Titre: Title:	Conception d'outils d'aide à la décision visant à améliorer l'efficacité opérationnelle de la chaîne d'approvisionnement dans un contexte de modernisation technologique
Auteur: Author:	Zineb Aboutalib
Date:	2025
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Aboutalib, Z. (2025). Conception d'outils d'aide à la décision visant à améliorer l'efficacité opérationnelle de la chaîne d'approvisionnement dans un contexte de modernisation technologique [Thèse de doctorat, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/71010/

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Conception d'outils d'aide à la décision visant à améliorer l'efficacité opérationnelle de la chaîne d'approvisionnement dans un contexte de modernisation technologique

ZINEB ABOUTALIB
Département de mathématiques et de génie industriel

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
Génie industriel

Novembre 2025

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Cette thèse intitulée :

Conception d'outils d'aide à la décision visant à améliorer l'efficacité opérationnelle de la chaîne d'approvisionnement dans un contexte de modernisation technologique

présentée par **Zineb ABOUTALIB**
en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

Jean-Marc FRAYRET, président
Bruno AGARD, membre et directeur de recherche
Samira KEYVANPOUR, membre
Bernard PENZ, membre externe

DÉDICACE

À ma famille, spécialement à mon père Ahmed Aboutalib et à mon mari Félix Tremblay. Votre soutien inébranlable a été mon moteur pour achever cette thèse. Le soutien et la bienveillance de mon directeur de recherche ont été mes piliers. Merci du fond du cœur.

REMERCIEMENTS

Je commence par exprimer ma profonde reconnaissance à SID qui a toujours été là pour moi.

La réalisation de ma thèse de doctorat n'aurait pas été possible sans la présence de Bruno Agard en tant que directeur de recherche. Mon respect et ma gratitude envers lui sont immenses. Bruno a été à la fois un guide et un grand frère. Sa bienveillance, son expertise et sa confiance en moi ont été des éléments moteurs essentiels pour mener à bien ce projet.

Il est difficile de mettre en mots toute l'appréciation que je ressens envers Camélia Dadouchi, dont l'assistance a été cruciale pour mon doctorat. C'est elle qui m'a permis de rencontrer Bruno Agard, un moment pivot dans ma carrière académique.

Je tiens à exprimer ma profonde gratitude envers mes parents, car sans eux, je n'aurais pas évolué en la femme que je suis aujourd'hui. À ma mère et à mon père, je dois la leçon fondamentale que nous sommes les maîtres de notre destinée, en particulier en tant que femmes, où l'indépendance, la persévérance et la poursuite de nos rêves sont essentielles. Leur croyance inébranlable en moi a toujours été mon moteur.

Je tiens à adresser une reconnaissance spéciale à mon père et à mon mari, mes deux plus précieux amis. Leur soutien inconditionnel et leur écoute attentive m'ont accompagnée dans les moments tant aisés que difficiles.

Mes remerciements vont également à M. Jean-Marc Frayret, Mme. Samira Keyvanpour et M. Bernard Penz, qui ont honoré mes travaux de recherche en acceptant de faire partie du jury d'évaluation.

Je tiens à exprimer un remerciement tout particulier à Logistik Unicorp, ainsi qu'à Michel Ricard, Guillaume Thibault et Philippe St-Aubin. Votre soutien a joué un rôle décisif dans la faisabilité de ce projet. Votre appui financier et moral ont constitué des piliers cruciaux de mon parcours, et je suis profondément reconnaissante pour cette opportunité.

J'adresse également mes remerciements aux membres dévoués, chaleureux et souriants du personnel MAGI, notamment Marie-Carline, Melisa et Suzanne, pour leur disponibilité constante et leur support.

À mon frère et à ma sœur, je vous suis reconnaissante pour votre soutien et vos encouragements.

Enfin, un immense merci à toutes les personnes qui m'ont aidé de près ou de loin, ne serait-ce qu'avec un gentil mot, un sourire, un encouragement ou un conseil. Votre présence a été une

lumière bienveillante dans ce parcours.

RÉSUMÉ

Face aux défis contemporains tels que les perturbations des chaînes d'approvisionnement mondiales, la volatilité des marchés et la pression concurrentielle, les entreprises recherchent activement des solutions innovantes pour optimiser leurs opérations logistiques. L'émergence de technologies basées sur les données offre de nouvelles perspectives pour résoudre des problématiques logistiques traditionnellement abordées par des méthodes analytiques classiques. Cette thèse s'inscrit dans cette dynamique en proposant des méthodologies qui exploitent intelligemment les données pour créer des outils décisionnels adaptés aux réalités complexes des chaînes d'approvisionnement modernes.

