupt shifts in demand at specific machines [7]. To preserve CONWIP's simplicity while avoiding these adverse effects, admission and sequencing must account for job-level attributes rather than relying solely on counts.

A data-driven approach provides the necessary bridge. By capturing real-time job attributes (such as operation counts, estimated remaining processing and setup times, routing overlap, and recent processing history) control logic can make informed admission decisions that align incoming work with current capacity and recent completions [8,9]. Similarity-based admission rules, workload estimates and setup-aware selection reduce instantaneous workload swings, smooth machine utilization, and limit the frequency and severity of tardiness. Importantly, data-driven policies can be calibrated and adapted using historical traces and online performance metrics, enabling continuous improvement and robust behavior across varying demand regimes [10].

This paper investigates how augmenting CONWIP with data-driven admission and sequencing affects service levels in a job shop context. We compare the following classical dispatching rules: Earliest Due Date (EDD), First in First Out (FIFO), and Slack, with similarity-aware selectors that prioritize candidates whose workload profile better matches the job that just exited, and with hybrid rules that integrate due-date urgency. Experiments simulate realistic arrival patterns and processing variability, measuring both the incidence of late jobs and the magnitude of lateness. Results demonstrate that workload-aware, data-driven admission substantially improves robustness under moderate and heavy loads while preserving CONWIP's operational simplicity. The findings suggest a practical pathway for firms seeking to combine pull control with modern data capabilities to deliver customization without sacrificing predictability or throughput.

II. Background and related works

A. Mass Customization production

Customization in manufacturing involves adapting products, processes, and schedules to satisfy specific customer requirements. This can include design options, material choices, finishing treatments, and even bespoke routing through the shop floor. While customization enhances customer value and market differentiation, it also multiplies the number of product variants and invalidates many assumptions of standardized production [11]. Individual orders may require different operations, setups, and processing times, which complicates capacity planning, forecasting, and resource allocation [12].

Scheduling in highly customized environments is driven by this variability. Traditional scheduling methods assume repeatable tasks and predictable processing times; when every job differs, those assumptions break down [13]. Common problems include unpredictable routing, uneven workloads across machines, frequent setup changes, and uncertain lead times until order details are known. These conditions create bottlenecks, increase work-in-process variability, and raise the likelihood of missed due dates. Simple dispatching rules such as FIFO or EDD, or fixed WIP limits, can exacerbate the problem: a heavy job followed by a light one may leave some machines idle while others become overloaded, producing oscillating loads and unstable throughput [14].

Figure 1 maps manufacturing systems along four dimensions: production volume, number of variants, batch size, and process flexibility; and helps visualize where different paradigms sit. Job shops occupy the high-mix, low-volume quadrant, characterized by small batches and high routing flexibility. Mass production lies at the opposite extreme, with large volumes, few variants, and low flexibility. Mass customization appears between these extremes, aiming to deliver variety while preserving efficiency through modular design or flexible automation. The figure underscores how structural choices about volume, variety, batching, and flexibility directly influence scheduling complexity and control needs.

Practical approaches that perform well under customization blend pull logic with data-driven admission and sequencing. By using real-time job attributes (for example operation count, or recent processing history) systems can match incoming jobs to current capacity and recent completions. Adaptive, similarity-based admission rules smooth instantaneous demand, reduce setup churn, and balance utilization across machines. When combined with due-date awareness, these methods retain the simplicity and robustness of pull control while mitigating the adverse effects of routing heterogeneity in job shops.

