Operations Research Perspectives 15 (2025) 100367

Contents lists available at ScienceDirect

Operations Research Perspectives

journal homepage: www.elsevier.com/locate/orp

A systematic review of machine learning approaches in inventory control optimization

Ritsaart Bergsma [a,b,*], Corné de Ruijt [b], Sandjai Bhulai [a]

[a] Department of Mathematics, VU University Amsterdam, De Boelelaan 1111, Amsterdam, 1081HV, The Netherlands
[b] Professorship Supply Chain Finance, Windesheim University of Applied Sciences, Campus 2, Zwolle, 8017CA, The Netherlands

ARTICLE INFO

Keywords:
Inventory optimization
Machine learning
Reinforcement learning
Demand forecasting
Supply chain analytics
Multi-echelon systems

ABSTRACT

This systematic review investigates the applications of machine learning (ML) in inventory control, analyzing 122 articles to provide a comprehensive overview of the state of the art and identify future research directions. The study proposes a typology to classify the integration of ML into the inventory optimization framework, distinguishing three primary approaches: (1) separate estimation and optimization, where ML is applied to demand forecasting before optimization, (2) static ML-integrated optimization, where ML is directly embedded into optimization models, and (3) dynamic ML-integrated optimization, where reinforcement learning (RL) is employed to derive optimal inventory policies. The findings highlight that while RL applications are gaining prominence, significant research gaps remain, particularly in scaling algorithms to real-world problems, handling large action spaces, and developing RL algorithms that are tailored to inventory control. The review also assesses the operational dynamics of inventory systems addressed in the literature, such as single/multi-item models, lead time assumptions, and echelon structures. Underexplored areas include stochastic lead times, complementary items, quantity discounts, product obsolescence, and multi-echelon networks. The study concludes by outlining key research gaps and offering directions for future research to advance the integration of ML in inventory control.

1. Introduction

Artificial Intelligence (AI) has become a transformative force in supply chain management (SCM), enabling organizations to optimize operations in ways that were previously infeasible. According to surveys, business managers expect AI to save costs in SCM more than in other disciplines [1]. Supply chain managers have to make a range of decisions in order to satisfy a multitude of stakeholders. AI has the potential to impact many of these decisions, offering both assistance and the potential for fully autonomous decision making [2]. One of the key techniques within AI is machine learning (ML). This review examines the role of ML in inventory control, the practice of making the optimal inventory decisions.

Inventory control remains a critical challenge in modern supply chains. In today's consumer market, customers increasingly expect rapid fulfillment—driven in part by the growing availability of same-day delivery options [3]. At the same time, holding excess stock can lead to significant waste. In the United States alone, an estimated 31% of the food supply — valued at $382 billion — is discarded annually [4]. This persistent trade-off between minimizing stockouts and avoiding excess inventory underscores the need for more responsive and data-driven inventory systems. Machine learning offers promising solutions to better manage this trade-off by enabling more accurate forecasting, dynamic decision-making, and adaptive control under uncertainty.

Recent advances in ML have the potential to impact inventory control in various ways. Forecasting demand is an important aspect of inventory control, and ML has the potential to generate superior forecasting accuracy compared to statistical models [5]. Furthermore, there has been a flourishing line of research related to data-driven inventory models that leverage data-rich environments to make replenishment decisions [6,7]. In addition, there is growing interest in applying reinforcement learning to inventory control problems. In this method, an agent learns to make decisions, unlocking the potential to address complex inventory control scenarios that were previously considered intractable [8].

A growing body of literature explores the application of ML to inventory management, yet most existing reviews fall short in two key ways. First, they tend to group papers by ML algorithm (e.g., neural networks, decision trees) without examining how ML is functionally

* Corresponding author at: Department of Mathematics, VU University Amsterdam, De Boelelaan 1111, Amsterdam, 1081HV, The Netherlands.
E-mail address: rf.bergsma@windesheim.nl (R. Bergsma).

https://doi.org/10.1016/j.orp.2025.100367
Received 3 July 2025; Received in revised form 23 October 2025; Accepted 18 November 2025
Available online 19 November 2025

Table 1
Summary of related systematic reviews of machine learning in inventory control. * papers that do consider the ML-integration aspect but only focus on a single method of integrating ML such as reinforcement learning.

Article	Area of review	ML integration	Inventory system characteristics	Time span	Articles
Gutierrez et al. [9]	ML in inventory control	–	–	2014–2024	81
Albayrak Ünal et al. [10]	ML in inventory control	–	–	2012–2022	59
Rolf et al. [11]	Reinforcement learning in SCM	x*	–	2000–2021	103
de Castro Moraes and Yuan [12]	data-driven newsvendor problems	x*	–	all until 2021	24
Our review	**ML in inventory control**	x	x	1980–2024	122

integrated—whether as a forecasting tool or directly into the optimization technique. Second, they often overlook the underlying inventory system dynamics, such as stochastic lead times, shelf life, or multi-echelon structures, which are crucial for understanding the applicability and limits of ML models in practice. Table 1 contains an overview of other recent systematic literature reviews.

This review addresses these limitations by synthesizing existing research at the intersection of ML techniques and inventory system design. Rather than proposing an entirely new typology, we organize the literature using well-established inventory modeling dimensions (e.g., demand structure, lead time, product characteristics), and cross-analyze them with a structured classification of ML integration methodologies:

- We systematically analyze how machine learning (ML) techniques are integrated into inventory optimization frameworks (RQ1).
- We categorize and synthesize inventory-system characteristics along eight aspects: number of items and product interactions, time horizon (single- vs. multi-period), supply process (lead times and sourcing, including deterministic/stochastic and multi-supplier cases), procurement structure (e.g., fixed setup costs, quantity discounts), shortage treatment (backorders vs. lost sales), shelf-life dynamics (perishability/obsolescence), echelon structure (single vs. multi-echelon, including serial/divergent/network forms), and capacity constraints. This framing lets us pinpoint the operational contexts in which ML has been applied (RQ2).
- We identify key gaps and emerging opportunities in the literature, offering a roadmap for future research (RQ3).

The remainder is structured as follows. Section 2 describes the systematic review methodology. Section 3 reports publication trends and a meta-analysis of the corpus. Section 4 presents the classification framework. Section 5 synthesizes the literature by category and addresses RQ1 and RQ2. Finally, Section 6 concludes and outlines directions for future research (RQ3).

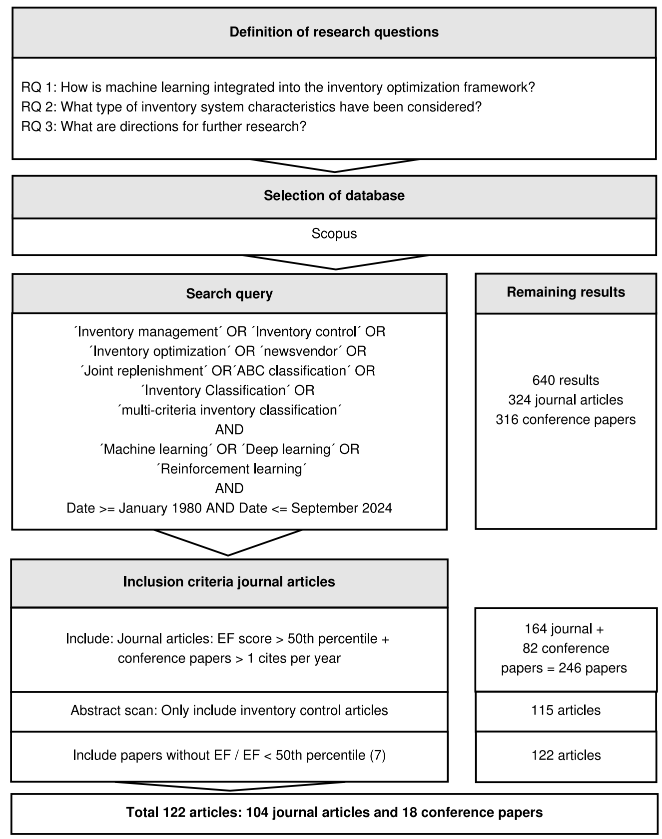

Fig. 1. Review methodology.

2. Review methodology

We conducted a *semi-systematic literature review* [13], structured around PRISMA-style stages—identification, screening, eligibility, and inclusion. This approach suits our goal of synthesizing a broad, multi-decade field by combining descriptive statistics with qualitative analysis [14]. Our review methodology is summarized in Fig. 1.

Research questions. We organized the review around three questions:

- **RQ1:** How is machine learning integrated into the inventory optimization framework?
- **RQ2:** What types of inventory system characteristics have been considered?
- **RQ3:** What are the key directions for further research?

Sources and search strategy. We searched Scopus for 1980–September 2024. Scopus was selected for its broad coverage of interdisciplinary research in management science, computer science, and engineering.

To address challenges in keyword selection, we prioritized terms commonly used in recent, high-quality publications to ensure alignment with current methodological trends and terminologies in the field. While this inevitably skews the search results toward more recent contributions, it enables a more accurate representation of contemporary research practices. To mitigate the risk of omitting foundational literature that uses older or alternative terminology, we employed a snowballing technique — a backward and forward citation search — on key papers identified in the initial dataset. This process led to the inclusion of important terms such as "newsvendor" and "joint replenishment", which were underrepresented in the original keyword set but are essential to the inventory control literature. The final query blends ML terms with inventory control-specific terminology (query details are displayed in Fig. 1).

Screening and quality filters. The initial search returned 324 journal records and 316 conference records (total 640). Because of the large number of articles, we applied filters: journals at or above the 50th percentile on the *Eigenfactor* score [15] and conference papers with more than 1 citation per year. We then removed duplicates and screened titles/abstracts. If needed, full texts were reviewed. During quality control, we noted that several papers, cited frequently by other key works in the corpus, had been excluded by the Eigenfactor (EF) filter because their journals lacked an EF score or were just below the cutoff.

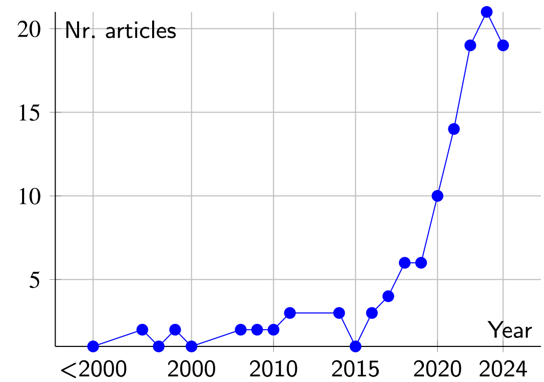

Fig. 2. Publications per year (n = 122). Articles before 2000 were grouped.

Table 2
Publications per journal.

Journal	Count
Expert Systems with Applications	15
European Journal of Operational Research	11
Computers and Industrial Engineering	7
International Journal of Production Research	6
International Journal of Production Economics	6
Management Science	4
Computers & Industrial Engineering	3
Journal of the Operational Research Society	3
Operations Research	3
Annals of Operations Research	2
Applied Soft Computing	2
Computers and Operations Research	2
Manufacturing and Service Operations Management	2
Decision Support Systems	2
Engineering Applications of Artificial Intelligence	2
Production and Operations Management	2
Journal of Intelligent Manufacturing	2
Transportation Research Part E: Logistics and Transportation Review	2
Other	28

These papers were verified and retained, yielding seven additional inclusions.

Eligibility criteria. We included studies that (i) address an *inventory control* decision problem and (ii) *apply machine learning* (forecasting, ML-integrated optimization, or RL). We excluded studies that (a) are purely managerial/behavioral without formal ML or optimization, (b) focus primarily on vehicle routing/ride-hailing/last-mile without an inventory decision, or (c) are domain-specific infrastructures (e.g., power, water, gas) where the "inventory" concept does not generalize to stock control.

Inclusion and additions. The process yielded 122 included studies (104 journal, 18 conference). During snowballing we identified three seminal papers not captured by the keyword query and added them.

3. Publication trend and analysis

The journal publications are published in the journals shown in Table 2. There is a clear upward publication trend as shown in Fig. 2. Although this could be attributed to the overall increase in scientific publications (see STM [16]), the magnitude of the growth suggests that ML is gaining interest among researchers in the field of inventory control.

Table 2 shows the publications per journal. Given the review's focus on inventory control, it is unsurprising that most of the journals fall within the field of operations research.

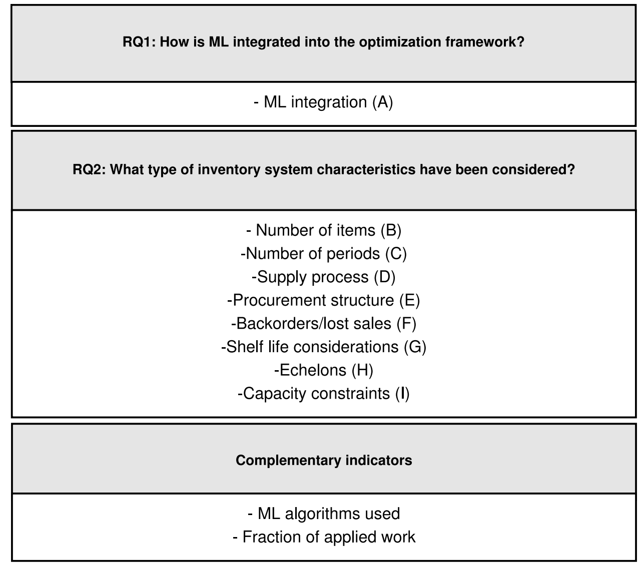

Fig. 3. Classification framework linking the research questions to the typology dimensions and complementary indicators used in this review.

4. Classification framework

The analysis follows a classification framework structured around the two research questions. It comprises (i) a typology that captures the conceptual dimensions of how ML is applied (RQ1) and which inventory systems are studied (RQ2), and (ii) two complementary indicators — *ML algorithms used* and *fraction of applied work* — that provide corpus-wide trend analyses related to these questions. The typology offers an interpretable mapping of the literature through nine letter-coded dimensions, while the complementary indicators summarize technical and empirical tendencies beyond the typology. Fig. 3 illustrates this framework and Fig. 4 presents the typology with each dimension described in detail.

4.1. Classification elements RQ 1: How is ML integrated into the inventory optimization framework?

ML integration (A)

A thorough reading of the literature revealed that ML techniques can be applied to inventory control in several main ways. Firstly, ML can be used to forecast demand that is then used in traditional inventory control optimization models. This is the most straightforward application of ML and we call this approach *separate estimation and optimization (A1)*. We also include in A1 papers that estimate not only point forecasts but also variances, quantiles, full distributions, or that select/weight historical scenarios conditional on features. In all such cases, the ML model is trained on statistical accuracy (e.g., MSE, likelihood, pinball) and its output serves as an input to a subsequent optimization model.

Articles that incorporate ML directly in the optimization step are grouped in the category of *static ML-integrated optimization (A2)*. In this category, we include the "data-driven inventory models" exemplified by works such as Bertsimas and Kallus [6] and Ban and Rudin [7]. Instead of separating forecasting and optimization, these models train with respect to the operational objective (e.g., underage/overage cost) or learn a mapping from state-features directly to decisions (including imitation of solver-optimal decisions) using a fixed ("static") dataset; the learned policy does not adapt via interaction after training. Our distinction between A1 and A2 is partly motivated by a prevailing trend in the literature, with researchers advocating for the integration

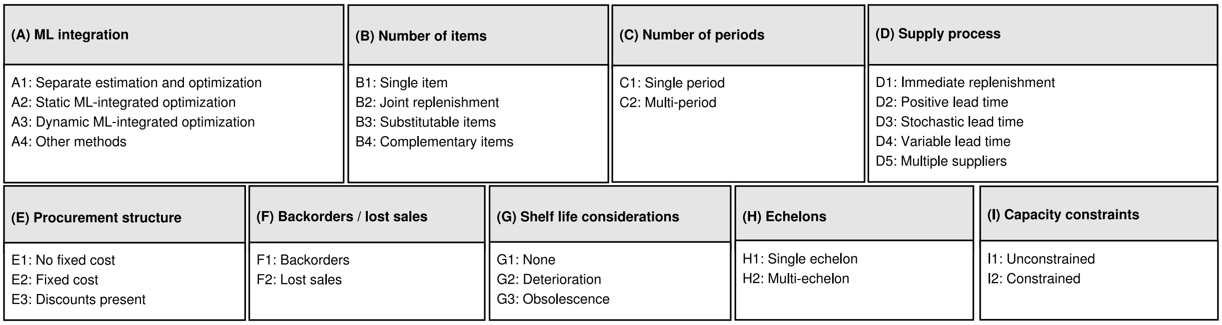

Fig. 4. Proposed inventory control typology.

of forecasting and optimization steps [17]. Additionally, several articles utilize this distinction, contrasting ML-integrated approaches with those in category A1 [18,19].

