

A DIGITAL-TWIN BASED CONWIP FRAMEWORK FOR WORK IN PROGRESS (WIP) MANAGEMENT IN MODULAR CONSTRUCTION

Nima Moghimi¹, Sahar H. Shamaee², Haitao Yu³, Qipei Mei⁴, Vicente A. Gonzalez⁵, & Farook Hamzeh⁶

ABSTRACT

Effective Work-in-Process (WIP) management is a principle of lean production, and pull-based control methods such as Kanban and CONWIP have been shown to regulate WIP levels and improve production flow. However, conventional implementations based on physical cards or visual signals are often impractical in modular construction environments, where high product customization, limited floor space, and heterogeneous WIP units complicate traditional control mechanisms. This paper presents a digital-twin-based CONWIP framework designed to overcome these challenges. The framework replaces physical signaling with automated, real-time WIP tracking enabled by Internet of Things (IoT) technologies and introduces a WIP harmonization method that standardizes diverse product configurations into a single reference unit for consistent system-wide control. A centralized CONWIP controller regulates job release based on the harmonized WIP level relative to a predefined cap. The feasibility of the proposed framework is demonstrated through a simulation of a wall assembly line in a panelized prefabrication factory in Edmonton, Canada. Performance evaluation using Factory Physics benchmarks shows that the framework effectively controls WIP, absorbs production variability, and achieves near-maximum throughput, constrained primarily by the system bottleneck. These results validate the framework's practicality and highlight the potential of digital twins to extend CONWIP to high-variability, space-constrained modular construction environments.

KEYWORDS

Modular construction, work-in-progress (WIP), CONWIP, digital twin, hybrid simulation.

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

² Graduate Research Assistant, Department of Civil and Environmental Engineering, University of Alberta, Canada, hamedsha@ualberta.ca

³ R&D Lead, Landmark Homes, Edmonton, Canada, haitaoy@landmarkgroup.ca

⁴ Assistant Professor, Department of Civil and Environmental Engineering, University of Alberta, Canada, qipei@ualberta.ca

⁵ 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

⁶ Professor, Department of Civil and Environmental Engineering, University of Alberta, Canada, hamzeh@ualberta.ca, orcid.org/0000-0002-3986-9534

INTRODUCTION

Work-in-Process (WIP) inventory refers to partially finished goods and materials at various stages of production, including items under transformation and components awaiting further processing (Hopp & Spearman, 2011). In manufacturing systems, WIP plays a dual role: it acts as a necessary buffer to absorb demand variability and production disruptions, while excessive levels degrade cash flow, operational efficiency, and system responsiveness (Hemalatha et al., 2021). The fundamental relationship between WIP, throughput (TH), and cycle time (CT) is formalized by Little's Law ($WIP = TH \times CT$), which provides a critical lens for understanding production performance and the consequences of inventory accumulation (Little, 1961). From a lean perspective, excess WIP constitutes one of the seven forms of waste (*muda*), representing capital tied up in incomplete work that masks underlying operational inefficiencies (Ohno, 2019). Accordingly, achieving an appropriate balance between insufficient WIP—which leads to starvation and underutilized capacity—and excessive WIP—which increases cycle times and hides systemic problems—remains a central challenge in production management (Hopp & Spearman, 2011). To address this challenge, a range of production control methods has been developed, broadly categorized as push, pull, and hybrid systems (Hirakawa et al., 1992). Traditional push systems, such as Material Requirements Planning (MRP), rely on demand forecasts and tend to drive higher WIP levels, longer lead times, and reduced responsiveness (Spearman & Zazanis, 1992). In contrast, pull-based systems authorize production based on downstream consumption and have consistently demonstrated superior performance in reducing WIP, cycle time variability, and inventory costs while sustaining or improving throughput and revealing hidden inefficiencies (Spearman & Zazanis, 1992). Empirical studies report WIP reductions of 38–67% and substantial improvements in value-added time and lead time through the application of pull principles in combination with lean tools such as Value Stream Mapping (VSM) (Belokar et al., 2012; Saraswat et al., 2015).

Among pull-based production control systems, Kanban and CONWIP (Constant Work-In-Process) are the two most widely studied and implemented approaches to WIP regulation (Pettersen & Segerstedt, 2009). Kanban systems, rooted in the just-in-time (JIT) philosophy, use product-specific visual or electronic signals at individual workstations to authorize production based on downstream consumption, thereby controlling local WIP levels for each product type (de Treville et al., 2023; Ohno, 2019; Soliman, 2020). In contrast, CONWIP employs a single set of generic authorization signals to enforce a global WIP cap across the entire production line, which simplifies implementation and adjustment in high-mix, low-volume environments where maintaining dedicated Kanban inventories for numerous product variants becomes impractical (Onyeocha et al., 2015). Comparative studies consistently show that CONWIP achieves higher throughput with fewer cards (lower WIP) than Kanban in serial production lines, although Kanban may outperform CONWIP in settings characterized by high setup times or highly variable product routings (Huang, 2011; Pettersen & Segerstedt, 2009). These characteristics make CONWIP particularly well suited to modular and off-site construction factories, where production is shifted from variable site conditions to controlled, sequential factory workflows with largely predictable routings (Björnfot & Stehn, 2004; Lennartsson et al., 2009). Although such facilities produce multiple product variants to meet project-specific requirements, their production processes remain fundamentally serial across configurations (Jensen et al., 2009). By enabling global WIP control without product-specific parameterization, CONWIP facilitates the management of engineered-to-order components while balancing customization and manufacturing predictability, making it more appropriate than Kanban for modular and off-site construction environments (Bataglin et al., 2025; Mohamad et al., 2013).