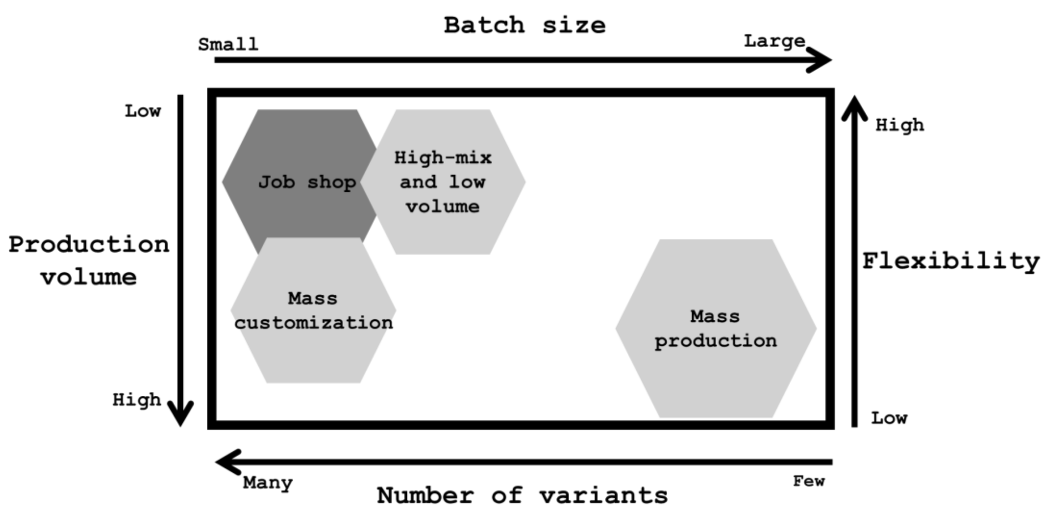


Fig. 1 Positioning of manufacturing systems according to production volume, product variety, batch size, and process flexibility. Job shops operate in the high-mix/low-volume region, mass production in the low-mix/high-volume region, and mass customization occupies the intermediate space between variety and scale.

B. Conwip concepts and related work

CONWIP is a pull control mechanism that maintains a fixed total WIP level by circulating a limited number of authorization cards: whenever a job departs the shop floor, its card is released and permits a new job to enter. Unlike Kanban, which allocates cards to individual stations, CONWIP governs the aggregate volume of work in progress across the entire shop, a feature that makes it simple, robust, and particularly suitable for environments with high product mix. In practice, orders still originate from planning systems, like Material Requirement Planning or Sales and Operations Planning (MRP or S&OP), but they are only released to the floor when a CONWIP card is available; this coupling of planning and controlled release helps limit global WIP and stabilize flow [6].

Figure 2 schematically depicts the CONWIP control loop and the information architecture used in this study. Jobs arrive at the sorter and wait in the release queue until the control loop verifies whether current WIP is below the authorized CONWIP size. If the condition is met, the sorter releases the next order for processing; otherwise, the order remains queued. The diagram also shows the shop floor topology (machines M1, M2 and M3) and the information flow connecting readers, a processing unit, and a central database that stores online order specifications. Dashed and solid lines indicate the transition from queued to active execution, while the readers and processing unit provide real-time status that the control loop uses to decide admissions. This figure highlights how CONWIP integrates physical flow and digital information to regulate admissions across heterogeneous routings.

CONWIP's main impacts on production planning are straightforward but significant. By fixing the card count, it reduces excessive queues and makes lead times more predictable, which simplifies capacity estimation and improves on-time performance. Because it controls total WIP rather than station inventories, CONWIP adapts well to high variety: aggregate planning tools (forecasts, MRP) remain useful while shop-floor releases are governed by the authorized WIP [16]. Its operational simplicity often yields productivity gains and fewer bottlenecks when combined with synchronization practices. However, CONWIP does not replace demand planning: forecasts, lot sizing, and prioritization remain necessary, and the chosen card count is critical, too many cards inflate WIP and lead times, while too few cause underutilization and reduced throughput. Finally, CONWIP can support make-to-order (MTO), make-to-stock (MTS), or hybrid mixes on a single release list, but doing so requires clear priority rules and operational discipline to preserve service objectives [17].

