

Article

Reinforcement Learning for Congestion Mitigation in Inland Freight Terminals: A Simulation-Based Serial Mediation Analysis of Operational Learning Stability and Logistics Efficiency

Md. Mizanur Rahman ^{1,*}, Jianqiang Fan ^{1,*}, Edvard Tijan ² and Neven Grubišić ²

¹ School of Economics and Management, Chang'an University, Xi'an 710064, China

² Faculty of Maritime Studies, University of Rijeka, 51000 Rijeka, Croatia; edvard.tijan@pfri.uniri.hr (E.T.); neven.grubisic@pfri.uniri.hr (N.G.)

* Correspondence: mdmizanurr@chd.edu.cn or dr.md.mizanur.rahman998@gmail.com (M.M.R.); fanjianq@126.com or fanjianq@chd.edu.cn (J.F.)

Abstract

This study explains how reinforcement learning (RL) contributes to congestion mitigation in inland freight terminal operations by testing a serial process model in which RL strengthens operational learning stability (OLStab), OLStab improves logistics efficiency, and logistics efficiency lowers congestion. Rather than presenting RL as a stand-alone congestion-reduction instrument, the paper examines a distinct inland-terminal application in which congestion emerges from interacting gate, yard, transfer, and dispatch frictions. Using a simulation-based explanatory design calibrated to a realistic macro-logistics context, and interpreting the results as simulation-informed evidence rather than direct field proof, the study analyzes 500 episode-level observations representing complete terminal runs under varying control conditions. The results show that RL positively affects OLStab, OLStab positively affects logistics efficiency, and logistics efficiency negatively affects congestion. The serial indirect pathway from RL through OLStab and logistics efficiency to congestion is statistically significant, whereas the direct effect of RL on congestion becomes non-significant once the mediators are introduced. Decision latency sensitivity does not significantly moderate the RL-to-OLStab relationship, suggesting that latency-related boundary conditions are more context-specific than the main capability pathway. The article contributes by offering a cautious simulation-based and mechanism-centered explanation of RL-enabled congestion mitigation in inland terminals, by treating OLStab as a simulation-grounded intermediate operational stability index, and by showing that the empirical pattern is better explained by theory-ordered simulator-level mechanism than by a residual direct RL effect.

Keywords: reinforcement learning; inland freight terminals; dry ports; operational learning stability; logistics efficiency; congestion; serial mediation

Academic Editor: Emilio Larrodé

Received: 22 April 2026

Revised: 3 June 2026

Accepted: 8 June 2026

Published: 26 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Inland freight terminals and dry-port-type interfaces play a critical coordinating role in hinterland logistics because they connect road flows, yard handling, storage buffers, and rail or onward dispatch operations [1,2]. Congestion in these facilities is therefore not

simply a matter of local queue length. It is an emergent outcome of interacting arrival surges, staging constraints, dispatch timing, equipment availability, and uneven work-load propagation across terminal subsystems [3,4]. Even when digital tools improve visibility, inland nodes remain vulnerable to coordination breakdowns because operational decisions must be updated repeatedly under variability rather than solved once under stable conditions [5–7].

Reinforcement learning offers a promising control logic for such settings because it adapts decisions through repeated interaction with changing system states and reward feedback [8–10]. Reviews of RL in logistics and transportation emphasize its suitability for stochastic, high-dimensional, and operationally complex environments, while recent freight studies show that learned policies can improve dispatching, replanning, and yard decisions under uncertainty [11–15]. What remains less well developed, however, is a process explanation of how learning-based control translates into lower congestion inside terminal operations. Much of the existing literature reports performance gains at the outcome level, but leaves the intermediate operational pathway undertheorized [9,10].

That gap is especially important in inland terminal environments, where congestion is often generated by unstable internal adjustments rather than by a single visible bottleneck [2–4]. A learned controller may improve performance not because it directly suppresses queues at all times, but because it gradually stabilizes operational responses, reduces oscillatory adjustments, and enables more reliable execution across coupled activities [16–18]. The present study therefore shifts attention from headline performance claims to the internal capability sequence through which learning affects congestion.

To capture that sequence, the paper introduces operational learning stability (OLStab) as a simulation-grounded intermediate operational stability index. OLStab refers here to the consistency, manageability, and reduced volatility of operational adjustment across repeated terminal decisions. It is treated as conceptually distinct from logistics efficiency: OLStab reflects whether the operating pattern becomes more stable, whereas logistics efficiency reflects how well the terminal converts that stability into throughput, turnaround quality, coordination, and resource use. This distinction makes it possible to test whether RL influences congestion indirectly through a staged operational mechanism rather than through a dominant stand-alone direct effect.

The study also examines decision latency sensitivity (DLS) as a boundary condition. In principle, learning-based control should be harder to translate into performance when observations, decisions, or execution responses lag behind the current state of operations [19,20]. Yet latency effects in terminals are likely to be context-dependent because some operating environments absorb timing frictions through buffering or standardized routines more effectively than others [19,20]. For that reason, DLS is treated as a secondary contextual factor rather than the central contribution of the article.

The broader supply-chain, logistics, and terminal-operations literature offers contextual support for understanding learning-enabled decision systems, operational risk, and congestion-sensitive logistics coordination [19–23]. However, these studies are used here only as contextual support. They do not substitute for direct terminal evidence, and the present analysis remains limited to the inland-terminal simulation setting examined in this paper.

Accordingly, the contribution of this paper is threefold, and each contribution is limited to the designed simulation environment. First, it provides a simulation-based mechanism explanation of congestion mitigation in an inland freight terminal application, rather than treating RL as a generic performance booster. Second, it cautiously positions OLStab as a simulation-grounded intermediate metric linking adaptive learning to realized logistics efficiency. Third, it presents the study as a related but distinct application within the broader family of RL-based terminal simulations: the analytical focus,

simulator organization, and empirical pattern of effects are specific to the inland-terminal setting examined here. On that basis, the paper tests, within this designed simulation environment, a serial model in which RL enhances OLStab, OLStab improves logistics efficiency, and logistics efficiency reduces congestion, while DLS is evaluated as a secondary boundary condition.

The remainder of the paper is organized as follows. Section 2 develops the conceptual background and hypotheses. Section 3 explains the simulation-based methodology, data-generation procedure, construct operationalization, and model equations. Section 4 reports the empirical results. Sections 5–7 discuss the findings, theoretical contributions, and practical implications. The paper then presents limitations and future research directions before the conclusion.

2. Conceptual Background and Hypothesis Development

2.1. Reinforcement Learning as Adaptive Control in Inland Logistics

Within logistics research, reinforcement learning is best understood as an adaptive control approach for repeated decision problems under uncertainty. Instead of optimizing a single static decision, an RL agent observes operational states, selects actions, receives reward feedback, and updates future choices through experience. This logic is particularly attractive for logistics systems in which congestion, equipment utilization, demand, and routing options evolve over time. Reviews by [8–10] consistently show that RL becomes most valuable when state spaces are large, feedback is sequential, and repeated re-optimization by exact methods is difficult or too slow.

For terminal-related applications, the relevance of RL lies in its capacity to manage continual adjustment rather than one-off optimization [9,10]. Studies on synchromodal replanning, rail-yard control, AGV dispatching, and container stacking show that RL can improve operational decisions under uncertainty, disturbance, and incomplete foresight [11,14,21,22]. These studies support the view that RL is suited to inland freight terminals, where dispatch coordination, gate admission, yard handling, and outbound release decisions recur throughout an operating episode rather than appearing as isolated optimization tasks [9,10].

The implication for the present paper is that RL should be theorized as a learning mechanism embedded in an operating process. Its value must therefore be assessed not only by whether final congestion falls, but also by whether the controller creates a more stable and effective pattern of operational adjustment along the way.

2.2. Why Congestion Persists in Digitally Enabled Terminal Systems

Digitalization improves sensing, information sharing, and decision support, but it does not automatically remove congestion from terminal environments [5,6]. In practice, digitally enabled terminals still confront interacting queues, limited buffers, unbalanced workloads, and synchronization failures across operational units [3,23]. Research on digital twins and intelligent terminal control shows that better data can improve representation and responsiveness, yet performance remains highly sensitive to how decisions are coordinated across interdependent subsystems and how quickly those decisions can be executed in practice [6,7,15].

This problem is especially visible when local disturbances propagate beyond their point of origin. In automated terminal and truck-dispatch settings, seemingly efficient local rules can produce downstream conflicts, idle time, or spillback when task assignment, routing, and release decisions are not aligned [15,23,24]. Likewise, inland and cross-docking studies show that fluctuating arrivals and asymmetric traffic conditions destabilize scheduling even when real-time information is available [3,25]. Congestion persistence

should therefore be treated as a process problem rooted in coupled operational variability, not simply as an information deficit [3,25].

For inland freight terminals, the challenge is amplified by the role of the facility as a transfer and buffering node [2,3]. Gate inflows, yard movements, storage decisions, and dispatch releases interact tightly, so operational instability in one part of the node can quickly appear elsewhere as longer waiting times, blocking, and accumulated dwell [3,23]. This motivates a mechanism-oriented framework that explains how learning affects the internal operating pattern before it affects the final congestion outcome.

Seen from this perspective, digital transformation is a necessary but insufficient condition for congestion relief [5,7]. What matters is whether digital and learning-enabled control produces a more coherent and stable operating rhythm across the terminal [6,15].

2.3. Operational Learning Stability as an Intermediate Stability Index

The concept of operational learning stability was introduced in related simulation-based work by Rahman, Md.M 2026 [25,26]. In the present paper, OLStab is used as an episode-level intermediate stability index that captures whether adaptive control is associated with steadier operational behavior across repeated simulation episodes. The index draws on the operations literature on learning-by-doing and routine refinement, where repeated execution can improve performance when knowledge is retained, implemented, and translated into more predictable processes [15,16,26]. In this study, however, OLStab is not treated as a field-validated organizational capability. It is a simulation-grounded formative index constructed from normalized indicators of decision volatility, queue-evolution volatility, and adjustment magnitude. The idea draws on the operations literature on learning-by-doing and routine refinement, where repeated execution can improve performance only when knowledge is retained, implemented, and translated into more predictable processes [15,16,26]. Zhang et al. (2024) and Glock et al. (2019) both show that learning dynamics matter because they shape whether operational trajectories become progressively more stable or remain vulnerable to disruption and reversal [15,16].

In the present context, OLStab does not refer to a perceptual or survey-based construct. It refers to an episode-level operating property reflected in lower adjustment volatility, smoother queue evolution, and more consistent controller behavior. This framing is important because inland terminal congestion is rarely driven by a lack of activity; it is driven by unstable activity that generates repeated corrections, uneven loads, and avoidable spillovers.

The construct is also theoretically useful because it separates adaptive learning from realized performance. Studies on organizational learning and logistics service effectiveness suggest that a learning-oriented capability base can support better execution outcomes, but capability and performance are not the same phenomenon [18,27]. By modeling OLStab as an intermediate operational stability index, the paper clarifies the step between learning-based control and logistics efficiency instead of collapsing them into a single outcome.

This distinction further helps position the paper as a mechanism study. The theoretical question is not merely whether RL works, but whether it works by stabilizing internal adjustments in a way that later improves execution quality.

To address the conceptual proximity between OLStab and logistics efficiency, the two constructs are separated by role and indicator meaning. OLStab describes the stability of the adjustment process itself: lower decision volatility, smoother queue evolution, and more manageable operational changes. Logistics efficiency describes realized execution quality: throughput, turnaround, coordination, and resource utilization. Therefore, the high OLStab-LE correlation reported later indicates strong empirical association in the simulator, but the interpretation does not assume that the two constructs are identical. For

that reason, the paper frames the serial pathway as simulation-informed evidence consistent with staged operational development, not as universal proof of organizational capability formation. Although OLStab and logistics efficiency are separated conceptually and operationalized with different component signals, they should not be interpreted as fully independent field-validated constructs in this simulation-based design. Both indices are generated within the same episode-level simulator and therefore may share part of the same underlying operational structure. The high correlation between OLStab and logistics efficiency indicates that smoother adjustment and better execution quality are closely connected in the simulated terminal environment. Therefore, the serial pathway should be interpreted as a theory-ordered decomposition of related simulator-derived indices, rather than as definitive evidence that two fully independent constructs have been empirically separated.

2.4. Logistics Efficiency and Congestion as Linked Performance Outcomes

Logistics efficiency is modeled as the quality of realized terminal execution, expressed through throughput, turnaround, coordination, and resource use [23,28]. In tightly coupled freight nodes, these dimensions are inseparable from congestion because inefficient handling, poor synchronization, and repeated rework quickly translate into waiting, blocking, and dwell-time accumulation [3,23]. Studies on AGV dispatching, yard stacking, and dry-port forwarding all point to the same operational logic: better execution quality reduces self-generated pressure inside the system [3,11,13,22].

For inland freight terminals, this means that congestion should be interpreted as a downstream consequence of how well the terminal converts available capacity into coordinated flow [2,3]. Faster handling alone is not enough. Efficiency must also reflect whether activities are aligned across gate processing, yard handling, and release decisions so that queue pressure is not simply shifted from one subsystem to another [13,23,28].

