

A flow-generator platform to assess a data-driven design of material handling and production systems

I. Battarra*, R. Accorsi, G. Lupi, R. Manzini, G. Sirri

Alma Mater Studiorum - University of Bologna, Bologna, Italy

(*ilaria.battarra2@unibo.it, riccardo.accorsi2@unibo.it, giacomo.lupi2@unibo.it,
riccardo.manzini@unibo.it, gabriele.sirri4@unibo.it)

Abstract. This study introduces a flow-generator platform designed to enable data-driven approaches, e.g., models, decision support systems, and digital twins, in production and storage systems. It aims to address the challenges of providing realistic operational data when actual data are unavailable as in greenfield scenarios. The platform simulates inbound and outbound material flows by capturing variability in production rates, shipping schedules, and storage processes. Users configure parameters such as production line capacity, items diversity, batch sizes, and material handling strategies through a visual dashboard, facilitating detailed analysis of operational production peaks, trends, and variabilities. By enhancing decision-making processes through robust and realistic data, the platform complements simulators, digital twins, and advanced modeling methods. A numerical example set in the food processing industry demonstrates its applicability, showcasing its ability to support robust performance modeling and optimization tailored to dynamic supply chain environments.

Copyright © 2025 The Authors. This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: Data-driven; Data-driven model; Digital platform; Data handling; Supply Chain; Industrial ecosystem; Industry 5.0.

1. INTRODUCTION

Supply chain networks interconnect different actors such as suppliers, manufacturers, logistics providers, distributors, and retailers, all working collaboratively to move materials and goods from production sites to final consumers. Each actor within the chain plays a unique role that contributes to the efficient flow of materials, products, and related information. In this complex ecosystem, data has emerged as a strategic asset, enabling each actor to make informed real-time decisions and enhance their responsiveness to market shifts (Ghadge et al., 2020). The evolution towards Industry 5.0 further emphasizes the role of data by integrating human-centric technologies and collaborative systems that complement advanced automation.

Data-driven modelling approaches facilitate seamless communication across supply chain nodes, offering insights that help decision makers to mitigate risks, anticipate disruptions and optimize the flow of goods. The strategic use of data extends beyond operational efficiency enabling the generation of alternative supply chain scenarios and identifying potential bottlenecks (Lu et al. 2020). Additionally, it supports the implementation of proactive solutions to enhance resilience, adaptability (Kamble et al. 2022; Singh et al. 2023) and sustainability (Samayamantri and Vaddy 2024). Nevertheless, many supply chains face challenges obtaining accurate, real-world data to fuel data-driven modelling approaches. These challenges arise for several reasons, including issues such as data fragmentation, lack of standardization, and difficulties in integrating information from diverse sources, and, in the case of greenfield projects, the complete absence of data.

As Industry 5.0 technologies, such as digital twins and artificial intelligence, become integral, they can help address these gaps by enabling virtual simulations and predictive models that mirror physical operations in real time. By generating synthetic data and simulating various scenarios, these technologies provide valuable insights even in contexts where historical data is unavailable.

Recent studies show a growing interest in leveraging digital twins and artificial intelligence to address data challenges in supply chain management. Jackson et al. (2024) showed how simulation models of logistics systems can be produced automatically from verbal descriptions in natural language and how human experts and artificial intelligence (AI)-based systems can collaborate in the simulation modelling domain. Aron et al. (2024) explored how cyber-physical systems can enhance interoperability in logistics networks by leveraging cloud computing, machine learning, and real-time data to dynamically allocate material handling resources, improving efficiency, resilience, and sustainability.

A key area where these advancements are particularly relevant concerns the automated production and storage systems, where real-time data processing and dynamic resource allocation play a crucial role in optimizing operations (Sell et al. 2022, Pratik et al. 2023, Lupi et al. 2024, Cimino et al. 2024).