However, while CONWIP has proven effective in prefabrication environments involving small and relatively homogeneous elements (e.g., rebar and precast components) (Li et al.,

2018), its application to modular construction—particularly volumetric and panelized production—faces several structural challenges (Alsakka et al., 2022). Modular production lines routinely process WIP units that vary significantly in size, configuration, and labor content, causing traditional card-based caps to inadequately reflect true resource demand and buffer occupancy (Barkokebas et al., 2023; Mohamad et al., 2013). In addition, the large physical footprint of volumetric and panelized elements introduces severe space constraints that limit the feasibility of maintaining conventional buffer stocks or deploying visual signaling mechanisms along the line (Bataglin et al., 2025; Niu et al., 2019). These challenges are compounded by high product customization and stochastic project demand, which lead to frequent changes in WIP composition and production parameters, rendering static card-based control increasingly ineffective (Bataglin et al., 2025; Brown et al., 2019; Pourrahimian et al., 2023). Finally, the use of conveyors and automated transport systems in modular factories further constrains physical signaling and motivates the adoption of digital or hybrid production-control mechanisms (Gastermann et al., 2012).

To address these challenges, this study proposes a digital twin (DT)–enabled CONWIP production-control framework tailored to modular construction factories and developed following a Design Science Research approach. In this paper, the DT is defined at the control level as a digital state-representation layer that maintains system-wide WIP visibility through event-based updates from the production system. This digital representation supports two key control functions: harmonization of heterogeneous work units into a common reference unit and centralized CONWIP-based release control at the line entry point. Accordingly, the core contribution of the study is not a full cyber-physical implementation of sensing and communication infrastructure, but a formally specified and simulation-evaluated DT-enabled control artifact for WIP regulation in heterogeneous, space-constrained modular production. The proposed artifact is evaluated through a hybrid simulation case study of a wall assembly line in a penalized prefabrication factory in Edmonton, Canada. Simulation results demonstrate the framework’s effectiveness in regulating WIP and production flow, identify the critical WIP region, and benchmark system performance against Factory Physics best-case and practical worst-case curves (Hopp & Spearman, 2011).

The proposed framework is complementary to other digital lean approaches in construction, such as BIM-integrated planning, visualization, and production-monitoring systems. While such tools improve coordination, transparency, and decision support, they do not typically provide a direct mechanism for harmonized, system-wide WIP regulation under pull control (Moghimi et al., 2026). In contrast, the present study focuses specifically on digital WIP visibility and centralized CONWIP-based release control for heterogeneous, space-constrained modular production.

METHODOLOGY

This study adopts a Design Science Research (DSR) methodology (Hevner et al., 2004) to address work-in-process (WIP) management challenges in modular construction. Following the DSR logic outlined in the Introduction, the research proceeds through three stages: (1) problem identification, which establishes limitations of conventional CONWIP under heterogeneous and space-constrained production conditions; (2) artifact design, which specifies a digital twin–based CONWIP framework through formal control logic and reference-unit WIP harmonization; and (3) artifact evaluation, which assesses the framework through hybrid simulation and Factory Physics benchmarking. This section focuses on the second stage by detailing the design and operational specification of the proposed artifact.

From a design-science perspective, the identified problem implies three main artifact requirements. First, the control approach must provide system-wide WIP visibility without relying on physical cards or local visual signals, which are difficult to maintain in modular

factories with conveyors, limited floor space, and heterogeneous products. Second, it must represent diverse work units on a common control basis so that the WIP cap reflects actual production effort rather than raw component counts. Third, it must support centralized pull-based release decisions that remain operational under stochastic, space-constrained production conditions. These requirements directly motivate the main design choices of the proposed artifact: an event-based digital state layer for WIP tracking, reference-unit harmonization for heterogeneous jobs, and a centralized CONWIP release mechanism based on total harmonized WIP.

The present study reports an initial design and simulation-based assessment of this artifact in a representative modular production context. The artifact development process moved from a broad concept of digitally enabled pull control toward a more focused control-oriented specification tailored to the identified problem. Evaluation through hybrid simulation and benchmark comparison was used to assess the adequacy of this refined artifact scope and its behavioural performance under realistic variability, rather than to document multiple formal design–evaluation cycles. Figure 1 illustrates the structure of the DT-based CONWIP framework and the interaction between the production system, the sensing layer, and the centralized CONWIP controller. The figure should be read from left to right: segment-level WIP states are first captured in native units, then converted into harmonized reference units, aggregated into total system WIP, and finally used by the centralized CONWIP controller to determine whether the next job can be released.

