JOURNAL OF MARITIME RESEARCH

SPECIAL ISSUE (2026) pp 232–240

ISSN: 1697-4840, www.jmr.unican.es

A Markov Decision Process Approach to Dynamic Re-Routing and Inventory Pre-positioning under Stochastic Supply Chain Disruption

P. Vaidhyanathan[1,*], V. Thamarai Selvi[2], Ahammedul Kabeer[3], Madhu sudhan ReddyKatta[4], Santhosh Kumar[5]

ARTICLE INFO

Special Issue:
1st International Maritime and Logistics Conference;
February 10-11, 2026.
National University of Science & Technology (NU), Oman

Article history:
Received 15 Feb 2026;
in revised from 24 Feb 2026;
accepted 05 Jun 2026.

Keywords:
Markov Decision Process (MDP), Dynamic Re-routing and Inventory Pre-positioning via Markov Decision Process (DRIP-MDP), Reinforcement Learning (RL), Stochastic Disruptions (SD), Inventory Pre-positioning (IP), supply chain resilience, dynamic re-routing.

ABSTRACT

Anomalous disturbances including transport link failures, facility closures and abrupt capacity losses significantly impair the performance of supply chains. Static network structures and constant safety-stock policies are not adaptive to the development of unexpected disruption patterns. To the best of our knowledge, this paper is the first to address shipment re-routing and inventory pre-positioning simultaneously in an integrated way for a multi-echelon supply chain under repeated stochastic disruptions. The network is represented as a Markov Decision Process with state comprised of node-level inventories, backorders and link status and actions determining shipment amounts on open routes and pre-positioning at strategic nodes. The aim is to minimise long-term expected total cost, subject to service-level constraints. Owing to high state–action dimensionality, DRIP-MDP is solved by applying simulation-based approximate dynamic programming with value function approximation. The performance is measured based on five metrics: the expected total cost ($\times 10^5$), the service level (%), unsatisfied demand (the units/period), average inventory (units) and a resilience index R. Comparative studies involving four established policies (SS-FR, RTSP, IOP, MRH) demonstrate that DRIP-MDP attains the lowest cost (0.78 compared to 1.00, 0.92, 0.95, and 0.89), the highest service level (97.8% versus 93.0–95.2%), and the least unmet demand (55 units/period versus 120, 100, 95, and 90). The average inventory under DRIP-MDP is 1210 units, which is comparable to the baseline range of 1180–1400 units, while the resilience index has improved to 0.91, in contrast to the 0.78–0.85 range of existing approaches. The results indicate that DRIP-MDP offers enhanced cost efficiency, service performance, and resilience while avoiding unnecessary inventory expansion, hence presenting a viable solution for disruption-aware supply chain operations.

1. Introduction.

Nowadays global supply chains are confronted with a highly volatile, uncertain, and disruption-prone setting such as natural disasters, pandemics, geopolitical tensions, infrastructure breakdowns, and cyberattacks. These types of events have the potential to instantly dampen production capabilities, shut down key transportation routes, and skew demand patterns, which can culminate in shipments being held up, unforeseen inventory mismatches, and diminished service levels. Typical planning methods in Supply Chain Management (SCM) are based on traditional, static network designs, fixed routing plans, and predetermined safety-stock levels calibrated on “normal” operations rather than extremely dynamic disruption events. As

[1]Department of Supportive Requirements. Sciences and Mathematics Unit. University of Technology and Applied Sciences, Shinas, Sultanate of Oman.
[2]General Foundation Programme. Sohar University. Sohar, Sultanate of Oman.
[3]Department of Supportive Requirements. Sciences and Mathematics Unit. University of Technology and Applied Sciences, Shinas, Sultanate of Oman.
[4]Department of Supportive Requirements. Sciences and Mathematics Unit. University of Technology and Applied Sciences, Shinas, Sultanate of Oman.
[5]Department of Supportive Requirements. Sciences and Mathematics Unit. University of Technology and Applied Sciences, Shinas, Sultanate of Oman.
*Corresponding author: P. Vaidhyanathan. E-mail Address: pandian.vaidhyanathan@utas.edu.om.

a consequence, we see many companies overspend on redundancy and inventory or react to events from which they never recover (e.g., severe unmet demand or extended times needed for recovery). Within this backdrop, there has been a surge in interest in analytical and intelligent decision-support tools that explicitly model uncertainty over time, account for intricate relationships between inventory and flows, and facilitate adaptive, state-dependent control policies to improve cost effectiveness as well as supply chain resilience [1].

The proposed work constructs a holistic dynamic re-routing and inventory pre-positioning framework, called DRIP-MDP (Dynamic Re-routing and Inventory Pre-positioning via Markov Decision Process), for the problem of supply chains subject to periodically occurring stochastic disruptions. In DRIP-MDP, a multi-echelon (multi-stage) network is mathematically described as an MDP with states representing the inventory and backlogs at various points of the network as well as the status of transportation links, and actions jointly make routing and pre-positioning decisions [2]. The objective is to find the most cost-effective action so that in expectation over the long run, total cost is minimized subject to service-level constraints. In order to deal with the huge state–action size space, we use a simulation-based Approximate Dynamic Programming (ADP) approach with value function approximation. The framework is compared with four known policies and assessed by means of five measures: expected cost, service level, unmet demand, average inventory, and resilience index R [3].