Category A3, dynamic ML-integrated optimization, also integrates ML into the optimization step but differs fundamentally in how the learning process is conducted. These articles employ reinforcement learning (RL) to derive optimal policies through sequential interaction with a simulated environment. The term dynamic refers here to the agent–environment feedback loop: the policy is not fixed after training on a static dataset (as in A2) but evolves through iterative decision–outcome cycles. This characteristic makes RL-based approaches qualitatively different from the data-driven models in A2 and justifies their treatment as a separate category. Furthermore, RL has emerged as a major research stream within operations research over the past decade, warranting special attention [8].

Category A4 — other methods contains articles that do not fit neatly into the aforementioned categories. These contributions are reviewed on a per-topic basis (e.g., inventory classification, backorder prediction, or hybrid heuristic/metaheuristic approaches assisted by ML).

4.2. Classification elements: RQ2: What type of inventory system characteristics have been considered?

Inventory control has its origin in the economic order quantity (EOQ) model formulated by Harris in 1913 [20]. The model aims to optimize the order quantity by taking into account order costs and holding costs, assuming deterministic and constant demand. In addition, this model is considered single-item, meaning that it optimizes for the inventory of each item separately, not taking into account inter-item dependencies. Inventory theory has since expanded to include a multitude of models that account for stochastic demand, variable lead times, single- or multi-echelon inventory, perishable items, and production planning. Each of these models provides added value depending on a company's specific inventory characteristics.

In this review, we will examine the model characteristics along various dimensions that are incorporated in our proposed typology (Fig. 4). In order to identify relevant dimensions, we draw upon earlier typologies such as Prasad [21], de Kok et al. [22], and Silver [23]. In addition, a recently compiled research handbook featuring many leading academics in the field of inventory control was used to inform the selection of relevant dimensions [24]. All dimensions (B through I) are incorporated into our proposed typology (Fig. 4).

The next subsections will discuss each of the dimensions (B through I) in more detail.

4.2.1. Number of items (B)

Categories B1-B4 are taken directly from Silver's taxonomy [23]. Most of the literature in inventory control is devoted to modeling a single item in separation of all other items (B1). However, cost might be saved when ordering items together (B2). Common approaches to doing this include cyclic ordering [25]. Category B3 contains models with product substitution. When items are out of stock, a customer might opt for a different item. It is relevant for inventory models to consider this substitution effect because a stockout will not necessarily result in a lost sale [26]. Complementary items, denoted as B4, refer to multi-item models in which the service level depends on multiple items being in stock. This may be encountered in, for example, assemble-to-order systems, where the absence of one component can result in the inability to manufacture the entire end product [27]. Spare part inventory problems may also involve multi-item models, as parts are required to support the availability of a capital good, although single-item models also exist [28].

4.2.2. Number of periods (C)

Inventory models can either consider a single period or multiple periods. In some situations (newspapers, fashion), there is a short selling period and excess stock cannot be used to cover the demand in the following period. Single-period models (C1) decouple adjacent periods and do not account for leftover inventory from previous periods. This simplifies the analysis. An example of such a model is the classic newsvendor problem [29]. Multi-period models (C2) do not consider excess inventory as lost. There is, however, a holding cost associated with inventory.

4.2.3. Supply process (D)

When replenishing stock, the speed and manner in which replenishment arrives are important considerations. Longer lead times require a higher optimal base stock level [30]. Many models assume immediate replenishment ($L = 0$)(D1). Other papers assume lead times to be known and fixed on a certain number of periods ($L \geq 1$)(D2). Stochastic lead times are sometimes also considered (D3). Category D4 includes models in which lead time varies by product. Dual sourcing models typically involve two suppliers per stock keeping unit, with one usually offering a shorter lead time at a higher cost, while the other provides a longer lead time at a lower cost [31]. We group these articles into D5.

4.2.4. Procurement structure (E)

Most inventory models separate a fixed, quantity -independent cost per order from a variable per-unit purchase cost. We write the procurement cost for an order of size z as $C(z) = K + c \cdot z$, where K is the fixed ordering/setup cost and c the unit price. When $K = 0$ (E1) there is no incentive to batch orders; when $K > 0$ (E2) each order triggers a fixed charge, capturing, for example, administrative/handling fees for external purchasing or a production setup for in-house manufacturing. Category E3, following Prasad [21], covers discount schemes such as quantity discounts or sudden discounts introduced by the supplier.

4.2.5. Backorders/lost sales (F)

When a product is out of stock, various outcomes are possible. Some models incorporate backorder costs (F1), meaning the sale is not lost but incurs an additional delivery-related fee. Other models treat excess demand as lost sales — the newsvendor problem is a

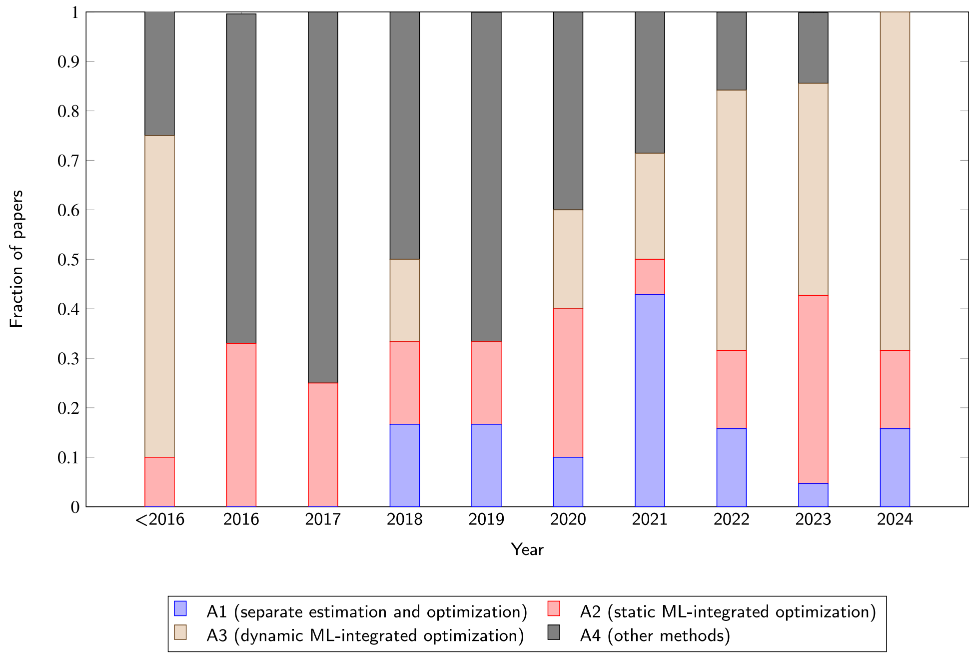

Fig. 5. Fraction of papers using different ML integration approaches per year.

common example. The newsvendor model is a simple single-period model in which shortage is charged at a "shortage cost" and excess inventory is charged at an overage cost. All excess inventory is therefore salvaged at the end of the period and is not carried over. Although the newsvendor model is relatively tractable, multi-period lost sales systems are generally more complex and less tractable [32].

4.2.6. Shelf life considerations (G)

Most classic inventory models do not take into account shelf life considerations. These models either assume unlimited shelf life or simply do not consider any inventory to carry over into the next period (such as the newsvendor model). We will group these articles into category G1. Category G2 includes models with deterioration schemes, which are especially relevant for fresh food inventory systems. Category G3 considers inventory subject to obsolescence. These models often assume unlimited shelf life, but demand declines as the product ages. Obsolescence is common, for example, in electronics: although older devices remain functional and degrade slowly, they can quickly become obsolete due to technological innovation.

In this review, articles employing the classical newsvendor model will be classified under the category G1. Although these articles do account for perishable inventories by salvaging excess inventory at the end of the period, they are grouped as G1 to distinguish them from more complex deterioration schemes.

4.2.7. Number of echelons (H)

Single echelon systems (H1) are the most common in the literature. They focus on optimizing the inventory level at a single stock point per stock-keeping unit. In multi-echelon systems (H2), stock-keeping units move through a network of installations toward the end customer. These models seek to determine the optimal inventory levels at each installation point.

These installations may be configured in various ways. One such configuration is the serial system, in which installation 1 orders from installation 2, which in turn orders from installation 3, and so on [33]. Other configurations also exist, such as one warehouse restocking multiple retail locations. Scenarios involving multiple installations introduce additional decision variables for each location. These problems are thus considered more complex. Notable approaches include Clark and Scarf's method, which first solves for installation 1 and then for subsequent installations [34].

4.2.8. Capacity constraints (I)

Inventory systems are often subject to capacity constraints. For example, a fashion retailer's product assortment is constrained by limited shelf space. In manufacturing systems, constraints often involve limited storage space for semi-finished products. It is known that inventory systems modeled without capacity limits are often a poor proxy for systems that are subject to these limits [35]. As environmental considerations like carbon budgets gain relevance, capacity constraints may play a more critical role — making this a promising area for future research. This review distinguishes between papers that do not consider capacity constraints (I1) and papers that do (I2).

4.3. Complementary indicators

In addition to the typology dimensions discussed above, two complementary indicators were tracked to contextualize the corpus and to enable trend analyses over time: (i) the ML algorithms used, which summarizes the technical methods applied across studies and relates broadly to RQ1, and (ii) the fraction of applied work, which distinguishes between empirical and simulation-based analyses and complements both research questions. These indicators provide quantitative context — showing how algorithmic choices and empirical validation have evolved — but are not assigned letter codes, as they summarize corpus-wide tendencies rather than typology categories.

4.3.1. ML algorithms used

ML is a subfield of artificial intelligence (AI) that focuses on the development of algorithms and models that enable computers to learn from data and make predictions or decisions without these being explicitly programmed. The fundamental idea behind ML is to allow systems to automatically improve their performance over time by learning from experience. A wide variety of algorithms exist in the field of ML. Broadly speaking, these techniques fall into supervised, and unsupervised learning techniques. In supervised learning, the algorithm learns from a labeled dataset (such as historic demand), whereas unsupervised techniques find patterns in data that are not labeled. Within these categories, there are a multitude of algorithms such as Artificial Neural Networks (ANNs), linear regression, decision trees, and others.

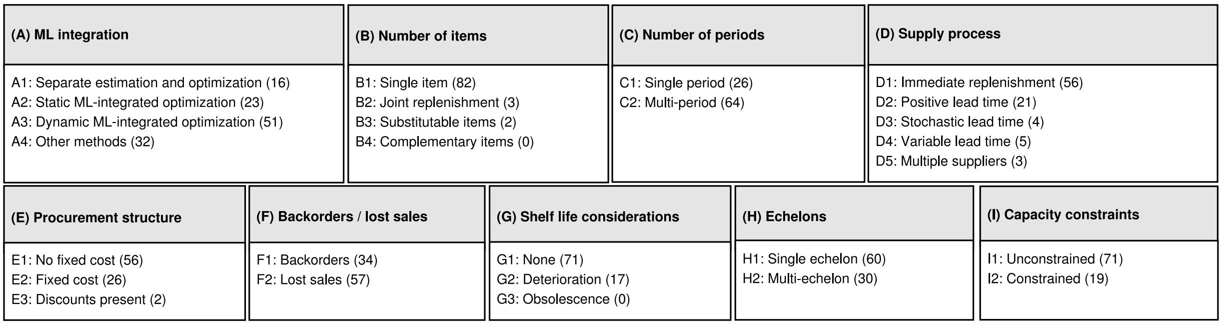

Fig. 6. Proposed typology of inventory system characteristics. The numbers in parentheses indicate the count of papers that consider each characteristic. Since some papers analyze multiple settings, they may be counted in more than one category.

Beyond supervised and unsupervised learning, we recognize reinforcement learning (RL) as a distinct category. While deep reinforcement learning (DRL) may employ neural network architectures similar to those used in supervised learning, RL fundamentally differs in its learning paradigm: it focuses on learning optimal actions through interaction with an environment, guided by trial-and-error and feedback in the form of rewards. Because RL algorithms are designed to address sequential decision-making problems — where the goal is not just to predict or classify but to optimize long-term outcomes — they warrant separate treatment from traditional supervised and unsupervised approaches.

This review assesses the frequency of each technique's usage by recording how often it appears in the literature. Since most papers employ multiple techniques, double counting is permissible. The primary objective is to identify the most prevalent techniques in the literature.

4.3.2. *Fraction of applied work*

This review examines the fraction of papers that use real-world data as opposed to synthetic datasets. Other authors have noted that ML applications in the literature only consider stylized problems with synthetic data instead of realistic real-world problems [11]. The applicability of newly developed techniques is important to the inventory control literature. Therefore, it is essential that enough empirical studies evaluate their practical effectiveness. Accordingly, we assess how many studies utilize real-world data obtained from companies.

5. Analysis

Fig. 5 shows the field's gradual shift away from *separate estimation and optimization* (A1) toward ML-integrated methods—first to *static ML-integrated optimization* (A2) and, more recently, to *dynamic ML-integrated optimization* (A3, RL). Two forces explain this movement. First, forecast accuracy does not guarantee decision quality: the training loss for a predictor can be misaligned with inventory costs, especially under positive lead times, capacity limits, or asymmetric overage/underage trade-offs [17]. A2 addresses this by embedding the inventory objective directly in a supervised loss. Second, when actions materially shape future states — multi-echelon flows, perishability, or stochastic lead times — endogeneity and feedback effects become important. A3 addresses this by learning from interaction with a simulated environment.

Fig. 6 displays the variety of inventory characteristics that are considered in the literature across all methods. We see that most aspects are considered by the literature, with the notable exception of complementary items (B4) and obsolescence (G3). In addition we see that the literature is skewed toward simpler model assumptions, especially on number items (B) and lead time (D).

Sections 5.1–5.3 examine *how* machine learning is integrated into inventory optimization (A1–A3) and *where* each approach is used across inventory characteristics B–I (number of items; number of periods; supply process; setup/ordering costs; backorders vs. lost sales; shelf life; echelons; capacity). We treat RQ1 and RQ2 jointly: for each approach we summarize the modeling idea, then analyze its fit to specific problem features and note underexplored combinations.

Section 5.4 briefly covers applications that fall outside A1–A3 (e.g., inventory classification). Section 5.5 summarizes algorithmic trends. Section 5.6 quantifies the share of studies using real-world versus simulated data.

5.1. *A1: Separate estimation and optimization*

The articles in this category use ML to produce inputs for established inventory models. This includes point forecasts, estimates of variance, quantiles or full conditional distributions, and feature-conditioned scenario sets. These ML components are trained for statistical accuracy and then fed into standard optimization (e.g., newsvendor, Wagner–Whitin, stochastic/robust programs). In total, 16 articles were classified under this category, displayed in Table 3. Notably, there has been a gradual decline in new contributions, which may reflect a growing consensus that integrating forecasting and optimization can lead to improved outcomes, as noted earlier.

5.1.1. *Methodological contributions*

Recent years have shown that ML can achieve superior forecasting accuracy compared to traditional statistical approaches [5]. This has encouraged researchers to apply ML techniques to inventory control. For example, Shi [46] employs ML to estimate both the mean and standard deviation of demand (the latter based on historical forecast errors), before applying a standard newsvendor model to determine the optimal order quantity. Another advantage of ML models is their ability to seamlessly incorporate additional data sources available within organizations. For instance, Abolghasemi et al. [52] utilize promotional data to enhance forecast accuracy. However, they also find that for products with high volatility, conventional methods can sometimes outperform ML models both in forecasting accuracy and inventory performance.

ML can also be integrated with stochastic and robust optimization approaches. López Lázaro et al. [36] integrate ML into robust optimization to optimize cash inventories of banks. In their robust optimization setup, they use the training set errors to build a confidence interval around demand. Galli et al. [43] use a similar approach, also for blood inventory optimization. A powerful aspect of these approaches is that they do not assume a standard distribution of demand, but let the ML model generate this distribution, potentially providing a more realistic fit to the data.

Other approaches that focus on distributional forecasting are also present in the literature. Ulrich et al. [44] apply several models that belong to the class of Generalized Additive Models for Location, Scale and Shape (GAMLSS), to forecast the distribution of demand and use this to select the optimal order quantity in the newsvendor model. They show that most of the models that belong to this class outperform standard regression models in terms of inventory cost. Cao and Shen [37]

Table 3
Articles that use the separate estimation and optimization approach (A1) with number of items (B), number of periods (C), lead time assumptions (D), setup cost (E), shelf life assumptions (G), single/multi-echelon (H), and capacity constraints (I).