The link between efficiency and congestion is therefore more than a cost argument. It is a systems argument. When resources are used more coherently and processing becomes more reliable, fewer queues are formed, fewer blockages persist, and the terminal is better able to absorb variability without widespread spillover [22,23,28].

This makes logistics efficiency the natural second-stage mediator in the present model: it captures the performance consequences of stabilized operational learning and provides the most immediate pathway to lower congestion.

2.5. Hypothesis Development: RL to Stability, Stability to Efficiency, Efficiency to Congestion

H1 proposes that reinforcement learning positively affects operational learning stability. The logic is straightforward: repeated reward-guided adjustment should reduce ad hoc decision oscillation and generate more consistent operational responses than non-learning control that does not systematically incorporate feedback [10,14]. Evidence from truck dispatching, AGV control, and other logistics applications supports the expectation that learning-based policies can produce more orderly operating patterns under uncertainty [11,12,14].

In inland-terminal terms, a controller that learns from repeated interaction should become better at balancing interventions across coupled queues and service points, thereby creating smoother adjustment trajectories across episodes [11,12].

Thus,

H1. *Reinforcement learning positively affects operational learning stability (OLStab).*

H2 links operational learning stability to logistics efficiency. Once operational adjustment becomes more stable, execution should become easier to coordinate because fewer

abrupt corrections, priority reversals, and workload imbalances disturb the process. The organizational learning literature suggests that stable learning routines improve effectiveness by making coordination more predictable and execution less wasteful [16–18].

Applied to inland freight terminals, this implies that steadier operating behavior should support better throughput, turnaround, and resource utilization [17,18]. OLStab is therefore expected to improve logistics efficiency because stability creates the conditions under which efficient execution can be sustained rather than repeatedly interrupted [16,18].

Therefore,

H2. *Operational learning stability (OLStab) positively affects logistics efficiency.*

H3 states that logistics efficiency negatively affects congestion. This follows both queueing logic and terminal operations logic. When handling becomes more coordinated and resource use becomes more balanced, fewer delays accumulate and fewer local bottlenecks escalate into broader congestion [3,13]. Research on rehandling reduction, forwarding design, and conflict-aware dispatching supports this expectation by showing that more efficient logistics execution reduces pressure on the broader system [3,13,22].

In the inland-terminal setting examined here, higher logistics efficiency should therefore correspond to lower congestion because the node can process variability with less blocking and less spillover across coupled activities [3,22].

Consequently,

H3. *Logistics efficiency negatively affects congestion.*

2.6. Serial Mediation and Decision Latency Sensitivity

H4 integrates the preceding claims into a serial mediation mechanism. The argument is that RL contributes to congestion reduction primarily by first stabilizing operational adjustment and then converting that stability into better logistics execution [10,29]. This sequencing is theoretically important because it distinguishes among learning capability, execution performance, and final system outcome [16,18]. It also speaks directly to a recurring weakness in the RL-for-logistics literature, where policy success is often reported without an explicit account of the operating pathway through which success is produced [9,10].

Accordingly,

H4. *Operational learning stability (OLStab) and logistics efficiency serially mediate the relationship between reinforcement learning and congestion. Because the focal variables are exported as episode-level summary indices after each simulator run, the term “serial mediation” is used here to describe a theory-ordered statistical specification rather than direct temporal observation of separate operational stages. Thus, H4 tests whether the covariance pattern among RL, OLStab, logistics efficiency, and congestion is consistent with the proposed mechanism. It does not imply that OLStab, logistics efficiency, and congestion were observed as temporally separated outcomes within each episode.*

Consequently,

H5. *Decision latency sensitivity (DLS) moderates the relationship between reinforcement learning and operational learning stability (OLStab). H5 introduces decision latency sensitivity (DLS) as a secondary contextual condition in the RL–OLStab relationship. Conceptually, delayed observation, decision execution, or feedback may disturb the translation of learning-enabled control into stable operational adjustment. However, latency effects may vary across operating contexts because some*

terminal environments may absorb timing frictions through buffering, standardized routines, or slack capacity. Therefore, DLS is tested cautiously as a possible moderator of the RL–OLStab relationship rather than as a general moderator of all downstream outcomes.

Accordingly, H5 proposes that decision latency sensitivity moderates the relationship between reinforcement learning and operational learning stability. The interaction term between RL and DLS is statistically non-significant, indicating that DLS does not meaningfully change the strength of the RL–OLStab relationship in the simulated inland-terminal environment. The positive direct coefficient of DLS should also be interpreted cautiously. This coefficient is inconsistent with the initial expectation that greater latency sensitivity would impair operational stability. Therefore, the result should not be interpreted as evidence that latency sensitivity improves OLStab. A more cautious interpretation is that, within this simulator, the DLS index may partly capture context-specific buffering, standardized routines, or latency-management conditions that coexist with more stable adjustment. Consequently, DLS is not supported as a theoretically robust boundary condition in this study.

3. Methodology

This study uses a simulation-based explanatory design to examine the mechanism through which reinforcement learning relates to congestion mitigation in inland freight terminals [10,23,29]. The empirical context is an inland logistics node characterized by interacting gate arrivals, yard handling, transfer buffering, and dispatch release decisions [2,3]. Rather than representing a generic port environment, the simulator is organized around inland-terminal flow coordination and the propagation of congestion across linked service points [3,29]. Secondary indicators from the World Bank calibration were used to calibrate plausible operating conditions at the macro-logistics level, while the analytical dataset itself consists of simulated episodes rather than observations from real terminals [23,29]. Each episode corresponds to one complete run of the inland-terminal simulator under a defined control setting and environmental condition, after which the model exports a summary record for reinforcement learning strength, operational learning stability, logistics efficiency, decision latency sensitivity, and congestion [14,29]. The unit of analysis is therefore the simulated episode, and the claims of the paper are limited to the mechanism produced within this distinct simulator setting rather than to direct causal proof for all real-world terminals [22,23].

3.1. Data-Generation Procedure

The test dataset was generated by running the inland-terminal simulator for 500 complete episodes. Each episode began with a defined operating environment, including gate arrival pressure, yard-handling conditions, transfer-buffering conditions, dispatch-release logic, and decision-latency conditions. During an episode, the RL controller observed the current terminal state, selected actions, received reward feedback, and updated its policy. At the end of each run, the simulator exported one episode-level row containing reinforcement learning strength, operational learning stability, logistics efficiency, decision latency sensitivity, and congestion. Therefore, OLStab, logistics efficiency, DLS, and congestion should be understood as same-episode summary outputs generated from the completed simulator run. They are analytically ordered according to the proposed mechanism, but they are not time-stamped field observations collected at separate temporal stages. This distinction is important for interpreting the mediation model as a simulator-level statistical pathway rather than as direct evidence of temporal causal sequencing. The raw analytical workbook contains these episode-level records and the statistical output sheets used for the tables reported in this paper.

3.2. Calibration and Interpretation

World Bank macro-logistics indicators served only as contextual calibration benchmarks for establishing plausible operating ranges in the simulated inland-terminal environment. The calibration used the six Logistics Performance Index dimensions—customs efficiency, infrastructure quality, international shipment reliability, logistics competence and quality, tracking and tracing capability, and timeliness—to define low, medium, and high scenario settings. Because these indicators are macro-logistics descriptors rather than terminal-level observations, they were not treated as field validation data. The conversion logic was transparent and rule-based: LPI-type scores below 2.50 were mapped to a low operating condition, scores from 2.50 to 3.50 were mapped to a medium operating condition, and scores above 3.50 were mapped to a high operating condition. The thresholds were used as design anchors for scenario construction, not as empirical estimates from a named terminal. The low, medium, and high LPI thresholds should therefore be interpreted as heuristic scenario anchors rather than empirically validated calibration cutoffs. They were used to structure plausible operating conditions for the simulator, not to estimate terminal-specific parameters or validate the simulated environment against observed terminal records. Accordingly, the calibration procedure supports contextual plausibility, but it does not constitute field calibration or external validation. Equal contextual weight was assigned to the six LPI dimensions because the purpose was to create plausible simulation ranges rather than to estimate a country-level logistics performance model. The resulting episode-level observations are therefore repeated outputs from a designed simulator, and the coefficients, mediation estimates, *p*-values, and confidence intervals are interpreted as simulation-based evidence about the internal mechanism generated by this design, not as direct causal estimates from a field implementation. Table 1 provides an overview of how the World Bank Logistics Performance Index (LPI) dimensions were translated into low, medium, and high simulation settings for the inland freight terminal model. The mapping illustrates how customs efficiency, infrastructure quality, international shipment reliability, logistics competence, tracking and tracing, and timeliness were employed solely as contextual calibration benchmarks. These dimensions informed plausible parameter ranges for factors such as gate-processing reliability, yard and buffer capacity, arrival variability, dispatch responsiveness, observability, and delay sensitivity. Furthermore, the table emphasises that the LPI indicators were not regarded as direct observations at the terminal level or as field-validation data; instead, they functioned as rule-based design anchors for the development of realistic simulation scenarios. Consequently, the episode-level observations resulting from this process should be understood as outputs from a carefully designed simulation environment, rather than as independent records from real-world terminals.

Table 1. Calibration mapping used for World Bank LPI-based simulation context setting.

LPI Dimension	Simulation Parameter Affected	Low Setting (<2.50)	Medium Setting (2.50–3.50)	High Setting (>3.50)
Customs efficiency	Gate-processing reliability and clearance delay	Service reliability = 0.70; delay sensitivity = 1.30	Service reliability = 0.85; delay sensitivity = 1.00	Service reliability = 1.00; delay sensitivity = 0.75
Infrastructure quality	Yard/buffer capacity and service-rate variability	Capacity factor = 0.75; variability multiplier = 1.30	Capacity factor = 1.00; variability multiplier = 1.00	Capacity factor = 1.20; variability multiplier = 0.80
International shipment reliability	Arrival-pressure variability and schedule reliability	Arrival variability = 1.30; schedule reliability = 0.70	Arrival variability = 1.00; schedule reliability = 0.85	Arrival variability = 0.80; schedule reliability = 1.00

Logistics competence and quality	Service adaptation and dispatch responsiveness	Adaptation factor = 0.70; dispatch responsiveness = 0.75	Adaptation factor = 0.85; dispatch responsiveness = 0.90	Adaptation factor = 1.00; dispatch responsiveness = 1.00
Tracking and tracing	State observability and information-delay condition	Observability = 0.70; information delay = 1.30	Observability = 0.85; information delay = 1.00	Observability = 1.00; information delay = 0.75
Timeliness	Release adherence and dispatch-delay condition	Adherence = 0.70; dispatch delay = 1.30	Adherence = 0.85; dispatch delay = 1.00	Adherence = 1.00; dispatch delay = 0.75

To improve consistency between Figure 1 and the statistical model, the variables and parameters used in the figure are defined as follows. In the mediation framework, X denotes the independent variable, reinforcement learning strength (RL); M_1 denotes the first mediator, operational learning stability ($OLStab$); M_2 denotes the second mediator, logistics efficiency (LE); Y denotes the outcome variable, congestion; and W denotes the moderator, decision latency sensitivity (DLS). To avoid duplication, all formal equations are presented once in the methodology section, while Figure 1 is used only as a visual summary of the analytical workflow. In descriptive terms, the main mechanism is $RL \rightarrow OLStab \rightarrow LE \rightarrow Congestion$, and the moderation path is $RL \times DLS \rightarrow OLStab$.