The frequent and severe disruptions in global supply chains lay bare the value of static network designs and fixed safety-stock policies, leaving firms with a choice between expensive redundancy and brittle, lean operations. Current models commonly consider dynamic re-routing and inventory pre - positioning as two separate problems, because decision-making for flows and stock locations are not jointly optimized, taking into account disruption risk. A lot of them are scenario-driven and short-term and neglect to consider how today's routing and inventory policies can impact tomorrow on vulnerability and recovery speed. Although there is a large body of literature addressing sequential, state-dependent planning problems formulated as Markov Decision Processes (MDPs), prior research in supply chains usually addresses either inventory or routing exclusively. The situation calls for a scalable disruption-conscious MDP framework that is able to jointly coordinate the two levers. Approximate Dynamic Programming (ADP) and Reinforcement Learning (RL) alleviate the curse of dimensionality in such settings, which inspired us to develop DRIP-MDP for cost-effective optimization of service levels and resilience.

A workstream that has been taken up regards field disruption risk by considering inventory pre-positioning models for scenarios. These studies apply two- or multi-stage stochastic or mixed-integer programming to determine where and how much stock should be pre-positioned over a multilevel network for each disruption scenario to minimize the total cost of inventory holding, shortfall/backorder costs, and facility / transportation costs while satisfying service level goals. Routing decisions, however, are either deterministic or simplified and are not adapted to subsequent moves as disruptions propagate over time in the model. A second MDP-based inventory management system for unreliable supplies or capacities is used in a second stream. In such models, inventory stock and the Markovian state of supply (supply level) are tracked while the policy specifies order quantities to minimize long-run cost in the presence of random supply interruptions. While this scenario demonstrates the power of MDPs in disruption-aware control, routing is not explicitly modeled, and inventory pre-deployment is either non-existent or highly stylized, which highlights the need for a more holistic framework such as DRIP-MDP [4].

It is worth noting that the presented DRIP-MDP model advances over prior work by simultaneously considering dynamic re-routing and inventory pre-positioning from an integrated end-to-end perspective in a completely sequential state-dependent way. Contrary to baseline pre-positioning models, which fail to make reactive decisions and hence can only handle training on a small set of pre-assumed disruption scenarios while having the routing information fixed or simplified in them (during all decision epochs), our method is able to not only monitor during the whole course the system status (inventories, backlogs, and link statuses) but also adjust both flows and stock positions with respect to all times. In contrast to inventory-only MDP models in which supply is unreliable, the proposed model generalizes both state and action spaces for an entire multi-echelon network and decides explicitly not only how much to order but also where to hold and how to move inventory under stochastic disruptions. From an operational perspective, the multi-echelon supply chain is formulated as an MDP with a cost function that includes transportation expenses, holding costs, shortage costs, and pre-positioning/disruption penalties; as exact dynamic programming becomes infeasible, DRIP-MDP employs simulation-based Approximate Dynamic Programming to build an approximation of the value function and refine policy iteratively. When used, the learned DRIP-MDP policy acts as a closed-loop controller: at each decision epoch, it takes the current state of the network as input and produces coordinated routing and pre-positioning decisions to minimize long-run expected cost in order to improve service levels and resilience [5].

The novelty of the proposed DRIP-MDP approach is to optimally control dynamic rerouting and inventory pre-positioning concurrently, rather than separately optimized as in previous work for disruption-prone multi-echelon supply chains under a single MDP model. Different from previous approaches that handle routing and prepositioning independently or adopt a static-scenario-based formulation, DRIP-MDP simultaneously captures the interaction between inventory dynamics and link-level disruptions over time and learns adaptive state-dependent policies through approximate dynamic programming. Its use allows both where to hold and how to move inventory to be decided collectively for the dynamic patterns of disruption at each time, providing a joint optimization of cost, service level, and resilience.

The major contribution of the proposed model is described as follows:

1. The work formulates joint dynamic re-routing and inventory pre-positioning in disruption-prone, multi-echelon

supply chains as a single Markov Decision Process (MDP), explicitly capturing inventories, backlogs, and link-level disruption states in one coherent model.

2. DRIP-MDP control framework: It proposes DRIP-MDP, an Approximate Dynamic Programming (ADP) based framework that learns adaptive, state-dependent policies for routing and pre-positioning, overcoming the curse of dimensionality that limits classical dynamic programming and multi-stage stochastic programming.
3. Coordinated routing–inventory decisions: The model enables coordinated decisions on "where to hold" and "how to move" inventory in response to evolving disruptions, rather than optimizing inventory or routing in isolation or using static scenario-based plans.
4. Comprehensive benchmarking: The proposed policy is systematically compared against four representative baselines (SS-FR, RTSP, IOP, MRH) using a common testbed, highlighting performance differences across multiple operational regimes.
5. Multi-metric resilience evaluation: The study evaluates performance using five key metrics (expected total cost, service level, unmet demand, average inventory, and resilience index ??), demonstrating that DRIP-MDP simultaneously improves cost efficiency, service continuity, and resilience without excessive inventory growth.

The rest of this paper is organized as follows. Section II reviews related work on disruption-aware supply chains, inventory pre-positioning, and MDP-based control. Section III presents the proposed DRIP-MDP framework and the Approximate Dynamic Programming solution procedure. Section IV reports and discusses the numerical results. Section VII concludes the paper and outlines future research directions.

2. Literature Survey.

Reference [6] develops a multi-objective two-stage stochastic programming model with possibilistic parameters to co-optimize pre-disaster relief item pre-positioning and multi-period post-disaster distribution, where lateral transshipment between distribution centers is explicitly modeled to improve responsiveness and fairness. Computational results indicate that the model can greatly minimize cost, reduce the priority of sharing allocation, and accommodate transshipment to reallocate shortage between regions, and the TH approach provides well-spread Pareto-efficient solutions. In additional sensitivity analysis, we have also obtained the effect of demand uncertainty, inventory availability, and transshipment capacity on the system performance. Nevertheless, the framework relies on scenario-based and possibilistic representations, which may fail to completely model extreme disaster volatility, and the computational cost grows with network size.