This study addresses this gap by proposing a flow-generator platform specifically designed to generate realistic operational data across different production and storing systems. By capturing actual operational patterns, such as seasonal demand spikes, batch production variability, and operational fluctuations, the platform generates a list of inbound and

outbound tasks, enabling a realistic representation of complex industrial ecosystems when actual data are unavailable, as happens in greenfield scenarios. Utilizing an intuitive interface, users can configure critical parameters—such as the number of production or assembly lines, number of products, set-up times, or shipping rates—to model inbound and outbound material flows such as storage and retrieval tasks in a storage system. The output is a chronologically sorted task list that integrates seamlessly with PowerBI dashboards, facilitating data analysis and visualization to identify operational peaks, trend fluctuations, and performance insights.

The original contribution of this platform lies not only in its ability to generate synthetic yet realistic data but also in its scalability and flexibility. The platform is designed to integrate diverse options and customizable rules, allowing users to tailor it to their specific operational needs. This high level of customization ensures that the generated data accurately reflects the unique characteristics of different production and storage environments. By enabling a dynamic and user-driven configuration, the platform enhances the development of robust data-driven performance models, optimization methods, and decision support systems tailored to the evolving complexities of modern supply chains.

The remainder of this study is described as follows: Section 2 illustrates the proposed methodology, and three user interfaces adopted for setting production and shipping parameters. Section 3 introduces a numerical example that aims to replicate the flows within a facility operating in the pre-packed food industry. Finally, Section 4 presents some conclusions and outlines for further research.

2. METHODOLOGY

This section introduces the methodological approach behind the development of the proposed flow-generator platform designed to simulate materials flow in complex industrial ecosystems. Figure 1 illustrates the two following key sections: Parameters setting and Data visualization.

The parameter settings enable users to configure production line features, set-up times, shipping rates, and warehousing characteristics, ensuring a tailored representation of all production and storage processes. The aim is to replicate the flow of materials across production/assembly, warehousing, and shipping activities, focusing on accurately representing operational patterns rather than designing or dimensioning the system, thus providing a realistic dataset for analysis and optimization.

The platform consists of three user-friendly graphical user interfaces (GUIs): two of them are designed to capture essential parameters related to production line and warehouse (Production GUI), while the third grabs the parameters related to the shipping activity and trucks capacity (Shipping GUI). The Data visualization section integrates the generated data into a visual dashboard, providing insights into operational peaks, trends, and performance indicators.

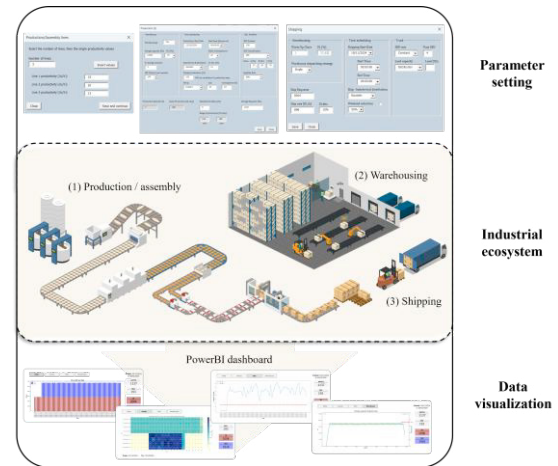


Figure 1. Methodological framework.

Figure 2 illustrates the Production (1) GUI used to collect the parameters of the production or assembly lines.

Figure 2. Production (1) GUI

At the top, the Number of Lines indicates how many lines are operating in the analyzed industrial environment. The form dynamically adjusts the input fields for line productivity based on the Number of Lines value, allowing the user to set the production rate (UL/h) for each line. These initial values influence batch sizes and the production or assembly rate. Figure 2 illustrates an example of a production plant with three lines. Additional key production parameters are configured through the Production (2) GUI, depicted in Figure 3.

Figure 3. Production (2) GUI.

This interface provides a detailed framework for managing warehousing operations, time scheduling, and the SKU

portfolio. The GUI displays the Production Rate of the system, representing the cumulative throughput of all operational lines. This metric sets the speed at which goods are produced. The Standard Deviation parameter enables users to introduce variability into production rates to effectively simulate real-world variations due to factors such as scheduled or unscheduled maintenance and minor delays. Lastly, the Production Requests defines the total number of unit loads (ULs) generated by the production or assembly lines. This value indicates the total amount of work to be managed, influencing the overall duration of activities and the scheduling of necessary resources.