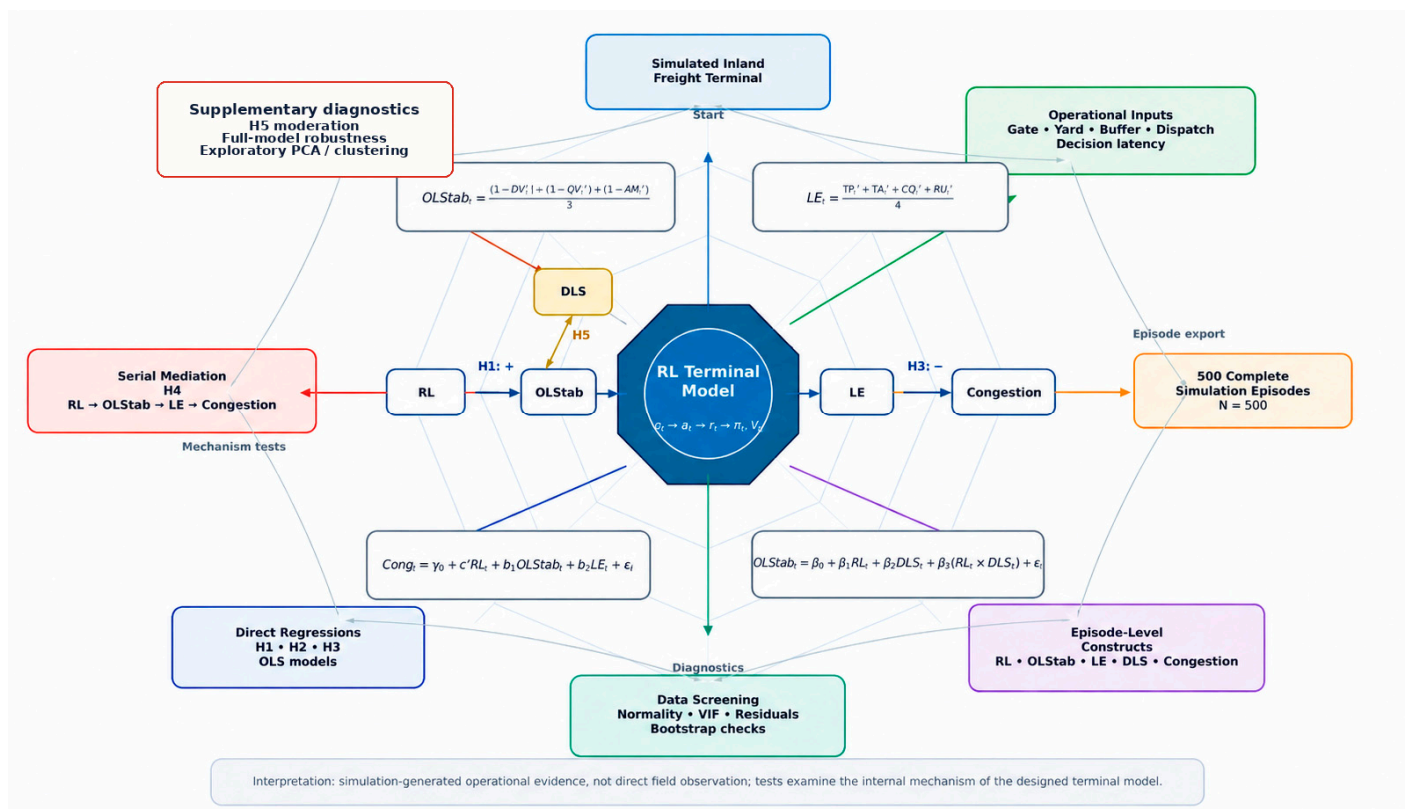


Figure 1. Experimental situation for reinforcement learning model comparison. Source: Authors' own creation using Python 3.13.

In the reinforcement learning terminal model, s_t denotes the terminal state at time step t , including queue status, yard condition, buffer condition, dispatch status, and decision-latency condition. a_t denotes the action selected by the RL controller at time step t , r_t denotes the reward feedback received after action execution, and π_t denotes the learned decision policy at time step t . V_t denotes the value estimate associated with the current terminal state or policy evaluation. The symbol ω_k denotes the weight assigned to the k -th normalized component in a formative construct index.

Where DV_i^* denotes normalized decision volatility, QV_i^* denotes normalized queue-evolution volatility, and AM_i^* denotes normalized adjustment magnitude. These three components are reverse-coded because lower volatility and lower adjustment magnitude indicate higher operational learning stability. TP_i^* denotes normalized throughput performance, TA_i^* denotes normalized turnaround performance, CQ_i^* denotes normalized coordination quality, and RU_i^* denotes normalized resource utilization. Thus, *OLStab* captures the stability of operational adjustment, whereas *LE* captures realized logistics execution quality. Table 2 summarizes the notation and variable definitions used in Figure 1 and the model equations. It clarifies the meaning of the main analytical symbols, including the independent variable, mediators, moderator, outcome variable, reinforcement learning terms, and simulation indexes. The table helps understand how the figure, reinforcement learning process, and statistical model are connected.

Table 2. Notation and variable definitions used in Figure 1 and model equations.

Symbol/Notation	Full Meaning	Role in This Study
X	Reinforcement learning strength (RL)	Independent variable
M1	Operational learning stability (OLStab)	First mediator
M2	Logistics efficiency (LE)	Second mediator
Y	Congestion	Dependent/outcome variable
W	Decision latency sensitivity (DLS)	Moderator
RL	Reinforcement learning strength	Captures the intensity and quality of learning-enabled control
OLStab	Operational learning stability	Captures stability of operational adjustment
LE	Logistics efficiency	Captures execution quality through throughput, turnaround, coordination, and resource utilization
DLS	Decision latency sensitivity	Captures sensitivity to delayed observation, decision, or execution
s _t	Terminal state at time step t	Represents the current operational condition, such as queue status, yard condition, buffer condition, dispatch status, and latency condition
a _t	Action selected at time step t	Represents the decision/action chosen by the RL controller
r _t	Reward feedback at time step t	Represents performance feedback received after action execution
pi _t	Learned decision policy at time step t	Represents the RL policy mapping states to actions
V _t	Value estimate at time step t	Represents the estimated value of the current state or policy evaluation
omega _k	Weight of the k-th component	Represents the weight assigned to a normalized component in a formative index
DV _i [*]	Normalized decision volatility	Component of OLStab
QV _i [*]	Normalized queue-evolution volatility	Component of OLStab
AM _i [*]	Normalized adjustment magnitude	Component of OLStab
TP _i [*]	Normalized throughput performance	Component of LE
TA _i [*]	Normalized turnaround performance	Component of LE
CQ _i [*]	Normalized coordination quality	Component of LE
RU _i [*]	Normalized resource utilization	Component of LE
i	Simulation episode index	Identifies each of the 500 simulated episodes

t	Time-step index	Identifies each decision step inside an episode
---	-----------------	---

Figure 1 provides an overview of the entire workflow of the study, commencing with the simulated inland freight terminal environment and concluding with statistical mechanism testing. The workflow illustrates the interaction between the reinforcement learning (RL) controller and various terminal operations, which encompass gate arrivals, yard handling, transfer buffering, dispatch release, and decision-latency conditions. Each complete simulation episode yields an episode-level record that includes RL strength, OLStab, logistics efficiency, DLS, and congestion. These outputs are subsequently employed for data screening, operational construct development, direct-effect regression analysis, serial mediation testing, and the H5 moderation test. Additional procedures, including full-model robustness checks, PCA, clustering, and related diagnostic outputs, are treated only as supplementary checks and are not interpreted as independent validation of the proposed mechanism. The figure emphasizes that the study relies on operational evidence generated through simulation rather than direct field observations.

To avoid confusion with ordinary least squares, operational learning stability is abbreviated as OLStab throughout the equations and results. All core constructs were operationalized as standardized episode-level indicators. Reinforcement learning strength captures the intensity and quality of learning-enabled control within an episode.

Reinforcement learning strength was operationalized as a normalized algorithm-internal control index rather than as a direct congestion or logistics-efficiency outcome. The index captures the extent to which the RL controller actively updated, improved, and stabilized its learned policy during a complete simulation episode. It was calculated as follows:

$$\text{RL strength} = (\text{normalized learning-update intensity} + \text{normalized value-improvement behavior} + \text{normalized policy-convergence quality})/3$$

In this calculation, normalized learning-update intensity refers to the magnitude of policy or value-function updates during each episode. Normalized value-improvement behavior reflects improvement in the controller's internal value estimate during the episode. Normalized policy-convergence quality refers to the reduction in policy instability or temporal-difference error across the episode. Higher values of RL strength therefore indicate stronger learning-enabled control. Importantly, the RL-strength index was computed from algorithm-internal learning signals and was not constructed from OLStab, logistics efficiency, DLS, or congestion components.

Reinforcement learning strength was operationalized as a normalized algorithm-internal control index rather than as a direct congestion, logistics-efficiency, DLS, or OLStab outcome. The index captures the extent to which the RL controller actively updated, improved, and stabilized its learned policy during a complete simulation episode. It was calculated as follows:

$$\text{RL Strength}_i = (\text{LUI}_i + \text{VIB}_i + \text{PCQ}_i^*)/3^{**}$$

where LUI* denotes normalized learning-update intensity, VIB* denotes normalized value-improvement behavior, and PCQ* denotes normalized policy-convergence quality for simulation episode *i*. All three components were normalized to the 0–1 range before aggregation, and equal weights were assigned because the purpose was to construct a transparent formative index of algorithm-internal learning strength rather than to estimate an externally validated latent construct. Higher values indicate stronger learning-enabled control within the simulated episode. Importantly, RL strength was derived only from algorithm-internal learning signals and was not constructed from OLStab, logistics efficiency, DLS, or congestion components.

This derivation clarifies that RL strength is a formative, simulator-specific index of learning-control behavior. It does not represent a direct operational-performance outcome and should not be interpreted as an externally validated measure of universal RL capability. Instead, it provides a reproducible episode-level indicator of how strongly the RL controller updated, improved, and stabilized its policy within the designed inland-terminal simulation.

OLStab captures the consistency of operational adjustment, including lower decision volatility, smoother queue evolution, and steadier execution patterns. Logistics efficiency is modeled as a realized execution construct incorporating throughput, turnaround speed, coordination quality, and resource utilization. Decision latency sensitivity reflects the extent to which delayed decisions erode operating performance, and congestion is treated as the system-level pressure outcome composed of queueing, blockage, dwell accumulation, and yard saturation. OLStab and logistics efficiency were intentionally kept separate in the measurement logic so that the analysis could distinguish stabilization of adjustment from performance of execution, while recognizing that both are generated within the same simulation environment.

To improve the transparency of the newly introduced operational learning stability construct, *OLStab* was operationalized as a formative, normalized, episode-level stability index. The index combines three simulation-derived components: decision volatility, queue-evolution volatility, and adjustment magnitude. Each component was normalized before aggregation. Because these components represent instability, they were reverse-coded so that higher *OLStab* values consistently indicate more stable learning behavior, lower operational fluctuation, and more manageable adjustment patterns.

The *OLStab* index was calculated as follows:

$$OLStab_i = \frac{(1 - DV_i^*) + (1 - QV_i^*) + (1 - AM_i^*)}{3}$$

Logistics efficiency was operationalized as a formative execution-quality index based on throughput performance, turnaround performance, coordination quality, and resource utilization. The logistics efficiency index was calculated as follows:

$$LE_i = \frac{TP_i^* + TA_i^* + CQ_i^* + RU_i^*}{4}$$

where i indexes simulation episodes; DV_i^* denotes normalized decision volatility; QV_i^* denotes normalized queue-evolution volatility; AM_i^* denotes normalized adjustment magnitude; TP_i^* denotes normalized throughput performance; TA_i^* denotes normalized turnaround performance; CQ_i^* denotes normalized coordination quality; and RU_i^* denotes normalized resource utilization.

This operationalization clarifies that *OLStab* captures the stability of operational adjustment, whereas logistics efficiency captures the quality of operational execution. In other words, *OLStab* reflects whether the learning-enabled control process becomes smoother and less volatile, while *LE* reflects whether the terminal converts that stability into better throughput, turnaround, coordination, and resource use.

The component-level implementation of these indices should remain aligned with the simulator code and the supplementary dataset supplied with the submission. Table 3 below outlines the definitions, measurements, and distinctions of each study construct within the simulation-based research design. It clarifies the model roles of the variables: reinforcement learning strength as the independent variable, *OLStab* as the first mediator, logistics efficiency as the second mediator, decision latency sensitivity as the moderator, and congestion as the outcome variable.

Table 3. Derivation of the RL Strength index.

RL Strength Component	Definition	Source Signal in Simulator	Normalization	Weight
Learning-update intensity (LUI [*])	Captures the magnitude of policy or value-function updating during the episode.	Episode-level update magnitude from the RL controller, such as value-update or temporal-difference update activity.	Min–max normalized to 0–1 across all simulated episodes.	1/3
Value-improvement behavior (VIB [*])	Captures the extent to which the controller’s value estimate improves during the episode.	Difference or improvement in value estimates between earlier and later learning steps within the episode.	Min–max normalized to 0–1 across all simulated episodes.	1/3
Policy-convergence quality (PCQ [*])	Captures the reduction in policy instability or learning error during the episode.	Decline in policy instability, action-value fluctuation, or temporal-difference error across the episode. Higher values indicate stronger convergence quality.	Reverse-coded where necessary and min–max normalized to 0–1 across all simulated episodes.	1/3

This construct mapping helps address potential circularity concerns by differentiating among five separate operational phenomena: learning-enabled control, adjustment stability, execution performance, latency sensitivity, and downstream congestion pressure.

The distinction remains important because OLStab and logistics efficiency are operationally adjacent and generated within the same simulator. For that reason, the study does not treat their association as proof that two fully independent field constructs have been validated; it treats the separation as a simulator-level measurement distinction that requires future external testing.

To reduce the risk of a mechanical or tautological interpretation, the congestion index was kept separate from the components used to construct RL strength, OLStab, logistics efficiency, and DLS. Congestion was represented by downstream pressure signals such as queueing, blocking, dwell accumulation, and yard saturation, whereas OLStab used volatility and adjustment signals and logistics efficiency used execution-quality signals. The estimated paths are therefore not definitional identities, although they remain simulator-internal associations that require cautious interpretation and external validation.