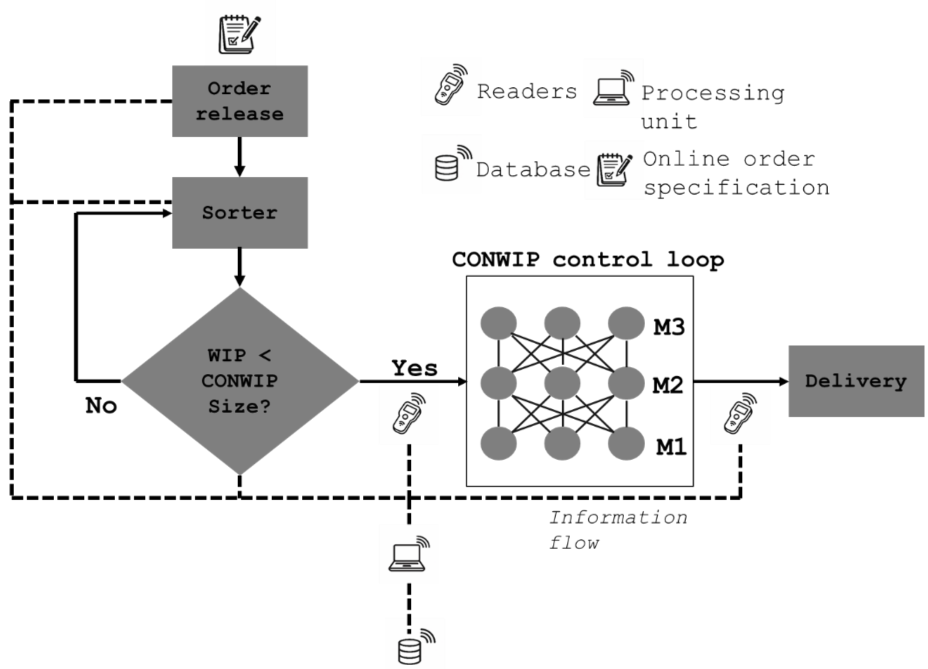


Fig. 2. CONWIP control loop for order release. When WIP is below the authorized CONWIP level, a new job is released from the sorter to the shop floor; otherwise, orders wait in the release queue. The diagram shows the interaction between machines, information readers, the processing unit, and the database that stores online order specifications.

C. Job shop related

Most CONWIP research has focused on flow-shop environments, where all jobs follow the same sequence of operations and process structure is highly regular. In flow shops the global WIP control that CONWIP provides maps naturally to the linear, predictable routing of parts, making analytical models and simulation studies tractable and results easier to generalize. Consequently, many empirical and theoretical papers evaluate CONWIP performance, sizing rules, and stability properties under flow-shop assumptions, producing a substantial body of evidence about throughput, lead-time reduction, and robustness to demand variability in those settings [15].

By contrast, literature addressing CONWIP in job shop contexts is comparatively sparse [7,18]. Job shops present heterogeneous routings, variable operation counts, and frequent setup changes, which complicate both modeling and practical implementation of a single, system-level WIP control. These complexities make it harder to derive closed-form results or to design universally applicable card-counting heuristics, so fewer studies have explored how CONWIP interacts with routing diversity, dynamic bottlenecks, and family-specific workloads. As a result, evidence about CONWIP's effectiveness, sizing, and integration with sequencing rules in job shops remains limited [6,19]. This gap motivates research that adapts CONWIP to job shop realities (by combining it with data-driven admission and sequencing policies, similarity metrics, or hybrid control schemes) to assess whether the benefits observed in flow shops can be realized under high routing variability.

III. PROBLEM DESCRIPTION AND SOLUTION APPROACH

In job-shop environments, a naïve CONWIP implementation (where the next waiting job is admitted whenever a card is freed) can produce counterproductive outcomes. CONWIP regulates only the count of jobs in the system, not the composition of their work. In flow-shop settings this coarse control often suffices because parts follow a common sequence and exhibit similar processing profiles; in job shops, however, routings, operation counts, processing times and setup requirements vary widely between orders. Consequently, a released CONWIP slot may be filled by a job whose routing and total workload differ substantially from the job that just exited. That mismatch can trigger abrupt shifts in demand at specific machines: a light job replacing a heavy one leaves downstream resources underutilized, while a heavy job replacing a light one can immediately overload particular centers and create new bottlenecks.

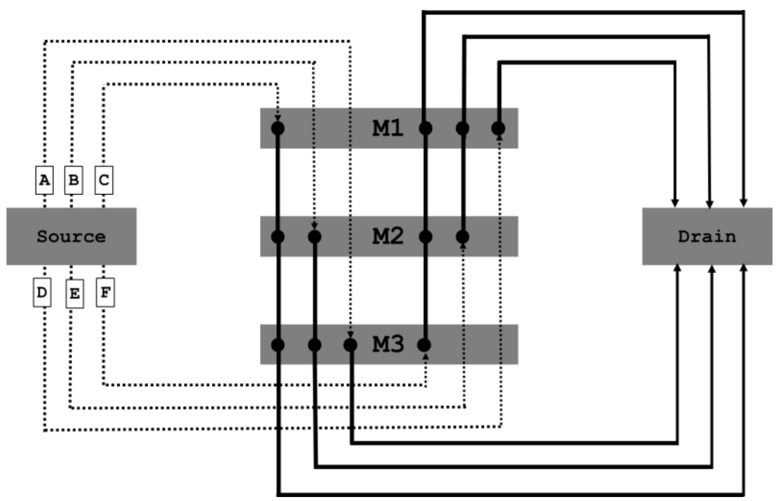


Fig. 3. Schematic representation of a job shop layout with multiple machines (M1–M3) and product routes (A–F). The diagram highlights the diversity of processing paths and the flow of jobs through different work centers before exiting the system.