Article	Year	B items	C periods	D supply process	E procurement structure	F back order	G shelf life	H echelon	I constraint
López Lázaro et al. [36]	2018	B1	C2	D1	E2	F2	G1	H1	I1
Cao and Shen [37]	2019	B1	C1	D1	E1	F2	G1	H1	I1
Abolghasemi et al. [38]	2020	B1	C1	D1	E1	F2	G1	H1	I1
Li et al. [39]	2021	B1	C2	D1	E2	F1	G2	H1	I1
Pereira and Frazzon [40]	2021	–	–	–	–	F1, F2	–	H2	I1
Deng and Liu [41]	2021	–	C2	–	–	–	–	H2	I1
Gonçalves et al. [42]	2021	B1	C1	D1	E1	F1	G1	H1	I1
Galli et al. [43]	2021	B1	C2	D1	E2	–	G1	H1	I1
Ulrich et al. [44]	2021	B1	C1	D1	E1	F1	G1	H1	I1
Shokouhifar and Ranjbarimesan [45]	2022	B1	C2	D1	E2	F2	G1	H2	I1
Shi [46]	2022	B1	C1	D1	E2	F2	G1	H1	I1
Li et al. [47]	2022	B2	C2	D1	–	F1	G2	H1	I1
Singh and Mishra [48]	2023	B1	C2	D1	E2	F1	G2	H1	I1
Fan et al. [49]	2024	B1	C2	D5	E1	F2	G1	H1	I1
Mete Ayhan and Kır [50]	2024	B1	C2	D1	E1	–	G1	H1	I1
Singh and Mishra [51]	2024	B1	C2	D2	E2	F1, F2	G2	H1	I2

propose a new neural network model for quantile forecasting and apply this to the newsvendor problem. These approaches offer an advantage over approaches that simply use the historical forecast's error standard deviation, in that they have the ability to also differentiate between forecasts that are more or less certain. For example, when a company puts certain products on sale, this not only increases expected demand but also increases uncertainty.

5.1.2. Inventory system characteristics

A1 appears beyond single-period settings whenever a forecast (point, quantile, or scenarios) can be plugged into an existing model and solved as usual: multi-period cash inventories [36], hospital drug replenishment via scenario inputs [43], plug-ins to classical procedures (e.g., Wagner–Whitin, Silver–Meal) [50], and multi-echelon contexts [41].

The strength of A1 is interoperability: ML outputs drop cleanly into linear/stochastic/robust formulations without redesigning constraints or solvers, keeping models explainable and auditable. The trade-off is that forecasts are not co-optimized with the policy, so errors can propagate, especially with positive/uncertain lead times. Consistent with this, most A1 studies in our sample assume immediate replenishment and single-item settings, with fewer cases involving uncertain lead times, strict capacities, or multi-echelon structures (see Table 3).

While theory suggests that end-to-end integration of forecasting and decision making can improve inventory performance, the interoperability with existing optimization procedures of A1 is often highly valuable in practice. We see value in applying A1 forecasting to richer settings — positive or stochastic lead times, capacity constraints, multi-echelon networks, perishables, and joint replenishment/substitution — and testing whether distributional/quantile models yield decision gains over simple baselines. A practical lever is to align forecasts with the decision context (e.g., forecast congruence [53]) to reduce order volatility without changing the optimization backbone.

5.2. A2: static ML-integrated optimization

In this category, the forecasting and optimization steps are *integrated* rather than treated separately (as in (A1) separate estimation and optimization). We refer to these as *static ML-integrated* methods because the decision rule is obtained by minimizing an empirical risk (a loss that embeds inventory costs) on a fixed dataset; the policy is not learned via interaction with a simulated environment or state transitions (as in A3 (dynamic ML-integrated optimization)). In practice, these "data-driven" models leverage feature-rich data and asymmetric cost structures (underage vs. overage) by optimizing a modified loss that targets the operational objective directly. We include 23 articles in this category (see Table 4).

5.2.1. Methodological contributions

There has been substantial development in the inventory control literature regarding data-driven newsvendor models. These models often extend the classic Sample Average Approximation (SAA) method [54]. In SAA, the unknown probability distribution of demand is replaced by the empirical distribution formed from sample data. The standard SAA formulation for the newsvendor problem is given by Cheung and Simchi-Levi [55]:

$$q^* = \arg\min_{q\geq 0} \frac{1}{n}\sum_{i=1}^{n}\left[c_u(d_i-q)^+ + c_o(q-d_i)^+\right],$$

where n is the sample size, $d_1,\dots,d_n$ are historical demand observations, c_u is the underage cost, c_o is the overage cost, q is the decision variable (order quantity), and $(x)^+ = \max(0,x)$.

Since the basic SAA does not incorporate feature or covariate information, several authors have extended this approach to include such data. Bertsimas and Kallus [6] propose a weighted SAA method, where the empirical distribution is replaced by a weighted empirical distribution based on a feature vector $\mathbf{x}$. The key idea is that, given a new feature vector $\mathbf{x}$, some historical demand scenarios are more relevant than others, and thus should be assigned higher weights. The modified SAA equation becomes:

$$q^*(\mathbf{x}) = \arg\min_{q\geq 0} \sum_{i=1}^{n} w_{n,i}(\mathbf{x})\left[c_u(d_i-q)^+ + c_o(q-d_i)^+\right],$$

where $w_{n,i}(\mathbf{x})$ is a weight assigned to each historical observation i, based on the similarity between the feature vector $\mathbf{x}$ and the observed feature vector $\mathbf{x}_i$. The weights are non-negative and sum to one: $\sum_{i=1}^{n} w_{n,i}(\mathbf{x}) = 1$. Bertsimas and Kallus [6] discuss several choices for the weight function, including k-nearest neighbors, random forests, and kernel regression. The subscript n in $w_{n,i}(\mathbf{x})$ refers to the sample size, and i indexes the historical data points (see Table 4).

Ban and Rudin [7] address the data-driven newsvendor problem from a different perspective by modeling the optimal order quantity as a function of the feature vector. Specifically, they formulate the problem as:

$$\min_{q(\cdot)\in\mathcal{Q},\, q:\chi\to\mathbb{R}} \sum_{i=1}^{n}\left[c_u(d_i-q(\mathbf{x}_i))^+ + c_o(q(\mathbf{x}_i)-d_i)^+\right],$$

where χ denotes the feature space, and $q(\cdot)$ is a function mapping feature vectors $\mathbf{x}_i \in \chi$ to order quantities. For example, a linear function is given by $q(\mathbf{x}_i) = \mathbf{q}'\mathbf{x}_i = \sum_{j=1}^{p} q_j x_{ij}$, where p is the number of features, q_j are the coefficients to be estimated, and x_{ij} is the jth feature of the ith observation. This formulation can be solved using quantile regression, which can be implemented as a linear program. Ban and Rudin [7] also consider kernel-based approaches for greater flexibility.

Table 4
Articles that integrate use *static ML-integrated optimization (A2)* with number of items (B), number of periods (C), lead time assumptions (D), setup cost (E), shelf life assumptions (G), single/multi-echelon (H), and capacity constraints (I1).

Article	A ML integration	B items	C periods	D supply process	E procurement structure	F back order	G shelf life	H echelon	I constraint
O'Neil et al. [56]	2016	B1	C1	D1	E1	F2	G1	H1	I1
Ban and Rudin [7]	2018	B1	C1	D1	E1	F2	G1	H1	I1
Huber et al. [19]	2019	B1	C1	D1	E1	F2	G1	H1	I1
Zhang and Gao [57]	2019	B1	C1	D1	E1	F2	G1	H1	I1
Oroojlooyjadid et al. [58]	2020	B1	C1	D1	E1	F2	G1	H1	I1
Bertsimas and Kallus [6]	2020	B1	C1	D1	E1	F2	G1	H1	I2
Punia et al. [59]	2020	B1	C1	D1	E1	F2	G1	H1	I2
Abbasi et al. [60]	2020	B1	C2	D1	–	F2	G2	H1	I1
Bertsimas and Koduri [61]	2021	B1	C1	D1	E1	F2	G1	H1	I1
Chen [62]	2021	B1	C1	D1	E1	F2	G1	H1	I1
Clausen and Li [63]	2022	B1	C2	D1	E1	F2	G1	H1	I1
Pirayesh Neghab et al. [64]	2022	B1	C1	D1	E2	F2	G1	H1	I1
Qi et al. [65]	2023	B1	C2	D4	E1	F1	G1	H1	I1
Tian and Zhang [66]	2023	B1	C1	D1	E1	F2	G1	H1	I1
Ren et al. [67]	2023	B1	C2	D1	E2	F1	G1	H1	I1
Bertsimas et al. [68]	2023	B1	C1	D1	E1	F1	G1	H1	I1
Forel et al. [69]	2023	B1	C1	D1	E1	F2	G1	H1	I1
Zhang and Tan [70]	2023	B1	C1	D1	E1	F2	G1	H1	I1
Kallus and Mao [71]	2023	B1	C1	D1	E1	F2	G1	H1	I1
Chen et al. [72]	2023	B3	C1	D1	E2	F2	G1	H1	I1
Bertsimas and Kim [73]	2024	B1	C1	D1	E1	F2	G1	H1	I1
Qi et al. [74]	2024	B1	C2	D1	E1	F2	G1	H1	I1
van der Haar et al. [75]	2024	B1	C2	D2, D5	E1	F1, F2	G1, G2	H1	I1

In addition to these methods, authors have applied neural networks to solve the newsvendor problem. Oroojlooyjadid et al. [18] modify the loss function of the deep learning algorithm to obtain the minimizer of the newsvendor cost function directly. They compare their approach to other benchmark approaches such as the methods introduced by Ban and Rudin [7] and Bertsimas and Kallus [6] and find that it outperforms these methods.

All of the aforementioned methods integrate the forecasting and optimization steps and leverage feature data. In principle, integrating these approaches is expected to outperform the separate estimation and optimization approach (A1), as it retains more of the dataset's information during optimization, rather than relying on assumed parameters like mean and standard deviation. However, Huber et al. [19] find that the out-performance of the integrated approaches only pertains to situations in which target service levels are below 0.8. In other words, if the holding costs become increasingly large compared to the stockout costs, integrating estimation and optimization has benefits. The literature would gain from additional studies to further substantiate these outcomes, given their significance for guiding research in this field.

5.2.2. Inventory system characteristics

Despite the methodological variety, most A2 papers utilize data-driven newsvendor–type formulations, where the inventory cost is embedded in a supervised loss function and minimized on historical samples (e.g., Bertsimas and Kallus [6],Ban and Rudin [7],Oroojlooyjadid et al. [18]). Single period models such as the newsvendor model are a very natural pairing to this modeling approach as the loss function of these methodologies does not capture multi-period/dynamic effects. Capacity does appear in these models by for example adding constraints via Lagrangian duality [6] or simple heuristics when quantile-based orders exceed shelf space [59]. In practice, these approaches shine where rich covariates can inform demand. For instance, [6] study a DVD retailer where features such as IMDB ratings and review signals augment the historical data.

Some papers do move beyond these simplified settings. Qi et al. [65] introduce an "end-to-end" inventory model, which combines dynamic programming for labeling optimal order quantities with neural network training for prediction. Their approach outperforms benchmark models and excels in accommodating complex inventory settings, including variable lead times for different products. Bertsimas and Koduri [61] introduce a method based on regression in reproducing kernel Hilbert spaces to solve optimization problems. Their approach takes the single-period newsvendor problem as one example, but is able to deal with sequential decision making as well, opening doors for further research in applying this model to more complex inventory settings. Van der Haar et al. [75] propose a supervised loss that replaces the Bellman value term with the portion of future cost that is an irrevocable, deterministic consequence of today's order over the lead-time/shelf-life window—capturing multi-period effects (and, for dual sourcing, via sequential expedite/regular losses) while preserving supervised-learning tractability.

A2 methods that encode multi-period effects directly in the loss can be far more sample- and compute-efficient than RL (A3), but they are still lightly tested outside stylized settings. A clear next step is to evaluate them systematically in contexts such as positive or stochastic lead times (D2–D3), capacity constraints (I2), multi-echelon systems (H2), and perishables (G2), using head-to-head benchmarks against the best-established methods in each problem family. Direct comparisons to RL to quantify the performance–compute trade-off are also valuable. Such evaluations would clarify whether this class can retain supervised-learning tractability while delivering competitive performance in richer inventory environments.

5.3. A3: dynamic ML-integrated optimization

Reinforcement learning (A3: dynamic ML-integrated optimization) learns a decision rule by interacting with a simulated environment, rather than minimizing a fixed, supervised loss as in A2 (static ML-integrated optimization). The inventory problem is cast as an MDP: the agent observes a state (e.g., on-hand, pipeline), takes an action (order), receives a cost/reward, and updates its policy to improve long-run performance. Because actions today change tomorrow's state, RL naturally captures endogeneity and multi-period trade-offs—a good fit for settings with positive or uncertain lead times, perishables, and multi-echelon flows. The next subsection surveys methodological contributions (algorithm choices, state/action design, and constraint handling) and then synthesizes where RL has been applied and what gaps remain. In our sample, A3 is the largest of the three integration families (51 papers, see Table 5), exceeding the combined counts of

A1 (separate estimation and optimization) and A2 (static ML-integrated optimization); see Fig. 6 for details.

5.3.1. Methodological contributions

In certain inventory problems, analytical solutions for finding the optimal policy are only possible when making simplifying assumptions. While these simplified policies may offer decision-makers useful insights, they can also be misleading under real-world complexity. Various inventory control settings, such as assemble-to-order systems, are considered intractable and the optimal policy structure remains largely unknown [76]. Reinforcement Learning (RL) offers a promising alternative in such settings.

Unlike supervised learning models — which are trained on fixed historical data — RL agents learn by interacting with their environment, updating their behavior through trial and error [77]. This interactivity enables RL to account for the long-term effects of actions, a key advantage in inventory settings where decisions today affect costs and availability in future periods. Traditional approaches often struggle to capture this endogeneity of decisions, particularly in multi-period environments.

To apply RL to inventory control, the problem must first be modeled as a Markov Decision Process (MDP). In this framework, the system transitions from one state to another based on the agent's actions, with each action yielding a reward according to a predefined function. Over time, the RL agent learns a policy — a mapping from states to actions — that maximizes cumulative reward [77].

The application of RL in inventory control has a long history, with some papers dating back to the 90 s. In that time, reinforcement learning was applied using case-based reasoning [78–80]. While some initial results were promising, this line of research did not result in widespread RL applications within the field. With the increase of computing power came Deep Reinforcement Learning (DRL), which utilizes neural network architectures. Famous algorithms are Deep Q-Learning and AlphaZero (both developed by Google DeepMind) [81, 82]. The widespread application to inventory control soon followed, with some papers even preceding this "reinforcement learning boom" (see Table 5).

A variety of RL algorithms have been applied to inventory control. These can be broadly categorized into,

- Value-based methods (e.g., Deep Q-Learning), where the agent estimates the expected return (Q-value) for each state–action pair and selects the action with the highest value,
- Policy-based methods, which learn a stochastic policy directly,
- Actor–Critic methods such as Proximal Policy Optimization (PPO), which combine both approaches by learning a policy (actor) and a value function (critic).

Most recent studies use either value-based methods such as Q-Learning [107,114] or actor–critic methods like PPO [112,121]. Deep Q-Learning is favored for its sample efficiency and suitability for discrete action spaces. PPO, on the other hand, offers improved stability and can handle high-dimensional or continuous action spaces — making it especially suitable for complex inventory scenarios. The use of pure policy-gradient methods remains limited, likely due to their instability and sample inefficiency in discrete, cost-sensitive inventory settings. The algorithms used within the papers are discussed more thoroughly in Section 5.5.

A noticeable trend in recent work is the adaptation of RL algorithms to better reflect inventory-specific challenges, including non-stationary demand, perishability, and coordination across multiple actors. For example, Mohamadi et al. [118] apply actor–critic methods in a vendor-managed inventory setting with perishables, while Kaynov et al. [121] introduce multi-output policy architectures to address action space complexity in multi-retailer systems. Enhancements in algorithmic scalability and convergence are proposed by Stranieri et al. [122], Tian et al. [123], and Luo et al. [125]. These studies collectively underscore a shift away from off-the-shelf algorithms toward domain-aware, customized RL approaches. A more detailed analysis of the algorithms used — including Q-learning variants, actor–critic models, and recent hybrid innovations — is presented in Section 5.5.

Many research gaps within the realm of reinforcement learning remain. Much of the work is primarily focused on stylized inventory control problems and does not use a real-world case study to verify the efficacy of DRL (see Section 5.6 for a further discussion of the fraction of applied studies). Scaling up DRL to real-world problems can be challenging, as they often involve substantially larger action spaces when the decision maker must make simultaneous and interdependent decisions (such as in multi-item models). Some recent studies have proposed ways to mitigate this issue within the inventory domain. For example, van Hezewijk et al. [112] reduce the action space of a multi-item EOQ model by allowing the agent to continue making production and switching decisions until it determines that no further actions are needed. Kaynov et al. [121] address a multi-retailer problem by letting the neural network output several probability distributions — one per retailer — instead of a single one, as is common in policy-based methods.