Nevertheless, this separation should be understood as a design-based reduction in tautology risk rather than a complete removal of interdependence. Because congestion, OLStab, and logistics efficiency are all generated within the same inland-terminal simulator, the constructs may still share underlying operational dynamics such as queue evolution, flow coordination, resource pressure, and blocking behavior. Therefore, the model does not claim that the variables are fully independent empirical constructs. Instead, it treats them as analytically separated but operationally related simulator-derived indices.

The linkage between the simulator and the structural model was handled explicitly. At the control layer, a single RL controller interacted sequentially with terminal states, selected actions, received rewards, and updated its policy over repeated iterations. The aim of this algorithmic layer was not to claim universal superiority of one RL family over all alternatives, but to generate repeated observations from a learning-enabled inland-terminal environment and then test whether the resulting operational pattern matched the proposed theoretical pathway. Because the design uses one primary RL controller together with simple benchmark rules, robustness is framed as within-design sensitivity rather than as an exhaustive cross-algorithm competition.

Before hypothesis testing, the episode-level dataset was screened for completeness, duplication, range validity, and anomalous cases. Given the simulator-based design, near-

complete data were expected after preprocessing, but checks were still performed to confirm that no episodes were duplicated and that normalized indicators remained within their intended bounds. Outliers were examined using visual inspection and Isolation Forest because rare congestion episodes may represent theoretically meaningful operating states rather than recording mistakes. Flagged cases were therefore not removed mechanically; instead, they were used as part of robustness re-estimation.

Diagnostic testing preceded formal hypothesis estimation. Descriptive statistics were complemented by skewness, kurtosis, Shapiro–Wilk tests, Q–Q assessment, Pearson and Spearman correlations, and variance inflation factor analysis. Because simulation-derived variables can depart from strict normality even when the underlying relationships are linear, assumption testing focused on both raw distributions and regression residuals. Residual independence, heteroskedasticity, and multicollinearity were evaluated through Durbin–Watson, Breusch–Pagan, and VIF statistics, and inferential emphasis was placed on confidence intervals and bootstrap procedures where appropriate.

The hypothesis tests were estimated using ordinary least squares regression, serial mediation, and moderated regression. To make the model specification reproducible and to avoid repeated equations, all substantive construct and hypothesis equations are presented once below.

$$x_{ij}^* = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (1)$$

$$OLStab_i = \frac{(1 - DV_i^*) + (1 - QV_i^*) + (1 - AM_i^*)}{3} \quad (2)$$

$$LE_i = \frac{TP_i^* + TA_i^* + CQ_i^* + RU_i^*}{4} \quad (3)$$

$$OLStab_i = \beta_0 + \beta_1 RL_i + \varepsilon_i \quad (4)$$

$$LE_i = \beta_0 + \beta_1 OLStab_i + \varepsilon_i \quad (5)$$

$$CONG_i = \beta_0 + \beta_1 LE_i + \varepsilon_i \quad (6)$$

$$OLStab_i = \alpha_0 + a_1 RL_i + \varepsilon_{1i} \quad (7)$$

$$LE_i = \beta_0 + a_2 RL_i + d_{21} OLStab_i + \varepsilon_{2i} \quad (8)$$

$$CONG_i = \gamma_0 + c' RL_i + b_1 OLStab_i + b_2 LE_i + \varepsilon_{3i} \quad (9)$$

$$OLStab_i = \beta_0 + \beta_1 RL_{ci} + \beta_2 DLS_{ci} + \beta_3 (RL_{ci} \times DLS_{ci}) + \varepsilon_i \quad (10)$$

Equations (1)–(3) define the normalized indicators and formative indices. Equations (4)–(6) test the direct paths for H1–H3. Equations (7)–(9) specify the serial mediation model used for H4, and Equation (10) tests the DLS moderation path for H5. In these equations, i indexes simulation episodes; RL denotes reinforcement learning strength; $OLStab$ denotes operational learning stability; LE denotes logistics efficiency; DLS denotes decision latency sensitivity; and $CONG$ denotes congestion. The terms β_0 , α_0 , and γ_0 denote intercepts; β , a , b , c' , and d_{21} denote estimated coefficients; the subscript c indicates mean-centered variables; and ε denotes the error term. Because H5 is a secondary boundary-condition test, DLS is interpreted cautiously and is not treated as the

defining claim of the study. All preprocessing and statistical analyses were conducted in a reproducible computational workflow suitable for simulation-generated tabular data, using Python-based data handling, diagnostics, regression estimation, and robustness routines. To keep the paper aligned with its central contribution, the main text emphasizes the direct-effect regressions, the serial mediation model, and the moderation test. Additional procedures such as SHAP, PCA, clustering, and related machine-learning checks are treated only as limited supplementary diagnostics; they are not presented as independent external validation of the simulator-generated mechanism.

Methodological rigor was further strengthened through construct mapping, distinctiveness checks, and sensitivity-oriented interpretation, especially for OLStab and DLS. Because these variables are composite episode-level indices rather than survey scales, the primary validity argument is conceptual and formative: the components were assigned to OLStab, logistics efficiency, DLS, and congestion on theoretical grounds and then examined empirically for redundancy and separation. The study therefore interprets the mediation results as evidence consistent with a simulator-level mechanism, not as proof that the same mechanism has been externally validated in field data.

The analytical workbook containing the 500 simulated episode-level records and the corresponding statistical outputs is available from the corresponding author upon reasonable request, subject to the journal's data-sharing policy. As the dataset was generated through simulation and contains no confidential terminal-level information, a de-identified version may be shared to support replication of the reported analyses.

4. Results

The results are reported in the same sequence as the analytical design: data screening and diagnostics, bivariate associations, direct-effect regressions, the serial mediation test, the moderation model, and supplementary robustness checks.

Data-quality screening indicated that the episode-level dataset was suitable for formal hypothesis testing. The final analytical sample contained 500 simulation episodes with no missing values or duplicate rows. The observed ranges were consistent with the intended construction of the indices, with RL, OLStab, logistics efficiency, and DLS bounded between 0 and 1 and congestion remaining non-negative. Although Shapiro–Wilk tests suggested departures from strict normality for some raw indicators, the regression residuals exhibited satisfactory behavior and the remaining diagnostics did not indicate serious threats to inference. VIF values showed overlap among the central constructs but not at a level that would invalidate interpretation. Overall, the diagnostic evidence supports the use of the episode-level dataset for the mechanism tests that follow.

Table 4 summarizes the descriptive statistics for the main variables used in the study, based on 500 simulated episodes. The results show that RL, OLStab, logistics efficiency, and DLS are all within the expected normalized range from 0 to 1, while congestion remains non-negative. The mean values indicate moderate levels of RL strength, operational learning stability, logistics efficiency, and decision latency sensitivity across the simulated episodes. Skewness and kurtosis values are generally within acceptable ranges, suggesting that the variables do not show extreme distributional problems. Overall, the descriptive statistics indicate that the simulation-generated dataset is suitable for the subsequent regression, mediation, moderation, and robustness analyses. The bivariate association pattern was consistent with the conceptual model. RL was strongly and positively associated with OLStab, and OLStab was strongly and positively associated with logistics efficiency. Congestion was negatively associated with RL, OLStab, and logistics efficiency, indicating that stronger learning, more stable operating adjustment, and better execution quality all coincided with lower congestion pressure. By contrast, DLS demonstrated negligible bivariate associations with RL and congestion, while its associations with OLStab

and logistics efficiency were minor. This pattern indicates that DLS is not strongly integrated within the core RL–OLStab–LE–congestion pathway at the bivariate level. However, the subsequent moderation model reveals that DLS possesses a small yet statistically significant positive direct association with OLStab, whereas the interaction between RL and DLS remains non-significant. Consequently, DLS is interpreted as a weak context-specific timing condition rather than a robust moderator of the RL–OLStab relationship.

Table 4. Construct-mapping check used to reduce circularity concerns.

Construct	Model Role	Main Signal Meaning	Separation Logic
RL strength	Independent variable	Learning-enabled control intensity/quality	Not defined from congestion outcome
OLStab	First mediator	Adjustment stability, lower volatility, smoother queue evolution	Stability process, not throughput performance
Logistics efficiency	Second mediator	Throughput, turnaround, coordination, resource utilization	Execution quality, not adjustment volatility itself
DLS	Moderator	Sensitivity to delayed observation/decision/execution	Contextual timing condition
Congestion	Outcome	Queueing, blocking, dwell accumulation, yard saturation	Downstream system pressure

Table 5 shows strong support for H1–H3. RL positively predicted OLStab, OLStab positively predicted logistics efficiency, and logistics efficiency negatively predicted congestion. The pattern is consistent with the proposed sequence from learning-based control to stability, from stability to execution quality, and from execution quality to congestion mitigation. On the other hand, Figure 2 presents the correlation results among the key study variables using both Pearson and Spearman correlation tests. The figure shows that RL is strongly and positively associated with OLStab, and OLStab is strongly and positively associated with logistics efficiency. Congestion is negatively associated with RL, OLStab, and logistics efficiency, supporting the proposed direction of the study model. In contrast, DLS shows only weak relationships with the main variables, suggesting that decision latency sensitivity plays a limited role in the core mechanism. Overall, the correlation results provide preliminary support for the serial pathway from RL to OLStab, from OLStab to logistics efficiency, and from logistics efficiency to lower congestion. The strong OLStab–logistics efficiency correlation indicates that the two constructs are closely related in the simulator. This is theoretically plausible because stable adjustment can support better execution quality; however, it also requires caution. The high correlation may partly reflect construct proximity or shared simulator structure. Therefore, the subsequent mediation results are interpreted cautiously as simulation-level evidence consistent with the proposed mechanism, not as conclusive proof of complete discriminant separation between OLStab and logistics efficiency.

Table 5. Descriptive statistics for key study variables.

Variable	N	Mean	SD	Min	Max	Skewness	Kurtosis
RL	500	0.492	0.287	0.002	0.999	0.062	−1.193
OLStab	500	0.574	0.204	0.016	1.000	0.009	−0.769
LE	500	0.507	0.169	0.040	0.987	0.156	−0.352
DLS	500	0.514	0.293	0.004	0.999	−0.006	−1.266
Congestion	500	66.258	6.154	46.008	82.849	−0.152	−0.257

Table 6 and Figure 3 provide the direct-effect evidence for H1–H3, while the serial mediation results are summarized in Table 7 and Figure 4. Within the simulated design,

RL significantly increased OLStab, and OLStab significantly increased logistics efficiency. When RL, OLStab, and logistics efficiency were entered jointly, OLStab and logistics efficiency remained significant negative predictors of congestion, whereas the direct RL effect became non-significant. Bootstrap estimates show a significant indirect effect through OLStab and a significant negative serial indirect effect through OLStab and logistics efficiency, while the path through logistics efficiency alone was not significant. Overall, the results are consistent with the interpretation that, in this designed simulator, RL is associated with lower congestion chiefly through improved operational stability and logistics execution rather than through a strong residual direct effect. However, this interpretation remains limited to a theory-ordered statistical pathway among simulator-derived episode-level indices. Because OLStab, logistics efficiency, and congestion are exported from the same completed simulation episode, the mediation result should not be read as proof of a temporally separated causal sequence. Instead, it indicates that the simulated data are statistically consistent with the proposed ordering of adjustment stability, execution quality, and congestion pressure.

