

IDENTIFYING DYNAMIC BOTTLENECK IN MODULAR CONSTRUCTION USING SIMULATION AND PROBABILISTIC ANALYSIS

Nima Moghimi¹, Müge Tetik², Qipei Mei³, Vicente A. Gonzalez⁴, & Farook Hamzeh⁵

ABSTRACT

Bottleneck identification is fundamental to Lean production, as system constraints limit throughput and generate waste through waiting, excess Work-in-Process (WIP), and flow interruptions. Lean practices such as Value Stream Mapping and takt-cycle time analysis, together with indicators including utilization, queue length, and active-period methods, are commonly used to diagnose flow restrictions. However, these approaches assume relatively stable conditions and sufficient buffering—assumptions that rarely hold in modular construction factories. High product variability, reliance on manual labor, limited space, and stochastic task interactions cause bottlenecks to shift dynamically, making conventional diagnostics inconsistent. To address this challenge, this study introduces a probabilistic diagnostic principle for dynamic production environments. The approach employs Interdeparture Time (IDT) analysis and Cumulative Distribution Function (CDF) matching to distinguish true system constraints from stations experiencing temporary flow disruptions. The diagnostic logic is implemented within a hybrid Discrete-Event and Agent-Based simulation framework and applied to a timber wall assembly line. Results show that traditional Lean and utilization-based indicators produce conflicting bottleneck locations, while the probabilistic approach consistently identifies predominant constraints. The study further introduces a decision-support mechanism using sequential sensitivity analysis to quantify capacity thresholds and anticipate bottleneck migration.

KEYWORDS

Dynamic bottleneck, modular construction, probabilistic analysis, simulation, production controls.

INTRODUCTION

A bottleneck in production systems is any process, machine, or resource whose output rate lags its input rate, leading to work accumulation, delays, and productivity losses (Mahmoodi et al.,

¹ Corresponding Author, Ph.D. Student, Department of Civil and Environmental Engineering, University of Alberta, Canada, nmoghimi@ualberta.ca, orcid.org/0009-0008-4733-1276

² Postdoctoral Researcher, Department of Civil Engineering, School of Energy Systems, Lappeenranta-Lahti University of Technology, Lahti, Finland, muge.tetik@lut.fi

³ Assistant Professor, Department of Civil and Environmental Engineering, University of Alberta, Canada, qipei@ualberta.ca

⁴ Hal Kvisle Professor and Tier 1 Canada Research Chair in Digital Lean Construction, Infrastructure Human Tech (IHT) Lab, Strategic Projects Insight Centre in Engineering (SPICE), Department of Civil and Environmental Engineering, Faculty of Engineering, University of Alberta, Canada, vagonzal@ualberta.ca, orcid.org/0000-0003-3408-3863

⁵ Professor, Department of Civil and Environmental Engineering, University of Alberta, Canada, hamzeh@ualberta.ca, orcid.org/0000-0002-3986-9534

2022). Within Lean philosophy, bottlenecks constrain flow and limit system capacity through the buildup of non-value-adding work-in-progress (Koskela, 2000). The bottleneck station, characterized by the highest long-term utilization, ultimately dictates system throughput (Hopp & Spearman, 2011). However, real production systems rarely operate under steady-state conditions. Stochastic variability from equipment breakdowns, demand fluctuations, and process-time variation causes bottlenecks to shift dynamically along the production flow (Roser et al., 2002). Such dynamic bottlenecks introduce persistent control challenges efforts to relieve congestion at one workstation often transfer the constraint elsewhere, undermining traditional scheduling and capacity-balancing approaches (Chen et al., 2023; Roser et al., 2002). This volatility disrupts smooth flow and continuous improvement, making reliable production planning and takt control difficult to sustain (Bortolotti et al., 2015). Understanding, detecting, and managing dynamic bottlenecks is therefore critical for enhancing flow reliability and system resilience.

Modular construction entails the prefabrication of volumetric units within a controlled factory environment for subsequent on-site assembly (Kamali & Hewage, 2016). While this method enhances quality, safety, and schedule reliability through standardized workflows (Zohourian et al., 2025), it exhibits far less stable production dynamics than traditional manufacturing. High product customization, manual labor intensity, and frequent design changes introduce substantial variability in process times and sequencing, making system constraints highly transient (Panahi et al., 2023; Xiaosheng & Hamzeh, 2020). Buffer and space constraints, combined with the large footprint of volumetric modules, amplify delay propagation through blocking and starving effects (Altaf et al., 2018). This variability manifests as multiple forms of *muda*—waiting, overproduction, excess work-in-progress, and idle labor—arising from continually shifting constraint points (Darwish et al., 2022). The stability assumed in Lean planning systems, becomes difficult to sustain when dynamic bottlenecks migrate in response to changes in product mix, crew performance, or rework. Additionally, high product variety driven by differences in module geometry, framing layout, and service configuration produces distinct processing-time distributions across stations (Brissi & Debs, 2019; Moghadam & Al-Hussein, 2013). Consequently, modular production lines exhibit inherently dynamic bottleneck behavior, where constraint locations shift over time due to the complex interplay among capacity, variability propagation, and spatial limitations.