Figure 3 illustrates this heterogeneity. Six product families (A–F) originate near the source and traverse a shop floor composed of workstations M1, M2 and M3. Each station hosts multiple processing resources, allowing several products (possibly from different families) to be processed concurrently. Crucially, each family follows a distinct processing route: some families visit all stations, others only a subset. In the figure, dashed lines indicate the start of processing for each family, while solid lines denote active execution. Because resource usage depends on family-specific routings, the instantaneous workload profile of the shop floor is highly sensitive to which families are in process.

These abrupt workload swings increase queueing variability, raise the frequency and cumulative duration of setups at critical machines, and render lead times less predictable despite a constant WIP count. In effect, the system's nominal capacity becomes uneven over time, eroding the principal practical benefits attributed to CONWIP, stable throughput and predictable cycle time. The problem is not the CONWIP mechanism per se but its blind admission policy:

controlling quantity without regard to job-level workload or routing can amplify variability rather than dampen it.

To address this, CONWIP should be augmented with admission and sequencing rules that incorporate job attributes. Practical enhancements include similarity-based admission (matching routing or operation counts to the recently completed job), workload estimates (total remaining processing and setup times), and setup-aware selection. Such policies aim to fill released slots with jobs whose profiles better align with current capacity and recent completions, thereby smoothing instantaneous demand across machines. When combined with priority adjustments for due dates or expedited orders, workload-aware CONWIP can preserve the simplicity and robustness of pull control while mitigating the adverse effects of routing heterogeneity in job shops.

A. Solution approach: Workload aware CONWIP admission rule

This rule extends standard CONWIP by choosing the next job to occupy a freed WIP slot according to its workload similarity with the job that has just exited. The objective is to stabilize the instantaneous workload profile so that machines are neither abruptly overloaded nor left idle when a new job replaces a completed one. Rather than admitting the next waiting order blindly, the admission logic evaluates candidate jobs in the pull queue and favors those whose workload descriptors best match the departing job, while still respecting due-date urgency as a secondary priority.

To implement this idea, real-time control loops were established to capture and compare the attributes of the job that just finished with those of jobs waiting in the release queue. This comparison enables the system to select the queued job that most closely resembles the outgoing job, thereby keeping the shop-floor workload constant not only in terms of the number of active orders but also in terms of the actual number of operations that must be processed.

Formally, let O_ℓ denote the set of operations composing job ℓ , and let $|O_\ell|$ be the cardinality of that set (i.e., the number of operations). When job j completes, the controller inspects O_j and compares $|O_j|$ with $|O_i|$ for each candidate i in the waiting set Q. The Simulation-based Algorithm A (Sim-A) selection criterion prioritizes candidates satisfying

$$|O_i| = |O_j|, \ \ i \in Q, i \neq j$$

placing those jobs at the front of the dispatch order. If multiple candidates meet this equality, tie-breaking is resolved by a secondary priority term that accounts for due-date urgency (for example, EDD or a slack-based score). In practice, the similarity metric can be extended beyond operation count to include routing overlap, estimated remaining processing time, and setup implications; these attributes can be combined into a compact workload descriptor and scored using a weighted similarity function.

By matching incoming jobs to the recent completion profile, the Sim-A rule reduces abrupt shifts in machine demand, lowers setup frequency and variability, and improves predictability of lead times while preserving the simplicity of CONWIP's global WIP control.

IV. EXPERIMENTS AND RESULTS

In the following section we present firstly the conception and design of the experiments, and then, the obtained results.

A. Experiments design and details

The experiments were conducted following an OKP-type (one-of-a-kind production) logic, which reflects environments characterized by high product variability and low repetition. To model this type of system, we adopted a parameterization similar to that used by [20,21], adapting it to the job-shop configuration analyzed in this study. The processing times p_{ij} for each operation of job j on machine i were generated from a uniform distribution $U \sim [500{,}1000]$. This wide interval captures the substantial variability typically observed in customized manufacturing, where operations may differ significantly in duration depending on product specifications. Jobs arrive dynamically throughout the planning horizon according to an arrival rate λ. Due dates for each job j, denoted d_j, were calculated by considering both the workload generated by the customer order and its arrival time r_j. Formally:

$$d_j = r_j + k \cdot \sum_{i \in M} p_{ij}$$

the parameter $k > 1$ acts as a design factor that controls the tightness of due dates. Lower values of k produce more restrictive deadlines, close to the total processing time of the job, while higher values allow more generous planning windows. In the experiments, k was varied within the range $\{1, \dots ,20\}$, enabling the evaluation of scheduling strategies under different levels of due-date pressure.