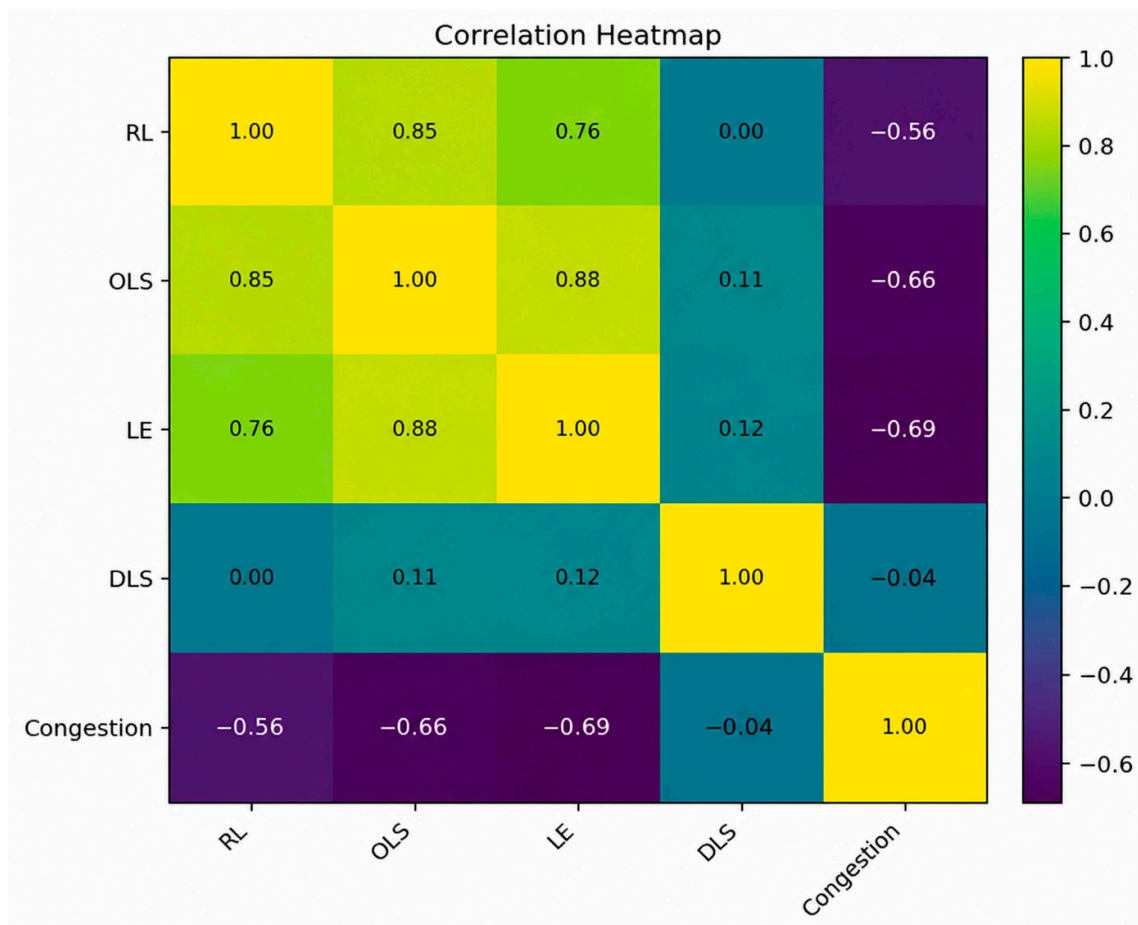


Figure 2. Pearson correlations and *p*-values and Spearman correlations and *p*-values. Source: Authors' own elaboration using Python 3.13.

Table 6. Correlation analysis (Pearson above diagonal; Spearman below diagonal).

Variable	RL	OLStab	LE	DLS	Congestion
RL	1.000	0.852 ***	0.760 ***	0.001	−0.557 ***
OLStab	0.860 ***	1.000	0.880 ***	0.107 *	−0.662 ***
LE	0.771 ***	0.883 ***	1.000	0.116 *	−0.688 ***
DLS	0.003	0.101 *	0.117 **	1.000	−0.036
Congestion	−0.564 ***	−0.652 ***	−0.677 ***	−0.030	1.000

Note: Upper triangle = Pearson; lower triangle = Spearman. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 7. Direct-effect regressions for H1–H3.

Hypothesis	Dependent Variable	Predictor	Coefficient	SE	t	p	95% CI	R ²
H1	OLStab	RL	0.606	0.017	36.382	<0.001	[0.573, 0.639]	0.727
H2	LE	OLStab	0.728	0.018	41.248	<0.001	[0.694, 0.763]	0.774
H3	Congestion	LE	−25.096	1.185	−21.183	<0.001	[−27.424, −22.768]	0.474

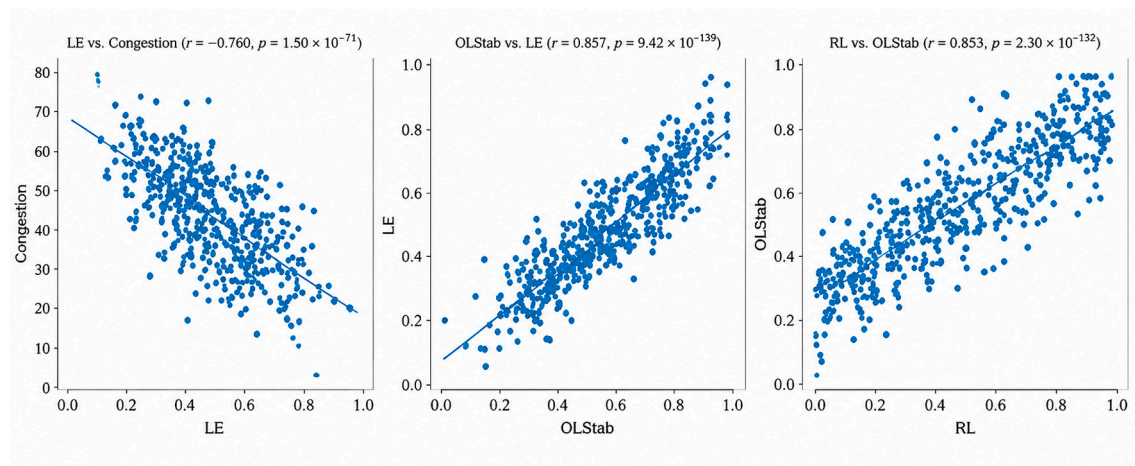


Figure 3. Direct-effect relationships among the core study variables. Source: Authors' own conception using Python 3.13.

However, this finding should be interpreted cautiously. Because RL, OLStab, and logistics efficiency are strongly correlated simulator-derived indices, the disappearance of the direct RL effect after including OLStab and logistics efficiency may partly reflect shared variance, multicollinearity, or variance absorption among closely related episode-level constructs. Therefore, the non-significant direct RL coefficient should not be treated as definitive proof of full mediation. Instead, it indicates that, within the designed simulator, the data are statistically consistent with a theory-ordered pathway in which RL-related variation is largely captured by operational stability and logistics execution indicators.

Serial Mediation Path Diagram (H4)

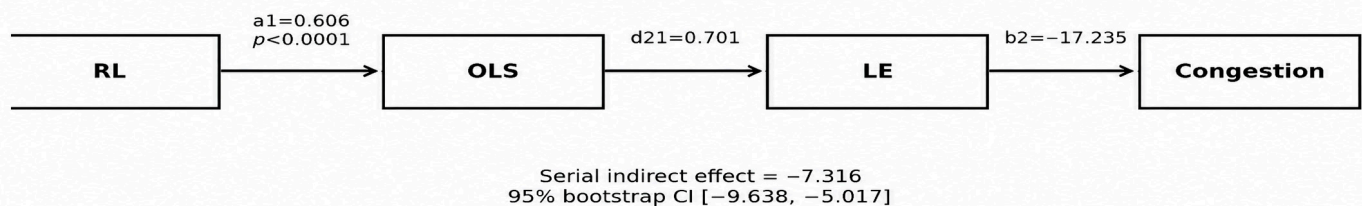


Figure 4. Serial mediation model for H4. Source: Authors' own elaboration using Python 3.13.

Figures 4 and 5 collectively summarize the simulation-level evidence for serial mediation in relation to H4. Figure 4 illustrates the pathway through which RL is associated with congestion: RL enhances OLStab, OLStab improves logistics efficiency, and logistics efficiency is associated with lower congestion. The direct path from RL to congestion becomes non-significant once OLStab and logistics efficiency are included, but this should be interpreted as evidence consistent with serial mediation within the designed simulator, not as definitive proof of full mediation in real terminal operations. Figure 5 displays the bootstrap distribution of the serial indirect effect; the confidence interval does not include zero, supporting the statistical presence of the proposed indirect path in the simulated data. The moderation analysis did not support H5. Although the model was significant overall, the $RL \times DLS$ interaction was small and non-significant, indicating that the RL-to-OLStab relationship did not vary meaningfully across the observed range of decision latency sensitivity. In this inland-terminal simulation, DLS therefore appears to be too weak or too context-specific to function as a reliable boundary condition. Table 8 reports the moderation results for H5 and confirms that H5 is not supported.

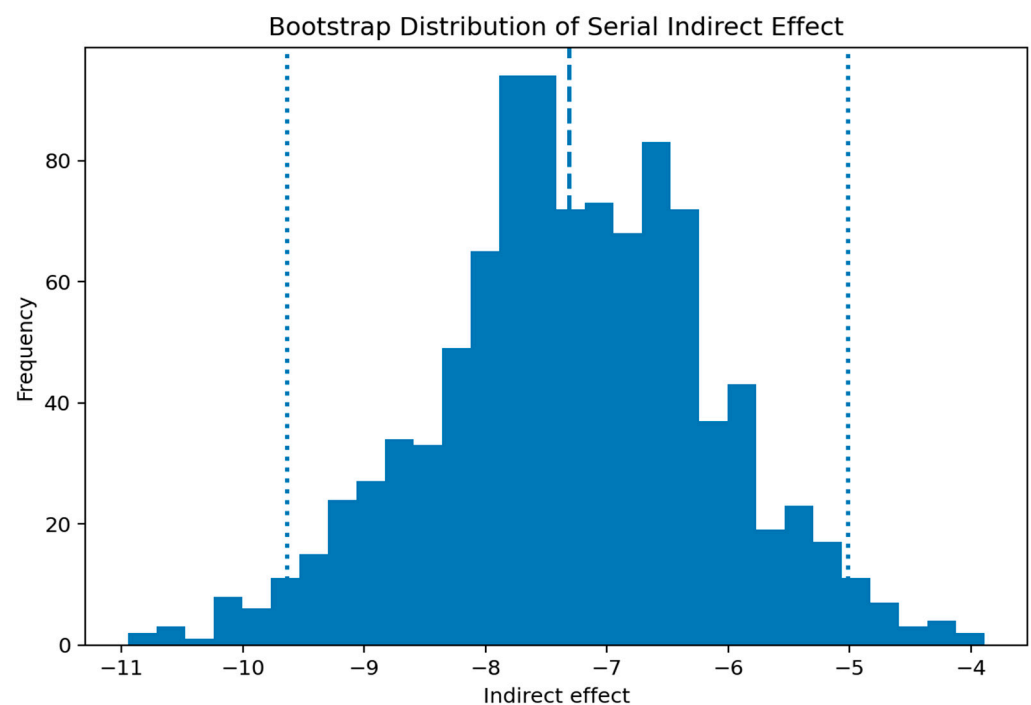


Figure 5. Bootstrap distribution of the serial indirect effect. Source: Authors' own elaboration using Python 3.13.

Table 8. Serial mediation results for H4.

Path/Effect	Coefficient	<i>p</i>	95% Bootstrap CI
a1: RL → OLStab	0.606	<0.001	—
a2: RL → LE (controlling for OLStab)	0.023	0.335	—
d21: OLStab → LE (controlling for RL)	0.701	<0.001	—
b1: OLStab → Congestion (controlling for RL, LE)	−8.516	<0.001	—
b2: LE → Congestion (controlling for RL, OLStab)	−17.235	<0.001	—
c': RL → Congestion (controlling for OLStab, LE)	0.916	0.488	—
c: RL → Congestion (total effect)	−11.958	<0.001	—
Indirect via OLStab	−5.159	—	[−7.970, −2.046]
Indirect via LE	−0.399	—	[−1.217, 0.362]
Serial indirect RL → OLStab → LE → Congestion	−7.316	—	[−9.638, −5.017]

As indicated in Table 8, the moderation analysis did not support H5. Although the overall moderation model was statistically significant, the interaction term between RL and DLS was small and statistically non-significant. This indicates that decision latency sensitivity did not meaningfully change the strength or direction of the RL–OLStab relationship within the observed range of the simulated inland-terminal environment. In other words, the results do not show that the effect of reinforcement learning on operational learning stability becomes stronger or weaker under different levels of DLS. Therefore, the empirical evidence does not support the original moderation expectation.

The direct coefficient of DLS, however, was small and positive. This finding requires careful interpretation because it does not align with the initial theoretical expectation that greater latency sensitivity would impair operational stability. The positive direct association should therefore not be interpreted as evidence that latency sensitivity improves OLStab. Such an interpretation would be theoretically misleading because the original logic of DLS was based on the assumption that delayed observation, delayed decision execution, or delayed feedback should make operational adjustment more difficult. Instead, the result suggests that the present operationalization of DLS may not purely capture harmful latency exposure.

A more cautious interpretation is that, within the present simulator, the DLS index may partly reflect context-specific operating conditions in which latency-sensitive environments are accompanied by buffering capacity, standardized dispatch routines, conservative adjustment behavior, or latency-management mechanisms. These conditions may coexist with more stable operational adjustment even when latency sensitivity is present. Therefore, the positive DLS coefficient is best understood as a simulator-specific direct association rather than as confirmation of a beneficial latency effect.

Taken together, the non-significant $RL \times DLS$ interaction and the small positive direct DLS coefficient indicate that DLS does not function as a theoretically robust boundary condition in this study. DLS appears to be too weak, too context-dependent, or too broadly operationalized to reliably explain variation in the RL–OLStab relationship. Accordingly, H5 is not supported. The finding should be treated as evidence that the DLS construct requires further refinement, particularly through more direct measures of observation delay, decision delay, execution lag, and feedback delay in future simulation or field-based studies.