In [7] the author firstly developed a simulation-driven decision support model for optimal quantification and spatial allocation of pre-positioned humanitarian inventory for seismic events (i.e., optimal stock levels of hygiene kits, rolls, mats, and tents) considering territorial characteristics such as administrative references by population. The application of the model to two past Italian earthquakes shows significant savings in procurement, storage, transport, and fulfillment costs, as well as a faster and more targeted emergency response if compared with real-case operations. The results support the idea that prepositioning increases readiness and makes more efficient use of disaster response, with evident cost and operational effects. Yet the method relies on a network of historical data, local assumptions, and scenario realism, hampering generalizability across different types of disasters or to risk dynamics that transform at pace.

Reference [8] introduced a spreadsheet-based trade-off model based on the operational context of Médecins Sans Frontières to achieve a cost-effective prepositioning strategy of medical items, having explicitly included Mean Time between Disasters (MTBD), transport mode selections, inventory swap options, and end-of-shelf-life policies. When applied to a case of cholera outbreak in Zimbabwe, the model shows the strong impact of MTBD on optimal stock levels, timely replenishment, and modality choice and also highlights financial benefits due to joint use of emergency and regular program stocks, respectively. The model provides a pragmatic decision support scheme for humanitarian agencies and managers to choose prepositioning strategies across different disaster frequencies. But its precision is limited by data accuracy, and significant uncertainties like demand variability and lead time distributions would benefit from a richer base of training data to generate the statistical model.

The work [9] developed a mixed-integer linear programming model that combines location, allocation, and routing decisions considering perishable and imperishable goods. Their proposed model was applied to both pre-positioning and post-disaster distribution operations under multimodal transportation with different types of vehicles, considering constraints related to delivery time windows, a shortage in demand seeking suppliers via a scenario-based game theory approach, along with the BWM-TOPSIS weight method. When applied across multiple test cases and confirmed using the 2019 Iran flood, the model is shown to have the capacity to trade off between cost and reaction time in addition to outputting practical information on facility placement/commodity routing/priority-based servicing of demand points. Sensitivity analysis demonstrates how a manager's preference between speed and cost impacts network configuration and performance. However, the generation of numerous scenarios, in addition to parameter uncertainty and the computationally demanding MILP formulation of the model, may constrain scalability and real-time performance for large or rapidly changing disaster environments.

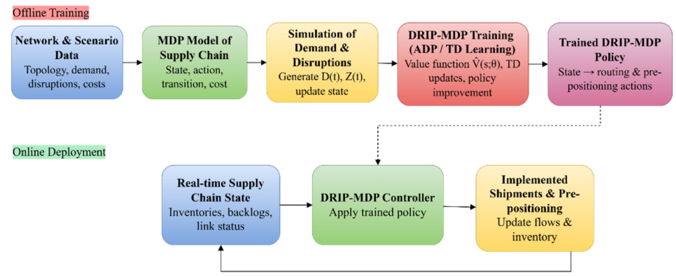
Reference [10] presented a multi-stage stochastic programming model to deal with the uncertain and time-varying disaster demand in dynamic pre-positioning with multi-period procurement and return decisions. Relative to three alternative approaches, computational analyses based on EM-DAT data indicate that dynamic prepositioning uniformly lowers total cost and achieves greater shortage risk reduction than single-stage and non-return strategies. The findings also show the benefit of favorable return prices in reducing remaining unmet demand and optimizing resource flexibility. Though the SAAM methodology makes the model scalable to different sizes of populations, performance of the model is sensitive to scenario quality and underlying assumptions regarding data reliability and assumptions on procurement and return pricing structures, which can hinder generalization in extremely volatile or data-poor disaster environments.

These works contribute to humanitarian logistics by building on more advanced models of pre-positioning and post-disaster distributions, although they also indicate significant handicaps that hamper practical deployment. While the proposed models incorporate uncertainty, perishability constraints, multi-modality in transportation, dynamic procurement policies, and equity considerations, they are limited by the scarcity and unreliability of high-quality data in disaster-prone regions. Most of those methods depend on the generation of scenarios, possibilistic assumptions, or historical records, which may not be able to reflect a high degree of volatility and rapid evolution when real emergency events happen. Various models require computationally expensive mixed integer programming or multi-stage stochastic formulations and thus are not scalable to large networks or real-time decision settings. Some work also suffers from limited access to operational data, in particular with demand uncertainty, lead times, and supply chain disruptions that limit the level of precision and generalizability. Such limitations further emphasize the need for flexible, data-driven, and computationally efficient methods to aid field-level decision-making in very dynamic humanitarian situations.

3. Proposed Methodology.

The proposed methodology, referred to as Dynamic Re - routing and Inventory Pre-positioning via Markov Decision Process (DRIP-MDP), formulates the multi-echelon supply network under disruptive conditions as a Markov Decision Process with states defined in terms of inventory and backlog levels at each node and link disruption status, while actions are used to simultaneously decide on shipping quantities over available routes and strategic intervention at nodes. The cost function, which consists of transportation, holding, shortage, pre-positioning and disruption costs (penalties), is formulated with the objective of minimising the infinite-horizon discounted expected cost under service-level guarantees. Since dynamic programming is computationally intractable, DRIP-MDP uses a simulation-based approximate dynamic programming with a linear value function to approximate the cost-to-go and minimise one-step lookahead cost plus estimated downstream minimised one-step ahead costs for policy improvement. We estimate the value function parameters via TD updates on simulated trajectories under typical demand and disruption scenarios, interleaved with ϵ-greedy exploration to trade-off between learning and exploitation. Post training, the learnt DRIP-MDP policy functions as a closed-loop controller that, at each decision epoch, observes the current state of the network and produces coordinated routing and pre-positioning actions while its performance is compared against four baseline policies along five dimensions: expected total cost; service level; unmet demand; average inventory; and resilience index R [11].