La performance logistique repose sur l'optimisation simultanée de multiples composantes interconnectées formant un système cohérent. Notre recherche aborde trois défis majeurs qui, ensemble, déterminent l'efficacité opérationnelle globale de la chaîne d'approvisionnement : la consolidation de marchandises, l'allocation des articles pour l'exécution des commandes (OFA), et le contrôle d'inventaire. Pour chacun de ces domaines, nous développons des approches novatrices qui tirent parti des avancées en fouille de données et en apprentissage automatique, en choisissant systématiquement l'outil méthodologique le mieux adapté à la nature spécifique de chaque problème.

Notre première contribution présente une méthodologie novatrice de consolidation de marchandises basée sur la fouille de données. La consolidation (un processus consistant à regrouper plusieurs petits envois en unités plus importantes) représente un levier essentiel pour la réduction des coûts logistiques. Contrairement aux modèles traditionnels qui s'appuient généralement sur des données agrégées, notre approche exploite les règles d'association pour identifier des patterns cachés dans des données non agrégées et intermittentes. Cette méthode comprend l'extraction automatique de règles d'associations, qui sont ensuite sélectionnées selon des critères de proximité spatiale et de force d'association. Appliquée à un cas d'étude sur la consolidation de produits retournés, notre méthodologie surpasse les approches classiques d'optimisation, démontrant une réduction significative des coûts totaux tout en offrant une flexibilité accrue face aux fluctuations de la demande. Un avantage majeur de notre contribution réside dans sa capacité à traiter efficacement des volumes de données considérables sans compromettre la finesse de l'analyse, rendant la solution particulièrement pertinente pour les environnements logistiques actuels caractérisés par l'abondance informationnelle.

Notre deuxième contribution élargit le cadre conceptuel en intégrant la consolidation spatiale à l'allocation des items pour l'exécution des commandes (OFA). Cette problématique, cen-

trale dans les centres de distribution, consiste à déterminer quelles commandes prioriser et satisfaire lorsque les stocks sont limités. Nous développons une méthodologie multi-objectifs qui équilibre deux impératifs souvent contradictoires : la minimisation des coûts d'expédition et l'optimisation du taux de remplissage des commandes pondéré (WOFR), indicateur clé du respect des niveaux de service client. L'innovation centrale de cette contribution est un modèle de programmation linéaire en nombres entiers mixtes—WOFR-EC (Weighted Order Fill Rate Encouraging Consolidation)—qui intègre explicitement la consolidation spatiale dans son architecture décisionnelle. Ce modèle introduit un mécanisme de pénalisation proportionnel au nombre de points de collecte visités, encourageant ainsi le regroupement géographique des livraisons sans compromettre les niveaux de service client. La méthodologie intègre également une version améliorée de l'approche par règles d'association de notre première contribution, avec l'ajout d'un algorithme de sélection automatique des règles. Les expérimentations sur données réelles démontrent l'efficacité de notre approche : pour un niveau de service équivalent, notre méthodologie permet une réduction jusqu'à 21% du nombre d'arrêts dans les itinéraires de livraison comparativement aux méthodes conventionnelles. Cette contribution représente une avancée significative vers l'intégration des objectifs souvent contradictoires d'optimisation des coûts logistiques et de satisfaction client.

Notre troisième contribution, d'une portée technique et méthodologique plus étendue, aborde la problématique fondamentale du contrôle d'inventaire dans des environnements complexes. Nous développons un système novateur qui intègre l'apprentissage par renforcement multi-agents (MARL) avec des prévisions par quantiles et de la programmation linéaire en nombres entiers mixtes. Cette approche hybride répond aux limitations des modèles traditionnels de gestion des stocks qui reposent souvent sur des hypothèses simplificatrices concernant la distribution de la demande et les structures de coûts. Notre système adresse directement les complexités des environnements de distribution réels : multiples unités de gestion des stocks (Stock Keeping Units, SKUs) avec des profils de demande interdépendants, accords de niveau de service (SLA) hiérarchisés avec différentes exigences de priorité, demandes non-stationnaires et intermittentes, et délais d'approvisionnement dynamiques. Cette architecture permet d'optimiser des fonctions objectifs personnalisées directement alignées sur les priorités opérationnelles : respect des SLAs, minimisation des coûts d'inventaire et réduction des retards de livraison.

Pour établir une base comparative rigoureuse, nous avons d'abord développé et optimisé une heuristique sophistiquée incorporant des prévisions par quantiles et une gestion explicite des contraintes multi-SLA. Notre système basé sur l'apprentissage par renforcement démontre une amélioration normalisée de 9% par rapport à cette référence déjà optimisée, réalisée principalement grâce à une réduction de 16% des coûts de détention d'inventaire tout en maintenant

les niveaux de service contractuels. Le système excelle particulièrement dans l'équilibrage des objectifs concurrents entre différents niveaux de SLA, les groupes de priorité inférieure bénéficiant d'améliorations dépassant 75% pour certaines métriques, tout en préservant la performance pour les catégories prioritaires. Cette approche représente une avancée significative par rapport aux méthodes traditionnelles car elle permet d'adapter dynamiquement les décisions d'inventaire en fonction des conditions changeantes du marché, sans nécessiter de recalibration manuelle des paramètres comme c'est souvent le cas pour les modèles statiques.