The production system is modelled as a sequence of processing segments $S_1, \dots, S_n$, each associated with a local work-in-process (WIP) level. In this conceptual design, the sensing layer provides discrete, event-based WIP states $L_i(t)$ at each segment, corresponding to job arrivals, departures, and in-process status updates. Such data can be obtained, for example, from fixed-position computer vision systems for panel counting, RFID-based job identification at station entry and exit points, or manufacturing execution system (MES) event logs. Although different sensing technologies (e.g., computer vision versus LiDAR) generate distinct raw data streams, this study abstracts from low-level signal processing and data fusion, and assumes that sensor outputs are preprocessed into consistent, segment-level WIP counts in their native units. In the intended implementation, each WIP update is triggered by timestamped production events such as job entry into a segment, job exit from a segment, or a change in in-process status. These events are mapped to a common system clock and synchronized before aggregation so that the controller always evaluates release decisions using the most recent consistent snapshot of segment-level WIP states. Thus, reproducibility at the control level depends on the availability of event-based state updates and a consistent synchronization layer, rather than on any specific sensing hardware.

In the present study, the digital twin should be understood as a control-oriented digital representation of production state rather than as a full hardware-level replica of the factory. Its role is to maintain a synchronized, segment-level view of current WIP, convert heterogeneous local units into harmonized reference units, and provide the state information required by the centralized CONWIP controller. Thus, what is explicitly modelled is the evolution of digital WIP states and their use in release control, whereas sensing hardware, low-level signal processing, communication latency, and synchronization implementation are treated as enabling assumptions rather than detailed modelling elements.

Because production segments may operate in different native units (e.g., large modules upstream and small wall panels downstream), the sensed WIP values $L_i(t)$ are converted into reference units using segment-specific conversion factors α_i and then aggregated to compute the total harmonized system WIP $L^*(t)$. Here, α_i represents the conversion factor between segment i 's native unit and the chosen reference unit. The harmonized WIP values are summed to obtain the total system WIP. The CONWIP controller then determines whether the next job

can be released by comparing the current harmonized WIP, augmented by the reference-unit contribution $\omega(j^*)$ of the candidate job, with the predefined CONWIP limit W . If the condition $L^*(t) + \omega(j^*) \leq W$ is satisfied, the job is released into the production system; otherwise, release is withheld until sufficient WIP has exited the system.

The following formulations define the operational logic of the proposed DT-based CONWIP artifact and should be interpreted as a formal design specification of the control artifact, rather than as an analytical or optimization model intended for solution or parameter optimization. Let $L_i(t)$ denote the time-varying WIP count at segment i in its native unit, and let α_i be the conversion factor from the native unit of segment i to the reference unit (expressed as reference-units per native unit). The harmonized WIP count at segment i is:

$$L_i^*(t) = \alpha_i \cdot L_i(t) \quad (1)$$

$$\text{and the systemwide harmonized WIP is: } L^*(t) = \sum_{i=1}^n L_i^*(t) \quad (2)$$

When a deterministic bill of materials applies, α_i is constant; under mixed-model variability, α_i is estimated online (e.g., via exponentially weighted moving average over the most recent K job releases) to account for design-driven variations in component requirements.

Following WIP counting at time t , the CONWIP controller evaluates whether the next job can be released without violating the WIP cap. Let $\omega(j)$ denote the reference-unit contribution that job j introduces upon release (i.e., the equivalent reference units implied by its bill of materials or product specification). The controller releases job j^* if and only if:

$$L^*(t) + \omega(j^*) \leq W \quad (3)$$

where W is the CONWIP limit expressed in reference units. If the inequality holds, release job j^* ; otherwise, hold release. This mechanism prevents transient violations of the WIP cap when jobs have heterogeneous, unit-mismatched sizes, ensuring global production synchronization despite local unit variability.

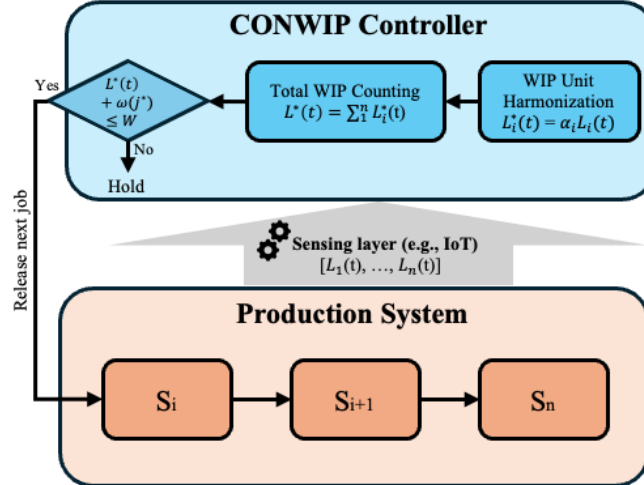


Figure 1: Schematic view of the DT-based CONWIP framework structure.