In addition to these primary parameters, three additional fields must be populated to further refine the system's operations and task scheduling. In the Warehousing section, users define key storage parameters relevant to industrial systems with integrated storage. The Storage Capacity specifies the maximum number of ULs the storage system can accommodate, while the Initial Fill [%] represents the warehouse's initial occupancy rate, simulating a realistic starting condition with pre-existing stock. The Number of Storage Modules divides the storage area into distinct sections, such as aisles serviced by dedicated shuttles or cranes, enabling modular storage strategies tailored to proximity or varying throughput rates. Lastly, the Material Handling Vehicle Load Capacity determines the capacity of vehicles transporting items from production lines to the warehouse, with options of 1, 2, or 4 ULs, directly influencing restocking efficiency and overall turnover.

In the Time Scheduling section, users define timelines and distributions for production-related events. The Production Start Date and Time sets the beginning of all activities, ensuring synchronization and enabling precise tracking of time-based metrics. The Daily Working Hours parameter, often set to 24 hours (as shown in Figure 3), reflects continuous operations typical of high-demand industries such as food processing, B2C, or third-party logistics, allowing simulations to explore the strain on warehousing and handling systems in a round-the-clock environment. The Prod-InterArrival Distribution models the statistical timing between UL arrivals at the end of the production lines, such as automated packing systems. Users can choose from Gaussian (normal), exponential, or constant distributions to simulate real-world production variability, with Gaussian reflecting regular deviations, exponential capturing irregular bursts, and constant representing stable intervals. Additionally, the Set-Up Time parameter defines the duration required to switch between SKUs on a production line, while the Working Time specifies the interval between set-ups. These parameters can follow fixed durations or probability distributions. For instance, a line may operate for 2 to 8 hours (working time) with varying probabilities, followed by set-ups lasting 2 to 35 minutes, as illustrated in Table 1.

Working time [h]	Probability	Set-up [min]
2	50%	2
4	10%	12
6	22%	25
8	18%	35

Table 1. Probability distribution of set-up and working times.

This flexibility reflects operational variability and complexity, enabling realistic scenario modeling particularly for systems with irregular set-up intervals. Whenever a set-up occurs, the production line switches to a new SKU and assigns a batch to the outgoing ULs. Finally, the Weekend Reduction parameter further enhances realism by allowing users to reduce the production rate on weekends as a percentage of the standard weekday rate. This enables modeling of operational differences between weekdays and weekends, offering insights into adjustments for staffing and storage during low-activity periods.

In the SKU Portfolio section, users define SKU properties and related production parameters. The Number of SKUs represents the total distinct products managed in the system, while the SKU Classification assigns production probabilities to each SKU using either a random approach, where all SKUs have equal probability, or an ABC classification. In ABC classification, SKUs are divided into three priority classes: Class A (high-frequency, 70–80% of production), Class B (moderate, 10–20%), and Class C (low, 10% or less). By default, the distribution is set to 10% (Class A), 20% (Class B), and 70% (Class C). Users can adjust these percentages as needed. Finally, the Expiration Date Rule governs expiration date assignment. Options include daily, weekly, or monthly expiration periods, where all SKUs produced within the same interval share the same expiration date. This rule is particularly relevant for shelf-life optimization and inventory control in systems managing perishable products as in the food and beverage industry.

The GUI shown in Figure 4 focuses on Shipping parameters, configuring essential settings for outbound flow, including dispatch rates, schedules, and load distribution. It mirrors the structure of Production (2) and integrates sections on warehousing, time scheduling, and shipping to manage the outbound logistics effectively.

Figure 4. Shipping GUI.

The Ship Requests parameter defines the total number of ULs expected to be shipped. To ensure workflow balance, this value should align with the Production Rate, except when a Warm-Up period is required to fill the warehouse to the predefined degree. By default, the Ship Requests value accounts for the warm-up, simplifying user input, although users can modify it to any non-negative integer. The Ship Rate [UL/h] specifies the hourly dispatch capacity, simulating

outbound flow, while the Standard Deviation introduces variability in the shipping rate.