To assess the performance of each production-control strategy, we focused on service-level metrics, consistent with the objectives of mass-customization systems. Two indicators were used:

- Percentage of late jobs, representing the proportion of orders not completed by their due date.
- Lateness, measuring the delay for those jobs that did finish late.

This dual perspective allows us to evaluate not only how often deadlines are missed but also the severity of those delays. The goal is to minimize both the frequency and magnitude of tardiness, as both dimensions directly affect customer satisfaction.

Another critical aspect of the analysis concerns the system's ability to effectively utilize available capacity. Planning and scheduling decisions strongly influence how well resources are exploited; poor scheduling may lead to idle machines, excessive queues, or inefficient use of processing capabilities. To capture these effects, we examined three different (λ) arrival-rate scenarios, each representing a distinct workload condition:

- **High-load scenario**: one job arriving, on average, every 5 hours.
- **Medium-load scenario**: one job arriving, on average, every 5.5 hours.
- **Low-load scenario**: one job arriving, on average, every 6 hours.

These scenarios allow us to evaluate how each scheduling rule behaves under demanding, balanced, and relaxed operating conditions.

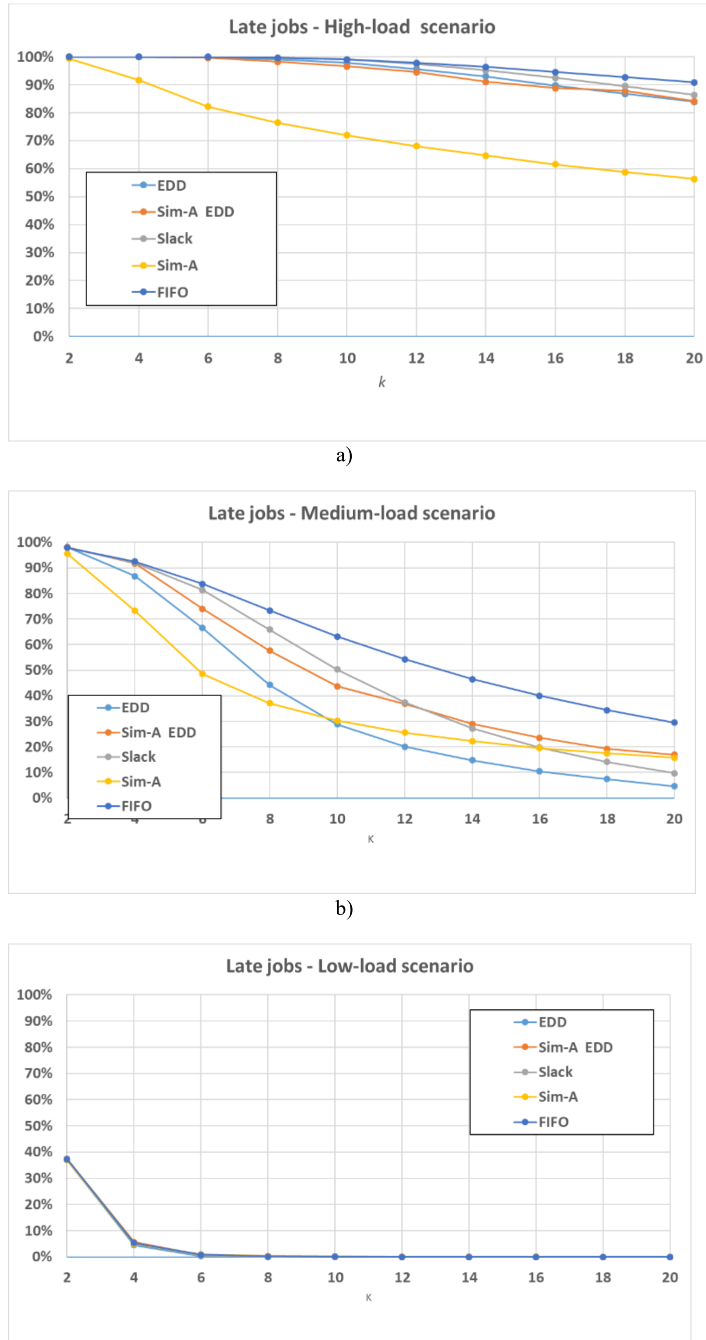
The scheduling strategies compared in the study include classical dispatching rules: EDD, Slack (or Critical Ratio, which relates remaining time to due date with remaining processing time), and FIFO, as well as two variants of the proposed Sim-A rule. Additionally, we evaluated a hybrid version that integrates operation-count matching with due-date prioritization, referred to as Sim-A-EDD.