Notre travail se distingue en proposant une combinaison judicieuse d'outils méthodologiques, chacun choisi pour ses forces spécifiques face au problème considéré. Pour les décisions d'OFA quotidiennes, nous privilégions la programmation mathématique car ces problèmes présentent des caractéristiques favorables : cadre bien délimité, absence de besoin de mémoire à long terme, et espace d'états relativement restreint (commandes actives et stocks disponibles). À l'inverse, pour le contrôle d'inventaire, l'apprentissage par renforcement s'avère supérieur grâce à sa capacité à incorporer la mémoire des événements passés, à s'adapter aux environnements changeants, et à gérer des espaces d'états volumineux tout en anticipant les conséquences à long terme des décisions actuelles. Cette approche reflète notre conviction qu'il n'existe pas de méthode universelle, mais plutôt un éventail d'outils dont l'efficacité dépend fondamentalement de la structure inhérente au problème traité.

Les applications pratiques de nos contributions offrent des bénéfices tangibles aux entreprises confrontées aux défis logistiques contemporains. La méthodologie de consolidation basée sur les règles d'association permet d'optimiser les stratégies de transport et de réduire les coûts opérationnels liés à la livraison, particulièrement dans des contextes où la demande présente une forte variabilité et intermittence. Le modèle WOFR-EC fournit aux centres de distribution un outil permettant de réduire substantiellement les coûts logistiques tout en respectant rigoureusement les niveaux de service établis avec les clients, grâce à l'intégration intelligente de la consolidation spatiale dans les décisions quotidiennes d'allocation des stocks. Quant au système de contrôle d'inventaire basé sur l'apprentissage par renforcement, il constitue une solution complète qui permet d'optimiser simultanément les niveaux de stock, la synchronisation des réapprovisionnements avec la demande anticipée, et l'allocation équilibrée des ressources entre différentes catégories de clients selon leurs priorités contractuelles, le tout en s'adaptant dynamiquement aux fluctuations du marché sans intervention manuelle récurrente.

Cette thèse démontre comment l'utilisation ciblée des technologies basées sur les données peut considérablement améliorer l'efficacité opérationnelle de la chaîne d'approvisionnement. Les méthodologies développées constituent des outils décisionnels qui répondent aux besoins

concrets des entreprises modernes, leur permettant de naviguer avec agilité dans un environnement commercial exigeant. Au-delà des contributions méthodologiques, notre travail ouvre de nouvelles perspectives pour la recherche en logistique, notamment vers le développement de systèmes encore plus intégrés qui orchestrent les différentes composantes de la chaîne d'approvisionnement, et l'exploration de méthodes d'apprentissage hybrides combinant les forces des différentes approches algorithmiques pour résoudre des problèmes logistiques de complexité croissante.

ABSTRACT

In the face of contemporary challenges such as global supply chain disruptions, market volatility, and competitive pressure, companies are actively seeking innovative solutions to optimize their logistics operations. The emergence of data-driven technologies offers new perspectives for addressing logistics problems traditionally approached through classical analytical methods. This thesis aligns with this dynamic by proposing methodologies that intelligently leverage data to create decision-support tools adapted to the complex realities of modern supply chains.

Logistics performance relies on the simultaneous optimization of multiple interconnected components forming a coherent system. Our research addresses three major challenges that together determine the overall operational efficiency of the supply chain: freight consolidation, order fulfillment allocation (OFA), and inventory control. For each of these domains, we develop innovative approaches that take advantage of advances in data mining and machine learning, systematically selecting the methodological tool best suited to the specific nature of each problem.

Our first contribution presents a novel methodology for freight consolidation based on data mining. Consolidation, the process of grouping multiple small shipments into larger units, represents a key lever for reducing logistics costs. Unlike traditional models that typically rely on aggregated data, our approach exploits association rules to uncover hidden patterns in non-aggregated and intermittent data. This method includes the automatic extraction of association rules, which are then selected based on criteria of spatial proximity and association strength. Applied to a case study on the consolidation of returned products, our methodology outperforms classical optimization approaches, demonstrating a significant reduction in total costs while offering increased flexibility in the face of demand fluctuations. A major advantage of our contribution lies in its ability to efficiently handle large volumes of data without compromising analytical granularity, making the solution particularly relevant for modern logistics environments characterized by data abundance.