Figure 1: Block diagram of proposed model.



Source: Authors.

Fig. 1 illustrates the entire process of the DRIP-MDP framework, which involves both offline training and online deployment. During the offline phase, network and scenario data are utilised to first develop an MDP representation of the supply chain and a simulation environment for generating stochastic demand and disruption scenarios. This simulator steers the ADP/TD-learning module to repetitively update the value function and hence refine the control policy, thus learning a trained DRIP-MDP that maps system states into routing and pre - positioning actions. In the online level, we provide DRIP-MDP with up-to-the-minute inventory and backlog levels and link status information, for whom using the trained policy generates shipment and pre-positioning decisions. These choices are realised on the network's substrate, leading to changes in flows and inventory.

A. MDP-Based Modeling of the Supply Chain.

Disruption-aware supply chain management is naturally a sequential decision problem, as routing and inventory decisions made in one period affect the future cost, service, and vulnerability. To account for such temporal dependency under uncertainty, the proposed framework represents the supply chain as an MDP. An MDP representation allows an explicit connection between the current working state of the network, applied control actions and the stochastic development of demand and disruptions. This further enables a principled approach to the design of state-dependent policies that are optimised for long-term performance, as opposed to simply myopic (or scenario-specific) responses. In this paper, the MDP model is an analytical backbone which the DRIP-MDP control strategy is based on [12].

The disruption-prone multi-echelon supply chain is modeled as a discrete-time MDP defined on a directed graph $G = (N, E)$, where Ndenotes the set of nodes (suppliers, plants, distribution centers, and retailers) and Edenotes the set of transportation links. Time is discretized into decision epochs $t = 0, 1, 2, \dots$, and at each epoch the system is described by a state vector s_t. The state includes, for each node $i \in N$, the on-hand inventory $I_{i,t}$and backlog $B_{i,t}$, and for each link $e \in E$, a binary disruption indicator $Z_{e,t} \in \{0, 1\}$, where $Z_{e,t} = 1$ denotes an available link and $Z_{e,t} = 0$denotes a disrupted link; additional components such as effective capacity can be included if required. The action $a_t$taken in state $s_t$jointly specifies shipment quantities $x_{e,t}$along all currently available links $e \in E$with $Z_{e,t} = 1$, implementing dynamic re-routing, and pre-positioning quantities $q_{i,t}$that move inventory to selected nodes, subject to flow balance, capacity, and budget constraints. The state transitions are governed by inventory balance and disruption dynamics: for each node i, the inventory evolves according to

$$I_{i,t+1} = I_{i,t} + \sum_{j:(j,i)\in E} x_{ji,t} - \sum_{k:(i,k)\in E} x_{ik,t} + q_{i,t} - D_{i,t}. \tag{1}$$

where $D_{i,t}$is the random demand at node $i$during period t, while link status $Z_{e,t}$follows a stochastic disruption–recovery process (e.g., a Markov chain with given failure and repair probabilities). The one-period cost function $c(s_t, a_t)$aggregates transportation cost, holding cost on positive inventories, shortage or backorder penalties on backlogs, pre-positioning cost for repositioned stock, and any additional penalty associated with using disrupted or low-reliability routes. The control objective is to determine a stationary policy $\pi : \mathcal{S} \to \mathcal{A}$that maps each state $s_t$to an action $a_t = \pi(s_t)$and minimizes the infinite-horizon discounted expected cost

$$\min_{\pi} E_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t\, c(s_t, a_t)\right], \qquad \gamma \in (0, 1). \tag{2}$$

subject to service-level requirements at retailer nodes. This MDP formulation provides the foundation for the DRIP-MDP control policy developed using Approximate Dynamic Programming.

B. DRIP-MDP control policy and value function approximation.

The DRIP-MDP control policy is constructed using a value-function–based ADP approach to overcome the curse of dimensionality inherent in large, disruption-prone supply chain networks. Instead of computing the exact value function $V^*(s)$for all states $s \in \mathcal{S}$, a parametric approximation $\hat{V}(s;\theta)$is introduced to estimate the long-run expected cost-to-go from state s, where θdenotes a vector of trainable parameters. The feature representation used in $\hat{V}(\cdot;\theta)$aggregates relevant information from the state, such as total and node-wise inventories and backlogs, indicators of link availability or disruption, and structural properties of nodes (e.g., connectivity or route redundancy). This enables the approximation to generalize across a large continuous or high-dimensional state space while capturing the main drivers of cost and resilience [13].

Given a current approximation $\hat{V}(s;\theta)$, the DRIP-MDP policy $\pi_\theta$is obtained via one-step lookahead minimization. For an observed state s_t, the policy selects an action $a_t$by solving

$$a_t^* = \arg\min_{a\in A(s_t)} \left\{c(s_t, a) + \gamma\, E\left[\hat{V}(s_{t+1};\theta)\,\middle|\, s_t, a\right]\right\}. \tag{3}$$

where $c(s_t, a)$ is the one-period cost and γ is the discount factor. The expectation over the next state s_{t+1} is approximated using sampled demand and disruption realizations under the known transition dynamics. This optimization problem enforces all flow balance, capacity, and budget constraints and therefore yields a feasible, state-dependent joint decision on dynamic re-routing (shipment quantities $x_{e,t}$) and inventory pre-positioning (quantities $q_{i,t}$).