Each experimental configuration was evaluated through 50 independent simulation replications, each using a different random-number seed using Plant Simulation Software. Results were analyzed with a 95% confidence level. Prior to data collection, preliminary experimentation was conducted to determine an appropriate warm-up period. A warm-up of 2000 simulated days was selected to ensure that the system reached steady-state conditions; all observations from this initial phase were discarded. After warm-up, performance metrics were recorded for an additional 2000 days, resulting in a total simulated horizon of 4000 days per replication [19]. This extensive horizon ensures statistically robust comparisons across all scheduling strategies and operating conditions.

B. Results

This section presents the experimental outcomes comparing dispatching and CONWIP admission rules across three arrival-rate scenarios. We report service-level metrics (percentage of late jobs and lateness magnitude) alongside utilization indicators and statistical pairwise comparisons (95% confidence).

Figure 4 shows how the different dispatching and admission rules perform across three load regimes. Under the high-load scenario (4a) the gap between policies is largest: the similarity-aware rules (Sim-A and Sim-A-EDD) clearly outperform classical dispatchers, producing the lowest percentage of late jobs. Sim-A-EDD combines workload matching with due-date sensitivity and therefore reduces both the frequency and severity of tardiness under stress, while FIFO and pure EDD exhibit the highest tardiness rates as they fail to prevent abrupt workload shifts that create transient bottlenecks. In the intermediate scenario (4b) differences narrow but remain meaningful. Sim-A variants still lead, indicating that workload matching improves robustness even when the system is not saturated. Slack (critical-ratio) performs better than FIFO and plain EDD here because it dynamically balances urgency and remaining work, but it does not match the consistency of the similarity-based selectors.



a)

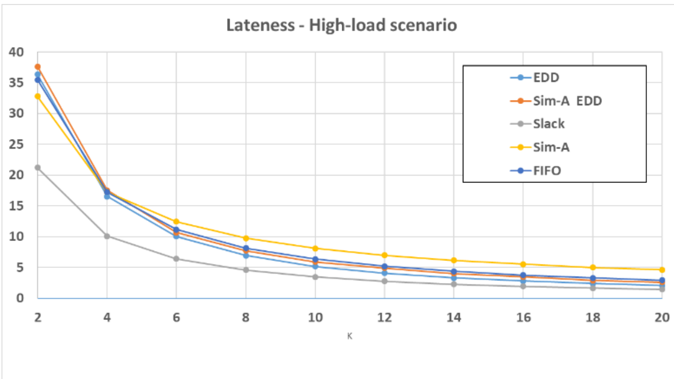
b)

c)

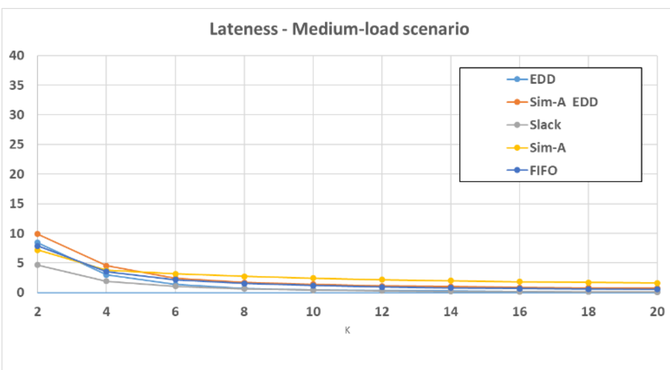
Fig. 4. Percentage of Late jobs results of all heuristics for the three arrival rates: a) exigent arrival rate, b) intermediate arrival rate and c) mild arrival rate.

Under the mild arrival rate (4c) all heuristics converge toward low tardiness levels; the advantage of workload matching diminishes because capacity is sufficient to absorb routing heterogeneity. However, Sim-A-EDD retains a small but consistent edge in reducing late deliveries and limiting extreme outliers, suggesting it improves worst-case behavior even in light load.