Bottleneck identification has long been a cornerstone of production improvement research. Theory of Constraints (TOC) by Goldratt (1984) formalized the continuous improvement process through a five-step cycle—identifying, exploiting, subordinating, elevating, and re-evaluating the system constraint (Moyo & Emuze, 2023; Şimşit et al., 2014). Lean manufacturing extended this foundation by emphasizing continuous flow and waste elimination through systematic constraint recognition. Tools such as Value Stream Mapping visualize interdependencies and delays across processes (Moghadam & Al-Hussein, 2013; Rother & Shook, 2003), while takt-time and cycle-time analyses align production pace with customer demand (Linnik et al., 2013; Moghadam & Al-Hussein, 2014). Gemba walks operationalize these principles by facilitating real-time observation of constraints directly at the workplace (Liker, 2020). From manufacturing perspective, bottlenecks are typically identified through utilization, queue length, and throughput variance analyses (Hopp & Spearman, 2011); however, such static indicators often fail to capture the transient nature of constraints in complex and stochastic systems.

To address this limitation, several analytical and data-driven methods have emerged. Roser et al. (2002) introduced the *active period method*, identifying bottlenecks based on the longest uninterrupted activity durations of machines or workstations. Li et al. (2009) proposed the *turning point method*, focusing on transitions between blocked and starved states to distinguish capacity-related from flow-related constraints—an essential distinction in buffer-constrained

system. Skoogh et al. (2023) reviewed fourteen major detection methods and categorized them according to information source (queue, process, or system state), identifying the turning-point approach as particularly robust in dynamic contexts. Simulation-based studies have also been used to integrate multiple indicators (e.g., queue time, utilization, and idle duration) to identify transient bottlenecks more holistically (Larsson & Rudberg, 2020). More recently, advanced data analytics and AI techniques have been adopted, including computer vision for spatio-temporal constraint tracking (Panahi et al., 2023) and deep learning models such as LSTM and graph neural networks for predictive bottleneck identification (Lai et al., 2023). Together, these developments reflect a transition from static diagnostics toward real-time, data-driven approaches capable of supporting continuous flow improvement in Lean-oriented production systems.

While bottleneck detection has progressed from manual observation to data-driven and predictive techniques, current methods remain limited in their ability to support decision-making *under* variable production conditions (Moghimi et al., 2026). Most approaches provide only static or predictive views of performance, without indicating which constraints exert the greatest system-wide influence or at what point local improvements lose effectiveness (Lai et al., 2021). These limitations are particularly acute in modular construction, which represents a qualitatively different production context from high-volume manufacturing. Unlike automated assembly lines, modular factories face severe spatial constraints due to the large physical footprint of volumetric units, restricting buffer capacity, while relying heavily on manual labor that introduces distinct behavioral variability. Under such conditions, conventional indicators such as throughput and utilization often produce conflicting signals, obscuring which stations truly limit flow. Consequently, a critical gap remains: existing approaches may detect or predict a bottleneck, but they fail to indicate which constraint exerts the dominant influence on system throughput or how improvement efforts should be prioritized in a dynamic environment.

To investigate and address this limitation, this study poses the following question: What diagnostic principle can reliably distinguish true system-governing constraints from temporary flow disruptions in stochastic, buffer-limited environments? In response, this study introduces a probabilistic diagnostic principle which posits that in variable systems, the true bottleneck is identified by the synchronization of its interdeparture time (IDT) distribution with the system's output, rather than by static utilization metrics. To operationalize this principle, an integrated simulation-based framework is developed combining Discrete Event Simulation (DES)—to model process flow, blocking, and starving—with Agent-Based Simulation (ABS) to capture dynamic crew collaboration (Moghimi & Kashani, 2025, 2026). This framework is explicitly scoped for production environments characterized by high product mix variability, labor-intensive operations, and rigid buffer limitations. The framework is applied to a wall assembly line in a prefabricated timber factory in Edmonton, Canada, demonstrating that conventional indicators yield inconsistent diagnoses due to stochastic interactions. Monte Carlo simulation is then employed to generate the probabilistic IDT and Cumulative Distribution Function (CDF) profiles required to validate the diagnostic principle. Subsequently, a sequential sensitivity analysis formalizes the decision-support capability, quantifying the capacity thresholds required to relieve constraints. This research thereby offers a dual contribution to the Lean Construction toolkit: (1) a diagnostic principle utilizing probabilistic flow synchronization to detect dynamic constraints, and (2) a decision-support mechanism for prioritizing capacity investments and predicting bottleneck migration. In this study, utilization- and throughput-based indicators are treated as preliminary diagnostic signals only. The adopted criterion for identifying the governing bottleneck in a dynamic, buffer-constrained environment is probabilistic flow synchronization, that is, the degree to which a station's interdeparture-time distribution matches the system output distribution.

METHODOLOGY

To address the practical challenge of dynamic constraints in modular construction, this research adopts the Design Science Research (DSR) methodology, which focuses on the creation and evaluation of innovative artifacts to solve relevant organizational problems (Hevner et al., 2004). Following the DSR process model defined by Peffers et al. (2007), the study proceeds through problem identification (inadequacy of static metrics), objective definition, artifact design, and demonstration. The primary artifact developed is a simulation-based framework for identifying and prioritizing bottlenecks, which integrates four analytical phases. First, processing time variability is quantified by fitting probability distributions to observed job attributes (e.g., module type, component count, finishing), capturing how product mix drives station-level bottlenecks. To ensure meaningful cross-station comparison in systems with heterogeneous jobs, all bottleneck metrics are harmonized by converting each job's workload ($q_{j,s}$) to a dimensionless reference unit, $u_{j,s} = q_{j,s}/q_{\text{ref}}$, where q_{ref} is the standard workload, following aggregate flow analysis principles (Hopp & Spearman, 2011). Throughput, WIP, and interdeparture time distributions are then computed on a consistent reference-unit basis, enabling direct comparison between stations regardless of job mix or module size. Although the final framework is presented in four analytical phases, its development was iterative: insights from simulation behavior, probabilistic diagnosis, and sensitivity-based decision support were used to refine the artifact toward a practically applicable bottleneck-identification logic.