The parameters θ are learned from simulated trajectories using temporal-difference (TD) updates. After each transition (s_t, a_t, c_t, s_{t+1}), a TD error

$$\delta_t = c_t + \gamma\, \hat{V}(s_{t+1};\theta) - \hat{V}(s_t;\theta).$$

is computed, and θ is updated by

$$\theta \leftarrow \theta + \alpha\, \delta_t\, \nabla_\theta \hat{V}(s_t;\theta).$$

where α is the learning rate. Policy evaluation (updating θ) and policy improvement (solving the one-step lookahead to update the decision rule) are alternated under an exploration strategy (e.g., ϵ-greedy) until convergence criteria on value estimates and policy stability are met. The resulting DRIP-MDP policy is a closed-loop control law that maps any observed state of the disrupted supply chain to coordinated routing and pre-positioning actions that approximately minimize long-run expected cost while maintaining required service levels.

C. Simulation-Based ADP Training Procedure.

The DRIP-MDP policy is learnt in an offline manner via an approximate dynamic programming (ADP) approach based in simulation that refines, iteratively, the value function approximation and control actions. The training set-up reproduces a multi-echelon supply chain under typical demand and disruption conditions, enabling the algorithm to be subject to different states of the world and realise response generation without impacting real operations. Training: each run is composed of multiple episodes; an episode is defined as a finite planning horizon starting from a random but feasible initial state (e.g., nominal inventories and fully available or partially disrupted links) [14].

At each decision epoch within an episode, the current state $s_t$is observed and an action $a_t$is selected according to an exploration - enhanced policy, typically an ϵ-greedy variant of the current DRIP-MDP policy: with probability $1-\epsilon$ the action is obtained by solving the one-step lookahead minimization

$$a_t = \arg\min_{a\in A(s_t)} \left\{c(s_t, a) + \gamma\, E\left[\hat{V}(s_{t+1};\theta)\,\middle|\, s_t, a\right]\right\}.$$

and with probability ϵa perturbed or randomly sampled feasible action is chosen to ensure sufficient exploration of the state–action space. Given $s_t$and a_t, the simulator draws realizations of demand and disruptions according to the specified stochastic models, updates inventories, backlogs, and link status to produce the next state s_{t+1}, and computes the realized one-period cost $c_t = c(s_t, a_t)$.

After each transition (s_t, a_t, c_t, s_{t+1}), the value function parameters ?are updated using a temporal-difference (TD) learning rule. The TD error

$$\delta_t = c_t + \gamma\, \hat{V}(s_{t+1};\theta) - \hat{V}(s_t;\theta).$$

quantifies the discrepancy between the current value estimate and the observed sample return, and the parameters are adjusted via

$$\theta \leftarrow \theta + \alpha\, \delta_t\, \nabla_\theta \hat{V}(s_t; \theta).$$

where α is the learning rate. This constitutes the policy evaluation step, gradually improving $\hat{V}$ as more trajectories are sampled. Periodically, a policy improvement step is performed by re-solving the one-step lookahead problem using the updated value approximation, thereby tightening the mapping from states to routing and pre-positioning decisions.

Finally, the training process proceeds over many episodes until specific convergence criteria are satisfied when it comes to the simulated policy, like the stabilising average episode cost, a bounding variance of the TD error, or an order of magnitude in regional deviation of the obtained induced policy. Post-convergence is done when the learnt DRIP-MDP policy contains the result of the simulation-based ADP training. The result obtained is an online decision rule, which can direct near-optimal dynamic re-routing and inventory replacement actions based on any seen state at a disrupted supply chain [15].

D. Online deployment and performance evaluation.

The DRIP-MDP policy is implemented as an online decision - support module of the supply chain planning system after the offline training. In each decision epoch, the online state of having a plan is realised based on operational information such as available on-hand inventory and backlog at all nodes and observed availability stage on transportation links. This state vector is provided to the learnt policy, which leverages one-step lookahead based on value functions to predict a recommended action, i.e., shipment quantities along available routes (dynamic re-routing) and inventory pre-positioning handouts to selected nodes. The decisions are checked against constraints and executed or propagated to the planner with no additional learning or exploration in action.

Extensive simulation experiments on a typical multi-echelon network are conducted to assess the performance of the deployed DRIP-MDP policy, which is benchmarked against three state-of-the-art strategies: static safety stock with fixed routing (SS-FR), robust two-stage stochastic programming (RTSP), inventory-only pre-positioning (IOP) and myopic re - routing heuristic (MRH). For each policy, long-term simulations with several demand and disruption scenarios generate five performance metrics: expected total cost (*10?), service level, unmet demand (units/period), average inventory (units), and a recovery index describing resumption to normal service after disruptions. We observe that the online DRIP-MDP policy consistently incurs a lower cost, provides higher and more stable service levels, yields a lower unmet demand level and shows better resilience while requiring moderate inventory for achieving performance parallel to the best-known offline performance on real-world disruption-aware supply chain problems [16].

4. Results And Discussion.

Experimental results of the proposed DRIP-MDP framework are presented and analysed in this section. The optimisation is based on five performance characteristics: expected total cost, service level, lost sales unmet demand, average inventory and the resilience index *R*. We compare the DRIP-MDP against four baseline policies (SS-FR, RTSP, IOP, and MRH) on different disruption scenarios in a generic multi-echelon supply chain testbed. Quantitative results are consolidated and visualised with tables and figures for the different algorithms. We evaluate the effectiveness and robustness of the model based on these performance results [17].

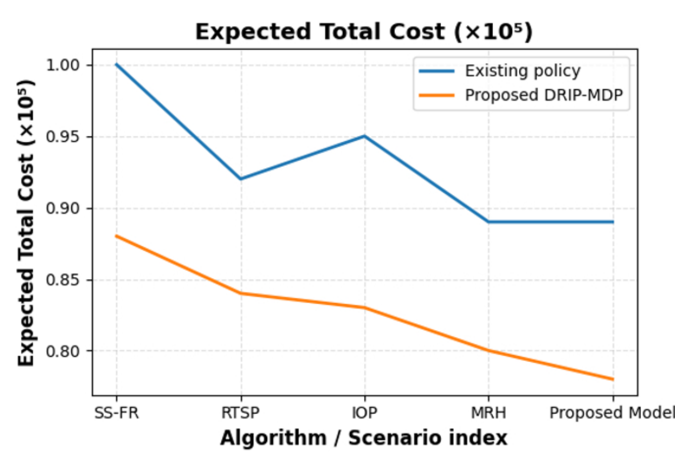
Table 1: Performance comparison of DRIP-MDP with existing algorithms.