To address similar scalability issues more broadly, a wide line of RL research has explored how to make learning efficient in large or structured discrete action spaces. Earlier approaches have used *factorization or hierarchical decomposition* of the action space, such as binary or tensor factorizations (see [131–133] for non-deep RL examples, and [134] for a DRL example). Others have relied on *nearest-neighbor or embedding-based selection* among predefined feasible actions [135–138], while additional strategies employ *symbolic representations or hierarchical and multi-agent formulations* to decompose decision spaces [139–142]. Each of these families of methods improves computational efficiency but often requires extensive parameter tuning or prespecified action structures.

Building on this stream of work, recent advances have proposed more flexible solutions directly applicable to inventory settings. Akkerman et al. [143] introduce *Dynamic Neighborhood Construction*, which exploits the structure of discrete action spaces through adaptive neighborhood search guided by the critic's Q-values, scaling to problems with up to 10^{73} feasible actions. Vanvuchelen et al. [144] propose a *continuous action representation* approach, in which continuous network outputs are mapped to feasible discrete actions via a direct mapping function that does not require the feasible-action set to be specified beforehand. Both methods advance earlier work on action-space reduction by achieving higher computational efficiency—for instance, by avoiding the explicit storage or enumeration of all feasible actions. The literature would benefit from comparative studies of action-space reduction and representation approaches to help researchers assess their relative effectiveness.

In addition to the issue of action spaces, there is the substantial computational expense involved in applying DRL compared to other models and heuristics. Deploying these models in companies would require periodic retraining across many SKUs, which is likely unfeasible at present for most firms. Batsis and Samothrakis [130] develop a method in which an agent is trained offline using data pertaining to different supply chain configurations. The agent was then deployed to specific supply chain contexts and quickly adapted, achieving performance similar to if it had known the context beforehand. Approaches such as these offer promising directions for further research that could enhance the scalability of DRL in inventory control.

Unlike the literature in categories A1 (separate estimation and optimization) and A2 (static ML-integrated optimization), the RL literature has a limited focus on forecasting and the use of exogenous variables. Typically, studies assume a fixed demand distribution and train agents on simulated data, rather than integrating explicit forecasting models. A few recent exceptions are emerging. For example, Wang et al. [111] incorporate ARIMA and LSTM demand forecasts directly

Table 5
Articles that use *dynamic ML-integrated optimization (A3)* with number of items (B), number of periods (C), lead time assumptions (D), setup cost (E), shelf life assumptions (G), single/multi-echelon (H), and capacity constraints (I1).

Article	Year	B items	C periods	D supply process	E procurement structure	F back order	G shelf life	H echelon	I constraint
Anagun [83]	1997	B1	C2	D1	E2	F1	G1	H1	I1
Giannoccaro and Pontrandolfo [84]	2002	B1	C2	D3	E1	F1	G1	H2	I1
Rao et al. [85]	2003	B1	C1	D1	E1	F2	G1	H2	I1
Emerson and Piramuthu [86]	2004	B1	C2	D2	E3	F2	G1	H2	I1
Ravulapati et al. [87]	2004	B1	C1	D1	E1	F2	G1	H2	I1
Piramuthu [88]	2005	B1	C2	D3, D5	E1	F2	G1	H2	I1
Kwon et al. [78]	2008	B1	C2	D2	E1	F2	G1	H2	I2
Kim et al. [89]	2008	B1	C2	D2	E1	F1	G1	H2	I1
Jiang and Sheng [79]	2009	B1	C2	D2	E1	F2	G1	H2	I1
Kwak et al. [90]	2009	B1	C2	D1	E1	F1	G1	H2	I1
Kim et al. [80]	2010	B1	C2	D2	E1	F1	G1	H2	I1
Sui et al. [91]	2010	B1	C2	D1	E2	F2	G1	H2	I2
Katanyukul et al. [92]	2011	B1	C2	D1	E2	F2	G1	H1	I1
Katanyukul and Chong [93]	2014	B1	C2	D1, D2	E2	F1	G1	H1	I1
Kara and Dogan [94]	2018	B1	C2	D2	E2	F1	G2	H1	I1
Vanvuchelen et al. [95]	2020	B2	C2	D1	E2	F1	G1	H1	I1
Bharti et al. [96]	2020	B1	C2	D2	E1	F2	G1	H2	I2
Perez et al. [97]	2021	B1	C2	D4	E1	F1, F2	G2	H2	I1
Wang and Lin [98]	2021	B1	C2	D4	–	–	G1	H2	I1
Kiyaei and Kiaee [99]	2021	–	–	–	E2	–	–	–	I1
Fallahi et al. [100]	2022	B1	C2	D1	E1	F2	G1	H1	I2
Oroojlooyjadid et al. [58]	2022	B1	C2	D2	E1	F1	G1	H2	I1
Preil and Krapp [101]	2022	B1	C2	D3	E1	F1	G1	H2	I1
Gijsbrechts et al. [102]	2022	B1	C2	D3, D5	E1	F1, F2	G1	H1,H2	I1
Zhou et al. [103]	2022	B1	C2	D1	E1	F1	G1	H1	I1
De Moor et al. [104]	2022	B1	C2	D2	E2	F2	G2	H1	I1
Meisheri et al. [105]	2022	B1	C2	D4	E2	F1	G2	H1	I2
Gioia et al. [106]	2022	B3	C2	D2	E1	F2	G2	H1	I2
Shakya et al. [107]	2022	B1	C2	D2	E1	F2	G1	H1	I2
Agrawal and Jia [108]	2022	B1	C2	D5	E2	F2	G1	H1	I1
Cuartas and Aguilar [109]	2023	B1	C2	D2	–	–	G1	H1	I1
Demizu et al. [110]	2023	B1	C1	D1	E1	F2	G1	H1	I1
Wang et al. [111]	2023	B1	C2	D5	E2	F1	G2	H1	I2
van Hezewijk et al. [112]	2023	B1	C2	D1	E2	F1	G1	H1	I2
Mo et al. [113]	2023	B1	C2	D2	E1	F1	G1	H1	I1
Lu and Meyn [114]	2023	B1	C2	D1	E1	F1	G1	H1	I1
Cheung et al. [115]	2023	B1	C2	D1	E2	F2	G1	H1	I1
Zhou et al. [116]	2023	B1	C2	D1	E1	F1	G1	H1	I1
Li et al. [117]	2023	B1	C2	D1	E1	F2	G2	H2	I2
Mohamadi et al. [118]	2024	B1	C2	D1	E1	F2	G2	H2	I1
Dehaybe et al. [119]	2024	B1	C2	D2	E2	F1, F2	G1	H1	I1
Liu et al. [120]	2024	B1	C2	D1	E1	F2	G1	H2	I2
Kaynov et al. [121]	2024	B1	C2	D2	E1	F1, F2	G1	H2	I1
Stranieri et al. [122]	2024	B1	C2	D1	E2	F1	G1	H2	I1
Tian et al. [123]	2024	B1	C1	D1	E1	F1	G1	H1	I1
Lee et al. [124]	2024	B1	C2	D2	E1	F2	G2	H2	I2
Luo et al. [125]	2024	B1	C2	D1	E1	F2	G1	H1	I2
Yavuz and Kaya [126]	2024	B1	C2	D1	E1	F2	G2	H1	I1
Rizqi and Chou [127]	2024	B1	C2	D4	E3	F2	G1	H2	I2
Saha and Rathore [128]	2024	B2	C2	D1	E2	F2	G2	H2	I2
Stranieri et al. [129]	2024	B1	C2	D2	E1	F1	G1	H2	I2
Batsis and Samothrakis [130]	2024	B1	C2	D2	E2	F2	G1	H2	I2

into the state representation of their RL agent, enabling the agent to adapt to predicted trends in demand. Liu et al. [120] combine a maskable LSTM model with PPO. While promising, such approaches remain rare. Most reinforcement learning studies still treat demand as an exogenous stochastic process and focus more on policy learning than demand modeling. A more systematic integration of forecasting into RL architectures could significantly improve decision quality in real-world inventory systems.

5.3.2. Inventory system characteristics

Reinforcement learning (RL) has been applied across inventory settings with diverse — and often more complex — system dynamics than those typically seen in A1 (separate estimation–optimization) and A2 (static ML-integrated) work. Accordingly, below we review RL papers along all typology dimensions except C (number of periods). Because RL is built for sequential decision-making, nearly all studies are multi-period (with only rare single-period demonstrations). Hence, a discussion of this dimension is omitted.

Number of items (B). Deep RL has been applied to *joint replenishment* (B2) [95] and scaled to large multi-agent settings [128]. For *substitutable items* (B3), Gioia et al. [106] model stockout-driven demand switching. By contrast, complementary-item settings (B4) in spare parts and assemble-to-order systems are notably absent.[1] The

[1] Wang et al. [111] study spare-part replenishment path optimization, but their reward treats stocking points independently and does not model item dependencies; we therefore do not classify it as B4.

Table 6
Articles with supply process assumptions other than immediate replenishment (D1).

Category	Year	Reference
D2 Deterministic lead time ≥ 1	2004	Emerson and Piramuthu [86]
	2008	Kwon et al. [78]
	2008	Kim et al. [89]
	2009	Jiang and Sheng [79]
	2010	Kim et al. [80]
	2014	Katanyukul and Chong [93]
	2018	Kara and Dogan [94]
	2020	Bharti et al. [96]
	2022	Oroojlooyjadid et al. [58]
	2022	De Moor et al. [104]
	2022	Gioia et al. [106]
	2022	Shakya et al. [107]
	2023	Cuartas and Aguilar [109]
	2023	Mo et al. [113]
	2024	Dehaybe et al. [119]
	2024	Kaynov et al. [121]
	2024	Lee et al. [124]
	2024	Stranieri et al. [129]
	2024	Batsis and Samothrakis [130]
D3 stochastic lead time	2002	Giannoccaro and Pontrandolfo [84]
	2005	Piramuthu [88]
	2022	Preil and Krapp [101]
	2022	Gijsbrechts et al. [102]
D4 product specific lead times	2021	Perez et al. [97]
	2021	Wang and Lin [98]
	2022	Meisheri et al. [105]
	2024	Rizqi and Chou [127]
D5 Multiple suppliers	2022	Gijsbrechts et al. [102]
	2023	Wang et al. [111]

omission of work considering complementary items (B4) is likely due to number of inter-dependent ordering decisions that have to be made each period (e.g., five items with three order levels already yield 3^5 actions per period). Promising directions include action-space structuring (e.g., factorized or hierarchical policies) (see [112]) and other dimensionality-reduction techniques tailored to cross-item interactions.

Supply process (D). Most papers assume zero lead time and a single supplier (D1). In this setting, a replenishment decision is immediately followed by inventory delivery, so only next-period demand needs to be considered. This greatly simplifies the problem. Given the inherently multi-period nature of reinforcement learning (RL), it is somewhat surprising that so much of the literature is restricted to immediate replenishment. While the zero–lead-time assumption may be realistic in certain environments, it does not capture many practical situations. For this reason, a substantial body of work has considered positive deterministic lead times ($L \geq 1$, D2) as well.

By contrast, stochastic lead times are rarely modeled. Preil and Krapp [101], for instance, employ a bandit-based RL approach to solve a multi-echelon problem with random lead times between echelons. This demonstrates the potential of RL to capture the added uncertainty of non-deterministic settings. Given that supply chain resilience is increasingly emphasized [145], the lack of work on stochastic lead times (D3) represents an important research gap.

A smaller set of studies also examine heterogeneous lead times across products (D4). Meisheri et al. [105] consider different product-specific lead times, while Rizqi and Chou [127] extend this by analyzing a multi-echelon system with multiple delivery options and uncertain discounts. These examples highlight the flexibility of RL in handling complex, heterogeneous supply settings.

Finally, some work has addressed dual sourcing (D5), where multiple suppliers exist for the same item [102,111]. Similar to stochastic lead times, dual sourcing is directly linked to resilience considerations, as it provides firms with redundancy and flexibility.

All contributions that move beyond the D1 (immediate replenishment) baseline are summarized in Table 6. Although the literature has begun to explore richer lead time and sourcing settings, the majority of research still relies on deterministic lead times. More attention to stochastic lead times (D3) and dual sourcing (D5) is therefore encouraged, as these settings better reflect the challenges faced in resilient supply chain management.

Procurement structure (E). Approximately half of the papers consider fixed order costs, while the other half do not. We found almost no papers that consider discounts (E3). Rizqi and Chou [127] consider uncertain discounts in their multi-echelon inventory optimization problem. When a supplier offers a sudden discount, the agent can decide to acquire extra inventory, thereby reducing costs.

We do not find any papers which consider quantity discounts. Considering that these discounts are prevalent in practice, this could be a potentially interesting area of research—especially since reinforcement learning enables more complex inventory settings.

Backorders/lost sales (F). Lost-sales models are generally harder to analyze than backorder models. Bijvank et al. [32] survey the classical lost-sales literature and note that backorders are often assumed for analytical tractability. In B2B contexts, that assumption may be realistic because customers are willing to wait. In many consumer settings, however, stockouts translate into demand that is not recovered, i.e., lost sales. Given how common these dynamics are in practice, analyzing them remains important.

In our sample, 28 of the 51 RL papers assume lost sales (F2). Relative to traditional analytical methods, implementing lost sales in RL is straightforward — lost demand can be encoded via immediate penalties and no carry-over in the state — though the choice still affects learning targets and stability.

Despite this prevalence, relatively few studies isolate the classical lost-sales problem and test against well-established benchmarks. An exception is Gijsbrechts et al. [102], who show that an A3C agent produces strong policies yet retains an optimality gap of about 6.7% in a controlled setting (lead time 4, underage cost 4). When moving to settings where the optimal policy is unavailable (e.g., longer lead times), their RL approach outperforms the benchmark heuristics in some instances.

Given the practical importance of lost sales, it is encouraging that many RL studies model them. Part of this prevalence likely reflects modeling convenience: in RL, lost sales are easy to encode via the reward, which is a genuine advantage of the approach. Still, more work on the classical lost-sales benchmark — with transparent, like-for-like comparisons to established policies — is needed to assess whether recent RL advances systematically close the remaining optimality gaps.

Shelf life considerations (G). A substantial number of papers consider perishable products. Modeling perishable products is known to introduce substantial complexity, especially when multiple products are involved. Perishability models can be categorized into those that consider fixed lifetimes, stochastic lifetimes, and time-dependent lifetimes [146]. As shown in Table 7, many of these shelf life assumptions have been used.

Most papers assume fixed shelf lives. Some papers consider more complex deterioration schemes. Li et al. [117] apply Q-learning to find a joint markdown, freshness, and ordering policy. In their study, the freshness of products is influenced by the policy. RL demonstrates the capability to handle such complex deterioration schemes, presenting an opportunity for further research in this direction. Meisheri et al. [105] apply RL to a system with perishable products. Their problem includes additional complexities, such as different lead times for each product and transportation constraints, clearly demonstrating the effectiveness of RL in handling complex inventory settings.

Papers that consider obsolescence (G3) are not present in our sample. In obsolescence, it is not the product itself that deteriorates, but

Table 7
Papers that consider perishable products (G2) split up by the type of shelf life assumption.

Category	Year	Reference
all inventory is lost after 30 days	2021	Perez et al. [97]
	2022	Meisheri et al. [105]
fixed deterioration rate	2018	Kara and Dogan [94]
fixed lifetime	2022	De Moor et al. [104]
	2022	Gioia et al. [106]
	2023	Wang et al. [111]
	2024	Mohamadi et al. [118]
	2024	Lee et al. [124]
	2024	Yavuz and Kaya [126]
	2024	Saha and Rathore [128]
policy dependent stochastic lifetime	2023	Li et al. [117]

rather the demand. This occurs, for example, in spare parts systems where certain capital goods are no longer produced, leading to a decline in demand for the spare parts. This context is studied in inventory control literature, but without applying RL [147,148], pointing toward a gap in the literature.

Number of echelons (H). In multi-echelon systems, two main setups are common: convergent/serial and divergent [149]. Table 8 groups the RL papers by echelon type. Recent work clusters around divergent systems, while convergent/serial cases remain present but fewer in the last four years. The table also includes a network variant where inventory can be relocated across nodes [98], and a mixed serial/divergent case [127]. Convergent settings remain useful for benchmarking coordination with a clear flow structure (e.g., beer-game–type serial chains [58]).

A practical challenge in divergent systems is the growth of the action space as decisions are coupled across many downstream nodes. Two recent strategies stand out: Kaynov et al. [121] infer a multi-discrete action distribution with output nodes that scale linearly in the number of retailers, and Saha and Rathore [128] use multi-agent RL to handle large, real-world deployments.

In addition to the serial and divergent multi-echelon systems, we noticed another type of system more akin to a routing problem. Wang and Lin [98] consider a distribution network where spare parts are transported from one node to another to meet demand at various locations. The inventory does not flow in serial fashion to a single end node, nor does it diverge to multiple end nodes; it can move in any direction in the network. Their approach optimizes the replenishment path of inventory, trying to minimize replenishment times. RL in this context demonstrates a capability to reduce replenishment time by 40%. The efficacy of RL in this context clearly encourages other researchers to further explore this approach.