The second phase employs hybrid discrete-event and agent-based simulation (DES-ABS) to jointly model material flow and adaptive crew behaviors. DES captures buffering, blocking, and starvation; ABS encodes (1) diminishing productivity from crew crowding, (2) self-organized task division, and (3) priority-driven multi-tasking responsive to downstream starvation. This layered approach separates capacity limits from structural constraints induced by buffers or staffing. Third, Monte Carlo sampling generates interdeparture time distributions across production scenarios, with CDFs applied to station and system-level flows to reveal bottleneck patterns. Finally, sensitivity analysis tests incremental capacity improvements, confirming bottleneck candidates, quantifying the threshold for diminishing returns, and sequencing investments for optimal system performance.

CASE STUDY DESCRIPTION

This section demonstrates the framework's practical implementation in a prefabricated timber factory wall assembly line in Edmonton, Canada, specializing in residential modular buildings. The factory operates two adjacent plants producing walls, roofs, floors, stairs, decks, and finishing components. The wall assembly line was selected because it exhibits three characteristics typifying the framework's target domain: (1) hybrid automation-manual operations combining machines with labor-intensive tasks, (2) severely constrained buffers between consecutive stations, and (3) high product variety driven by customization creating specification-dependent processing time variability. By modeling dynamic bottlenecks and variability, the framework helps eliminate waste and supports continuous flow, which are core Lean principles.

The wall assembly line comprises five interconnected stations: Framing, Sheeting, Computer Numerical Control (CNC) Nailing, Packaging, and Doors & Windows (DW) Installation, followed by Loading. Framing uses a single CNC machine operated by one crew to construct multiwall (MW) panel frames, which transfer via single-panel buffer to Sheeting, where two crews work on a two-panel conveyor. Following one-panel buffer, panels enter automated CNC Nailing (single-panel capacity), then Packaging where three crews subdivide MW panels at a single-table constraint (single-panel capacity). Interior panels route to storage;

remaining panels proceed through four parallel DW lanes (four-panel capacity each), where five crews perform specialized finishing. Panels then transfer via single-capacity transfer cart to a loading buffer (six lanes, four-panel capacity per lane), where two crews load panels using a single-capacity loading cart. Figure 1 illustrates the schematic layout of the wall assembly line, showing station configurations, buffer capacities, and material flow paths.

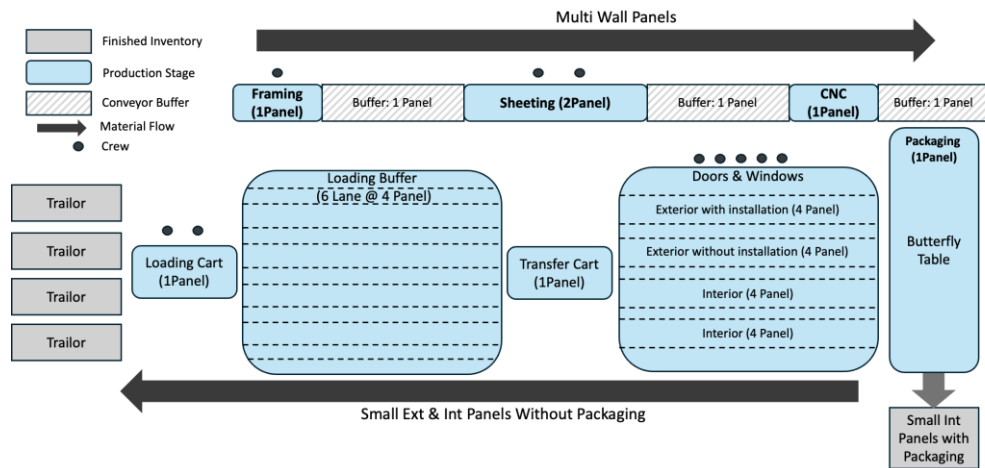


Figure 1: Schematic Layout of the Wall Assembly Line.

PROCESSING TIME DATA COLLECTION AND DISTRIBUTION FITTING

DATA COLLECTION APPROACH

Processing time data were collected through field observation from July to September 2025 for each panel, recording duration and job specifications (e.g., interior/exterior classification, installation requirements). A sample of 100–150 panel records per station was achieved, capturing both quantitative times and contextual factors (e.g., collaboration patterns, resource constraints) influencing variability.

JOB CATEGORIZATION

Analysis identified station-specific variability drivers. Framing processing times were uniform across all jobs, supporting a single distribution. Sheeting and CNC Nailing showed clear distinction: exterior MW panels required significantly longer times due to additional material and finishing steps, leading to separate interior/exterior models. Packaging variability was driven by subdivided panels routing (storage only, DW only, or both), affecting processing complexity. DW was categorized by panel type (interior/exterior) and, for exterior panels, installation requirements. Transfer Cart used a single distribution; Loading times were influenced by downstream trailer logistics. At Sheeting and CNC Nailing, mixed MW assemblies were classified as exterior if any exterior panel was present, based on operational practice where exterior requirements dictate timing.