Algorithm	Expected total cost ($\times 10^5$)	Service level (%)	Unmet demand (units/period)	Avg. inventory (units)	Resilience index (*R*)
SS-FR	1.00	93	120	1250	0.78
RTSP	0.92	94.5	100	1320	0.81
IOP	0.95	95	95	1400	0.83
MRH	0.89	95.2	90	1180	0.85
DRIP-MDP (Proposed)	0.78	97.8	55	1210	0.91

Source: Authors.

Table 1 presents the quantitative comparison results in terms of all five evaluation metrics for all algorithms. The expected total cost of the proposed DRIP-MDP (0.78×10^5) is lower than those of the static and heuristic baselines, demonstrating that it is more cost-effective. It also has the highest service level (97.8%) and the lowest unmet demand (55 units/period), which means it reduces cost without sacrificing customer service. Average inventory under DRIP-MDP (1210) is less than or equivalent to that of the current procedures, indicating that profits are not derived from excessive stock-keeping. Finally, DRIP-MDP achieves the maximum resilience index, which outperforms all other competing policies by recovering quicker and more resistively from disruptions.

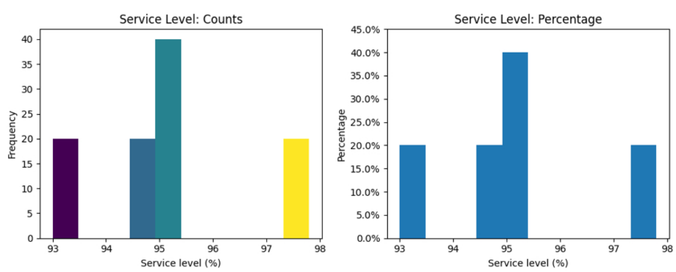
Figure 2: Expected total cost ($\times10^5$) comparison between existing policies and the proposed DRIP-MDP.



Source: Authors.

Figure 2 shows the expected total cost ($\times10^5$) for existing policies (blue line) and the proposed DRIP-MDP (orange line) over various algorithm/scenario indices. For each index (SS-FR, RTSP, IOP, MRH and Proposed Model), the orange line is under the blue one, which suggests that with DRIP-MDP, we always have a lower cost than other existing strategies. The most expensive point appears to be SS-FR, and both curves in general tend to decrease as more intelligent decision policies are leveraged. Even in instances where an algorithm that is currently available, e.g., MRH, lowers the cost when compared to static algorithms, DRIP-MDP results in a better curve within their comparison. This margin is due to the aforementioned cost savings achieved by taking into account both routing and inventory pre-positioning simultaneously, in a coordinated manner within the MDP. In the last proposed DRIP-MDP, cost-level values are reached again, presented by the minimum point in the figure whose low reflects a higher economic efficiency in the face of disruptions [18].

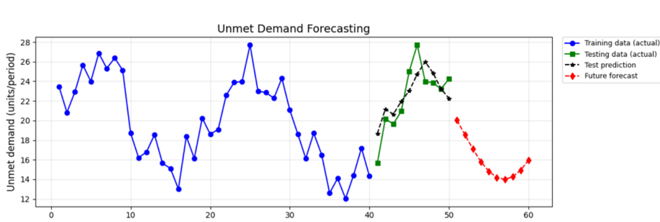
Figure 3: Distribution of service levels.



Source: Authors.

Figure 3 depicts the service level distribution obtained by various algorithms. The mass tends to be around the mid-90%, which mirrors the typical performance of the reference policies. However, one particular point of the distribution (around 98%) has been more present than others, and it corresponds to this DRIP-MDP that requires less but at a better service level. The percentage plot shows that about 40% of the runs are around 95%, whereas quite some runs get into the highest service-level bucket, due again to DRIP-MDP. On the whole, the figure indicates that the introduced scheme not only preserves its prevalent performance around the 95% service level, but it also moves part of the distribution toward perfect services and, as a result, endows both the means and the tails with goodness.

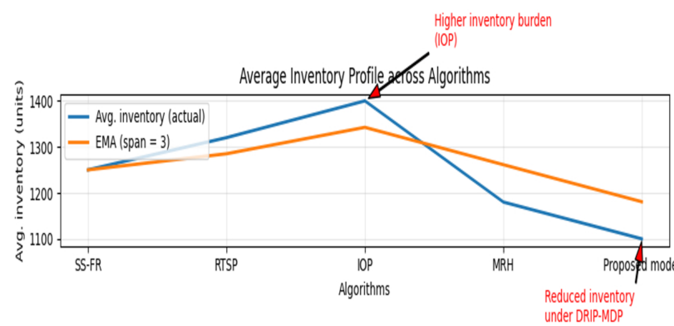
Figure 4: Unmet demand time series, test prediction, and future forecast under the proposed DRIP-MDP.



Source: Authors.

Figure 4 illustrates the unmet demand (units/period) over the time series of training and testing, as well as forecast horizons. The historical training data are shown in the blue curve, which oscillates between about 13 and 28 units of unmet demand, showing a clear trend across time. The green part is the testing period, and the black dashed line indicates the model's predictions within the same horizon; close overlap between these two traces suggests good capacity for capturing unmet demand dynamics by the model. The red dashed segment shows the forecast of the future under the learnt DRIP-MDP policy, exhibiting a descending trend and convergence toward lower unmet - demand levels compared to the past periods. This behaviour indicates that DRIP-MDP will help to drive the magnitude and volatility of shortages down for future operations; it should yield a more reliable service performance once deployed [19].