FRAMEWORK APPLICATION

To validate the proposed DT-based CONWIP framework, the approach is applied to a hybrid simulation of a wall assembly line in a prefabricated timber modular construction facility in Edmonton, Canada, which produces volumetric residential modules as well as wall, roof, floor, and finishing components. The wall line was selected because it exhibits the core challenges motivating the framework, including unit-of-work heterogeneity, severe space constraints, and high product customization. Multi-wall (MW) frames enter the system as large units and are later subdivided into smaller panels, creating variable WIP units that complicate traditional card-based CONWIP control, while buffer capacities are tightly constrained—often limited to

single panels—leaving little opportunity for visual signaling or physical cards. Specification-driven processing-time variability further necessitates dynamic rather than static WIP parameterization. The modeled production line consists of six interconnected stages (Figure 2): Framing, Sheeting, CNC Nailing, Packaging, Doors & Windows (DW) installation, and Loading. Framing employs a single CNC station and crew to produce MW frames, which advance through a one-panel buffer to Sheeting, where two crews operate on a two-panel conveyor. After a one-panel buffer, panels enter CNC Nailing with single-panel capacity, before Packaging subdivides MWs into smaller panels on a space-constrained table modeled as single-panel capacity. Interior panels requiring wrap are routed to inventory storage, while remaining panels proceed to one of four parallel DW lanes, each with four-panel buffer capacity and staffed by specialized crews. Finished panels are transferred via a single-capacity cart to a loading area comprising six lanes with four-panel capacity each and operated by two crews using a single-capacity loading mechanism. This configuration—characterized by heterogeneous work units, tight buffers, pooled resources, and shifting bottlenecks—provides a rigorous testbed for evaluating whether harmonized WIP accounting and cardless CONWIP release can maintain global flow control under local space constraints.

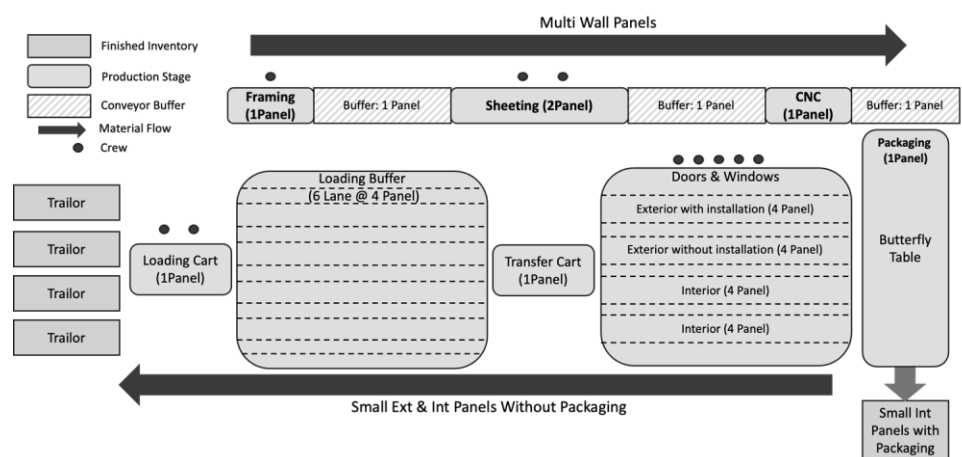


Figure 2: Schematic representation of the wall assembly line configuration.

MODEL DEVELOPMENT

HYBRID DISCRETE-EVENT AND AGENT-BASED SIMULATION

The simulation model was developed in Python using SimPy as a hybrid framework that integrates discrete-event simulation (DES) and agent-based simulation (ABS). This approach combines process-flow modeling with adaptive workforce behavior, enabling realistic representation of production dynamics under variability. Python was selected for its flexibility in hybrid modeling and its compatibility with analytical libraries (e.g., pandas, numpy) supporting reproducible Monte Carlo analysis and future model extensions (Moghimi & Kashani, 2026). The DES layer captures material flow, routing logic, resource contention, and buffer constraints using capacity-limited SimPy Store objects that enforce FIFO discipline. Single-capacity stores represent transfer elements, while multi-capacity stores model buffers (e.g., Doors & Windows and Loading). Blocking and starvation emerge endogenously as downstream constraints restrict upstream flow. Dynamic routing at the Packaging stage assigns panels to downstream processes based on panel type and queue states, replicating observed load-balancing behavior. The ABS layer represents workforce dynamics and adaptive task prioritization. Crew collaboration effects at Sheeting are modeled through diminishing-return productivity multipliers, while at Doors & Windows, hierarchical rules govern crew allocation, job prioritization, and helper assignment under capacity constraints. Preemption logic enables

dynamic reassignment of crews to higher-priority tasks, including material handling by the transfer cart, while loading operations follow global FIFO sequencing with time adjustments reflecting trailer-filling logistics. All behavioral parameters were calibrated using field observations to ensure fidelity to real factory operations.