DISTRIBUTION FITTING

Triangular distributions were selected for modeling processing times due to practical advantages in data-limited environments, requiring only minimum, mode, and maximum values (AbouRizk et al., 1994; Law et al., 2007). Parameters were derived from empirical data and validated with shopfloor experts. The fitted distributions for each station-job category combination are presented in Figure 2, spanning 6 distinct panel groups with differentiated time profiles. Simple mode-value analysis reveals dynamic bottleneck behavior—packaging

dominates for exterior panels without installation; DW dominates for exterior panels with installation. However, these deterministic assignments do not account for stochastic duration sampling, buffer constraints, or real-time crew allocation explored in simulation. It is to be noted that Multi-crew station parameters represent single-crew times; actual simulated times are reduced when multi-crew collaboration occurs.

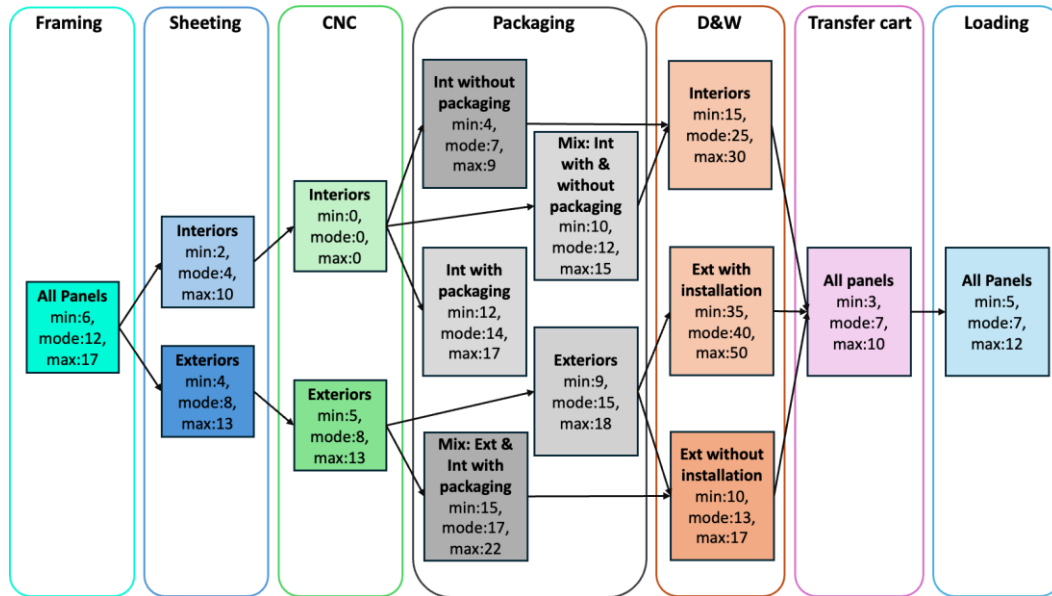


Figure 2: Triangular processing time distributions (in minutes) for each station-job category combination.

SIMULATION MODEL DEVELOPMENT

The hybrid DES-ABS framework was developed in Python using SimPy and deployed as a web-based application. Python was selected for three strategic reasons: (1) flexibility to integrate future advanced techniques for further research, (2) native support for hybrid modeling enabling seamless DES-ABS interaction, and (3) direct compatibility with analytical libraries (pandas, numpy) for reproducible Monte Carlo and sensitivity workflows.

DISCRETE-EVENT SIMULATION (DES)

The DES layer models process flow, resource contention, and buffer limitations using capacity-limited SimPy Store objects enforcing FIFO discipline. Single-capacity stores represent transfer elements; multi-capacity stores model buffers (e.g., DW: 4 lanes $\times$ 4 panels; Loading: 6 lanes $\times$ 4 panels). Blocking and starving arise naturally from resource constraints. Routing logic dynamically assigns panels at Packaging based on panel type and queue lengths, replicating observed load-balancing.

AGENT-BASED SIMULATION (ABS)

The ABS layer models dynamic crew collaboration and task prioritization (Abou-Ibrahim & Hamzeh, 2022). At Sheeting, crews reduce processing time through time multipliers reflecting productivity gains with diminishing returns. DW employs hierarchical decision logic: (1) add crews to under-staffed panels, (2) prioritize job starts (exterior with installation $\rightarrow$ exterior without $\rightarrow$ interior), (3) assign helper crews up to capacity limits. Preemption rules allow crew reassignment to higher-priority tasks. Transfer cart operations preempt DW crews at maximum priority, replicating real disruptions. Loading crews follow global FIFO rules with 3% per-panel time increase as trailers fill. Time adjustment factors were derived based field observations.

VERIFICATION AND VALIDATION

The model underwent structured walkthroughs with production managers and event-trace analyses confirming correct blocking/starving/preemption behavior. Validation against real production data (84 MW panels, 135 small panels) demonstrated high accuracy: Framing ($R^2 = 0.95$), DW ($R^2 = 0.89$), Loading ($R^2 = 0.88$), with mean absolute percentage errors of 16–18% (acceptable for complex construction simulations). Slight negative biases reflect unmodeled micro-delays. The model reliably reproduces production dynamics and variability, providing a validated foundation for bottleneck analysis (Law et al., 2007; Sargent, 2010).