In the Warehousing section, the Warm-Up Days parameter sets the time needed to reach the initial warehouse fill level before measurement begins. The UL [%] field specifies the warehouse's fill percentage during this period. The Warehouse Dispatching Strategy offers two options: Single, where shipments are consolidated from one storage module, or Multi, allowing collection from multiple modules using a round-robin approach.

The Time Scheduling section allows users to define shipping timelines, including the Shipping Start Date, Time, and End Time, with a default 24-hour operational window. The Ship-InterArrival Distribution governs intervals between shipments, offering Gaussian, Exponential, or Constant options to reflect dispatch rate variability. The Weekend Reduction parameter enables adjustments to weekend shipping volumes.

In the Truck section, the Truck Loading Capacity defines the number of palletized ULs per shipment, using either a fixed value (e.g., 33 pallets, as shown in Figure 4) or a probability distribution (as illustrated in Table 2).

Truck load [pallets]	Probability
32	5%
33	65%
34	18%
38	12%

Table 2. The probability distribution for truck loading

The Truck SKU parameter specifies the number of SKUs per shipment, either as a constant or a customizable probability distribution (Table 3). Users can modify or extend the default probability reported in Tables 2 and 3 as needed.

SKU-mix	Probability
1	22%
2	25%
3	7%
4	22%
5	8%
6	16%

Table 3. The probability distribution for Truck SKU.

These GUIs collectively support the configuration of production and shipment parameters, reflecting real industrial workflows. Table 4 exemplifies the resulting inbound and outbound tasks, each uniquely identified and chronologically ordered. Incoming tasks (type 1) represent ULs needing storage, while outgoing tasks (type -1) indicate ULs for shipment. Each entry includes details such as date, time, SKU, batch (formatted as XX_AA_BB, where XX is the SKU, AA is the current day, and BB is the production line), and the assigned storage module (SM).

ID	T	DateTime	SKU	Batch	ExpDate	SM
1	1	30/04/2024 00:00:00	19	19_0_1	30/04/2025	S1
2	1	30/04/2024 00:00:20	10	10_0_1	30/04/2025	S2
3	1	30/04/2024 00:00:40	74	74_0_1	30/04/2025	S1
4	1	30/04/2024 00:00:58	19	19_0_1	30/04/2025	S2
5	-1	30/04/2024 00:01:12	19	19_0_1	30/04/2025	S1

6	-1	30/04/2024 00:01:28	74	74_0_1	30/04/2025	S1
---	----	---------------------	----	--------	------------	----

Table 4. List of inbound and outbound tasks.

The generated request data is analyzed through an interactive PowerBI dashboard. This tool visualizes trends, peaks, and variability in production and shipping over different timeframes, providing critical insights into system performance.

Finally, the proposed platform has been validated to test the accuracy and reliability of the generated data. A real-world case study application was simulated and fed into the analytical-simulative data-driven model developed by Battarra et al. (2022) which monitors and assesses the performance of shuttle-based storage systems. By comparing the behavior of the real system with that of the system modeled using the platform's data, the validation process confirmed that the flow-generator could effectively replicate realistic operational conditions. This demonstrated its potential as a valuable tool for supporting decision-making, optimizing logistics processes, and enhancing data-driven performance analysis in supply chain and warehouse management.

3. CASE STUDY

The proposed flow-generator platform is applied to a production and distribution system operating in the beverage industry. The products moving within the system are carbonated drinks, juices, and other beverages packaged in cans and bottles. Figure 5 illustrates the main actors involved in the proposed beverage manufacturing and distribution plant: the production lines, warehousing system, shipping gates, and the material handling vehicles.

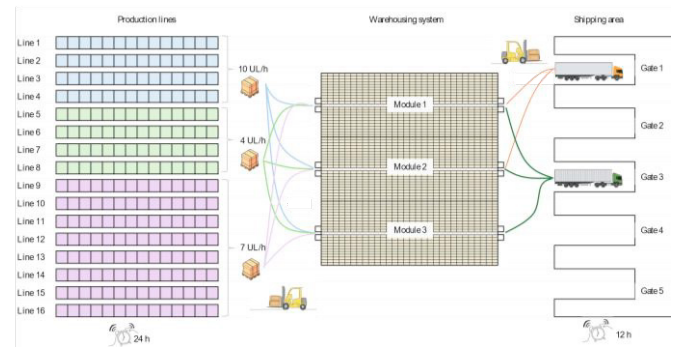


Figure 5. The beverage manufacturing and distribution plant.