Overall, Table 8 shows a clear tilt toward divergent applications of RL, with convergent and network variants also represented. Given the relevance to the practice of spare part management, we encourage further exploration of network-type echelon settings, including shared benchmarks and reporting that make results comparable across studies.

Capacity constraints (I). Capacity constraints are common in practice—retailers face shelf-space limits; manufacturers face finite buffers and workstation capacities. Such limits complicate inventory models and are often ignored, yet doing so can be a poor proxy for optimal decisions [35]. In the RL literature, capacity appears in several forms, including production limits (e.g., capacitated lot sizing with PPO [112]), combined storage and ordering limits [125], and storage caps in single- or multi-agent settings (e.g., [96,107,127–129]).

From an implementation standpoint, authors typically encode capacity directly into the decision process using a few recurring patterns:

Table 8
Papers that consider multi-echelon settings.

Structure	Year	Reference
Convergent/ serial	2002	Giannoccaro and Pontrandolfo [84]
	2004	Emerson and Piramuthu [86]
	2005	Piramuthu [88]
	2009	Kwak et al. [90]
	2020	Bharti et al. [96]
	2021	Perez et al. [97]
	2022	Oroojlooyjadid et al. [58]
	2022	Preil and Krapp [101]
	2023	Li et al. [117]
	2024	Mohamadi et al. [118]
	2024	Batsis and Samothrakis [130]
Divergent	2003	Rao et al. [85]
	2004	Ravulapati et al. [87]
	2008	Kwon et al. [78]
	2008	Kim et al. [89]
	2009	Jiang and Sheng [79]
	2010	Kim et al. [80]
	2010	Sui et al. [91]
	2022	Gijsbrechts et al. [102]
	2024	Liu et al. [120]
	2024	Kaynov et al. [121]
	2024	Stranieri et al. [122]
	2024	Lee et al. [124]
	2024	Saha and Rathore [128]
	2024	Stranieri et al. [129]
Network	2021	Wang and Lin [98]
Serial, Divergent	2024	Rizqi and Chou [127]

Table 9
Papers that consider capacity constraints.

Constraint category	Year	Reference
Max customers per day	2022	Gioia et al. [106]
Ordering capacity	2010	Sui et al. [91]
	2024	Liu et al. [120]
	2024	Lee et al. [124]
Production capacity	2008	Kwon et al. [78]
	2023	van Hezewijk et al. [112]
	2024	Batsis and Samothrakis [130]
Storage and order capacity	2022	Meisheri et al. [105]
	2024	Luo et al. [125]
Storage capacity	2020	Bharti et al. [96]
	2022	Shakya et al. [107]
	2023	Wang et al. [111]
	2023	Li et al. [117]
	2024	Rizqi and Chou [127]
	2024	Saha and Rathore [128]
	2024	Stranieri et al. [129]
Storage capacity and budget	2022	Fallahi et al. [100]

- *Invalid-action masking:* remove infeasible choices before sampling so the policy only selects from feasible actions (e.g., feasibility masks in [112]; masking for transshipments in [124]).
- *Hard bounds via the action set:* define state-dependent caps so actions cannot exceed remaining capacity (e.g., capped order sets in [107], explicit upper limits in [124]).

Although many RL papers include capacity (see Table 9), most do not study *capacitated systems* as a topic in their own right. As a result, head-to-head comparisons with classical capacitated OR benchmarks remain limited, and work on canonical multi-echelon capacitated systems (e.g., assembly-type structures) is still sparse (see Table 9).

5.4. A4: Other applications

Some papers do not neatly fall into our proposed typology. These papers still have close ties with the inventory control literature but approach inventory problems from different angles, often without directly optimizing parameters such as order quantity or review period. The most prominent among these applications is that of ABC-style inventory classification, which we discuss first, followed by a set of miscellaneous applications. The papers that consider the topic of classification (ABC or related), are shown in Table 10.

Table 10
Machine learning approaches for multi-criteria inventory classification (sorted by year within topic).

Topic	Year	Authors
ABC Unsupervised inventory classification	2011	Xiao et al. [159]
	2017	Hu et al. [160]
	2018	Balugani et al. [168]
	2019	Zowid et al. [157]
	2020	Zhang et al. [158]
	2021	Rengasamy and Murugesan [169]
	2021	Wang and Gao [170]
ABC Supervised classification (expert labels)	2002	Partovi and Anandarajan [161]
	2011	Yu [162]
	2023	Khanorkar and Kane [171]
ABC Supervised classification (simulation labels)	2016	Kartal et al. [172]
	2016	López-Soto et al. [173]
	2017	Lolli et al. [174]
	2017	López-Soto et al. [175]
	2019	Lolli et al. [164]
	2019	Sundar and Punniyamoorthy [176]
non-ABC Supervised classification (simulation labels)	2019	Priore et al. [165]
	2022	Svoboda and Minner [166]
	2022	Badakhshan et al. [167]
Other MCIC approaches	2014	Lolli et al. [177]

Table 11
Other applications of ML in inventory control.

Topic	Authors
Third Party Logistics (3PL)	Ren et al. [178] Kmiecik [179]
Backorder prediction	Ntakolia et al. [180] Islam and Amin [181] Ahmed et al. [182] de Santis et al. [183]
Dynamic buying and selling of inventory depending on price	Namir et al. [184]
Integrated inventory and scheduling framework	Guo et al. [185]
4.5cmDigital Twin	Badakhshan et al. [167]
Pareto optimal frontiers	Bandaru et al. [186]
Reorder point prediction based on historic reorder points	Inprasit and Tanachutiwat [187]
Blood discard prediction	Singha and Panse [188]
Learning dominance relations	Yu and Wah [189]

Inventory classification

Inventory classification papers extend the ABC analysis that is widely used in practice. Standard ABC analysis divides SKUs into the categories A, B, and C based on historic demand and dollar usage [150]. The appeal of this approach lies in its simplicity: rather than assigning unique parameters to thousands of SKUs, firms can apply a single replenishment policy to each category.

Because of its prevalence, researchers have long sought to improve ABC classification. Early multi-criteria approaches such as AHP [151, 152] or weighted optimization models [153–155] allowed for the inclusion of additional factors like lead time or criticality. However, these approaches relied on subjective weights and did not always align with cost performance. Teunter et al. [156] showed that common criteria such as demand value or demand volume can perform poorly from a cost perspective and proposed more objective cost-oriented criteria.

This critique opened the door to machine learning approaches, which seek to replace subjective weighting by data-driven classification. A first strand uses unsupervised learning. Zowid et al. [157] apply Gaussian Mixture Models to cluster SKUs into ABC classes, while Zhang et al. [158] combine clustering with a backpropagation neural network to improve spare parts classification. Other work uses association-based clustering: Xiao et al. [159] exploit cross-selling relationships to derive lost-profit-based groupings. Hu et al. [160] introduce a dominance-based rough set approach that learns if–then decision rules from historical data.

A second strand adopts supervised learning. Early studies such as Partovi and Anandarajan [161] trained neural networks on expert-labeled ABC categories to automate classification, while Yu [162] compared AI classifiers (SVMs, BPNs, k-NN) against traditional multiple discriminant analysis on benchmark datasets. More recently, research has shifted away from human-provided labels toward simulation-derived ground truth. Lolli et al. [163] and Lolli et al. [164] propose frameworks in which optimal (R, S) parameters are first determined via simulation, then grouped into categories, and finally used to train classifiers such as decision trees, random forests, SVMs, and ANNs. This approach demonstrates that ML models can approximate cost-minimizing classifications efficiently, especially in settings with intermittent demand.

In addition to these approaches, there are some studies that use this simulation-classification framework, but do not restrict themselves to the ABC categories widely used in practice. [165] study a multi-echelon setting and train a classification model that selects the best policy (such as continuous review or periodic review) using a simulation framework and achieve an accuracy of 88% in classifying the cost minimizing policy. Svoboda and Minner [166] use a genetic algorithm to train cost-minimizing decision trees and find that their approach only increases the cost 1% over cost-optimal allocation. Badakhshan et al. [167] extend this approach by embedding it within a digital twin for joint inventory and cash management. These methods are related to reinforcement learning, since policies are derived from simulated environments, but differ in that the policy space is pre-structured (e.g., restricted to (R, S) policies) and the learning is supervised rather than sequential (see Table 10).

Together, these contributions illustrate an evolving line of research: from subjective weighting models, through cost-based critiques, to fully data-driven ML methods for inventory classification. The continued prevalence of ABC-classification in industry underscores the potential practical impact of these advances. However, implementing simulation-intensive procedures to derive optimal classes poses significant hurdles for practitioners, including the need for domain-specific simulation models, and substantial computational resources. This may impede real-world adoption.

A promising direction for further research would be to develop general-purpose or "zero-shot" models capable of accurately predicting ABC classes without requiring firm-specific training data. Such models could leverage transfer learning or meta-learning approaches and would be especially valuable given the diversity of inventory environments and the scarcity of openly available labeled datasets.

Moreover, concepts from this research stream could enrich other ML approaches, particularly A3 (dynamic ML-integrated optimization). Notably, reinforcement learning (RL) has yet to be applied to multi-item inventory classification. Most RL studies to date focus on single-item settings with constant demand, where the agent incrementally learns the demand process. An exciting avenue for future work would be to

Table 12
ML Techniques in inventory control papers by year.

Type	Technique	<2016	2016	2017	2018	2019	2020	2021	2022	2023	2024	Total
Supervised	Multilayer Perceptron	2	1	2	1	2	3	2	4	3	0	20
	Random Forest	1	0	0	0	0	2	4	2	3	1	13
	Decision Tree	2	0	0	1	1	1	1	3	2	2	13
	Linear Regression	0	0	0	0	1	0	2	3	0	2	8
	k-Nearest Neighbors	3	1	0	0	1	1	2	1	4	0	13
	Long Short-Term Memory	0	0	0	0	0	1	2	2	2	1	8
	Extreme Gradient Boosting	0	0	0	0	0	0	3	1	0	2	6
	Kernel Regression	0	0	0	1	0	0	0	1	2	0	4
	Quantile Regression	0	0	0	1	0	1	0	0	2	0	4
	Support Vector Regression	0	0	0	1	0	1	1	0	1	0	4
	Other	2	4	0	1	2	4	6	2	4	6	31
	Total	**10**	**6**	**2**	**6**	**7**	**14**	**23**	**19**	**23**	**14**	**124**
Reinforcement Learning	Deep Q-Learning	1	0	0	1	0	1	2	5	4	2	16
	PPO	0	0	0	0	0	1	1	1	2	6	11
	Non-DRL	10	0	0	0	0	0	0	0	0	0	10
	SAC	0	0	0	0	0	0	0	1	0	2	3
	SARSA	2	0	0	1	0	0	0	0	0	0	3
	A3C	0	0	0	0	0	0	0	1	0	1	2
	Other	1	0	0	0	0	0	0	3	6	4	14
	Total	**14**	**0**	**0**	**2**	**0**	**2**	**3**	**11**	**12**	**15**	**59**
Unsupervised Learning	Clustering	1	0	0	1	0	1	2	0	0	0	5

design RL agents that observe and dynamically recategorize SKUs based on evolving demand patterns—potentially incurring switching costs. This could extend to settings with multiple items, intermittent demand, or more complex inventory networks, offering both methodological challenges and opportunities.

Other methods

Beyond classification, several other themes emerge (Table 11). One identified theme is the prediction of product backorders. These papers build on a Kaggle dataset of historical demand and inventory records [190]. Supervised models are trained to predict the likelihood of backorders, with approaches ranging from random forests to gradient boosting [180–183]. While these methods do not optimize control policies, they can provide useful early-warning systems to anticipate shortages.

Another line of work focuses on third-party logistics. Ren et al. [178] propose a hybrid deep learning model that integrates LSTM layers to capture temporal demand patterns and CNN layers to capture spatial dependencies, improving capacity allocation for a cross-border e-commerce logistics provider. Kmiecik [179], in turn, evaluates the use of forecasting tools from the perspective of a 3PL company, highlighting challenges of implementation.

Finally, we observe several stand-alone applications: dynamic inventory policies that react to price changes [184], integrated scheduling and inventory management frameworks [185], and dominance-learning approaches [189]. Together, these highlight the breadth of ML applications beyond direct inventory control.

5.5. ML techniques used

Table 12 shows the techniques used in the literature included in the review. There is a wide variety of ML techniques and hence it was chosen only to display the most important techniques grouped by three overarching categories: supervised, unsupervised, and RL.

In our exploration of neural networks, we observed a common occurrence where papers employ distinct terminology to describe identical techniques. It is apparent that numerous neural network models can be regarded as iterations or variations of overarching concepts. Therefore, we distinguish between three categories of neural networks: Multi-layer perceptrons (MLPs), Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs).[2] We split the RNN category into three sub-categories: Long Short-Term Memory (LSTM), Gated Recurrent Units (GRUs) and Other RNNs. Note that the LSTM is a special version of the Recurrent Neural Network (RNN) [191]. The 'Other RNNs' category contains papers that use different configurations of RNNs.

Within the category of reinforcement learning, a range of algorithmic approaches has been applied to inventory control problems, each with different implications for scalability, convergence, and sample efficiency. To provide a clearer overview of this landscape, Table 13 summarizes the specific RL algorithms used in the literature, grouped by time period and algorithm type. This breakdown illustrates the evolution from early non-deep reinforcement learning (Non-DRL) methods to more recent actor–critic and tailored approaches, reflecting the growing algorithmic sophistication in the field.

Q-learning remains the most commonly applied RL method across the literature. It is a value-based approach in which a neural network approximates the action-value function. The advantage of this method is that it allows for off-policy learning: an agent can select actions that are not constrained by the current best policy, as determined by the maximum Q-values. This makes Q-learning sample efficient. The downside of these algorithms is that they do not converge well and are unstable. According to Boute et al. [8] value-based methods are most appropriate when sample size matters.

In addition to Q-learning, Proximal Policy Optimization (PPO) is a popular method in the reinforcement learning literature. PPO is an actor–critic method, which combines value-based and policy-based approaches. In PPO, an agent learns a policy but is evaluated by the critic, which estimates the Q-values. One advantage is that it supports continuous action spaces, which is beneficial in inventory control problems, where ordering decisions — while discrete — often span a wide range of values. When we look at the trend, we see that PPO is growing in popularity compared to DQN.

Few papers compare different reinforcement learning algorithms. Meisheri et al. [105] apply both PPO and Q-learning and find that Q-learning outperforms once the sample size is large enough. Similar comparative studies could guide researchers in identifying the most promising algorithms for inventory control.

Recent work has begun to tailor reinforcement learning algorithms to the specific challenges of inventory control, moving beyond standard

[2] Extreme Learning Machine (ELM) is grouped in with MLPs. Radial Basis Function Neural Networks are grouped in 'other'.

Table 13
Reinforcement Learning algorithms Used in inventory control studies by time period.

RL category	<2014	2015–2020	2021–2022	2023–2024
Non-DRL	Giannoccaro and Pontrandolfo [84], Rao et al. [85], Emerson and Piramuthu [86], Ravulapati et al. [87], Piramuthu [88], Kim et al. [89], Kwon et al. [78], Jiang and Sheng [79], Kwak et al. [90]	Kim et al. [80]	–	—
Q-Learning	–	Sui et al. [91]	Kara and Dogan [94], Bharti et al. [96], Kiyaei and Kiaee [99], Wang and Lin [98]	De Moor et al. [104], Fallahi et al. [100], Oroojlooyjadid et al. [58], Shakya et al. [107], Zhou et al. [103], Cuartas and Aguilar [109], Li et al. [117], Lu and Meyn [114], Mo et al. [113], Saha and Rathore [128], Yavuz and Kaya [126]
SARSA	–	Katanyukul et al. [92], Katanyukul and Chong [93]	Kara and Dogan [94]	—
PPO	–	–	Vanvuchelen et al. [95], Perez et al. [97]	Meisheri et al. [105], Zhou et al. [116], van Hezewijk et al. [112], Batsis and Samothrakis [130], Dehaybe et al. [119], Kaynov et al. [121], Liu et al. [120], Stranieri et al. [129], Tian et al. [123]
SAC	–	–	–	Gioia et al. [106], Lee et al. [124], Yavuz and Kaya [126]
A3C	–	–	–	Gijsbrechts et al. [102], Tian et al. [123]
Other	–	Katanyukul and Chong [93] (Ruminative)	Preil and Krapp [101] (MAB)	Mohamadi et al. [118] (A2C), Demizu et al. [110] (BNN, TRPO, MML), Cheung et al. [115] (BORL, SWUCRL2-CW), Stranieri et al. [122] (DRLBD), Zhou et al. [103] (Double Q-learning, TN-DDQN), Luo et al. [125] (MARS), Rizqi and Chou [127] (NERL), Wang et al. [111] (RL4LS)

approaches like vanilla Q-learning or PPO. Several studies introduce algorithmic innovations aimed at improving performance under uncertainty, variance, and real-world constraints. For example, Zhou et al. [103] adapt Double Q-learning with a target network to reduce overestimation bias in joint pricing and inventory decisions. Cheung et al. [115] propose dynamic exploration methods like Bandit-over-Reinforcement Learning (BORL) and Sliding Window Upper-Confidence bound for Reinforcement Learning with Confidence Widening (SWUCRL2-CW) to better handle time-varying non-stationary environments. Rizqi and Chou [127] develop a neuroevolutionary RL method for sourcing decisions in multi-echelon systems, while Luo et al. [125] introduce a model-adaptive actor–critic algorithm (MARS) with provable convergence guarantees, directly addressing earlier-mentioned convergence issues.