RESULTS

THROUGHPUT AND UTILIZATION ANALYSIS (SINGLE SIMULATION RUN)

A baseline simulation was conducted to evaluate real-time throughput and utilization patterns across all stations under typical operating conditions. The production schedule, based on 80 multiwall (MW) panels with varying specifications, was derived from actual plant data for operational accuracy. Throughout the analysis, small panels—defined as the downstream units remaining in the system until final loading—served as the reference unit across all stations. At stations working on MW modules, each MW was converted to an integer count of reference units based on how many final small panels it produced. Throughput and work-in-progress were thus measured in reference units per hour, ensuring comparable metrics across stations with different job configurations. Utilization was calculated as the percentage of active crew working time relative to total available capacity, excluding waiting due to blocking or starvation. For multi-crew stations such as Doors & Windows (five crews) and Sheeting (two crews), aggregate utilization was normalized based on combined crew time (total active crew-minutes divided by the product of available minutes and crew count).

Figure 3 presents throughput and utilization time-series analysis across all stations. Results demonstrate that no single station consistently dominated as the bottleneck throughout the simulation period. Instead, throughput rates fluctuated dynamically, with alternating constraints shifting among stations. For instance, while Sheeting and CNC occasionally experienced idle phases during upstream delays, both Transfer Cart and Loading frequently reached full utilization when downstream flow was impeded. This pattern indicates that bottleneck emergence results not only from intrinsic station capacity but also from system-level interactions created by limited buffering and dynamic crew collaboration.

It should be noted that Figure 3 is presented as an initial diagnostic view of real-time throughput and utilization behavior rather than the final bottleneck identification criterion. In this study, these indicators are used to reveal candidate constraints, while the governing bottleneck is ultimately identified through the IDT/CDF synchronization analysis reported in the following subsection.

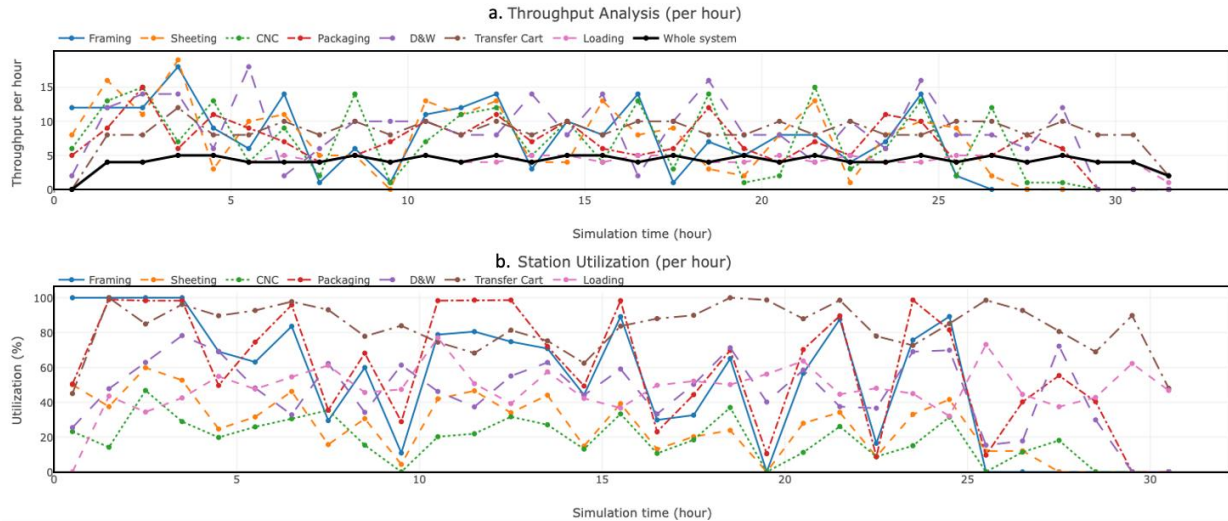


Figure 3: Time-series analysis of (a) throughput and (b) utilization across stations from a single-run simulation.

Monte Carlo Simulation and Inter-Departure Time (IDT) Analysis

To capture the stochastic nature of system variability, a Monte Carlo experiment with 200 iterations was executed, consistent with established guidelines for achieving convergence reliability in manufacturing simulations (Law et al., 2007). Each replication simulated the same 80-panel schedule but randomized processing durations according to the triangular distributions defined. IDT, defined as the time interval between consecutive completions of reference panels at each station, was selected as the primary performance metric.

After aggregating all replications, IDT data for each station and the overall system were fitted to best-fitting probability distributions using the Anderson-Darling criterion (Anderson & Darling, 1952), and CDFs were generated. As illustrated in Figure 4, the CDF comparison reveals distinct behavioral patterns. Several stations (Framing, Sheeting, CNC) exhibit IDT distributions noticeably shifted to the right of the system CDF. Statistically, this signifies that their interdeparture times are longer than the system output cycle, despite these stations not being the primary constraints. Conversely, the CDFs for the Transfer Cart and Loading stations closely track the system CDF, particularly up to the 90th percentile. This statistical synchronization identifies them as the governing constraints that dictate overall system output in the current configuration.