Figure 6 illustrates the moderating effect of decision latency sensitivity (DLS) on the relationship between reinforcement learning (RL) and operational learning stability (OLStab). The figure indicates a positive association between RL and OLStab across various levels of DLS. However, the slopes do not demonstrate a meaningful difference, which corroborates the statistical finding that the interaction between RL and DLS is not significant. Consequently, the figure visually confirms that DLS does not notably weaken or strengthen the RL–OLStab relationship in the simulated inland-terminal context. This suggests that DLS serves more as a context-specific condition rather than as a strong moderator in this study. The moderation analysis and figure generation were performed using Python 3.13. As a robustness check, a full-model regression was estimated with RL, OLStab, logistics efficiency, and DLS entered simultaneously as predictors of congestion. The overall model remained significant, but only OLStab and logistics efficiency retained significant negative effects, whereas RL and DLS were not significant. This pattern is consistent with the mediation results and indicates that, once the internal capability pathway is modeled explicitly, RL no longer shows a distinct residual direct effect on congestion.

Table 9 presents the full-model regression conducted as a within-simulator sensitivity check, with RL, OLStab, logistics efficiency, and DLS included as predictors of congestion. The findings indicate that the overall model is statistically significant. However, only OLStab and logistics efficiency emerge as significant negative predictors of congestion,

while RL and DLS do not exhibit significant direct effects. This result is consistent with the primary mediation findings, but it should be interpreted as internal robustness evidence rather than independent validation. Consequently, the full-model result supports a cautious reading that congestion in this simulation is mainly explained by operational learning stability and logistics efficiency. The analysis was performed using Python 3.13.

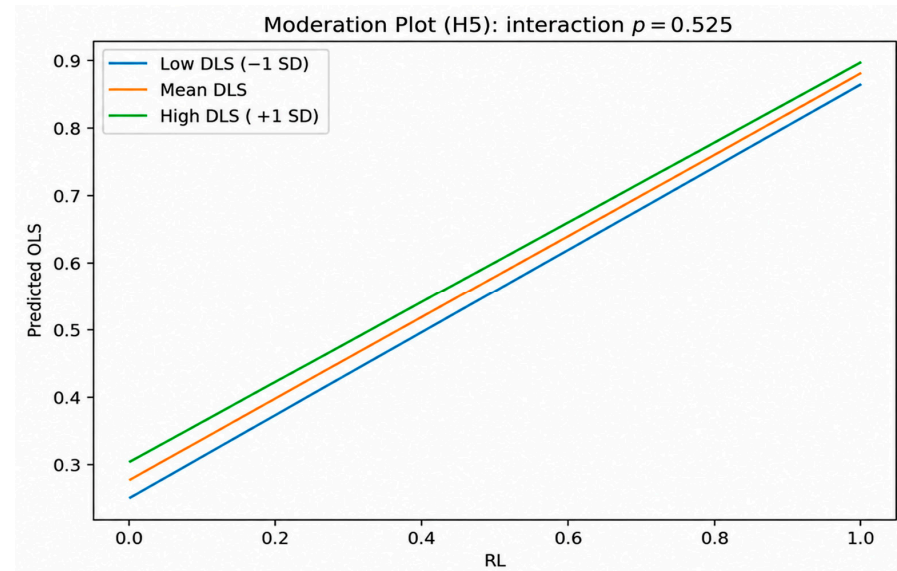


Figure 6. Moderation effect of decision latency sensitivity on the RL–OLStab relationship (H5). Source: Authors’ own creation using Python 3.13 output.

Table 9. Moderation results for H5.

Term	Coefficient	SE	t	p	95% CI
Constant	0.574	0.005	122.640	<0.001	[0.565, 0.583]
RL (centered)	0.605	0.016	36.939	<0.001	[0.573, 0.637]
DLS (centered)	0.073	0.016	4.578	<0.001	[0.042, 0.105]
RL × DLS	−0.036	0.057	−0.637	0.525	[−0.148, 0.075]

Model summary: $R^2 = 0.738$; adjusted $R^2 = 0.737$; $F = 466.098$; $p < 0.001$. Decision: H5 not supported.

To further examine the stability of the proposed mechanism, a full-model regression was estimated as a robustness check by including the main predictors jointly in the congestion model. As reported in Table 10, logistics efficiency remains a significant negative predictor of congestion, while the direct effect of reinforcement learning becomes non-significant after OLStab and logistics efficiency are included. This pattern supports the interpretation that the relationship between reinforcement learning and congestion is mainly explained through the proposed intermediate mechanism rather than through a residual direct effect. Overall, the empirical pattern is coherent and internally consistent within the simulation design. The diagnostics indicate that the episode-level data are usable, the correlation structure aligns with the theory, the direct-effect models support H1–H3, and the serial mediation analysis is consistent with H4 in this designed environment. By contrast, H5 is not supported. The most stable and substantively important predictors of congestion are the internal process variables—OLStab and logistics efficiency—rather than RL or DLS as stand-alone direct drivers. This pattern is consistent with, but does not prove, a sequential capability-development process in inland-terminal congestion mitigation.

Table 10. Full-model regression for robustness.

Term	Coefficient	SE	t	p	95% CI
Constant	79.069	0.723	109.355	<0.001	[77.649, 80.490]
RL	1.282	1.338	0.958	0.338	[−1.347, 3.912]
OLStab	−8.973	2.546	−3.524	<0.001	[−13.976, −3.970]
LE	−17.439	2.463	−7.081	<0.001	[−22.278, −12.600]
DLS	1.072	0.689	1.555	0.120	[−0.282, 2.425]

Model summary: $R^2 = 0.491$; adjusted $R^2 = 0.487$; $F = 119.289$; $p < 0.001$.

4.1. Limited Supplementary Diagnostics

As a limited supplementary diagnostic, SHAP values were used to examine the relative predictive contribution of the four focal predictors in the congestion model. Logistics efficiency produced the largest mean absolute SHAP value, followed by OLStab, whereas DLS and RL contributed much less to model output. This ordering is consistent with the full-model regression and with the mechanism logic of the paper, but it remains a description of the simulator-generated data rather than independent evidence that the mechanism holds in field settings.

These supplementary results are therefore interpretively supportive rather than constitutive. They do not replace the theory-driven mediation analysis and should not be read as independent validation; they only show that the predictive profile of the simulated data is aligned with the proposed process explanation.

4.2. Descriptive Latent-Structure Checks: PCA and Clustering

As a limited supplementary assessment, principal component analysis and clustering were used to describe the internal structure of the episode-level dataset beyond the main regression-based results. The PCA results show that the first principal component explained 64.4% of total variance, while the second explained an additional 20.0%, so that the first two components jointly captured 84.5% of the overall variation. When the third component was added, cumulative explained variance increased to 94.0%. These results indicate that the simulated data have a compact internal structure, but they do not provide external validation of the constructs or of the mediation mechanism.

Figure 7 illustrates the PCA scatter plot for the simulated episode-level dataset. The plot displays the distribution of 500 simulation episodes across the first two principal components, which summarize the primary variation among RL, OLStab, logistics efficiency, DLS, and congestion. The observed clustering pattern suggests internal structure in the simulated data, but it should be interpreted descriptively rather than as independent confirmation of the theoretical model. Under the K-means ($k = 2$) specification, the silhouette coefficient was 0.375, indicating a moderate degree of separation between clusters rather than a sharply polarized structure. In parallel, DBSCAN identified a noise proportion of 0.01, showing that only a very small share of episodes behaved as isolated or anomalous observations. Taken together, the PCA-cluster evidence is treated as a supplementary description of the simulation output, not as a substitute for external validation or real terminal evidence.

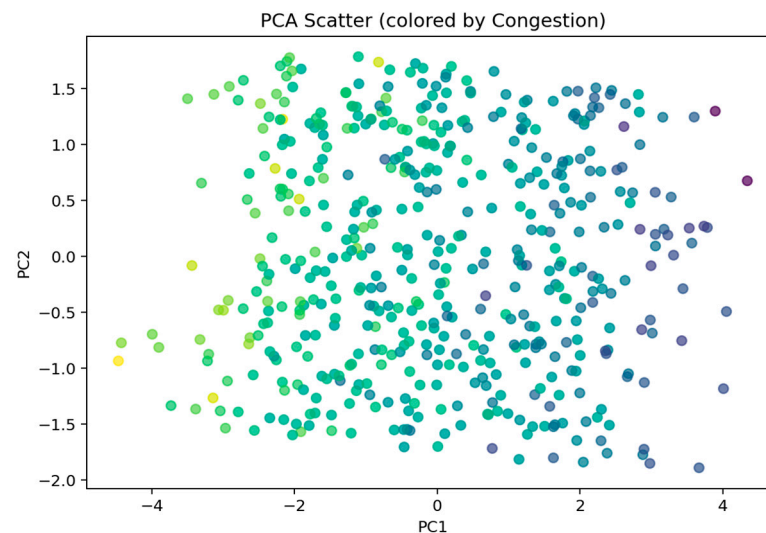


Figure 7. PCA scatter plot of observations based on PC1 and PC2. Each point is colored according to the normalized Congestion value, with darker colors indicating higher congestion and lighter colors indicating lower congestion. Source: Authors' own creation using Python 3.13 production.

4.3. Construct Clarification: Operational Learning Stability (OLStab)

Operational Learning Stability (OLStab) is conceptualized in this study as an episode-level stability index derived from observable controller behavior and performance trajectories rather than as a latent perceptual construct. The purpose of this operationalization is pragmatic: to capture the extent to which a learned policy remains consistent, low-volatility, and operationally manageable under conditions of variability and disruption. Accordingly, OLStab should be interpreted as a simulation-grounded operational metric developed for the present study, rather than as a universally established theoretical construct.

This approach enables OLStab to be measured using routine runtime signals generated within the simulation environment, including action-change frequency, adjustment magnitude, and variability in congestion-related cost. It also enhances transparency and reproducibility by linking the construct directly to observable system behavior. At the same time, the broader construct validity of OLStab should be interpreted cautiously, as its generalizability beyond the present simulation setting remains to be established. Future research should therefore validate the construct using real terminal operational logs, alternative formulations of stability, and independent datasets from learning-enabled logistics environments.

Further clarification concerns the relationship between OLStab and logistics efficiency. Although OLStab is operationalized as an adjustment-stability index and logistics efficiency is operationalized as an execution-quality index, both are generated within the same simulation environment and may therefore share part of the same underlying operational dynamics. For this reason, the distinction between OLStab and logistics efficiency should be interpreted as an analytical and simulator-level separation, not as definitive evidence of full discriminant validity between two independently validated field constructs. The high association between the two constructs may reflect the operational reality that smoother adjustment and better execution quality are closely connected in freight-terminal processes, but it may also partly reflect common simulator structure. Therefore, OLStab is best understood as a cautious, simulation-derived stability indicator that requires further validation through independent operational records, time-separated measures, and alternative component definitions.

5. Discussion

The findings are consistent with the interpretation that reinforcement learning is associated with congestion mitigation in the simulated inland freight terminal primarily through an indirect operating pathway. RL improves OLStab, OLStab improves logistics efficiency, and logistics efficiency reduces congestion; once both mediators are modeled, the direct RL-to-congestion path is no longer significant. This is an informative simulation-based result because it suggests that the value of RL in this setting may lie less in immediate queue suppression and more in the development of a steadier and more manageable operating pattern. However, the result remains internal to the designed simulator and should not be interpreted as field-validated proof of full mediation. In addition, the disappearance of the direct RL-to-congestion path should be interpreted cautiously because it may partly reflect shared variance, construct overlap, or variance absorption among closely related simulator-derived indices rather than a fully independent causal mediation process.

This process-based reading sharpens the interpretation of learning-enabled logistics control within the modeled inland-terminal environment. Much of the transportation RL literature emphasizes improved outcomes such as lower delay, better dispatch quality, or reduced dwell time. The present study complements that literature by clarifying a possible simulator-level sequence through which such gains may emerge inside terminal operations. In the inland-terminal setting modeled here, RL appears to matter because it first stabilizes internal adjustment and only then translates into stronger execution performance. The paper therefore advances a mechanism-oriented simulation explanation rather than an outcome-only or field-general claim. At the same time, this explanation should be understood as a theory-ordered statistical pathway among episode-level simulator outputs, not as direct evidence that OLStab, logistics efficiency, and congestion occur as temporally separated stages in real terminal operations.

The discussion also helps differentiate this study from superficially similar simulation papers. Even when related studies share a broad interest in RL and congestion, the present article is organized around a distinct inland-terminal application, a simulator centered on inland-node coordination frictions, and an empirical pattern in which the serial pathway is dominant while the DLS moderation effect is not. These context-specific features are central to the interpretation of the results and help position the paper as a complementary application rather than a duplicate presentation. Nevertheless, because OLStab and logistics efficiency are strongly related and generated within the same simulation environment, the study does not claim that they are fully independent field-validated constructs. Their relationship should instead be interpreted as an analytically separated but operationally adjacent simulator-level distinction between adjustment stability and execution quality.