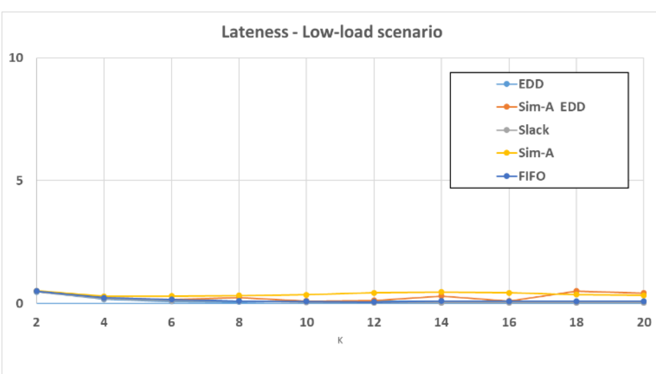
Overall, the results indicate that incorporating job-level similarity into CONWIP admission materially improves delivery performance in moderate and heavy traffic, while classical rules suffice only under light load. The findings also highlight the importance of tuning the similarity bonus and tie-breaking logic to balance on-time performance and fairness.



a)



b)



c)

Fig. 5 Lateness results of all heuristics for the three arrival rates: a) exigent arrival rate, b) intermediate arrival rate and c) mild arrival rate.

In the lateness counterpart, Figure 5 reports lateness magnitudes across the three arrival regimes, revealing how each policy affects not just the frequency of tardiness but the severity of delays. In the high-load case (5a) the differences are most pronounced: similarity-aware rules (Sim-A and Sim-A-EDD) produce substantially lower median lateness and noticeably shorter upper tails than classical dispatchers. Sim-A-EDD, by combining workload matching with due-date sensitivity, reduces extreme delays and compresses the distribution of lateness, indicating better protection against catastrophic backlog events when capacity is strained. FIFO and plain EDD show both higher medians and more frequent extreme outliers, reflecting their inability to prevent abrupt workload shifts that cascade into long delays.

In the intermediate regime (5b) the advantage of similarity-based admission persists but is reduced. Medians for Sim-A variants remain lower and their interquartile ranges tighter, suggesting improved consistency; Slack (critical-ratio) narrows the gap relative to FIFO and EDD by accounting for remaining work, yet it still yields heavier tails than Sim-A-EDD. This pattern indicates that workload matching improves worst-case performance even when the system is not saturated.

Under the mild arrival rate (5c) overall lateness shrinks for all heuristics, but Sim-A-EDD continues to show the smallest spread and fewest extreme delays. The results imply that similarity-aware admission primarily improves robustness (reducing both typical and extreme lateness) while classical rules only suffice when spare capacity is ample. Tuning the similarity bonus and tie-breaking logic is therefore critical: a well-calibrated bonus minimizes severe lateness without unduly penalizing urgent jobs.

V. Conclusions

The results demonstrate that augmenting CONWIP with data-driven, workload-aware admission significantly improves delivery performance in job shop settings characterized by high routing heterogeneity. Similarity-based selectors (Sim-A and Sim-A-EDD) consistently reduced both the percentage of late jobs and the magnitude of lateness under moderate and heavy arrival rates, while preserving CONWIP's simplicity under light load. These gains arise because matching incoming jobs to the profile of recently completed work smooths instantaneous demand across machines, lowers setup churn, and mitigates transient bottlenecks that blind count-based admission can create. The hybrid Sim-A-EDD rule further balances workload stability with due-date urgency, offering improved worst-case behavior without sacrificing average performance.

The study also highlights practical considerations for implementation. Card count remains a critical tuning parameter: it must be chosen to reflect target cycle times and arrival intensity, and similarity weights require calibration to avoid unfairly delaying urgent orders. Real-time data capture (operation counts, routing overlap, and estimated remaining processing and setup times) is essential for effective admission decisions, and the approach benefits from historical calibration and online adaptation. While simulation results are robust across the tested scenarios, empirical validation in live shop floors would strengthen external validity and reveal integration challenges such as data latency, measurement noise, and human factors.

Future work should explore richer similarity metrics, dynamic card-count adjustment, and integration with predictive maintenance and capacity planning. Extending the framework to larger, multi-cell layouts and mixed MTO/MTS portfolios will clarify scalability and economic trade-offs. Overall, combining pull control with data-driven admission offers a practical pathway for manufacturers to deliver customization with improved predictability and throughput.

References

[1] Wan, J., Li, X., Dai, H. N., Kusiak, A., Martinez-Garcia, M., & Li, D. (2020). Artificial-intelligence-driven customized manufacturing factory: key technologies, applications, and challenges. *Proceedings of the IEEE*, *109*(4), 377-398.