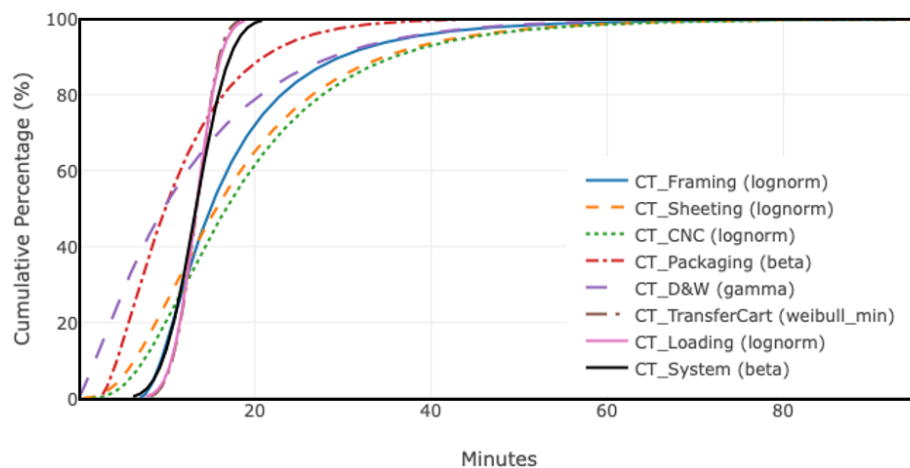


Figure 4: Best-fit CDFs of station and system interdeparture times from Monte Carlo

simulation results.

SENSITIVITY ANALYSIS

The final phase of the framework implemented a sequential, CDF-guided sensitivity procedure. Rather than applying sensitivity analysis to all stations in parallel, the approach operated in rounds: at the start of each round, candidate bottlenecks were prioritized based on their distribution's similarity (in shape and tail alignment) to the system CDF. For the prioritized bottleneck, incremental 10% reductions in processing time were applied until marginal throughput gains diminished.

This sequential process continued for four rounds, yielding the following outcomes (Figure 5): Round 1 identified the Transfer Cart as the primary bottleneck, requiring 6 steps (factor = $0.531\times$ baseline) to achieve a throughput of 6.19 panels/hour (gain +1.88). Round 2 identified Loading as the next constraint, requiring 4 steps ($0.656\times$), resulting in 6.66 panels/hour (gain +0.47). Round 3 identified Packaging, requiring 6 steps ($0.531\times$), yielding 7.32 panels/hour (gain +0.66). Round 4 identified Framing, requiring 3 steps ($0.729\times$), yielding 7.73 panels/hour (gain +0.41).

The cumulative throughput improvement across all four rounds totaled +3.41 panels/hour. Figure 6 illustrates the evolution of CDFs across these rounds. After each improvement phase, the CDFs for all stations became increasingly similar, quantitatively demonstrating the synchronization of flow.

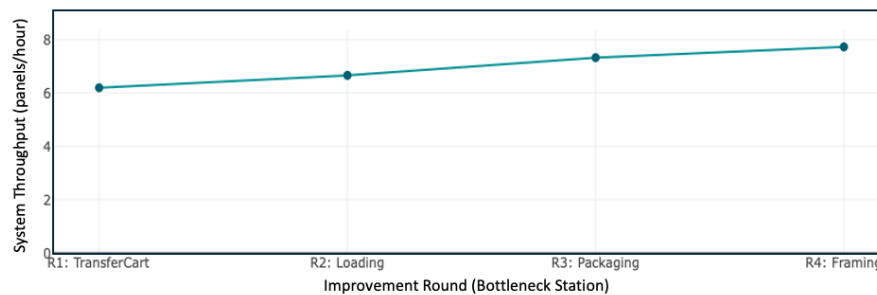


Figure 5: Cumulative throughput improvement across four sequential sensitivity analysis rounds.

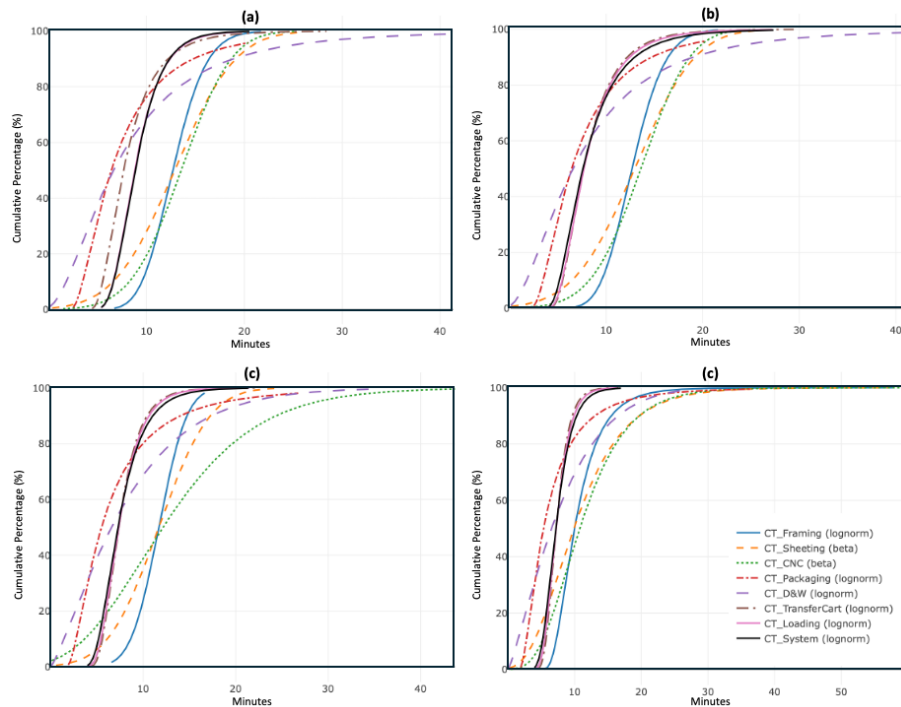


Figure 6: CDF evolution across sequential bottleneck improvements: (a) Transfer Cart, (b) Loading, (c) Packaging, (d) Framing.