PROCESSING TIME CHARACTERIZATION AND DISTRIBUTION FITTING

Processing time data were collected through field observations conducted between July and September 2025, yielding approximately 100–150 panel records per station. For each panel, processing duration was recorded along with key job attributes (e.g., interior/exterior classification and installation requirements) and selected contextual factors affecting variability, such as temporary crew interference, local congestion, material handling interruptions, and short waiting periods caused by upstream or downstream conditions. These data were used to parameterize station-level processing behaviour in the simulation model. Analysis of the collected data revealed distinct variability drivers across stations. Framing exhibited relatively uniform processing behaviour, meaning that observed durations showed limited sensitivity to panel specification compared with downstream stations and could therefore be represented adequately using a single distribution. In contrast, Sheeting and CNC Nailing required separate distributions for interior and exterior panels, as exterior multiwall (MW) panels involved additional handling and finishing activities; mixed MW assemblies containing any exterior panel were classified as exterior to reflect operational practice. Packaging times varied by routing option (storage, Doors & Windows installation, or both), while Doors & Windows processing depended on panel type and installation requirements. Transfer Cart operations followed a single distribution, whereas Loading times reflected variability associated with trailer-filling logistics. Triangular distributions were adopted to model processing times due to their suitability for data-limited environments and minimal parameter requirements (AbouRizk et al., 1994; Law et al., 2007). Distribution parameters were derived from empirical data and validated through consultation with production-floor experts to ensure consistency with observed operations.

MODEL VALIDATION AND TEST PRODUCTION SCHEDULE

The simulation model was validated to ensure behavioral fidelity to actual factory operations through a combination of qualitative and quantitative methods. Qualitative validation included structured walkthroughs with production managers to confirm logical correctness and event-trace analyses to verify the emergence of blocking, starvation, and crew preemption behaviors across all stages. Quantitative validation was performed using real production data comprising 84 multiwall (MW) panels and 135 small panels from representative runs. The model achieved strong agreement with observed performance at key stations, with coefficients of determination of $R^2 = 0.95$ for framing, $R^2 = 0.89$ for Doors & Windows, $R^2 = 0.88$ for Loading, and mean absolute percentage errors between 16–18%, which are acceptable for complex construction production simulations (Law et al., 2007; Sargent, 2010). Minor negative biases were attributed to unmodeled micro-delays inherent in factory operations. Overall, the validated model reliably reproduces production dynamics, variability, and resource contention. To evaluate CONWIP implementation, a production schedule of 80 MW panels with varying specifications was derived from actual plant scheduling data. The schedule reflects realistic design diversity, including interior and exterior panels, different packaging requirements, and installation specifications. This heterogeneous mix captures the unit-of-work variability and customization complexity typical of modular construction and provides a representative basis for assessing the performance of the proposed DT-based CONWIP framework.

CONWIP FRAMEWORK IMPLEMENTATION

The DT-based CONWIP control approach was implemented in the simulation model following the artifact design described above. The reference unit for WIP tracking was defined as the "small panel" that remains in the main production line until reaching the trailer loading station. Each simulation run commences with a queue of multiwall (MW) panels behind the Framing station, representing the job release interface for the CONWIP controller. As each multiwall (MW) enters the system, its specific configuration—including the number of constituent small panels and job specifications (e.g., interior/exterior, packaging requirements)—is recognized. Processing times for all operations at each station are sampled from the appropriate station- and specification-specific distributions, as previously parameterized. Throughout the simulation, the total system WIP (expressed in small-panel reference units) is continuously tracked based on the instantaneous state of jobs undergoing processing and those waiting in downstream buffers. This simulated WIP state represents the information availability that would be provided by real-time sensing and data acquisition in an actual factory, rather than explicitly modeling the underlying data collection, communication, or synchronization mechanisms. The core WIP feedback and job release logic follows the formulas defined in methodology section. At any simulation time t : $L'(t)$ = total system WIP (in small panel equivalents), $\omega(j^*)$ = small panel count for the next MW to release, W = specified CONWIP cap (in small panel equivalents). Release job j^* into the system therefore happens based on equation 3 in the methodology.

PERFORMANCE METRICS AND BENCHMARK ANALYSIS

To evaluate the framework's effectiveness, the simulation model calculates key performance indicators (KPIs) and compares them against theoretical benchmarks derived from factory physics principles (Hopp & Spearman, 2011). Because the system is characterized by stochastic processing times and a heterogeneous job mix, a single simulation run is insufficient to determine performance. Therefore, a Monte Carlo approach is employed, where multiple independent replications of the simulation are executed for each CONWIP level. This method allows for the calculation of expected values and confidence intervals for each KPI, providing statistically robust insights into system behavior under uncertainty (Law et al., 2007; Moghimi & Kashani, 2025).

The primary performance metrics, averaged over all Monte Carlo replications, are:

- **Throughput (TH):** The rate at which finished small panels exit the system, calculated as the total number of panels loaded onto trailers divided by the simulation makespan.

$$TH = \frac{N_{exit}}{T_{makespan}} \quad (4)$$

- **Cycle Time (CT):** The **average** time a small panel spends in the system, measured from its parent MW's release into Framing until the panel is loaded onto a trailer.

$$CT = \frac{1}{N_{exit}} \sum_{i=1}^{N_{exit}} (t_{exit,i} - t_{entry,i}) \quad (5)$$

- **Work-in-Process (WIP):** The average number of units present in the system at any given time, defined as the time integral of the instantaneous WIP divided by the makespan:

$$WIP = \frac{1}{T_{makespan}} \int_0^{T_{makespan}} L'(t) dt \quad (6)$$

where $L'(t)$ is the total harmonized WIP (in small panel equivalents) present in the system at time t .