Beyond these bespoke methods, researchers are increasingly exploring alternative actor–critic frameworks such as A3C [102,123] and A2C [118], which offer scalability and parallelization benefits compared to PPO. Others are combining multiple RL components within hybrid frameworks—for instance, Demizu et al. [110] integrate TRPO with Bayesian neural networks and meta-learning, while Yavuz and Kaya [126] fuse Q-learning and Soft Actor–Critic (SAC) to manage pricing and perishability jointly. Soft Actor–Critic (SAC), while not yet widely adopted in inventory control, has begun to appear in recent studies [106,124,126], likely due to its robustness in high-dimensional settings and its entropy-regularized exploration strategy. Its off-policy nature and support for continuous action spaces make it a theoretically attractive candidate for scaling RL to real-world inventory systems.

In line with this trend, recent work outside the scope of our review has introduced Deep Controlled Learning (DCL) [192]. DCL addresses variance in simulation-heavy environments — like inventory control — by comparing actions across shared exogenous demand trajectories and allocating simulation effort using multi-armed bandit principles. While not included in our dataset due to its recency, it illustrates the broader potential for domain-specific RL innovations in operations research.

5.6. *Fraction of applied work*

Fig. 7 shows the fraction of studies that use real-world data to verify model performance. Even though the field of inventory control benefits greatly from the development of new mathematical models and procedures, the end goal is to improve inventory system performance in real-world settings. The fraction of papers that actually use real-world data can be used as a proxy for how close the field is to the practice of inventory control.

Within category A3, we see that simulation-only studies are still dominant. The lack of real-world studies using RL is also noted by other authors [8,11] and is considered a significant gap in the research. RL studies such as Kara and Dogan [94], De Moor et al. [104], and van Hezewijk et al. [112] serve the purpose of showing in a controlled setting that these models tend to outperform their statistical or heuristic counterparts. While simulation-only studies are still most prevalent, there is an upward trend in the fraction of work that utilizes real-world data [99,105,110,111,193].

The articles reviewed contained only one actual implementation study. Qi et al. [65] study an end-to-end (E2E) inventory control model

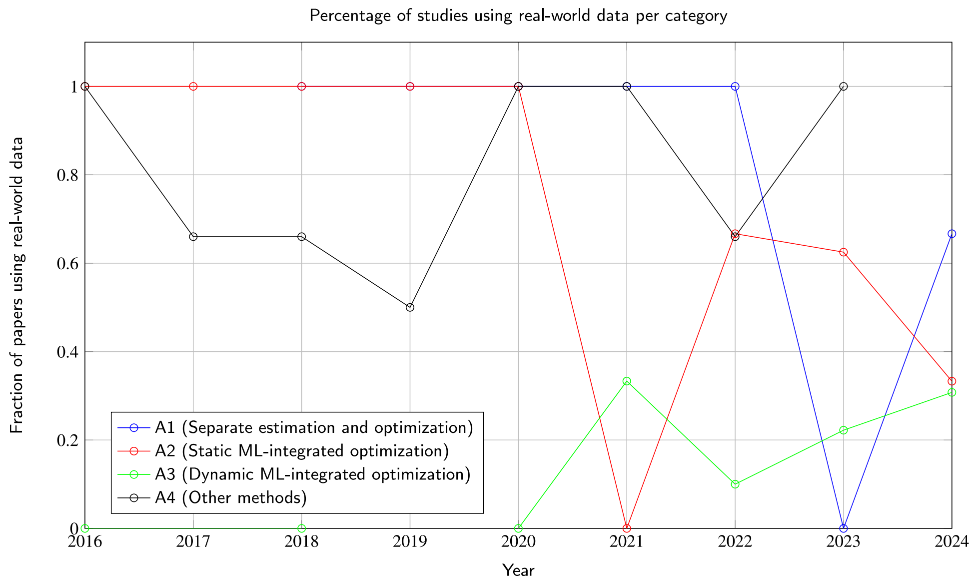

Fig. 7. Line graph showing the fraction of studies using real-world data for the different ML Integration categories A1, A2, A3 and A4 over the years. Lines are connected only when there are publications in adjacent years.

and implement the model for a Chinese e-commerce firm (JD.com). Their model achieves substantial cost savings. Although implementation studies are difficult to conduct due to the need for substantial cooperation from companies, they represent the most reliable method of verifying real-world performance.

6. Conclusions (integration, applications, and outlook)

This review provided a comprehensive overview of machine learning applications in inventory control, analyzing 122 articles based on a framework that classifies studies by the way ML is integrated into the inventory optimization framework (RQ1) and the specific inventory problem characteristics (RQ2) addressed. In this final section, we summarize our findings and discuss the research gaps (RQ3), with the latter also being included in Table 15. A central finding was the clear "division of labor" among ML approaches: simpler, static methods are applied to problems with basic characteristics, while more complex, dynamic methods like Reinforcement Learning (RL) are tackling a new frontier of challenging inventory problems. This cross-analysis provided insights into where future research is most needed.

6.1. *(RQ1) How is machine learning integrated into the inventory optimization framework?*

We identified three primary ways in which ML has been integrated into inventory optimization, grouped into categories A1 (separate estimation and optimization), A2 (static ML-integrated optimization), and A3 (dynamic ML-integrated optimization). In addition, a fourth category (A4) was used to capture contributions that did not fit neatly into the main three. In recent years we see a clear shift away from the predict-then-optimize paradigm toward ML-integrated approaches especially reinforcement learning (A3).

A1 (separate estimation and optimization). Forecasts (points, quantiles, scenarios) are produced by ML and then plugged into established optimization models. The strength is interoperability: A1 can be used with linear, stochastic, or robust formulations without changing policy structure or constraints, which helps in domains with governance requirements and complex constraint sets. The limitation is lack of co-optimization: when lead times, capacities, or network couplings make current decisions shape future states, forecast errors can propagate directly into costs.

Table 14
Mini crosswalks by ML approach (A1–A3). Columns are inventory dimensions (B–I); rows 1–5 denote the subcategories within each dimension. Legend: — = absent (0 studies), ∘= rare (1–2 studies), ⊚= emerging (3–5 studies), •= common (>5 studies).

A1 Separate estimation & optimization

	B	C	D	E	F	G	H	I
1	•	⊚	•	•	•	•	•	•
2	∘	•	∘	•	•	⊚	⊚	–
3	–		–	–		–		
4	–		–					
5			∘					

A2 Static ML-integrated

	B	C	D	E	F	G	H	I
1	•	•	•	•	⊚	•	•	•
2	–	•	∘	⊚	•	∘	–	∘
3	∘		–	–		–		
4	–		∘					
5			∘					

A3 Dynamic (RL)

	B	C	D	E	F	G	H	I
1	•	⊚	•	•	•	•	•	•
2	∘	•	•	•	•	•	•	•
3	∘		⊚	∘		–		
4	–		⊚					
5			⊚					

Row keys (by column): Items (B): (1) single item, (2) joint replenishment, (3) substitutable, (4) complementary; Periods (C): (1) single period, (2) multi-period; Supply process (D): (1) LT = 0, (2) deterministic LT≥1, (3) stochastic LT, (4) product-specific LT, (5) multi-supplier; Procurement (E): (1) no fixed cost, (2) fixed/setup, (3) discounts; Backorders/lost sales (F): (1) backorders, (2) lost sales; Perishability (G): (1) none, (2) perishables, (3) obsolescence; Echelons (H): (1) single echelon, (2) multi-echelon; Capacity (I): (1) unconstrained, (2) constrained.

A2 (static ML-integrated optimization). Here the inventory cost is embedded in a supervised loss and minimized on historical samples. This directly targets operational objectives and typically outperforms pure forecast-then-optimize in single-period settings. Most work to date adopts newsvendor-type losses. Capacity appears in simple forms (e.g., storage capacity limits). Other variants — which shape the supervised loss to approximate downstream costs — are a promising

Table 15
Key gaps and near-term opportunities by ML–integration approach.

Approach	Key gaps and near-term opportunities
A1: Separate estimation and optimization	• Go beyond point forecasts: use distributional forecasting (e.g., variance/quantiles) • Test in complexer settings: stochastic/positive lead times, capacity limits, multi-echelon networks, perishables, joint replenishment/substitution. • Calibrate forecasts for decisions: adopt *forecasting congruence* ideas to reduce order volatility
A2: Static ML-integrated optimization	• Test models across multiple service levels to assess outperformance over A1 approaches • Loss designs that approximate downstream (multi-period) costs are promising—validate them across regimes (service levels, lead times, perishability, capacity) and compare to RL on the performance–compute trade-off.
A3: Dynamic ML-integrated optimization (RL)	• Scale and stability: large action spaces, convergence, and training cost remain bottlenecks—evaluate reduction/structuring strategies head-to-head. • Data integration: few papers fuse forecasts/auxiliary features into the state; develop “data-driven RL”. • External validity: more real-world datasets, standardized environments, and like-for-like comparisons across algorithms/policies. • Explore multi-item settings (B2–B4), stochastic deterioration rates (G2), product obsolescence (G3), capacitated multi-echelon systems (I2, H2) • Treat canonical systems as subjects in their own right — lost-sales (F2) and capacitated (I2) settings — and run head-to-head comparisons against state-of-the-art benchmarks to test whether RL can close known optimality gaps.
A4: Other methods (inventory classification)	• General-purpose “zero-shot” ABC classifiers via transfer-learning • RL for multi-item, time-varying ABC assignments (with switching costs and service-level constraints)

bridge toward multi-period effects while retaining supervised-learning tractability.

A3 (dynamic ML-integrated optimization). RL optimizes by interacting with a simulated environment, making it natural when multi-echelon structures, lead-time uncertainty, or perishability couple decisions across time and space. Recent papers adapt architectures to inventory specifics (e.g., handling large action spaces), but scaling and systematic, like-for-like benchmarking remain open.

A4 (Other methods). Within this class we found multiple applications of ML in inventory management that are separate from A1-A3. The most prominent stream was Multi-criteria/ABC inventory classification. This application of ML is promising because of the widespread usage of these classification schemes in practice. In addition we found other applications such as, predicting risks (e.g., stockouts/spoilage), Third Party Logistics (3PL) among others.

6.2. (RQ2) What types of inventory system characteristics have been considered?

We categorized the literature along eight dimensions: number of items (B), number of periods (C), supply process/lead time (D), procurement/setup (E), backorders vs. lost sales (F), perishability/obsolescence (G), single vs. multi-echelon (H), and capacity constraints (I). For each dimension, we coded the model assumptions used in each paper. Table 14 crosswalks these inventory-system dimensions with the ML integration approaches (A1–A3), summarizing our research and highlighting coverage and remaining gaps.

RL (A3) was used most where actions change future states in meaningful ways: multi-echelon networks (H2), perishables with age dynamics (G2), positive or stochastic lead times (D2/D3), and dual/multi-sourcing (D5). Lost-sales assumptions (F2) and capacity limits (I2) also appear frequently. A persistent gap is complementary multi-item systems (B4) (e.g., assemble-to-order, spare parts), where the joint action space grows quickly and remains a practical barrier.

By contrast, A1 (separate estimation–optimization) and A2 (static ML-integrated) were used mainly with single-item, single-period, immediate-replenishment settings (B1, C1, D1). A1’s strength is interoperability: ML forecasts (points/quantiles/scenarios) plug into existing linear, stochastic, or robust models, so constraints (e.g., storage/budget, I2) are handled without changing policy structure. A2 embeds cost asymmetry directly in a supervised loss and performs well in data-rich newsvendor-type problems; capacity typically enters in simple forms (e.g., shelf-space limits). Early loss designs that approximate downstream (multi-period) costs are beginning to push A2 beyond single-period settings, but evidence here is still emerging.

Finally, although recent A2 work hints at handling richer dynamics (e.g., D2–D5 lead times/sourcing, H2 multi-echelon, G2 perishables), we found that actual applications in these settings remain sparse, and when they do appear they are rarely benchmarked head-to-head against well-known heuristics for the same problem class. Table 14 gives an overview of the neglected inventory system characteristics for each individual ML integration approach.

6.3. (RQ3) What are the key directions for further research?

Our review identified significant research gaps both in methodology (RQ1) and at the intersection of inventory system dynamics and methodological approaches (RQ2). Table 15 provides a comprehensive overview of these gaps.

Although integrating ML into optimization directly theoretically offers performance improvement over A1 (separate estimation and optimization). We see that A1 approaches may offer an advantage when applied in practice because they easily integrate with current inventory optimization procedures. In order to increase performance from an inventory cost perspective, further work is needed to incorporate forecast congruence — the stability of forecast traces across time — into model selection. Additionally, exploring distributional forecasting remains an open opportunity. A2 (static ML-integrated optimization) studies would benefit from evaluation across diverse operating regimes to assess whether preserving empirical distributions during optimization leads to tangible improvements. In A3 (dynamic estimation and optimization), there is a notable lack of applied studies using real-world data and actual demand time series.

Algorithmically, promising developments within A3 address the unique requirements of operations research (OR) problems. However, the absence of comparative studies among new algorithms, and the scalability of RL approaches to real-world contexts, remain open challenges. The emerging use of pre-trained agents is a step forward, but more research is warranted.

Within inventory classification (A4) we see some opportunities: develop general-purpose (“zero-shot”) ABC models that do not need firm-specific training, using transfer-learning to adapt across contexts. Insights here can also inform A3: design RL agents that dynamically reclassify multiple SKUs over time under intermittent demand and networked inventories.

Regarding inventory system dynamics, gaps remain across all methodologies. Multi-item systems, commonplace in production and spare parts management, are underrepresented in the literature. Given RL’s strengths in learning within large state spaces, its application to multi-item inventory systems is a promising area. This hinges upon the development of models that are able to deal with large action spaces.

The prevailing assumption of immediate replenishment (D1) limits practical relevance. Future research should extend to stochastic lead times (D3). Other areas needing attention include quantity discounts (E3), stochastic deterioration (G2), product obsolescence (G3), and complex network-like supply chain networks (H2).

Other observed trends (outside RQ1–RQ3)

In analyzing the articles, we noted several cross-cutting observations outside the scope of our typology:

- **Reproducibility.** Many papers do not share code. For ML applications, releasing code is natural and would greatly improve replication and extension of results.
- **Inventory–pricing interface.** Several RL studies couple inventory and pricing decisions [103,116,120,126]. Our review focuses on inventory control, but given the complexity of the interface with revenue management, RL appears particularly promising here.

CRediT authorship contribution statement

Ritsaart Bergsma: Writing – original draft. **Corné de Ruijt:** Writing – review & editing. **Sandjai Bhulai:** Writing – review & editing.

Funding sources

This work was supported by Regieorgaan Praktijkgericht Onderzoek (SIA), The Netherlands.

Declaration of competing interest

No potential conflict of interest was reported by the authors.

Data availability

Data will be made available on request.