The DLS finding requires caution. The study expected DLS to weaken the RL–OLStab relationship, but the interaction effect was not significant. In addition, the small positive direct association between DLS and OLStab does not align with the original theoretical expectation. This suggests that the DLS index may not purely represent harmful latency exposure in the current simulator. Instead, it may also reflect operating contexts in which latency-sensitive conditions are accompanied by buffering, standardized dispatch routines, or compensating control logic. For this reason, DLS should be interpreted as a weak, simulator-specific contextual indicator rather than as a confirmed moderator. The positive DLS coefficient should therefore not be interpreted as evidence that latency sensitivity improves operational learning stability. Rather, it indicates that the current DLS operationalization may require further refinement before it can be used as a robust theoretical boundary condition.

The systems' interpretation of the findings should also remain cautious. Although congestion is described conceptually as an emergent outcome of interacting gate, yard, buffer, and dispatch subsystems, the empirical analysis relies mainly on regression-based mediation and moderation models. Therefore, the present study does not directly test emergence through feedback loops, nonlinear adaptation, dynamic interactions, or system-dynamics behavior. The systems' contribution is consequently limited to a simulation-informed mechanism interpretation of related operational indices, not a full dynamic-systems explanation of congestion emergence.

Taken together, the simulation-informed findings refine the theoretical role of RL only within the modeled inland-terminal setting. RL should be viewed here as a potential contributor to a simulator-level process sequence whose operational value depends on whether it creates stable routines and better execution quality. Congestion reduction is interpreted as the downstream consequence of that simulated sequence, not simply the direct footprint of algorithm adoption and not a universal claim about all terminal environments. Overall, the discussion supports a cautious interpretation: the model provides a theory-ordered explanation that is internally consistent within the designed simulator, but it requires external validation with real terminal data, time-separated operational measures, and alternative simulator designs before stronger theoretical or practical claims can be made.

6. Theoretical Contributions

This study offers four simulation-informed theoretical contributions, all bounded by the designed inland-terminal simulator. First, it extends the RL-for-logistics literature by shifting the analytical focus from performance comparison alone to mechanism explanation. The paper shows that, within this simulation architecture, congestion mitigation can be examined as a staged operating process rather than as a simple direct outcome of algorithmic control. However, this contribution should be interpreted as a theory-ordered simulator-level explanation, not as externally validated evidence that the same mechanism necessarily operates in real inland freight terminals.

Second, it advances operational learning stability (OLStab) as a simulation-grounded intermediate metric in learning-enabled logistics analysis. By distinguishing OLStab from logistics efficiency, the paper identifies a possible stability layer linking adaptive learning to realized execution quality, while acknowledging that broader construct validity requires external validation. At the same time, the contribution is limited by the operational proximity between OLStab and logistics efficiency. Because both constructs are generated within the same simulation environment and are strongly correlated, the study should not be interpreted as proving full discriminant validity between adjustment stability and execution quality. Rather, it proposes OLStab as an analytically useful simulator-derived stability index that future research should validate using independent operational data, time-separated measures, and alternative component definitions.

Third, it contributes to inland freight terminal research by showing that, in the modeled environment, congestion is more usefully examined as the downstream result of internal process stability and execution quality than as an isolated bottleneck phenomenon. The serial mediation structure therefore adds a context-specific explanatory model for the inland-terminal simulation examined here. Nevertheless, this contribution should be understood cautiously because the serial pathway is estimated from same-episode simulator-derived indices. The disappearance of the direct RL-to-congestion effect after including OLStab and logistics efficiency may partly reflect shared variance, construct overlap, or variance absorption among related simulator outputs. Therefore, the contribution lies in proposing a theory-ordered statistical pathway for further testing, not in establishing definitive causal mediation in real terminal operations.

Fourth, the unsupported DLS moderation result contributes mainly by narrowing rather than confirming the boundary-condition argument. It indicates that timing-related contextual effects should be specified more narrowly and tested under stronger variation, rather than assumed to operate uniformly across learning-enabled logistics environments. Importantly, the small positive direct association between DLS and OLStab does not support the original expectation that latency sensitivity should weaken operational stability. This result suggests that DLS, as operationalized in the present simulator, may partly capture buffering, standardized dispatch routines, or latency-management capacity rather than harmful delay sensitivity alone. Therefore, the theoretical contribution of DLS is not the confirmation of a robust moderator, but the identification of a measurement and conceptual boundary that future studies should refine using direct measures of observation delay, decision delay, execution lag, and feedback delay.

Overall, the theoretical contribution of this study is best understood as a cautious, simulation-informed framework for examining how learning-enabled control, operational stability, execution quality, and congestion pressure may be statistically related within a designed inland-terminal environment. The study does not provide a full systems-theoretical explanation of congestion emergence through feedback loops, nonlinear adaptation, or dynamic interaction effects. Future research should therefore extend this framework using field data, digital-twin experiments, agent-based models, system-dynamics models, or dynamic queueing approaches that can more directly evaluate emergent congestion behavior in freight-terminal systems.

7. Practical Implications

From a practical perspective, the simulation-informed results suggest that investments in reinforcement learning or other adaptive decision technologies should be justified not only on the basis of technological sophistication, but on their ability to support stable operational routines within a controlled decision environment. The findings suggest cautiously that congestion reduction is unlikely to arise simply because a learning algorithm is introduced. Benefits are more likely to emerge when adaptive systems reduce oscillatory dispatching, smooth queue evolution, and improve execution consistency across terminal processes. Managers may therefore evaluate learning systems according to whether they create operational stability first, because stability appears to be the immediate pathway through which downstream efficiency and congestion benefits are realized in the simulated inland-terminal setting. However, this implication should be interpreted as simulation-informed guidance rather than as field-validated managerial evidence from an operating terminal.

A second simulation-informed implication is that operational policy should explicitly reinforce process consistency and coordination discipline. In practice, this means strengthening the routine alignment of yard handling, gate scheduling, truck appointment adherence, dispatch coordination, and feedback-based replanning. It also means designing monitoring systems that detect instability early, such as abrupt workload shifts, repeated rescheduling cycles, or queue volatility across terminal subsystems. Where learning-based systems are deployed, managers may need to ensure that the surrounding operational architecture supports consistent execution, because unstable routines can dilute the performance gains of otherwise capable technologies. At the same time, because OLStab and logistics efficiency are closely related simulator-derived indices in this study, these managerial implications should not be interpreted as proving that adjustment stability and execution quality are fully separable field-level capabilities. Rather, they suggest that terminal managers should jointly monitor stability-related signals and execution-quality indicators when evaluating learning-enabled decision systems.

Third, the results suggest, within the limits of the simulation design, that managers should prioritize efficiency-building mechanisms rather than relying on technological adoption alone. Logistics efficiency emerged as the strongest direct driver of lower congestion, while the full-model results also showed that operational learning stability and logistics efficiency remained significant when reinforcement learning itself no longer had a distinct direct effect. This suggests that managerial emphasis should be placed on throughput reliability, turnaround performance, coordination quality, and resource utilization. Technology should therefore be treated as an enabler of these mechanisms, not as a substitute for them. A practical strategy is to sequence implementation efforts so that digital or learning-based tools are embedded within broader programs of process redesign, coordination discipline, and execution-quality improvement. Nevertheless, because the results are derived from one designed simulation environment, this recommendation should be viewed as a cautious implementation logic rather than a direct operational prescription.

Finally, decision latency sensitivity should still be monitored, but managers should avoid assuming that it will always operate as a decisive moderator. In the present study, the interaction effect was not significant, which implies that latency-related vulnerabilities may be contingent on the specific operating environment, the range of timing disruptions, and the degree of process standardization. For practitioners, this means latency indicators remain useful as risk-monitoring variables, especially in environments exposed to traffic uncertainty, weather-related arrival deviations, or delayed decision execution. However, managerial attention should remain centered on the more robust leverage points identified here: stabilizing operations and improving logistics efficiency. Importantly, the small positive direct association between DLS and OLStab in the simulation should not be interpreted as evidence that latency sensitivity improves operational stability. A more cautious interpretation is that the DLS index may partly reflect buffering, standardized dispatch routines, or latency-management capacity under certain simulated conditions. Therefore, future managerial application should distinguish harmful delay exposure from organizational or technological capacity to absorb delay before using DLS as a decision-support indicator.

8. Limitations and Future Research

This study has several limitations that should be acknowledged. First, the empirical dataset was generated from 500 simulation episodes rather than collected from real inland freight terminal operations. The results therefore provide simulation-informed evidence about a designed mechanism, not direct field proof from a named terminal or a universal estimate of reinforcement learning effects across all logistics nodes. Accordingly, the findings should be interpreted as evidence from a controlled simulation environment rather than as externally validated operational evidence. Future studies should strengthen external validity by using terminal operating logs, field experiments, digital-twin testbeds, or independently developed simulation benchmarks.

Second, operational learning stability (OLStab) is introduced as a simulation-grounded, formative, episode-level stability index. Although the paper clarifies its distinction from logistics efficiency and explains the operational logic behind its component construction, its broader construct validity remains open to further testing. Future research should examine OLStab using real terminal logs, alternative stability measures, independently validated datasets, and sensitivity checks based on different weighting and normalization schemes. The high correlation between OLStab and logistics efficiency also indicates that future studies should continue to examine discriminant validity, construct redundancy, and the extent to which adjustment stability and execution quality can be separated empirically in operational logistics systems. The high correlation between

OLStab and logistics efficiency remains an important limitation. Although the study separates OLStab as adjustment stability and logistics efficiency as execution quality, both variables are derived from the same simulator and may partially reflect overlapping operational dynamics. Therefore, the present analysis should not be interpreted as definitive evidence that OLStab and logistics efficiency are fully independent field-validated constructs. Rather, the study provides a theory-ordered analytical separation between two closely related simulator-derived indices. Future research should test discriminant validity using alternative component definitions, independent terminal datasets, time-separated operational measures, and robustness checks that reduce possible redundancy between stability and execution-quality indicators.

Third, all focal variables are generated within the same simulator and exported at the same episode-level aggregation stage. This creates a risk that part of the observed serial pathway reflects simulator structure, construct proximity, or model specification rather than fully independent empirical emergence. Therefore, the serial mediation results should be interpreted as a theory-ordered statistical pathway among simulator-derived indices, not as definitive evidence of temporally separated causal development in real terminal operations. Similarly, the non-significant direct RL-to-congestion path after including OLStab and logistics efficiency should not be described as definitive full mediation. It may also reflect shared variance, multicollinearity, or variance absorption among closely related simulator-derived constructs. A further limitation concerns residual tautology and operational interdependence among the simulator-derived constructs. Although the study separates RL strength, OLStab, logistics efficiency, DLS, and congestion by using different component signals and model roles, these variables are still generated from the same simulated operating system. Congestion is operationally related to queueing, blocking, dwell accumulation, coordination, and resource pressure, which are also indirectly connected to the stability and efficiency indicators. Therefore, the analysis reduces circularity concerns but cannot fully eliminate them. Future studies should test the proposed mechanism using independent field measures, time-separated operational records, alternative congestion definitions, and external validation datasets.

Fourth, the operationalization of RL strength remains a limitation that should be addressed in future research. In this study, RL strength is treated as an algorithm-internal learning-control index rather than as a direct congestion, logistics-efficiency, or terminal-performance outcome. Although the manuscript defines RL strength through learning-update intensity, value-improvement behavior, and policy-convergence quality, these components are still produced within the same simulation architecture. Therefore, the RL strength construct should be interpreted as a simulator-specific control index rather than as a universally validated measure of reinforcement learning capability. Future research should improve reproducibility by reporting more detailed algorithm logs, value-function update traces, convergence diagnostics, reward trajectories, policy-stability indicators, and sensitivity tests across alternative RL algorithms. Such work would help determine whether the observed RL-related effects remain stable when RL strength is operationalized through different learning metrics or independently observed algorithmic outputs.

Fifth, the role of decision latency sensitivity (DLS) requires cautious interpretation. The theoretical expectation was that greater decision latency sensitivity would weaken operational stability, but the simulated results show only a small positive direct association with OLStab and no significant $RL \times DLS$ interaction. This means that DLS does not function as a robust moderator in the present model. More importantly, the positive direct association between DLS and OLStab does not align with the expected negative theoretical logic of latency sensitivity. This result should not be interpreted as evidence that latency sensitivity improves operational stability. One possible explanation is that the simulated representation of latency-sensitive operating conditions may also capture more

conservative decision adjustment, buffering, standardized dispatch routines, or latency-management capacity under certain scenarios. Future studies should therefore test DLS using alternative latency specifications, stronger latency shocks, delayed-information structures, asynchronous decision timing, direct measures of observation delay, decision delay, execution lag, feedback delay, and field-based operational data.

Sixth, the World Bank Logistics Performance Index indicators were used only as contextual calibration benchmarks for plausible simulation settings. The low, medium, and high thresholds used in the simulation should be understood as rule-based scenario anchors rather than empirically validated terminal-level calibration parameters. Future research should improve calibration by using real terminal throughput records, arrival distributions, dwell-time data, yard utilization, truck appointment logs, dispatch delays, and congestion-cost observations. The World Bank Logistics Performance Index indicators were used only as contextual calibration benchmarks for plausible simulation settings. The low, medium, and high thresholds used in the simulation should be understood as heuristic, rule-based scenario anchors rather than empirically validated terminal-level calibration parameters. These thresholds were not derived from observed terminal operating logs, and therefore they support contextual plausibility rather than external validation. In addition, macro-logistics indicators have limited direct correspondence with terminal-level operational behavior such as queue evolution, dispatch timing, congestion formation, equipment utilization, yard blocking, and truck turnaround. Future research should improve calibration by using real terminal throughput records, arrival distributions, dwell-time data, yard utilization, truck appointment logs, dispatch delays, and congestion-cost observations.