The facility operates 16 production lines working 24 hours a day, continuously producing multi-can and multi-bottle packages. Each line features a palletization robot that stacks secondary packages into fully palletized ULs. The production rates differ among the lines: 4 lines achieve a high-speed production rate of 10 ULs/hour, 4 lines operate at 6 ULs/hour, and the remaining 8 lines produce at a rate of 7 ULs/hour.

The overall hourly production rate is 120 ULs/h, resulting in a daily throughput of 2,880 ULs. A standard deviation of 5% from the mean is introduced to simulate minor variability in production rates which reflects into a daily production fluctuation within a range of 2,736 to 2,880 ULs. The inter-arrival times for production tasks follow a Gaussian

distribution with a standard deviation of 10%. This distribution fits with the production rates of the automated lines which are relatively constant. This consistency is maintained throughout the week, as there are no variations in production rates during the weekend, ensuring a uniform output regardless of the day. With a total of 300 SKUs, the facility adopts an ABC classification approach for SKU prioritization. Class A SKUs represent 10% of the total, high-demand products, Class B covers 20%, moderate-demand items, and Class C accounts for 70%, representing niche or seasonal beverages. This classification determines production frequency and volume.

A fleet of double-unit Laser Guided Vehicles (LGVs) is employed to transport units from the end of the production line to the warehousing system that is an automated shuttle-based storage system. It is composed of three identical modules. Each module consists of 13 levels, with a cross-aisle double-sided aisle with 30 lanes on each side. Each lane has 11 ULs storage locations, resulting in a total capacity of 25,740 ULs within all modules (see Figure 5). At the beginning, the warehouse operates at almost 70% fill capacity, reflecting ongoing inventory turnover.

Shipping is aligned with retail demand and occurs daily within a reduced timeframe, from 6:00 a.m. to 6:00 p.m., with no reductions during weekends. Trucks with a fixed capacity of 33 palletized ULs transport the products, featuring a variable SKU mix per shipment, reported in Table 3. The inter-arrival times for shipping tasks follow an exponential distribution, capturing the variability driven by market demand and retailer fluctuations.

Figure 6 reports the Power BI dashboard which visually presents the daily tasks distribution resulting from analyzing 200,000 inbound tasks and 191,414 outbound tasks.

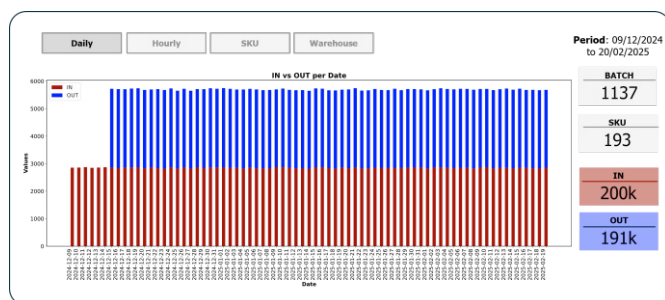


Figure 6. Case study: daily distribution.

The summary panel on the right aggregates key performance metrics, including the total number of SKUs and batches produced during the period of activity, offering insights into product variety and system capacity. During the 73 days of activity, 193 different SKUs out of the 300 available were produced. This is because the 16 production lines switch SKU every 24 hours, and the selection of SKUs follows an ABC logic.

The primary chart presents the daily distribution of inbound (red bars) and outbound (blue bars) tasks, facilitating the analysis of the workload distribution throughout the day. The initial six days focused exclusively on production, aimed at reaching the warehouse fill degree of 67%. Considering that the system operates automatically with consistent rates and low variability (with a standard deviation of 2%), no

significant fluctuations are observed, and the inbound and outbound tasks are well-balanced. The hourly distribution of inbound and outbound tasks is displayed in Figure 7 through a heat map which exemplifies the days from December 12th to December 15th, 2024.