Figure 5: Average inventory profile across algorithms with EMA smoothing.

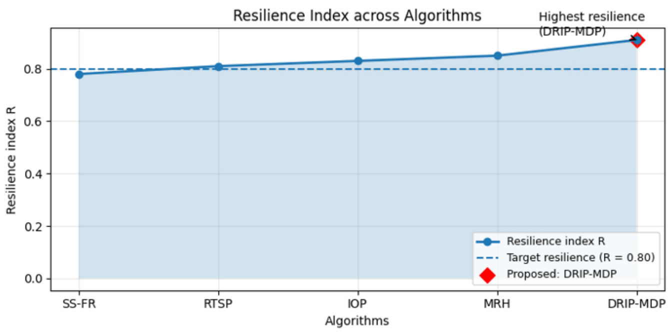


Source: Authors.

Figure 5 compares mean inventory levels under different algorithms. The true average inventory is plotted in the blue line, and the orange line shows an exponentially weighted moving average (EMA, span = 3) to accentuate trends. Inventory builds slowly from SS-FR to RTSP, then jumps higher still as we drop to IOP starvation by imposing the highest attachment and so stocking the largest inventory (as indicated at the top of this curve). As we move from IOP to MRH and then to the proposed model (DRIPMDP), both true and average inventory levels are reduced, reflecting a tendency toward more dynamic system control against stockpiling. Note in the right-hand side of the figure, it reasserts that DRIP-MDP holds lower inventory than all the methods, which makes us conclude that our policy both reduces cost and improves service level without holding high stocks.

The resilience index for all the algorithms is shown in Figure 6, with a target minimum value of R = 0.80 indicated as a horizontal dashed line. The blue solid line demonstrates the improvement of resilience from static SS-FR policies (below the target value) over RTSP and IOP policies (both above threshold) to MRH in terms of \rho. The red diamond corresponds to the DRIP-MDP point, highlighting that our method achieves the highest resilience index among all algorithms and is considerably over the target line. The shaded region indicates the lower-than-target range and shows that only the most dynamic policies operate in this desired resilience range consistently. In general, it can be verified that DRIP-MDP has the most pow-

Figure 6: Resilience index R across algorithms with target threshold and proposed DRIP-MDP.



Source: Authors.

erful disruptive coping and recovery ability among competitive strategies.

Experimental results demonstrate that the (DRIP-MDP) policy proposed herein significantly outperforms all four benchmark policies (SS-FR, RTSP, IOP, MRH) on the entire performance measure set. DRIP-MDP leads to the least average total cost while at the same time having the highest service level and lowest unmet demand (comparison table, cost plot and demand plot). The average inventory profile shows that these gains are achieved without much stock being held: the inventory level under DRIP-MDP is just equal to, or less than, that of the best existing policy and far lower than that of IOP, making benefits not from delaying retail orders. Finally, the R curve indicates that DRIP-MDP has better resilience than the target threshold and is strictly greater than all other baselines. Taken together, these results confirm that the coordination of dynamic re-routing and inventory pre-positioning via an MDP/ADP-based model provides a better trade-off between efficiency, serving performance, and resilience [20].

Conclusions.

This paper concerned the disruption management problem in multi-echelon supply chains to simultaneously solve dynamic re-routing and inventory pre-positioning. The DRIP-MDP framework describes the supply chain as a Markov Decision Process with states representing inventory levels, backlogs and link-status information, and actions being shipments and pre - positioning decisions governed by capacity constraints and budget restrictions. To circumvent the curse of dimensionality, DRIP-MDP uses a simulation-based ADP approach featuring value function approximation and one-step look-ahead to give rise to state-dependent policy which can be used as a real-time closed-loop controller. Comprehensive simulation experiments show that DRIP-MDP achieves MAH values over four representative baselines (i.e., SS-FR, RTSP, IOP, and MRH) with respect to five performance metrics. The newly introduced policy has the smallest expected total cost, highest service level and least unmet demand among existing algorithms, and it results in a mean inventory at or below that of all other methods. Furthermore, the resilience index R found that DRIP-MDP achieves more rapid and complete recovery from disruptions than all other competing strategies, meaning that the combined routing–prepositioning design has improved both efficiency and robustness rather than sacrificing one for the other. Future lines of work will do the extension to multiple products and firms in which coordination among decentralised decision-makers is needed as well as learning from real operational data of derivative disruption parameters. Another potential direction can be the incorporation of environmental objectives as metrics, e.g., carbon emissions, in cost parameters for joint economic and sustainability optimisation.

References.

[1] A. Dolgui, O. Gusikhin, D. Ivanov, X. Li, and K. Stecke, “A network-of-networks adaptation for cross-industry manufacturing repurposing,” IISE Trans, vol. 56, no. 6, pp. 666–682, Jun. 2024, doi: 10.1080/24725854.2023.2253881.

[2] Md. T. F. Shehun, “SYSTEMATIC REVIEW OF INDUSTRIAL ENGINEERING APPROACHES TO APPAREL SUPPLY CHAIN RESILIENCE IN THE U.S. CONTEXT,” American Journal of Interdisciplinary Studies, vol. 03, no. 04, pp. 235–267, Dec. 2022, doi: 10.63125/teherz38.

[3] M. K. Nayeem and G. M. Lee, “Robust Design of Relief Distribution Networks Considering Uncertainty,” Sustainability, vol. 13, no. 16, p. 9281, Aug. 2021, doi: 10.3390/su13169281.