[2] 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.

[3] Qin, Z., & Lu, Y. (2021). Self-organizing manufacturing network: A paradigm towards smart manufacturing in mass personalization. *Journal of Manufacturing Systems*, *60*, 35-47.

[4] He, C., Deng, J., Wu, J., Qin, B., Chen, J., Li, Y., & Huang, Q. (2025). Intelligent data-driven system for mold manufacturing using reinforcement learning and knowledge graph personalized optimization for customized production. *Scientific Reports*, *15*(1), 23970.

[5] Spearman, M. L., Woodruff, D. L., & Hopp, W. J. (1990). CONWIP: a pull alternative to kanban. *The International Journal of Production Research*, *28*(5), 879-894.

[6] Hopp, W. J., & Spearman, M. L. (2011). *Factory physics*. Waveland Press.

[7] Diaz, R., & Ardalan, A. (2023). Innovating in data-driven production environments: simulation analysis of Net-CONWIP priority rule. *Industrial Management & Data Systems*, *123*(5), 1569-1598.

[8] Tao, F., Qi, Q., Liu, A., & Kusiak, A. (2018). Data-driven smart manufacturing. *Journal of manufacturing systems*, *48*, 157-169.

[9] Rossit, D. A., Tohmé, F., & Frutos, M. (2019). A data-driven scheduling approach to smart manufacturing. *Journal of Industrial Information Integration*, *15*, 69-79.

[10] Ghasemi, A., Farajzadeh, F., Heavey, C., Fowler, J., & Papadopoulos, C. T. (2024). Simulation optimization applied to production scheduling in the era of industry 4.0: A review and future roadmap. *Journal of Industrial Information Integration*, *39*, 100599.

[11] Perez, A. T. E., Rossit, D. A., Tohme, F., & Vasquez, O. C. (2022). Mass customized/personalized manufacturing in Industry 4.0 and blockchain: Research challenges, main problems, and the design of an information architecture. *Information Fusion*, *79*, 44-57.

[12] Chen, X., Li, Y., Wang, L., Wang, K., Guo, J., & Liu, J. (2025). Multi-objective grey wolf optimizer based on reinforcement learning for distributed hybrid flowshop scheduling towards mass personalized manufacturing. *Expert Systems with Applications*, *264*, 125866.

[13] Gao, S., Daaboul, J., & Le Duigou, J. (2021). Process planning, scheduling, and layout optimization for multi-unit mass-customized products in sustainable reconfigurable manufacturing system. *Sustainability*, *13*(23), 13323.

[14] Vinci-Carlavan, G., & Rossit, D. A. (2022, March). Personalized production in Industry 4.0: a CONWIP approach. In *2022 International Conference on Decision Aid Sciences and Applications (DASA)* (pp. 670-675). IEEE.

[15] Jaegler, Y., Jaegler, A., Burlat, P., Lamouri, S., & Trentesaux, D. (2018). The ConWip production control system: a systematic review and classification. *International Journal of Production Research*, *56*(17), 5736-5757.

[16] Vespoli, S., Mattera, G., Marchesano, M. G., Nele, L., & Guizzi, G. (2025). Adaptive manufacturing control with Deep Reinforcement Learning for dynamic WIP management in industry 4.0. *Computers & Industrial Engineering*, *202*, 110966.

[17] Gosavi, A., & Gosavi, A. A. (2024). CONWIP control in the digitized world: The case of the cyber-physical jobshop. *International Journal of Production Economics*, *270*, 109169.

[18] Thürer, M., Fernandes, N. O., Ziengs, N., & Stevenson, M. (2019). On the meaning of ConWIP cards: an assessment by simulation. *Journal of industrial and production engineering*, *36*(1), 49-58.

[19] Vespoli, S., Guizzi, G., Gebennini, E., & Grassi, A. (2022). A novel throughput control algorithm for semi-heterarchical industry 4.0 architecture. *Annals of Operations Research*, *310*(1), 201-221.

[20] Huang, G., Chen, J., & Khojasteh, Y. (2021). A cyber-physical system deployment based on pull strategies for one-of-a-kind production with limited resources. *Journal of intelligent manufacturing*, *32*(2), 579-596.

[21] Arza, J,, Prieto, G., Prata, B. & Rossit, D. A. (2026). Resilient rescheduling in flow shop systems with heterogeneous production lines. *Communications in Computer and Information Science*, **in press**.