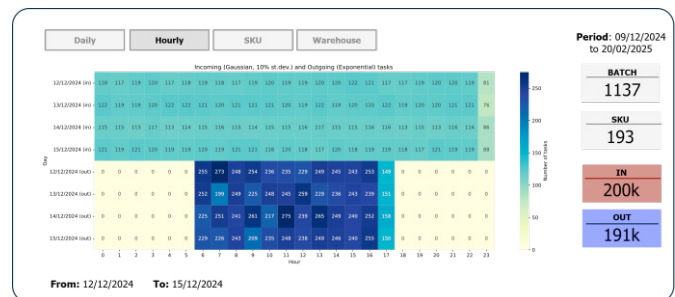


Figure 7. Case study: hourly distribution.

For inbound tasks, the maximum deviation between peak and minimum value is 4%, consistent with the Gaussian distribution adopted for inter-arrival times which leads to a more stable task flow. In contrast, outbound tasks show a significantly higher variability, with an average deviation of 20% reflecting the inherent nature of the exponential distribution which causes large fluctuations from day to day. Understanding the variability in inbound and outbound tasks is crucial for properly dimensioning fleets and workloads. By considering both the predictable patterns and fluctuations in task volume, companies can ensure resources are allocated efficiently, avoiding over-reliance on average values. Additionally, Figure 8 depicts the daily trends of inbound (SKU IN) and outbound (SKU OUT) SKUs.

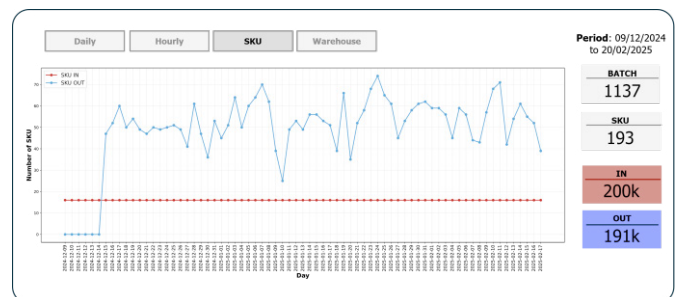


Figure 8. Case study: SKU distribution.

The number of SKU IN remains constant at 16 throughout the period, as production is standardized with 16 lines, each handling a single SKU per day. In contrast, SKU OUT fluctuates, starting at zero during the warm-up period and peaking at 74 on January 20th, 2025. This variability is influenced by shipment patterns. Increasing the number of SKUs per shipment (see Table 3) would involve more SKUs in the picking process, raising SKU OUT variability. A similar approach can be applied to SKU IN where variability is affected by production setup times. Finally, Figure 9 displays how the warehouse filling grade varies over time.

Initially, during the warm-up phase, the filling grade increases steadily as the system stabilizes, gradually reaching a certain value. After the warm-up, the filling grade follows a triangular wave pattern. This wave is characterized by a periodic increase

and decrease in a straight line, creating a "sawtooth" shape. The variation in the daily rates directly impacts how much the system's performance deviates from the average, making the shape of the wave more or less pronounced depending on these factors.

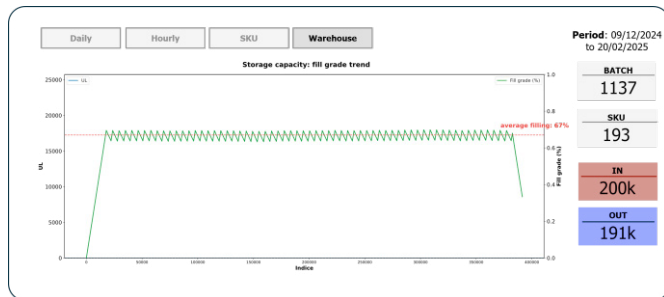


Figure 9. Case study: warehouse filling degree.

This case study illustrates how the proposed materials flow-generator platform functions as a dynamic tool for simulating and optimizing material handling and production systems. By allowing users to adjust key parameters, such as production rates and shipment schedules, the platform accurately replicates real-world operating conditions. This iterative approach enables simulations to be fine-tuned to reflect specific operational environments, as demonstrated by the dashboards tracking inbound and outbound flows, which highlight trends, variability, and peaks in production and material handling activities.