To contextualize these results, the system's performance is benchmarked against the Best Case and Practical Worst Case (PWC) scenarios as defined by Hopp & Spearman (2011). This requires estimating two fundamental line parameters from the simulation:

1. **Bottleneck Rate (r_b):** The maximum possible throughput of the system, estimated by running the simulation with a relaxed (very high) CONWIP limit and measuring the resulting throughput.
2. **Raw Process Time (T_0):** The theoretical minimum time a job could spend in the system if it never had to wait in a queue. This is estimated by running the simulation with a low CONWIP limit and measuring the average cycle time.

By using these parameters, the critical WIP level is calculated as $W_0 = r_b \times T_0$. This value distinguishes two fundamental performance regimes of the Best-Case production line (Hopp & Spearman, 2011). When WIP is below W_0 , the line operates in the flow-limited regime, where throughput increases proportionally with WIP because the bottleneck station is not always starved. Conversely, when WIP exceeds W_0 , the line enters the rate-limited regime, where throughput plateaus at the maximum bottleneck capacity r_b , and further WIP increases only add to cycle time without improving flow. The critical WIP thus represents the minimal WIP level required to keep the bottleneck fully utilized; it is the knee of the throughput curve, demarcating operational efficiency from wasteful excess inventory.

The benchmark formulas are then applied using the observed average WIP from the Monte Carlo simulation for a given CONWIP level. These benchmarks provide theoretical performance bounds under best-case and practical conditions, enabling rigorous evaluation of the CONWIP framework's effectiveness.

Best Case (BC) Benchmarks: The Best Case assumes zero variability and no blocking; throughput and cycle time depend on the WIP level relative to the critical threshold:

$$TH_{BC}(WIP) = \begin{cases} \frac{WIP}{T_0} & \text{if } WIP \leq W_0 \\ r_b & \text{if } WIP > W_0 \end{cases}, \quad CT_{BC}(WIP) = \begin{cases} T_0 & \text{if } WIP \leq W_0 \\ \frac{WIP}{r_b} & \text{if } WIP > W_0 \end{cases} \quad (7)$$

In the flow-limited regime ($WIP \leq W_0$), cycle time equals raw process time T_0 , and throughput grows linearly with WIP. In the rate-limited regime ($WIP > W_0$), throughput saturates at r_b , while cycle time grows linearly with excess WIP.

Practical Worst Case (PWC) Benchmark: The Practical Worst-Case accounts for realistic stochasticity and production variability while assuming serial processing with no product overlap beyond the bottleneck:

$$CT_{PWC}(WIP) = T_0 + \frac{WIP-1}{r_b} \quad (8)$$

This formula reflects that each unit beyond the first incurs an average delay of $1/r_b$ at the bottleneck. The corresponding throughput is:

$$TH_{PWC}(WIP) = \frac{r_b}{r_b(T_0 + \frac{1}{r_b})} = \frac{WIP}{T_0 + \frac{WIP-1}{r_b}} \quad (9)$$

Benchmark evaluation offers several key benefits for assessing the CONWIP framework in modular construction. First, performance contextualization places simulated system metrics within theoretical bounds; results observed between the Best Case and Practical Worst Case confirm both model realism and feasible operational behavior, while large gaps highlight where the CONWIP policy or system layout may limit throughput. Efficiency diagnosis then uses benchmark comparison to distinguish whether underperformance arises from intrinsic line design features or from a suboptimal CONWIP implementation—if simulated results consistently fall below the Practical Worst Case, it signals that the control policy may not be well-suited for the modular context. Critical WIP identification is facilitated by analyzing performance across different CONWIP settings; benchmark comparison helps pinpoint the optimal WIP cap that efficiently balances throughput, cycle time, and inventory—addressing a

core challenge in space-constrained modular construction. Finally, sensitivity and variability assessment uses these benchmarks to gauge the degree of performance loss due to stochastic factors and customization, yielding practical guidance for targeted process improvements and parameter adjustment.

RESULTS AND SYSTEM PERFORMANCE ANALYSIS

From a Design Science Research perspective, this section evaluates the performance of the proposed DT-based CONWIP artifact under varying WIP caps using Monte Carlo simulation and Factory Physics benchmarks. In this study, the digital twin is modelled at the control and state-representation level: it provides timely digital visibility of segment-level WIP states used to harmonize system inventory and drive release decisions. Communication latency, hardware-specific sensing, and sensor-fusion processes are abstracted from. The evaluation therefore focuses on the effectiveness of the DT-enabled control logic rather than on detailed implementation of cyber-physical infrastructure.

The framework was evaluated by sweeping the CONWIP cap from minimal values to levels well beyond the predicted critical WIP. For each cap, 50 independent Monte Carlo replications were performed, which were sufficient to obtain stable estimates of mean throughput and cycle time (Law et al., 2007). The resulting throughput–cycle time relationships were benchmarked against Best-Case (BC) and Practical Worst-Case (PWC) curves (Figure 3). As expected, throughput increases nearly linearly at low WIP levels before saturating at the bottleneck rate (≈ 4.2 panels/hour), while cycle time remains near its minimum around the critical WIP region and then increases rapidly with excess inventory. The simulated curves closely follow the BC envelope, indicating that variability is effectively absorbed and that system performance is dominated by bottleneck capacity rather than diffuse stochastic effects.