DISCUSSION

THEORETICAL IMPLICATIONS: THE FAILURE OF CLASSICAL DIAGNOSTICS

The probabilistic analysis yielded significant theoretical implications regarding the nature of dynamic constraints. As observed in Figure 4, several non-bottleneck stations exhibited IDT distributions shifted to the right of the system CDF. In classical queuing theory (Hopp & Spearman, 2011), such a shift would typically imply insufficient capacity. However, in this modular construction context, this shift serves as probabilistic evidence of blocking and starvation. These stations are not bottlenecks but "victims" of flow disruption. Put differently, these stations act as receivers of propagated variability: blocking and starving effects generated elsewhere distort their local interdeparture behavior, causing them to appear capacity-constrained even when they do not govern system throughput. This finding empirically demonstrates that finite buffer capacities fundamentally alter flow behavior, effectively invalidating diagnostic assumptions based on the classical, unconstrained-buffer manufacturing principles described by L. Li et al. (2009) and Roser et al. (2002). Therefore, relying solely on standard utilization metrics in this context is insufficient.

STRUCTURAL CONSTRAINTS AND BUFFER MANAGEMENT

While the sensitivity analysis demonstrated that capacity additions could theoretically synchronize flow, practical implementation is often limited by structural constraints. For instance, the simulation identified the Transfer Cart as a primary bottleneck, yet expanding its capacity is infeasible due to physical space restrictions on the factory floor. In such cases where capacity cannot be elevated, the focus must shift to buffer management. As observed in the factory, the downstream Doors & Windows station utilizes an inventory buffer of up to 16 panels to decouple itself from the Transfer Cart. This confirms that in modular construction, optimal control often requires a hybrid strategy: using CDF diagnostics to identify the constraint,

but relying on time-buffering strategies (similar to Drum-Buffer-Rope) rather than capacity investment when physical limits are reached (Hamzeh et al., 2007).

DECISION SUPPORT CAPABILITY

The proposed framework addresses the gap in prioritizing improvement efforts. By identifying which station is "synchronized" with the system output (matching CDFs), the method successfully differentiated true system-governing bottlenecks from stations that were merely inefficient due to disruptions. This distinction is critical for lean practitioners, ensuring that resources are applied only to the constraint that actually governs system performance. The sequential sensitivity approach provides a practical and scalable alternative to full optimization-based solutions, allowing decision-makers to predict bottleneck migration and determine the point of diminishing returns for capital investments.

CONCLUSION

This research addressed the limitations of conventional static indicators in modular construction by introducing a probabilistic diagnostic principle for bottleneck detection. Critically, the proposed approach supports Lean decision-making by distinguishing true capacity constraints from temporary flow disruptions under dynamic variability. This capability enables more effective prioritization of improvement efforts, proving that in buffer-constrained environments, true bottlenecks are identified by their synchronization with system output rather than by simple utilization metrics. The application to a timber wall assembly line demonstrated this practical value: while traditional diagnostics produced inconsistent signals, the probabilistic method reliably isolated the governing constraints (Transfer Cart and Loading) from stations merely suffering from blocking. The findings underscore that in modular construction, constraints are often structural rather than purely capacity-based, requiring a balance between capacity investment and strategic buffer management to synchronize flow. Furthermore, the sequential sensitivity analysis formalized a decision-support mechanism, enabling practitioners to quantify improvement thresholds and predict bottleneck migration. The findings should be interpreted in light of several limitations, including reliance on a single case-study line, simulation-based rather than field-implementation validation, and the focus on one modular production context. This study thereby expands the Lean Construction toolkit with a dual contribution: (1) a diagnostic principle that filters out stochastic noise to detect system-governing constraints, and (2) a predictive framework for targeting Lean interventions where they will yield the highest flow reliability.

ACKNOWLEDGEMENT

This research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Alliance grant ALLRP 581074–22, Discovery Grant RGPIN-2020-04420, and NSERC Alliance—Alberta Innovates Advance Program 232403286. All findings and conclusions expressed in this paper are those of the authors and do not reflect those of the contributors.

REFERENCES

- Abou-Idrahim, H., & Hamzeh, F. (2022). Understanding Stakeholders' impact on design workflow dynamics using agent-based modeling. *Automation in Construction*, 138, 104254.
- AbouRizk, S. M., Halpin, D. W., & Sawhney, A. (1994). Modelling Uncertainty in Construction Simulation. *MACHINE INTELLIGENCE AND PATTERN RECOGNITION*, 17, 287.