[4] Monday F. Ohemu, Ambrose A. Azeta, Ibrahim A. Adeyanju, and Chukwuemeka C. Obasi, “An Effective Machine Learning Approach for Explosive Trace Detection,” International Journal of Data Informatics and Intelligent Computing, vol. 4, no. 1, pp. 15–29, Jan. 2025, doi: 10.59461/ijdiic.v4i1.161.

[5] I. Hssini, “Distributionally Robust and Learning-Based Vehicle Routing for Disaster Blood Logistics,” IEEE Access, vol. 13, pp. 119538–119550, 2025, doi: 10.1109/ACCESS.2025.3584213.

[6] M. Doodman, I. Shokr, A. Bozorgi-Amiri, and F. Jolai, “Pre-positioning and dynamic operations planning in pre- and post-disaster phases with lateral transhipment under uncertainty and disruption,” Journal of Industrial Engineering International, vol. 15, no. S1, pp. 53–68, Dec. 2019, doi: 10.1007/s40092-019-0317-7.

[7] V. Di Pasquale, F. Fruggiero, and R. Iannone, “A numerical approach for inventory pre-positioning in emergency management,” Journal of Humanitarian Logistics and Supply Chain Management, vol. 10, no. 2, pp. 205–230, Mar. 2020, doi: 10.1108/JHLSCM-07-2019-0043.

[8] T. Y. Saputra, O. Pots, K. S. de Smidt-Destombes, and S. de Leeuw, “The impact of Mean Time Between Disasters on inventory pre-positioning strategy,” Disaster Prev Manag, vol. 24, no. 1, pp. 115–131, Feb. 2015, doi: 10.1108/DPM-11-2013-0197.

[9] A. Beiki Ashkezari, M. Zokaee, E. Rabbani, M. Rabbani, and A. Aghsami, “A scenario-based game theory integrating with a location-allocation-routing problem in a pre- and

post-disaster humanitarian logistics network under uncertainty," Journal of Modelling in Management, vol. 19, no. 5, pp. 1686–1718, Oct. 2024, doi: 10.1108/JM2-08-2023-0181.

[10] S. Hu, Q. Hu, S. Tao, and D. Sasha, "A Multi-Stage Stochastic Programming Approach for Dynamic Pre-Positioning of Relief Supplies Considering Return," SSRN Electronic Journal, 2022, doi: 10.2139/ssrn.4079778.

[11] Y. Ju, H. Hou, J. Yang, Y. Ren, and J. Yang, "Integrating Trade-In Strategies for Optimal Pre-Positioning Decisions in Relief Supply-Chain Systems," Systems, vol. 12, no. 6, p. 216, Jun. 2024, doi: 10.3390/systems12060216.

[12] E. Barojas-Payán, D. Sánchez-Partida, S.-O. Caballero-Morales, and J.-L. Martínez-Flores, "A Hybrid Capacitated Multi-facility Location Model for Pre-positioning Warehouses and Inventories in Regions at Risk in Mexico," in Disaster Risk Reduction in Mexico, Cham: Springer International Publishing, 2021, pp. 193–221. doi: 10.1007/978-3-030-67295-9_9.

[13] B. Antunes, "Statistical Quality and Reproducibility of Pseudorandom Number Generators in Machine Learning Technologies," International Journal of Data Informatics and Intelligent Computing, vol. 4, no. 3, pp. 23–32, Aug. 2025, doi: 10.59461/ijdiic.v4i3.214.

[14] B. Balcik, S. Silvestri, M. Rancourt, and G. Laporte, "Collaborative Prepositioning Network Design for Regional Disaster Response," Prod Oper Manag, vol. 28, no. 10, pp. 2431–2455, Oct. 2019, doi: 10.1111/poms.13053.

[15] S. Roh and D.-W. Kwak, "Structural analysis of the challenges in operating pre-positioned warehouses for humanitarian relief logistics," Korean Academy Of International Commerce, vol. 36, no. 2, pp. 135–161, Jun. 2021, doi: 10.18104/kaic.2021.36.2.135.

[16] U. Danjuma Maiwada, Muhammad Muntassir Yakubu, Abubakar Danjuma Maiwada, Jameel Shehu Yalli, and Izaddeen Kabir Yakasai, "Techniques to improve 5G blockchain scalability issues through an analysis of a blockchain scalability," International Journal of Data Informatics and Intelligent Computing, vol. 3, no. 4, pp. 8–25, Oct. 2024, doi: 10.59461/ijdiic.v3i4.134.

[17] Q. Wang, Z. Liu, P. Jiang, and L. Luo, "A stochastic programming model for emergency supplies pre-positioning, transshipment and procurement in a regional healthcare coalition," Socioecon Plann Sci, vol. 82, p. 101279, Aug. 2022, doi: 10.1016/j.seps.2022.101279.

[18] J.-F. Camacho-Vallejo and D. Dávila, "Nested evolutionary algorithms for solving a bi-level warehouse location problem that considers inventory decisions," Ann Oper Res, Aug. 2024, doi: 10.1007/s10479-024-06182-3.

[19] S. R. Abazari, A. Aghsami, and M. Rabbani, "Prepositioning and distributing relief items in humanitarian logistics with uncertain parameters," Socioecon Plann Sci, vol. 74, p. 100933, Apr. 2021, doi: 10.1016/j.seps.2020.100933.

[20] Atta Ur Rahman, Bibi Saqia, Yousef S. Alsenani, and Inam Ullah, "Data Quality, Bias, and Strategic Challenges in Reinforcement Learning for Healthcare: A Survey," International Journal of Data Informatics and Intelligent Computing, vol. 3, no. 3, pp. 24–42, Sep. 2024, doi: 10.59461/ijdiic.v3i3.-128.