Two structural insights emerge from the results. First, strict inter-station buffer limits cap average system WIP, causing excess jobs to queue at entry rather than accumulate on the floor once local buffers are full. Second, the rapid convergence toward BC behavior indicates that one or two dominant bottlenecks govern system performance, such that additional WIP beyond the critical level yields no throughput gains and only increases cycle time. Based on these findings, a CONWIP range of 7–9 reference units (≈ 8 preferred) is recommended for day-to-day control, as it achieves near-maximum throughput while limiting congestion and cycle time. Sustained output increases under high demand conditions require debottlenecking the dominant rate constraint—currently the transfer cart/loading segment—and targeted variability reduction at high-variance stations. Beyond demonstrating these principles, the results illustrate how the DT-based CONWIP framework can support continuous performance monitoring and adaptive decision-making through ongoing comparison with Factory Physics benchmarks.

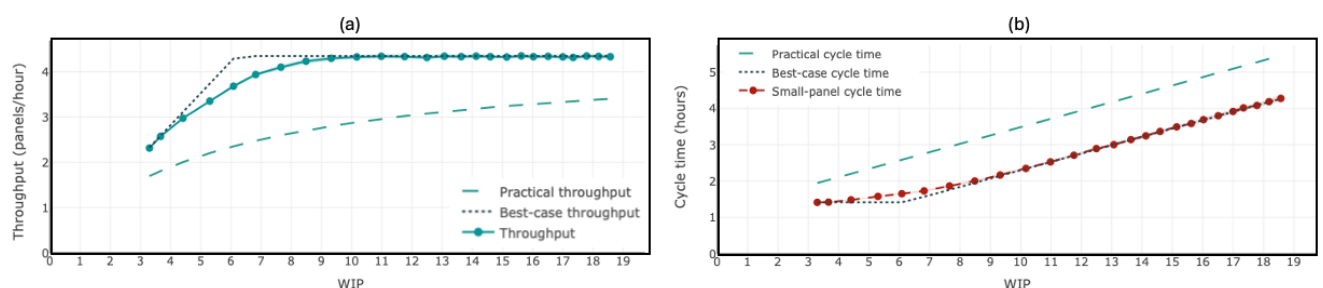


Figure 3: CONWIP-controlled system performance: (a) throughput and (b) cycle time versus WIP compared to theoretical benchmarks.

CONCLUSION

This study proposed a conceptual digital twin (DT)–based CONWIP framework to address work-in-process (WIP) management challenges in modular construction environments characterized by heterogeneous products and severe space constraints. By substituting conventional card-based signaling with real-time system-state visibility and a reference-unit WIP harmonization approach, the framework overcomes key barriers that limit the practical application of CONWIP in modular and panelized production systems. Using a Design Science Research approach, the framework was developed as a prescriptive production-control artifact and evaluated through a hybrid simulation case study of a wall assembly line. The results demonstrated the framework’s ability to enforce global WIP caps, regulate production flow, and achieve near bottleneck-limited throughput while avoiding the accumulation of non-value-adding WIP. Benchmarking against Factory Physics best-case and practical worst-case curves confirmed that the system operates as a bottleneck-limited line and that additional WIP beyond the critical threshold yields diminishing performance gains, consistent with lean production principles. In this study, the digital twin was modeled at the control and state-representation level, assuming timely and accurate observability of system WIP in order to evaluate the effectiveness of the proposed CONWIP logic. Communication latency, sensor-specific data fusion, and synchronization delays were intentionally abstracted from to isolate the impact of the control policy itself. Under this DT emulation setting, the results demonstrate that pull-based production control can be effectively implemented in modular construction without reliance on physical signaling mechanisms, provided that reliable system-state information is available. This research offers a practical, simulation-validated pathway for implementing digital, pull-based production control in factory-based construction, bridging lean manufacturing theory with the operational realities of modular construction. Future work will focus on empirical validation in industrial settings, extension to multi-line and multi-product systems, incorporation of production disruptions and imperfect observability, and integration of predictive and adaptive control mechanisms to further enhance the robustness and scalability of DT-enabled CONWIP systems.

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

- AbouRizk, S. M., Halpin, D. W., & Sawhney, A. (1994). Modelling Uncertainty in Construction Simulation. *MACHINE INTELLIGENCE AND PATTERN RECOGNITION*, 17, 287.
- Alsakka, F., Darwish, M., Yu, H., Hamzeh, F., & Al-Hussein, M. (2022). The impacts of lean implementation revealed in the course of building a digital twin of a construction manufacturing facility. *Proc. 30th Annual Conference of the International Group for Lean Construction (IGLC), Emonton, Canada*, 890–901.
- Barkokebas, B., Al-Hussein, M., & Hamzeh, F. (2023). Assessment of digital twins to reassign multiskilled workers in offsite construction based on lean thinking. *Journal of Construction Engineering and Management*, 149(1), 04022143.
- Bataglin, F. S., Formoso, C. T., & Viana, D. D. (2025). Model for Production System Design in Modular Construction Projects. In O. Seppänen, L. Koskela, & K. Murata (Eds.),