Seventh, although the manuscript discusses congestion as an emergent systems phenomenon arising from interacting terminal subsystems, the analytical strategy remains primarily regression-based and mediation-oriented. The present study does not directly model emergence through feedback loops, nonlinear adaptation, dynamic interaction effects, or system-dynamics simulation. Therefore, the systems-theoretical contribution should be interpreted cautiously as a simulation-informed mechanism analysis rather than as a full dynamic-systems explanation of congestion emergence. Future studies should extend the present framework by incorporating feedback-loop analysis, nonlinear simulation, agent-based modeling, system-dynamics modeling, dynamic queueing models, or digital-twin experiments that can more directly capture emergent congestion behavior across gate, yard, buffer, and dispatch subsystems.

Finally, the study's conclusions should be interpreted in line with its methodological boundaries. The findings suggest a theory-ordered simulator-level pathway in which RL-related variation is associated with OLStab, logistics efficiency, and congestion, but they do not establish universal causal evidence for real inland freight terminals. The contribution is therefore best understood as a cautious simulation-informed explanation of a designed inland-terminal scenario, not as direct operational validation of reinforcement learning-based congestion mitigation.

Finally, the World Bank Logistics Performance Index indicators were used only as contextual calibration benchmarks for plausible simulation settings. The low, medium, and high thresholds used in the simulation should be understood as heuristic, rule-based scenario anchors rather than empirically validated terminal-level calibration parameters. These thresholds were not derived from observed terminal operating logs, and therefore they support contextual plausibility rather than external validation. In addition, macro-logistics indicators have limited direct correspondence with terminal-level operational behavior such as queue evolution, dispatch timing, congestion formation, equipment utilization, yard blocking, and truck turnaround. Future research should improve calibration by using real terminal throughput records, arrival distributions, dwell-time data, yard

utilization, truck appointment logs, dispatch delays, and congestion-cost observations. Future work should also compare alternative simulator designs, use shared benchmark datasets where available, vary reward and congestion-cost formulations, and examine whether the RL–OLStab–logistics efficiency–congestion pathway remains stable across different inland, dry-port, seaport, and multimodal terminal contexts.

9. Conclusions

This study examined how reinforcement learning relates to congestion mitigation in inland freight terminal operations by testing a serial mechanism in which RL strengthens operational learning stability, OLStab improves logistics efficiency, and logistics efficiency reduces congestion, while DLS was examined as a secondary contextual factor. The simulation-based results support a clear pattern within the designed simulator: RL is positively associated with OLStab, OLStab is positively associated with logistics efficiency, logistics efficiency is negatively associated with congestion, and the serial indirect pathway is significant, whereas the direct RL effect becomes non-significant once the mediators are included. The moderation effect of DLS is not supported. Overall, the study suggests that, within this designed inland-terminal simulation, congestion mitigation can be understood as a process of sequential stability and execution development rather than as a strong stand-alone direct effect of RL. These conclusions remain simulation-informed and should be tested against alternative simulator designs and real-world operational data before broader field-level claims are made. Future research should test richer timing-related moderators, extend the framework to other inland and multimodal nodes, and examine whether the proposed pathway remains stable across independently validated data sources.

Author Contributions: Conceptualization, M.M.R. and J.F.; methodology, M.M.R. and J.F.; software, M.M.R.; validation, E.T., J.F. and N.G.; formal analysis, M.M.R.; investigation, J.F.; resources, E.T. and N.G.; data curation, M.M.R.; writing—original draft preparation, M.M.R.; writing—review and editing, E.T. and J.F.; visualization, M.M.R.; supervision, J.F. and E.T.; project administration, J.F.; funding acquisition, E.T. and N.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. The simulation-generated dataset and analytical workbook are available from the corresponding authors upon reasonable request. Further inquiries may be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Facchini, F.; Digiesi, S.; Mossa, G. Optimal Dry Port Configuration for Container Terminals: A Non-Linear Model for Sustainable Decision Making. *Int. J. Prod. Econ.* **2020**, *219*, 164–178. <https://doi.org/10.1016/j.ijpe.2019.06.004>.
2. Varese, E.; Marigo, D.; Lombardi, M. Dry Port: A Review on Concept, Classification, Functionalities and Technological Processes. *Logistics* **2020**, *4*, 29. <https://doi.org/10.3390/logistics4040029>.
3. Sciomachen, A.; Stecca, G. Forwarding Containers to Dry Ports in Congested Logistic Networks. *Transp. Res. Interdiscip. Perspect.* **2023**, *20*, 100846. <https://doi.org/10.1016/j.trip.2023.100846>.

4. Zhang, J.; Xin, X.; Liao, Z.; Dubey, R.; Nguyen, T.T.; Li, N.; Yang, Z. Analysis of the Ripple Effects of Disruptions on Multimodal Container Terminals Operations: A System Dynamics Approach. *Transp. Res. Part E Logist. Transp. Rev.* **2025**, *202*, 104264. <https://doi.org/10.1016/j.tre.2025.104264>.
5. Almeida, F. Challenges in the Digital Transformation of Ports. *Businesses* **2023**, *3*, 548–568. <https://doi.org/10.3390/businesses3040034>.
6. Kastner, M.; Saporiti, N.; Lange, A.-K.; Rossi, T. Insights into How to Enhance Container Terminal Operations with Digital Twins. *Computers* **2024**, *13*, 138. <https://doi.org/10.3390/computers13060138>.
7. Xu, H.; Liu, J.; Xu, X.; Chen, J.; Yue, X. The Impact of AI Technology Adoption on Operational Decision-Making in Competitive Heterogeneous Ports. *Transp. Res. Part E Logist. Transp. Rev.* **2024**, *183*, 103428. <https://doi.org/10.1016/j.tre.2024.103428>.
8. Parvez Farazi, N.; Zou, B.; Ahamed, T.; Barua, L. Deep Reinforcement Learning in Transportation Research: A Review. *Transp. Res. Interdiscip. Perspect.* **2021**, *11*, 100425. <https://doi.org/10.1016/j.trip.2021.100425>.
9. Lai, X.; Yang, Z.; Xie, J.; Liu, Y. Reinforcement Learning in Transportation Research: Frontiers and Future Directions. *Multimodal Transp.* **2024**, *3*, 100164. <https://doi.org/10.1016/j.multra.2024.100164>.
10. Yan, Y.; Chow, A.H.F.; Ho, C.P.; Kuo, Y.-H.; Wu, Q.; Ying, C. Reinforcement Learning for Logistics and Supply Chain Management: Methodologies, State of the Art, and Future Opportunities. *Transp. Res. Part E Logist. Transp. Rev.* **2022**, *162*, 102712. <https://doi.org/10.1016/j.tre.2022.102712>.
11. Chen, C.; Hu, Z.-H.; Wang, L. Scheduling of AGVs in Automated Container Terminal Based on the Deep Deterministic Policy Gradient (DDPG) Using the Convolutional Neural Network (CNN). *J. Mar. Sci. Eng.* **2021**, *9*, 1439. <https://doi.org/10.3390/jmse9121439>.
12. Jin, J.; Cui, T.; Bai, R.; Qu, R. Container Port Truck Dispatching Optimization Using Real2Sim Based Deep Reinforcement Learning. *Eur. J. Oper. Res.* **2024**, *315*, 161–175. <https://doi.org/10.1016/j.ejor.2023.11.038>.
13. Li, S.; Fan, L.; Jia, S. A Hierarchical Solution Framework for Dynamic and Conflict-Free AGV Scheduling in an Automated Container Terminal. *Transp. Res. Part C Emerg. Technol.* **2024**, *165*, 104724. <https://doi.org/10.1016/j.trc.2024.104724>.
14. Zhang, Y.; Negenborn, R.R.; Atasoy, B. Synchronomodal Freight Transport Re-Planning under Service Time Uncertainty: An Online Model-Assisted Reinforcement Learning. *Transp. Res. Part C Emerg. Technol.* **2023**, *156*, 104355. <https://doi.org/10.1016/j.trc.2023.104355>.
15. Zhang, L.; Yang, C.; Yan, Y.; Cai, Z.; Hu, Y. Automated Guided Vehicle Dispatching and Routing Integration via Digital Twin with Deep Reinforcement Learning. *J. Manuf. Syst.* **2024**, *72*, 492–503. <https://doi.org/10.1016/j.jmsy.2023.12.008>.
16. Glock, C.H.; Grosse, E.H.; Jaber, M.Y.; Smunt, T.L. Applications of Learning Curves in Production and Operations Management: A Systematic Literature Review. *Comput. Ind. Eng.* **2019**, *131*, 422–441. <https://doi.org/10.1016/j.cie.2018.10.030>.
17. Vörös, J. Production Dynamics in Case of Organizational Learning. *Comput. Ind. Eng.* **2021**, *157*, 107340. <https://doi.org/10.1016/j.cie.2021.107340>.
18. Yuan, Y.; Feng, B.; Lai, F.; Collins, B.J. The Role of Trust, Commitment, and Learning Orientation on Logistic Service Effectiveness. *J. Bus. Res.* **2018**, *93*, 37–50. <https://doi.org/10.1016/j.jbusres.2018.08.020>.
19. Batsis, A.; Samothrakis, S. Contextual Reinforcement Learning for Supply Chain Management. *Expert Syst. Appl.* **2024**, *249*, 123541. <https://doi.org/10.1016/j.eswa.2024.123541>.
20. Chang, W.-S.; Lin, Y.-T. The Effect of Lead-Time on Supply Chain Resilience Performance. *Asia Pac. Manag. Rev.* **2019**, *24*, 298–309. <https://doi.org/10.1016/j.apmr.2018.10.004>.
21. Zhang, H.; Lu, G.; Zhang, Y.; D'Ariano, A.; Wu, Y. Railcar Itinerary Optimization in Railway Marshalling Yards: A Graph Neural Network Based Deep Reinforcement Learning Method. *Transp. Res. Part C Emerg. Technol.* **2025**, *171*, 104970. <https://doi.org/10.1016/j.trc.2024.104970>.
22. Lee, W.; Cho, S.W. Reinforcement Learning Approach for Outbound Container Stacking in Container Terminals. *Comput. Ind. Eng.* **2025**, *204*, 111069. <https://doi.org/10.1016/j.cie.2025.111069>.
23. Bett, D.K.; Ali, I.; Gheith, M.; Eltawil, A. Simulation-Based Optimization of Truck Appointment Systems in Container Terminals: A Dual Transactions Approach with Improved Congestion Factor Representation. *Logistics* **2024**, *8*, 80. <https://doi.org/10.3390/logistics8030080>.
24. Jia, S.; Cui, H.; Chen, R.; Meng, Q. Dynamic Container Drayage with Uncertain Request Arrival Times and Service Time Windows. *Transp. Res. Part B Methodol.* **2022**, *166*, 237–258. <https://doi.org/10.1016/j.trb.2022.10.010>.
25. Torbali, B.; Alpan, G. A Multi-Agent-Based Real-Time Truck Scheduling Model for Cross-Docking Problems with Single Inbound and Outbound Doors. *Supply Chain Anal.* **2023**, *3*, 100028. <https://doi.org/10.1016/j.sca.2023.100028>.

26. Rahman, M.d.M.; Fan, J.; Tijan, E.; Fateha, U.A. Reducing Port Container Congestion with Reinforcement Learning: The Serial Mediation Role of Operational Learning Stability and Logistics Efficiency. *J. Mar. Sci. Eng.* **2026**, *14*, 687. <https://doi.org/10.3390/jmse14070687>.
27. Tortorella, G.L.; Cawley Vergara, A.M.; Garza-Reyes, J.A.; Sawhney, R. Organizational Learning Paths Based upon Industry 4.0 Adoption: An Empirical Study with Brazilian Manufacturers. *Int. J. Prod. Econ.* **2020**, *219*, 284–294. <https://doi.org/10.1016/j.ijpe.2019.06.023>.
28. Schepler, X.; Balev, S.; Michel, S.; Sanlaville, É. Global Planning in a Multi-Terminal and Multi-Modal Maritime Container Port. *Transp. Res. Part E Logist. Transp. Rev.* **2017**, *100*, 38–62. <https://doi.org/10.1016/j.tre.2016.12.002>.
29. Deineko, E.; Jungnickel, P.; Kehrt, C. Learning-Based Optimisation for Integrated Problems in Intermodal Freight Transport: Preliminaries, Strategies, and State of the Art. *Appl. Sci.* **2024**, *14*, 8642. <https://doi.org/10.3390/app14198642>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.