4. CONCLUSION

The proposed flow-generator platform is a robust and flexible tool for simulating data-inspired flows in complex industrial ecosystems. By leveraging adjustable parameters, the platform enables users to iteratively refine simulations, generating scenarios that closely replicate real-world operations. This novel approach addresses the frequent lack of real or realistic data, which often hampers assisted design of production systems and intralogistics. By filling this gap, the flow-generator enhances decision-making processes through robust and realistic data, complementing the use of simulators, digital twins, and advanced modeling methods.

Furthermore, the applicability of this platform extends beyond current systems, providing valuable insights for future growth scenarios. As industries face increasing challenges driven by globalization, e-commerce expansion, and shifting market demands, the ability to model and predict system behavior becomes critical. The platform offers customizable parameters for adapting to these changes such as parametric standard deviations, alternative inter-arrival distributions, set up distributions and fragmentation of shipments. By adjusting the input parameters, the managers can anticipate and accommodate future operational complexities.

Further research is expected on expanding the platform's functionalities to address more complex and dynamic industrial scenarios. Future enhancements may include the incorporation of additional inter-arrival and dispatch time distributions, such as Weibull or Gamma, to better reflect variability in specific operational environments. The material handling system might be expanded to accommodate a broader

range of vehicle capacities (e.g., 3 or 5 ULs per trip) and handling speeds to reflect a more comprehensive array of real-world configurations. Moreover, alternative dispatching strategies for warehousing, such as priority-based or demand-driven rules, might be introduced to optimize storage module utilization and reduce retrieval times.

REFERENCES

- Aron, C., Sgarbossa, F., Ballot, E., Dmitry I. (2024). Cloud material handling systems: a cyber-physical system to enable dynamic resource allocation and digital interoperability. *J Intell Manuf* 35, 3815–3836.
- Battarra, I., Accorsi, R., Manzini, R., Rubini, S. (2022). Hybrid model for the design of a deep-lane multisatellite AVS/RS. *International Journal of Advanced Manufacturing Technology*, 121 (1-2), pp. 1191-1217
- Cimino A., Longo F., Mirabelli G., Solina V., and Veltri P. (2024). Enhancing internal supply chain management in manufacturing through a simulation-based digital twin platform. *Computers and Industrial Engineering*, 198.
- Ghadge A, Er Kara M, Moradlou H, and Goswami M. (2020) The impact of Industry 4.0 implementation on supply chains. *Jou Manuf Tech Management* 31:669–686.
- Kamble S. S., Gunasekaran A., Parekh H., Mani V., Belhadi, A., and Sharma, R. (2022). Digital twin for sustainable manufacturing supply chains: Current trends, future perspectives, and an implementation framework. *Technological Forecasting and Social Change*, 176.
- Lu Y., Liu C., Kevin I., Wang K., Huang H., and Xu X. (2020). Digital Twin-driven smart manufacturing: Connotation, reference model, applications and research issues. *Robotics And Computer-Integrated Manufacturing*, 61.
- Lupi G., Accorsi R., Battarra I., Manzini R., Sirri G. (2024). Space efficiency and throughput performance in AVS/RS under variant lane depths. *JAMT*- 131.
- Jackson, I., Saenz, M., Ivanov, D. (2024). From natural language to simulations: applying AI to automate simulation modelling of logistics systems. *IJPR*, 62(4), 1434-1457.
- Pratik M., Kamble S., Belhadi A., Venkatesh M., and Abedin M.Z. (2023). Digital twin-driven real-time planning, monitoring, and controlling in food supply chains. *Technological Forecasting Social Change*, 195- 122799
- Samayamantri S. and Vaddy R.K. (2024). Sustainable Development Through Digital Twin Technology: Optimizing the Supply Chain. *Communications in Computer and Information Science*, 2196 CCIS, 92 - 101
- Sell R., Malayjerdi E., Malayjerdi M., and Baykara B. C. (2022). Safety Toolkit for Automated Vehicle Shuttle - Practical Implementation of Digital Twin. *ICCVE, USA, 2022*
- Singh G., Singh S., Daultani Y., and Chouhan M. (2023). Measuring the influence of digital twins on the sustainability of manufacturing supply chain: A mediating role of supply chain resilience and performance. *COMIND*, 186.