- Proceedings of the 33rd Annual Conference of the International Group for Lean Construction (IGLC 33)* (pp. 388–399). <https://doi.org/10.24928/2025/0165>
- Belokar, R. M., Kumar, V., & Kharb, S. S. (2012). An application of value stream mapping in automotive industry: a case study. *International Journal of Innovative Technology and Exploring Engineering*, 1(2), 152–157.
- Björnfot, A., & Stehn, L. (2004). Industrialization of construction: a lean modular approach. *Annual Conference of the International Group for Lean Construction: 03/08/2004-05/08/2004*.
- Brown, R., Barkokebas, B., Ritter, C., & Al-Hussein, M. (2019). Predicting performance indicators using BIM and simulation for a wall assembly line. *Proceedings of the 27th Annual Conference of the International Group for Lean Construction (IGLC 2019), Dublin, Ireland*, 3–5.
- de Treville, S., Browning, T. R., Marewski, J. N., & Weiss, J. (2023). : Toyota Production System practices as Fast-and-Frugal heuristics. In *Journal of Operations Management* (Vol. 69, Number 4, pp. 522–535). Wiley Periodicals, Inc. Boston, USA.
- Gastermann, B. C., Stopper, M., & Katalinic, B. (2012). Adapting conwip characteristics for conventional production planning. *DAAAM International Scientific Book, 2012*, 553–564.
- Hemalatha, C., Sankaranarayanan, K., & Durairaj, N. (2021). Lean and agile manufacturing for work-in-process (WIP) control. *Materials Today: Proceedings*, 46, 10334–10338.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 75–105.
- Hirakawa, Y., Hoshino, K., & Katayama, H. (1992). A Hybrid Push/Pull Production Control System for Multistage Manufacturing Processes. *International Journal of Operations & Production Management*, 12(4), 69–81.
- Hopp, W. J., & Spearman, M. L. (2011). *Factory physics*. Waveland Press.
- Huang, Y. (2011). *A comparative study on kanban, CONWIP and hybrid systems with illustration of a case study*. The University of Texas at San Antonio.
- Jensen, P., Hamon, E., & Olofsson, T. (2009). Product development through lean design and modularization principles. *Annual Conference of the International Group for Lean Construction: 13/07/2009-19/07/2009*, 465–474.
- Law, A. M., Kelton, W. D., & Kelton, W. D. (2007). *Simulation modeling and analysis* (Vol. 3). McGraw-hill New York.
- Lennartsson, M., Björnfot, A., & Stehn, L. (2009). Production control through modularisation. *Annual Conference of the International Group for Lean Construction: 13/07/2009-19/07/2009*, 453–464.
- Li, X., Li, Z., & Wu, G. (2018). Lean precast production system based on the CONWIP method. *KSCE Journal of Civil Engineering*, 22(7), 2167–2177.
- Little, J. D. C. (1961). A proof for the queuing formula: $L = \lambda W$. *Operations Research*, 9(3), 383–387.
- 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>

- Mohamad, A., Hickethier, G., Hovestadt, V., & Gehbauer, F. (2013). Use of modularization in design as a strategy to reduce component variety in one-off projects. *21th Conference of the International Group of Lean Construction (IGLC-21), Fortaleza, Brazil*.
- Niu, S., Yang, Y., & Pan, W. (2019). Logistics planning and visualization of modular integrated construction projects based on BIM-GIS integration and vehicle routing algorithm. *Modular and Offsite Construction (MOC) Summit Proceedings*, 579–586.
- Ohno, T. (2019). *Toyota production system: beyond large-scale production*. Productivity press.
- Onyeocha, C. E., Wang, J., Khoury, J., & Geraghty, J. (2015). A comparison of HK-CONWIP and BK-CONWIP control strategies in a multi-product manufacturing system. *Operations Research Perspectives*, 2, 137–149.
- Pettersen, J.-A., & Segerstedt, A. (2009). Restricted work-in-process: A study of differences between Kanban and CONWIP. *International Journal of Production Economics*, 118(1), 199–207.
- Pourrahimian, E., Shehab, L., & Hamzeh, F. (2023). Investigating and Simulating Collaboration among the Lps Phases. *31st Annual Conference of the International Group for Lean Construction (IGLC31)*, 929–941.
- Saraswat, P., Kumar, D., & Sain, M. K. (2015). Reduction of work in process inventory and production lead time in a bearing industry using value stream mapping tool. *International Journal of Managing Value and Supply Chains (IJMVSC)*, 6(2), 27–35.
- Sargent, R. G. (2010). Verification and validation of simulation models. *Proceedings of the 2010 Winter Simulation Conference*, 166–183.
- Soliman, M. H. A. (2020). *Kanban the Toyota Way: An Inventory Buffering System to Eliminate Inventory*. Mohammed Hamed Ahmed Soliman.
- Spearman, M. L., & Zazanis, M. A. (1992). Push and pull production systems: issues and comparisons. *Operations Research*, 40(3), 521–532.