- Altaf, M. S., Bouferguene, A., Liu, H., Al-Hussein, M., & Yu, H. (2018). Integrated production planning and control system for a panelized home prefabrication facility using simulation and RFID. *Automation in Construction*, 85, 369–383.
- Bortolotti, T., Boscari, S., & Danese, P. (2015). Successful lean implementation: Organizational culture and soft lean practices. *International Journal of Production Economics*, 160, 182–201.
- Brissi, S. G., & Debs, L. (2019). Lean, automation and modularization in construction. *27th Annual Conference of the International Group for Lean Construction, IGLC, 2019*, 711–722.
- Chen, G., He, C., Hsiang, S., Liu, M., & Li, H. (2023). A mechanism for smart contracts to mediate production bottlenecks under constraints. *31st Annual Conference of the International Group for Lean Construction (IGLC)*, 1232–1244.
- Darwish, M., Mohsen, O., Mohamed, Y., & Al-Hussein, M. (2022). Integrated simulation and lean approach for production line improvement in a prefabricated panelized homebuilding facility. *Proc. 28th Annual Conference of the International Group for Lean Construction (IGLC)*.
- Goldratt, E. M. ; C. J. (1984). *The Goal: A Process of Ongoing Improvement*. North River Press.
- Hamzeh, F. R., Tommelein, I. D., Ballard, G., & Kaminsky, P. (2007). Logistics centers to support project-based production in the construction industry. *Proceedings of the 15th Annual Conference of the International Group for Lean Construction (IGLC 15)*, 181–191.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 75–105.
- Hopp, W. J., & Spearman, M. L. (2011). *Factory physics*. Waveland Press.
- Kamali, M., & Hewage, K. (2016). Life cycle performance of modular buildings: A critical review. *Renewable and Sustainable Energy Reviews*, 62, 1171–1183. <https://doi.org/10.1016/j.rser.2016.05.031>
- Koskela, L. (2000). *An exploration towards a production theory and its application to construction*. VTT Technical Research Centre of Finland.
- Lai, X., Qiu, T., Shui, H., Ding, D., & Ni, J. (2023). Predicting future production system bottlenecks with a graph neural network approach. *Journal of Manufacturing Systems*, 67, 201–212.
- Lai, X., Shui, H., Ding, D., & Ni, J. (2021). Data-driven dynamic bottleneck detection in complex manufacturing systems. *Journal of Manufacturing Systems*, 60, 662–675.
- Larsson, R., & Rudberg, M. (2020). *A simulation-based approach for systematic analysis of workflow during the construction of in-situ concrete frames*.
- Law, A. M., Kelton, W. D., & Kelton, W. D. (2007). *Simulation modeling and analysis* (Vol. 3). Mcgraw-hill New York.
- Li, J., E. Blumenfeld, D., Huang, N., & M. Alden, J. (2009). Throughput analysis of production systems: recent advances and future topics. *International Journal of Production Research*, 47(14), 3823–3851.
- Liker, J. (2020). *The Toyota way: 14 management principles from the world's greatest manufacturer*. McGraw-Hill.
- Linnik, M., Berghede, K., & Ballard, G. (2013). An experiment in takt time planning applied to non-repetitive work. *21st Annual Conference of the International Group for Lean Construction*, 31–32.
- Mahmoodi, E., Fathi, M., & Ghobakhloo, M. (2022). The impact of Industry 4.0 on bottleneck analysis in production and manufacturing: Current trends and future perspectives. *Computers & Industrial Engineering*, 174, 108801.

- Moghadam, M., & Al-Hussein, M. (2013). Resource optimization for modular construction through value stream map improvement. *Proceedings of the 21th Annual Conference of the International Group for Lean Construction*, 503–512.
- Moghadam, M., & Al-Hussein, M. (2014). An enhanced scheduling technique for modular construction manufacturing. *Proceedings of the International Group for Lean Construction (IGLC-22)*.
- Moghimi, N., Arshang, O., & Hamzeh, F. (2026). Digital twins in offsite construction: Current implementations, challenges, and future pathways. *Automation in Construction*, 183, 106820.
- Moghimi, N., & Kashani, H. (2025). A probabilistic framework to design the financing scheme for resilience enhancement of infrastructure systems. *Reliability Engineering & System Safety*, 262, 111179. <https://doi.org/10.1016/j.ress.2025.111179>
- Moghimi, N., & Kashani, H. (2026). Probabilistic Modeling to Develop the Optimal Seismic Resilience Enhancement Strategy for Industrial Facilities. *Journal of Infrastructure Systems*, 32(1), 04025041. <https://doi.org/10.1061/JITSE4.ISENG-2754>
- Moyo, C., & Emuze, F. (2023). Building a lean house with the theory of constraints for construction operations in Zimbabwe: A conceptual framework. *Proceedings of the 31st Annual Conference of the International Group for Lean Construction (IGLC31)*, 870–881.
- Panahi, R., Louis, J., Podder, A., Swanson, C., & Pless, S. (2023). Bottleneck detection in modular construction factories using computer vision. *Sensors*, 23(8), 3982.
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77.
- Roser, C., Nakano, M., & Tanaka, M. (2002). Shifting bottleneck detection. *Proceedings of the Winter Simulation Conference*, 2, 1079–1086.
- Rother, M., & Shook, J. (2003). *Learning to see: value stream mapping to add value and eliminate muda*. Lean enterprise institute.
- Sargent, R. G. (2010). Verification and validation of simulation models. *Proceedings of the 2010 Winter Simulation Conference*, 166–183.
- Şimşit, Z. T., Günay, N. S., & Vayvay, Ö. (2014). Theory of constraints: A literature review. *Procedia-Social and Behavioral Sciences*, 150, 930–936.
- Skoogh, A., Thürer, M., Subramaniyan, M., Matta, A., & Roser, C. (2023). Throughput bottleneck detection in manufacturing: a systematic review of the literature on methods and operationalization modes. *Production & Manufacturing Research*, 11(1), 2283031.
- Xiaosheng, T., & Hamzeh, F. (2020). Precast concrete building construction process comparison. *Proceedings of the Annual Conference of the International Group for Lean Construction, Berkeley, CA, USA*, 6–12.
- Zohourian, M., Pamidimukkala, A., Kermanshachi, S., & Almaskati, D. (2025). Modular Construction: A Comprehensive Review. *Buildings*, 15(12), 2020.