References

[1] McKinsey B. The state of AI in 2022—and a half decade in review. 2022, http://ceros.mckinsey.com/mckinsey-commentary-ai-hall-desktop-1.
[2] Hendriksen C. Artificial intelligence for supply chain management: Disruptive innovation or innovative disruption? J. Supply Chain Manag. 2023;59(3):65–76.
[3] RaMResearch and Markets. Same Day Delivery Market Size, Share & Forecast to 2032. 2024, https://www.researchandmarkets.com/reports/5642299/same-day-delivery-market-report-by-service-and.
[4] ReFED. Food waste data—causes & impacts. 2025, https://refed.org/food-waste/the-problem/.
[5] Makridakis S, Spiliotis E, Assimakopoulos V. M5 accuracy competition: Results, findings, and conclusions. Int J Forecast 2022;38(4):1346–64.
[6] Bertsimas D, Kallus N. From predictive to prescriptive analytics. 2014.
[7] Ban G-Y, Rudin C. The big data newsvendor: practical insights from machine learning. Rochester, NY; 2018.
[8] Boute RN, Gijsbrechts J, van Jaarsveld W, Vanvuchelen N. Deep reinforcement learning for inventory control: a roadmap. Rochester, NY; 2021.
[9] Gutierrez JC, Polo Triana SI, León Becerra JS. Benefits, challenges, and limitations of inventory control using machine learning algorithms: literature review. OPSEARCH; 2024.
[10] Albayrak Ünal Ö, Erkayman B, Usanmaz B. Applications of artificial intelligence in inventory management: A systematic review of the literature. Arch Comput Methods Eng 2023;30(4):2605–25.
[11] Rolf B, Jackson I, Müller M, Lang S, Reggelin T, Ivanov D. A review on reinforcement learning algorithms and applications in supply chain management. Int J Prod Res 2023;61(20):7151–79.
[12] de Castro Moraes T, Yuan X-M. Data-driven solutions for the newsvendor problem: A systematic literature review, in advances in production management systems. In: Dolgui A, Bernard A, Lemoine D, von Cieminski G, Romero D, editors. Artificial intelligence for sustainable and resilient production systems. Cham: Springer International Publishing; 2021, p. 149–58.
[13] Snyder H. Literature review as a research methodology: An overview and guidelines. J Bus Res 2019;104:333–9.
[14] Tranfield D, Denyer D, Smart P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review, 14 (3). 2003, p. 207–22, [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8551.00375.
[15] eigenfactororg. Eigenfactor: About. 2023, http://eigenfactor.org/about.php.
[16] STM. STM global brief 2021–economics & market size. Tech. Rep., 2021.
[17] Goltsos TE, Syntetos AA, Glock CH, Ioannou G. Inventory–forecasting: Mind the gap. European J Oper Res 2022;299(2):397–419.
[18] Oroojlooyjadid A, Snyder LV, Takáč M. Applying deep learning to the newsvendor problem. IISE Trans 2020;52(4):444–63.
[19] Huber J, Müller S, Fleischmann M, Stuckenschmidt H. A data-driven newsvendor problem: From data to decision. European J Oper Res 2019;278(3):904–15.
[20] Erlenkotter D. Note—An early classic misplaced: Ford W. Harris’s economic order quantity model of 1915. Manag Sci 1989;35(7):898–900.
[21] Prasad S. Classification of inventory models and systems. Int J Prod Econ 1994;34(2):209–22.
[22] de Kok T, Grob C, Laumanns M, Minner S, Rambau J, Schade K. A typology and literature review on stochastic multi-echelon inventory models. European J Oper Res 2018;269(3):955–83.
[23] Silver EA. Operations research in inventory management: A review and critique. Oper Res 1981;29(4):628–45.
[24] Song J-SJ. Research handbook on inventory management. Edward Elgar Publishing; 2023.
[25] Roundy R. Rounding off to powers of two in continuous relaxations of capacitated lot sizing problems. Manag Sci 1989;35(12):1433–42.
[26] Song J-SJ, Song ZX, Shen X. Demand management and inventory control for substitutable products. 2021, Available at SSRN 3866775.
[27] DeValve L, Jeannette Song J-S, Wei Y. Assemble-to-order systems. In: Song J-SJ, editor. Research handbook on inventory management. Edward Elgar Publishing; 2023, p. 191–212.
[28] Basten R, van Houtum G-J. Chapter 19: Spare parts inventory planning. In: Research handbook on inventory management. Edward Elgar Publishing; p. 2023, ch. Research Handbook on Inventory Management.
[29] Arrow KJ, Harris T, Marschak J. Optimal inventory policy. Econ: J Econ Soc 1951;250–72.
[30] Song J-S. The effect of leadtime uncertainty in a simple stochastic inventory model. Manag Sci 1994;40(5):603–13.
[31] Xin L, Mieghem JAV. Dual-sourcing, dual-mode dynamic stochastic inventory models. In: Research handbook on inventory management. Edward Elgar Publishing; 2023, p. 165–90, ch. Research Handbook on Inventory Management.
[32] Bijvank M, Huh WT, Janakiraman G. Lost-sales inventory systems. In: Research handbook on inventory management. Edward Elgar Publishing; 2023, p. 2–26, ch. Research Handbook on Inventory Management.
[33] Axsäter S. Inventory control, ser. international series in operations research & management science. Cham: Springer International Publishing; 2015, p. 225.
[34] Clark AJ, Scarf H. Optimal policies for a multi-echelon inventory problem. Manag Sci 1960;6(4):475–90.
[35] Kapuściński R, Parker RP. Capacitated inventory systems. In: Research handbook on inventory management. Edward Elgar Publishing; 2023, p. 46–72, ch. Research Handbook on Inventory Management.

[36] López Lázaro J, Barbero Jiménez Á, Takeda A. Improving cash logistics in bank branches by coupling machine learning and robust optimization. Expert Syst Appl 2018;92:236–55.
[37] Cao Y, Shen Z-JM. Quantile forecasting and data-driven inventory management under nonstationary demand. Oper Res Lett 2019;47(6):465–72.
[38] Abolghasemi M, Beh E, Tarr G, Gerlach R. Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion. Comput Ind Eng 2020;142:106380.
[39] Li N, Chiang F, Down DG, Heddle NM. A decision integration strategy for short-term demand forecasting and ordering for red blood cell components. Oper Res Health Care 2021;29:100290.
[40] Pereira MM, Frazzon EM. A data-driven approach to adaptive synchronization of demand and supply in omni-channel retail supply chains. Int J Inf Manage 2021;57:102165.
[41] Deng C, Liu Y. A deep learning-based inventory management and demand prediction optimization method for anomaly detection. Wirel Commun Mob Comput 2021;2021:1–14.
[42] Gonçalves JN, Cortez P, Carvalho MS, Frazão NM. A multivariate approach for multi-step demand forecasting in assembly industries: Empirical evidence from an automotive supply chain. Decis Support Syst 2021;142:113452.
[43] Galli L, Levato T, Schoen F, Tigli L. Prescriptive analytics for inventory management in health care. J Oper Res Soc 2021;72(10):2211–24.
[44] Ulrich M, Jahnke H, Langrock R, Pesch R, Senge R. Distributional regression for demand forecasting in e-grocery. European J Oper Res 2021;294(3):831–42.
[45] Shokouhifar M, Ranjbarimesan M. Multivariate time-series blood donation/demand forecasting for resilient supply chain management during COVID-19 pandemic. Clean Logist Supply Chain 2022;5:100078.
[46] Shi J. Application of the model combining demand forecasting and inventory decision in feature based newsvendor problem. Comput Ind Eng 2022;173:108709.
[47] Li N, Arnold DM, Down DG, Barty R, Blake J, Chiang F, Courtney T, Waito M, Trifunov R, Heddle NM. From demand forecasting to inventory ordering decisions for red blood cells through integrating machine learning. Stat Model Inven Optim Transfus 2022;62(1):87–99.
[48] Singh R, Mishra VK. Inventory model using Machine Learning for demand forecast with imperfect deteriorating products and partial backlogging under carbon emissions. Ann Oper Res 2023.
[49] Fan L, Song Z, Mao W, Luo T, Wang W, Yang K, Cao F. Change is safer: A dynamic safety stock model for inventory management of large manufacturing enterprise based on intermittent time series forecasting. J Intell Manuf 2024.
[50] Mete Ayhan H, Kır S. Ml-driven approaches to enhance inventory planning: Inoculant weight application in casting processes. Comput Ind Eng 2024;193:110280.
[51] Singh R, Mishra VK. Machine learning based fuzzy inventory model for imperfect deteriorating products with demand forecast and partial backlogging under green investment technology. J Oper Res Soc 2024;75(7):1223–38.
[52] Abolghasemi M, Hurley J, Eshragh A, Fahimnia B. Demand forecasting in the presence of systematic events: Cases in capturing sales promotions. Int J Prod Econ 2020;230:107892.
[53] Pritularga K, Kourentzes N. Forecast congruence: a quantity to align forecasts and inventory decisions. Rochester, NY; 2024.
[54] Pagnoncelli BK, Ahmed S, Shapiro A. Sample average approximation method for chance constrained programming: Theory and applications. J Optim Theory Appl 2009;142(2):399–416.
[55] Cheung WC, Simchi-Levi D. Statistical learning in inventory management. In: Research handbook on inventory management. Edward Elgar Publishing; 2023, p. 307–32, ch. Research Handbook on Inventory Management.
[56] O'Neil S, Zhao X, Sun D, Wei JC. Newsvendor problems with demand shocks and unknown demand distributions. Decis Sci 2016;47(1):125–56.
[57] Zhang Y, Gao J. Assessing the performance of deep learning algorithms for newsvendor problem. In: Liu D, Xie S, Li Y, Zhao D, El-Alfy E-S M, editors. Neural information processing. Cham: Springer International Publishing; 2017, p. 912–21.
[58] Oroojlooyjadid A, Nazari M, Snyder LV, Takáč M. A deep Q-network for the beer game: Deep reinforcement learning for inventory optimization. Manuf Serv Oper Manag 2022;24(1):285–304.
[59] Punia S, Singh SP, Madaan JK. From predictive to prescriptive analytics: A data-driven multi-item newsvendor model. Decis Support Syst 2020;136:113340.
[60] Abbasi B, Babaei T, Hosseinifard Z, Smith-Miles K, Dehghani M. Predicting solutions of large-scale optimization problems via machine learning: A case study in blood supply chain management. Comput Oper Res 2020;119:104941.
[61] Bertsimas D, Koduri N. Data-driven optimization: A reproducing kernel Hilbert space approach. Oper Res 2022;70(1):454–71.
[62] Chen B. Data-driven inventory control with shifting demand. Prod Oper Manage 2021;30(5):1365–85.
[63] Clausen JBB, Li H. Big data driven order-up-to level model: Application of machine learning. Comput Oper Res 2022;139:105641.
[64] Pirayesh Neghab D, Khayyati S, Karaesmen F. An integrated data-driven method using deep learning for a newsvendor problem with unobservable features. European J Oper Res 2022;302(2):482–96.
[65] Qi M, Shi Y, Qi Y, Ma C, Yuan R, Wu D, Shen Z-JM. A practical end-to-end inventory management model with deep learning. Manag Sci 2023;69(2):759–73.
[66] Tian Y-X, Zhang C. An end-to-end deep learning model for solving data-driven newsvendor problem with accessibility to textual review data. Int J Prod Econ 2023;265:109016.
[67] Ren X, Gong Y, Rekik Y, Xu X. Anticipatory shipping versus emergency shipment: Data-driven optimal inventory models for online retailers. Int J Prod Res 2023;1–18.
[68] Bertsimas D, McCord C, Sturt B. Dynamic optimization with side information. European J Oper Res 2023;304(2):634–51.
[69] Forel A, Parmentier A, Vidal T. Explainable data-driven optimization: From context to decision and back again. In: Proceedings of the 40th international conference on machine learning. ICML'23, vol. 202, Honolulu, Hawaii, USA: JMLR.org; 2023, p. 10 170–87.
[70] Zhang Q, Tan Y. Data-driven E-commerce end-to-end inventory optimization algorithm. In: Tallón-Ballesteros AJ, Beltrán-Barba R, editors. Frontiers in artificial intelligence and applications. IOS Press; 2023.
[71] Kallus N, Mao X. Stochastic optimization forests. Manag Sci 2023;69(4):1975–94.
[72] Chen Z-Y, Fan Z-P, Sun M. Machine learning methods for data-driven demand estimation and assortment planning considering cross-selling and substitutions. INFORMS J Comput 2023;35(1):158–77.
[73] Bertsimas D, Kim CW. A machine learning approach to two-stage adaptive robust optimization. European J Oper Res 2024;319(1):16–30.
[74] Qi M, Shen Z-J, Zheng Z. Learning newsvendor problems with intertemporal dependence and moderate non-stationarities. Prod Oper Manage 2024;33(5):1196–213.
[75] van der Haar JF, Wellens AP, Boute RN, Basten RJ. Supervised learning for integrated forecasting and inventory control. European J Oper Res 2024;319(2):573–86.
[76] Atan Z, Ahmadi T, Stegehuis C, de Kok T, Adan I. Assemble-to-order systems: A review. European J Oper Res 2017;261(3):866–79.
[77] Li Y. Deep reinforcement learning: An overview. 2017, arXiv preprint arXiv:1701.07274.
[78] Kwon I, Kim C, Jun J, Lee J. Case-based myopic reinforcement learning for satisfying target service level in supply chain. Expert Syst Appl 2008;35(1–2):389–97.
[79] Jiang C, Sheng Z. Case-based reinforcement learning for dynamic inventory control in a multi-agent supply-chain system. Expert Syst Appl 2009;36(3):6520–6.
[80] Kim CO, Kwon I-H, Kwak C. Multi-agent based distributed inventory control model. Expert Syst Appl 2010;37(7):5186–91.
[81] Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, Riedmiller M. Playing Atari with deep reinforcement learning. 2013.
[82] Silver D, Hubert T, Schrittwieser J, Antonoglou I, Lai M, Guez A, Lanctot M, Sifre L, Kumaran D, Graepel T, Lillicrap T, Simonyan K, Hassabis D. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science 2018;362(6419):1140–4.
[83] Anagun A. Selecting inventory models using an expert system. Comput Ind Eng 1997;33(1–2):299–302.
[84] Giannoccaro I, Pontrandolfo P. Inventory management in supply chains: A reinforcement learning approach. Int J Prod Econ 2002;78(2):153–61.
[85] Rao JJ, Ravulapati KK, Das TK. A simulation-based approach to study stochastic inventory-planning games. Int J Syst Sci 2003;34(12–13):717–30.
[86] Emerson D, Piramuthu S. Agent-based framework for dynamic supply chain configuration. In: 37th annual Hawaii international conference on system sciences, 2004. proceedings of the, Big Island. HI, USA: IEEE; 2004, p. 9 pp.
[87] Ravulapati KK, Rao Jaideep, Das TK. A reinforcement learning approach to stochastic business games. IIE Trans 2004;36(4):373–85.
[88] Piramuthu S. Knowledge-based framework for automated dynamic supply chain configuration. European J Oper Res 2005;165(1):219–30.
[89] Kim CO, Kwon I-H, Baek J-G. Asynchronous action-reward learning for nonstationary serial supply chain inventory control. Appl Intell 2008;28(1):1–16.
[90] Kwak C, Choi JS, Kim CO, Kwon I-H. Situation reactive approach to Vendor Managed Inventory problem. Expert Syst Appl 2009;36(5):9039–45.
[91] Sui Z, Gosavi A, Lin L. A reinforcement learning approach for inventory replenishment in vendor-managed inventory systems with consignment inventory. Eng Manage J 2010;22(4):44–53.
[92] Katanyukul T, Duff WS, Chong EK. Approximate dynamic programming for an inventory problem: Empirical comparison. Comput Ind Eng 2011;60(4):719–43.
[93] Katanyukul T, Chong EKP. Intelligent inventory control via ruminative reinforcement learning. J Appl Math 2014;2014:1–8.
[94] Kara A, Dogan I. Reinforcement learning approaches for specifying ordering policies of perishable inventory systems. Expert Syst Appl 2018;91:150–8.
[95] Vanvuchelen N, Gijsbrechts J, Boute R. Use of proximal policy optimization for the joint replenishment problem. Comput Ind 2020;119:103239.
[96] Bharti S, Kurian DS, Pillai VM. Reinforcement learning for inventory management. In: Deepak Bbvl, Parhi D, Jena PC, editors. Innovative product design and intelligent manufacturing systems. Singapore: Springer Singapore; 2020, p. 877–85.

[97] Perez HD, Hubbs CD, Li C, Grossmann IE. Algorithmic approaches to inventory management optimization. Processes 2021;9(1):102.
[98] Wang F, Lin L. Spare parts supply chain network modeling based on a novel scale-free network and replenishment path optimization with Q learning. Comput Ind Eng 2021;157:107312.
[99] Kiyaei M, Kiaee F. Optimal ATM cash replenishment planning in a smart city using deep Q-network. In: 2021 26th international computer conference, computer society of Iran. Tehran, Iran: IEEE; 2021, p. 1–5.
[100] Fallahi A, Amani Bani E, Niaki STA. A constrained multi-item EOQ inventory model for reusable items: Reinforcement learning-based differential evolution and particle swarm optimization. Expert Syst Appl 2022;207:118018.
[101] Preil D, Krapp M. Bandit-based inventory optimisation: Reinforcement learning in multi-echelon supply chains. Int J Prod Econ 2022;252:108578.
[102] Gijsbrechts J, Boute RN, Van Mieghem JA, Zhang DJ. Can deep reinforcement learning improve inventory management? Performance on lost sales, dual-sourcing, and multi-echelon problems. Manuf Serv Oper Manag 2022;24(3):1349–68.
[103] Zhou Q, Yang Y, Fu S. Deep reinforcement learning approach for solving joint pricing and inventory problem with reference price effects. Expert Syst Appl 2022;195:116564.
[104] De Moor BJ, Gijsbrechts J, Boute RN. Reward shaping to improve the performance of deep reinforcement learning in perishable inventory management. European J Oper Res 2022;301(2):535–45.
[105] Meisheri H, Sultana NN, Baranwal M, Baniwal V, Nath S, Verma S, Ravindran B, Khadilkar H. Scalable multi-product inventory control with lead time constraints using reinforcement learning. Neural Comput Appl 2022;34(3):1735–57.
[106] Gioia DG, Felizardo LK, Brandimarte P. Inventory management of vertically differentiated perishable products with stock-out based substitution. IFAC-PapersOnLine 2022;55(10):2683–8.
[107] Shakya M, Ng HY, Ong DJ, Lee B-S. Reinforcement learning approach for multi-period inventory with stochastic demand. In: Maglogiannis I, Iliadis L, Macintyre J, Cortez P, editors. In: Artificial intelligence applications and innovations, vol. 646, Cham: Springer International Publishing; 2022, p. 282–91.
[108] Agrawal S, Jia R. Learning in structured MDPs with convex cost functions: Improved regret bounds for inventory management. Oper Res 2022;70(3):1646–64.
[109] Cuartas C, Aguilar J. Hybrid algorithm based on reinforcement learning for smart inventory management. J Intell Manuf 2023;34(1):123–49.
[110] Demizu T, Fukazawa Y, Morita H. Inventory management of new products in retailers using model-based deep reinforcement learning. Expert Syst Appl 2023;229:120256.
[111] Wang K, Long C, Ong DJ, Zhang J, Yuan X-M. Single-site perishable inventory management under uncertainties: A deep reinforcement learning approach. IEEE Trans Knowl Data Eng 2023;35(10):10 807–13.
[112] van Hezewijk L, Dellaert N, Van Woensel T, Gademann N. Using the proximal policy optimisation algorithm for solving the stochastic capacitated lot sizing problem. Int J Prod Res 2023;61(6):1955–78.
[113] Mo DY, Tsang YP, Xu W, Wang Y. Dynamic inventory replenishment with reinforcement learning in managing E-fulfilment centres. In: Wang T, Patnaik S, Jack WC Ho, Varela ML Rocha, editors. In: Applications of decision science in management, vol. 260, Singapore: Springer Nature Singapore; 2023, p. 313–9.
[114] Lu F, Meyn SP. Convex Q learning in a stochastic environment. In: 2023 62nd IEEE conference on decision and control. Singapore, Singapore: IEEE; 2023, p. 776–81.
[115] Cheung WC, Simchi-Levi D, Zhu R. Nonstationary reinforcement learning: The blessing of (more) optimism. Manag Sci 2023;69(10):5722–39.
[116] Zhou Q, Fu S, Yang Y, Dong C. Joint pricing and inventory control with reference price effects and price thresholds: A deep reinforcement learning approach. Expert Syst Appl 2023;233:120993.
[117] Li N, Qiao X, Wang Z, markdown Joint. Ordering and freshness-keeping policy under demand competition between new and old products. J Oper Res Soc 2023;74(9):2043–63.
[118] Mohamadi N, Niaki STA, Taher M, Shavandi A. An application of deep reinforcement learning and vendor-managed inventory in perishable supply chain management. Eng Appl Artif Intell 2024;127:107403.
[119] Dehaybe H, Catanzaro D, Chevalier P. Deep Reinforcement Learning for inventory optimization with non-stationary uncertain demand. European J Oper Res 2024;314(2):433–45.
[120] Liu S, Wang J, Wang R, Zhang Y, Song Y, Xing L. Data-driven dynamic pricing and inventory management of an omni-channel retailer in an uncertain demand environment. Expert Syst Appl 2024;244:122948.
[121] Kaynov I, van Knippenberg M, Menkovski V, van Breemen A, van Jaarsveld W. Deep reinforcement learning for one-warehouse multi-retailer inventory management. Int J Prod Econ 2024;267:109088.
[122] Stranieri F, Fadda E, Stella F. Combining deep reinforcement learning and multi-stage stochastic programming to address the supply chain inventory management problem. Int J Prod Econ 2024;268:109099.
[123] Tian R, Lu M, Wang H, Wang B, Tang Q. IACPPO: A deep reinforcement learning-based model for warehouse inventory replenishment. Comput Ind Eng 2024;187:109829.
[124] Lee J, Shin Y, Moon I. A hybrid deep reinforcement learning approach for a proactive transshipment of fresh food in the online–offline channel system. Transp Res Part E: Logist Transp Rev 2024;187:103576.
[125] Luo Y, Hu J, Gosavi A. A model-adaptive random search actor critic: Convergence analysis and inventory-control case studies. Ann Oper Res 2024.
[126] Yavuz T, Kaya O. Deep reinforcement learning algorithms for dynamic pricing and inventory management of perishable products. Appl Soft Comput 2024;163:111864.
[127] Rizqi ZU, Chou S-Y. Neuroevolution reinforcement learning for multi-echelon inventory optimization with delivery options and uncertain discount. Eng Appl Artif Intell 2024;134:108670.
[128] Saha E, Rathore P. A smart inventory management system with medication demand dependencies in a hospital supply chain: A multi-agent reinforcement learning approach. Comput Ind Eng 2024;191:110165.
[129] Stranieri F, Stella F, Kouki C. Performance of deep reinforcement learning algorithms in two-echelon inventory control systems. Int J Prod Res 2024;62(17):6211–26.
[130] Batsis A, Samothrakis S. Contextual reinforcement learning for supply chain management. Expert Syst Appl 2024;249:123541.
[131] Sallans B, Hinton GE. Reinforcement learning with factored states and actions. J Mach Learn Res 2004;5:1063–88.
[132] Pazis J, Parr R. Generalized value functions for large action sets. In: Proceedings of the 28th international conference on international conference on machine learning. ICML'11, Madison, WI, USA: Omni Press; 2011, p. 1185–92.
[133] Dulac-Arnold G, Denoyer L, Preux P, Gallinari P. Fast reinforcement learning with large action sets using error-correcting output codes for MDP factorization. 2012.
[134] Mahajan A, Samvelyan M, Mao L, Makoviychuk V, Garg A, Kossaifi J, Whiteson S, Zhu Y, Anandkumar A. Reinforcement learning in factored action spaces using tensor decompositions. 2021, arXiv.
[135] Dulac-Arnold G, Evans R, van Hasselt H, Sunehag P, Lillicrap T, Hunt J, Mann T, Weber T, Degris T, Coppin B. Deep reinforcement learning in large discrete action spaces. 2016.
[136] Chandak Y, Theocharous G, Kostas J, Jordan S, Thomas PS. Learning action representations for reinforcement learning. 2019.
[137] Whitney W, Agarwal R, Cho K, Gupta A. Dynamics-aware embeddings. 2020.
[138] Pritz PJ, Ma L, Leung KK. Jointly-learned state-action embedding for efficient reinforcement learning. In: Proceedings of the 30th ACM international conference on information & knowledge management. CIKM '21, New York, NY, USA: Association for Computing Machinery; 2021, p. 1447–56.
[139] Zhang T, Guo S, Tan T, Hu X, Chen F. Generating adjacency-constrained subgoals in hierarchical reinforcement learning. 2021.
[140] Kim M, Park J, Kim J. Learning collaborative policies to solve NP-hard routing problems. 2021.
[141] Peng B, Rashid T, Schroeder de Witt C, Kamienny P-A, Torr P, Boehmer W, Whiteson S. FACMAC: factored multi-agent centralised policy gradients. In: Advances in neural information processing systems, vol. 34, Curran Associates, Inc.; 2021, p. 12 208–21.
[142] Enders T, Harrison J, Pavone M, Schiffer M. Hybrid multi-agent deep reinforcement learning for autonomous mobility on demand systems. In: Proceedings of the 5th annual learning for dynamics and control conference. PMLR; 2023, p. 1284–96.
[143] Akkerman F, Luy J, van Heeswijk W, Schiffer M. Dynamic neighborhood construction for structured large discrete action spaces. 2024.
[144] Vanvuchelen N, De Moor BJ, Boute RN. The use of continuous action representations to scale deep reinforcement learning for inventory control. Rochester, NY; 2024.
[145] Wiel A, Durach CF. Two perspectives on supply chain resilience. J Bus Logist 2021;42(3):315–22.
[146] Chaudhary V, Kulshrestha R, Routroy S. State-of-the-art literature review on inventory models for perishable products. J Adv Manag Res 2018;15(3):306–46.
[147] van Jaarsveld W, Dekker R. Estimating obsolescence risk from demand data to enhance inventory control—A case study. Int J Prod Econ 2011;133(1):423–31.
[148] Teunter RH, Fortuin L. End-of-life service. Int J Prod Econ 1999;59(1):487–97.
[149] van Houtum G-J. Multiechelon production/inventory systems: Optimal policies, heuristics, and algorithms. In: Models, methods, and applications for innovative decision making. INFORMS TutORials in operations research, INFORMS; 2006, p. 163–99, ch. 7.
[150] Syntetos A, Keyes M, Babai M. Demand categorisation in a European spare parts logistics network. Int J Oper Prod Manage 2009;29(3):292–316.
[151] Flores BE, Olson DL, Dorai VK. Management of multicriteria inventory classification. Math Comput Modelling 1992;16(12):71–82.
[152] Gajpal PP, Ganesh LS, Rajendran C. Criticality analysis of spare parts using the analytic hierarchy process. Int J Prod Econ 1994;35(1):293–7.
[153] Ramanathan R. ABC inventory classification with multiple-criteria using weighted linear optimization. Comput Oper Res 2006;33(3):695–700.
[154] Ng WL. A simple classifier for multiple criteria ABC analysis. European J Oper Res 2007;177(1):344–53.
[155] Hadi-Vencheh A. An improvement to multiple criteria ABC inventory classification. European J Oper Res 2010;201(3):962–5.

[156] Teunter RH, Babai MZ, Syntetos AA. ABC classification: service levels and inventory costs. Prod Oper Manage 2010;19(3):343–52.
[157] Zowid F, Babai M, Douissa M, Ducq Y. Multi-criteria inventory ABC classification using Gaussian Mixture Model. IFAC-PapersOnLine 2019;52(13):1925–30.
[158] Zhang S, Qin X, Hu S, Zhang Q, Dong B, Zhao J. Importance degree evaluation of spare parts based on clustering algorithm and back-propagation neural network. Math Probl Eng 2020;2020:1–13.
[159] Xiao Y-y, Zhang R-q, Kaku I. A new approach of inventory classification based on loss profit. Expert Syst Appl 2011;38(8):9382–91.
[160] Hu Q, Chakhar S, Siraj S, Labib A. Spare parts classification in industrial manufacturing using the dominance-based rough set approach. European J Oper Res 2017;262(3):1136–63.
[161] Partovi FY, Anandarajan M. Classifying inventory using an artificial neural network approach. Comput Ind Eng 2002;41(4):389–404.
[162] Yu M-C. Multi-criteria ABC analysis using artificial-intelligence-based classification techniques. Expert Syst Appl 2011;38(4):3416–21.
[163] Lolli F, Gamberini R, Regattieri A, Balugani E, Gatos T, Gucci S. Single-hidden layer neural networks for forecasting intermittent demand. Int J Prod Econ 2017;183:116–28.
[164] Lolli F, Balugani E, Ishizaka A, Gamberini R, Rimini B, Regattieri A. Machine learning for multi-criteria inventory classification applied to intermittent demand. Prod Plan Control 2019;30(1):76–89.
[165] Priore P, Ponte B, Rosillo R, de la Fuente D. Applying machine learning to the dynamic selection of replenishment policies in fast-changing supply chain environments. Int J Prod Res 2019;57(11):3663–77.
[166] Svoboda J, Minner S. Tailoring inventory classification to industry applications: The benefits of understandable machine learning. Int J Prod Res 2022;60(1):388–401.
[167] Badakhshan E, Ball P, Badakhshan A. Using digital twins for inventory and cash management in supply chains. IFAC-PapersOnLine 2022;55(10):1980–5.
[168] Balugani E, Lolli F, Gamberini R, Rimini B, Regattieri A. Clustering for inventory control systems. IFAC-PapersOnLine 2018;51(11):1174–9.
[169] Rengasamy S, Murugesan P. PSO based data clustering with a different perception. Swarm Evol Comput 2021;64:100895.
[170] Wang A, Gao X. A variable-scale dynamic clustering method. Comput Commun 2021;171:163–72.
[171] Khanorkar Y, Kane PV. Selective inventory classification using ABC classification, multi-criteria decision making techniques, and machine learning techniques. Mater Today: Proc 2023;72:1270–4.
[172] Kartal H, Oztekin A, Gunasekaran A, Cebi F. An integrated decision analytic framework of machine learning with multi-criteria decision making for multi-attribute inventory classification. Comput Ind Eng 2016;101:599–613.
[173] López-Soto D, Yacout S, Angel-Bello F. Root cause analysis of familiarity biases in classification of inventory items based on logical patterns recognition. Comput Ind Eng 2016;93:121–30.
[174] Lolli F, Ishizaka A, Gamberini R, Balugani E, Rimini B. Decision trees for supervised multi-criteria inventory classification. Procedia Manuf 2017;11:1871–81.
[175] López-Soto D, Angel-Bello F, Yacout S, Alvarez A. A multi-start algorithm to design a multi-class classifier for a multi-criteria ABC inventory classification problem. Expert Syst Appl 2017;81:12–21.
[176] Sundar R, Punniyamoorthy M. Performance enhanced boosted SVM for imbalanced datasets. Appl Soft Comput 2019;83:105601.
[177] Lolli F, Ishizaka A, Gamberini R. New AHP-based approaches for multi-criteria inventory classification. Int J Prod Econ 2014;156:62–74.
[178] Ren S, Choi T-M, Lee K-M, Lin L. Intelligent service capacity allocation for cross-border-E-commerce related third-party-forwarding logistics operations: A deep learning approach. Transp Res Part E: Logist Transp Rev 2020;134:101834.
[179] Kmiecik M. Logistics coordination based on inventory management and transportation planning by Third-Party Logistics (3PL). Sustainability 2022;14(13):8134.
[180] Ntakolia C, Kokkotis C, Karlsson P, Moustakidis S. An explainable machine learning model for material backorder prediction in inventory management. Sensors 2021;21(23):7926.
[181] Islam S, Amin SH. Prediction of probable backorder scenarios in the supply chain using Distributed Random Forest and Gradient Boosting Machine learning techniques. J Big Data 2020;7(1):65.
[182] Ahmed F, Hasan M, Hossain MS, Andersson K. Comparative performance of tree based machine learning classifiers in product backorder prediction. In: Vasant P, Weber G-W, Marmolejo-Saucedo JA, Munapo E, Thomas JJ, editors. In: Intelligent computing & optimization, vol. 569, Cham: Springer International Publishing; 2023, p. 572–84.
[183] de Santis RB, de Aguiar EP, Goliatt L. Predicting material backorders in inventory management using machine learning. In: 2017 IEEE Latin American conference on computational intelligence. Arequipa: IEEE; 2017, p. 1–6.
[184] Namir K, Labriji H, Ben Lahmar EH. Decision support tool for dynamic inventory management using machine learning, time series and combinatorial optimization. Procedia Comput Sci 2022;198:423–8.
[185] Guo M, Kong XTR, Chan HK, Thadani DR. Integrated inventory control and scheduling decision framework for packaging and products on a reusable transport item sharing platform. Int J Prod Res 2023;61(13):4575–91.
[186] Bandaru S, Aslam T, Ng AH, Deb K. Generalized higher-level automated innovization with application to inventory management. European J Oper Res 2015;243(2):480–96.
[187] Inprasit T, Tanachutiwat S. Reordering point determination using machine learning technique for inventory management. In: 2018 international conference on engineering, applied sciences, and technology. phuket: IEEE; 2018, p. 1–4.
[188] Singha D, Panse C. Application of different machine learning models for supply chain demand forecasting: comparative analysis. In: 2022 2nd international conference on innovative practices in technology and management. Gautam Buddha Nagar, India: IEEE; 2022, p. 312–8.
[189] Yu C-F, Wah B. Learning dominance relations in combined search problems. IEEE Trans Softw Eng 1988;14(8):1155–75.
[190] Garcia J. Predict product backorders. 2020, [Online]. Available: https://www.kaggle.com/c/untadta/overview.
[191] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80.
[192] Temizöz T, Imdahl C, Dijkman R, Lamghari-Idrissi D, van Jaarsveld W. Deep controlled learning for inventory control. European J Oper Res 2025.
[193] Al Hajj Hassan L, Mahmassani HS, Chen Y. Reinforcement learning framework for freight demand forecasting to support operational planning decisions. Transp Res Part E: Logist Transp Rev 2020;137:101926.