Our second contribution expands the conceptual framework by integrating spatial consolidation into the order fulfillment allocation (OFA) process. This issue, central in distribution centers, involves determining which orders to prioritize and fulfill when inventory is limited. We develop a multi-objective methodology that balances two often conflicting imperatives: minimizing shipping costs and optimizing the weighted order fill rate (WOFR), a key indicator of customer service level adherence. The core innovation of this contribution is a

mixed-integer linear programming model—WOFR-EC (Weighted Order Fill Rate Encouraging Consolidation)—which explicitly incorporates spatial consolidation into its decision-making architecture. This model introduces a penalty mechanism proportional to the number of pickup points visited, thereby encouraging the geographic grouping of deliveries without compromising customer service levels. The methodology also includes an enhanced version of the association rule-based approach from our first contribution, with the addition of an automatic rule selection algorithm. Real-world data experiments demonstrate the effectiveness of our approach: for an equivalent service level, our methodology achieves up to a 21% reduction in the number of delivery stops compared to conventional methods. This contribution represents a significant advancement toward integrating the often-conflicting goals of logistics cost optimization and customer satisfaction.

Our third contribution, broader in technical and methodological scope, addresses the fundamental issue of inventory control in complex environments. We develop an innovative system that integrates multi-agent reinforcement learning (MARL) with quantile forecasting and mixed-integer linear programming. This hybrid approach addresses the limitations of traditional inventory management models, which often rely on simplifying assumptions regarding demand distribution and cost structures. Our system directly tackles the complexities of real-world distribution environments: multiple SKUs with interdependent demand profiles, hierarchical service level agreements (SLAs) with varying priority requirements, non-stationary and intermittent demand, and dynamic lead times. This architecture enables the optimization of custom objective functions aligned with operational priorities: SLA compliance, inventory cost minimization, and reduction of delivery delays.

To establish a rigorous comparative baseline, we first developed and optimized a sophisticated heuristic incorporating quantile forecasts and explicit multi-SLA constraint management. Our reinforcement learning-based system demonstrates a normalized improvement of 9% over this already optimized reference, achieved primarily through a 16% reduction in inventory holding costs while maintaining contractual service levels. The system excels particularly in balancing competing objectives across different SLA tiers, with lower-priority groups experiencing performance improvements exceeding 75% for certain metrics, all while preserving outcomes for high-priority categories. This approach represents a significant advancement over traditional methods, as it allows inventory decisions to be dynamically adapted to changing market conditions without requiring manual recalibration of parameters, as is often the case with static models.

Our work stands out by offering a judicious combination of methodological tools, each chosen for its specific strengths in addressing the problem at hand. For daily OFA decisions,

we favor mathematical programming due to the favorable characteristics of these problems: well-defined boundaries, absence of long-term memory requirements, and a relatively limited state space (active orders and available inventory). Conversely, for inventory control, reinforcement learning proves superior due to its ability to incorporate memory of past events, adapt to changing environments, and manage large state spaces while anticipating the long-term consequences of current decisions. This approach reflects our belief that no universal method exists, but rather a spectrum of tools whose effectiveness fundamentally depends on the inherent structure of the problem being addressed.

The practical applications of our contributions offer tangible benefits to companies facing contemporary logistics challenges. The association rule-based consolidation methodology optimizes transportation strategies and reduces operational delivery costs, particularly in contexts where demand is highly variable and intermittent. The WOFR-EC model provides distribution centers with a tool to substantially reduce logistics costs while rigorously meeting established customer service levels, through the intelligent integration of spatial consolidation in daily stock allocation decisions. As for the reinforcement learning-based inventory control system, it constitutes a comprehensive solution that simultaneously optimizes stock levels, aligns replenishment timing with forecasted demand, and balances resource allocation across different customer categories according to contractual priorities—all while dynamically adapting to market fluctuations without requiring ongoing manual intervention.

This thesis demonstrates how the targeted use of data-driven technologies can significantly enhance the operational efficiency of the supply chain. The developed methodologies constitute decision-support tools that address the concrete needs of modern enterprises, enabling them to navigate a demanding commercial environment with agility. Beyond methodological contributions, our work opens new perspectives for logistics research, particularly toward the development of more integrated systems that orchestrate the various components of the supply chain, and the exploration of hybrid learning methods combining the strengths of different algorithmic approaches to solve increasingly complex logistics problems.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES SIGLES ET ABRÉVIATIONS

AG	Algorithme Génétique
AI	Artificial Intelligence
AR	Association Rules
ATP	Disponible à la Promesse (*Available-to-promise)*
CP	Collection Point
CRC	Centralized Return Center
CRISP-DM	Cross Industry Process for Data mining
CTP	Capable de Promettre
DES	Discrete Event Simulation
DRL	Apprentissage par Renforcement Profond
EOQ	Economic Order Quantity
FCFS	First Come, First Served
FSA	Forward Sorting Area
FTL	Full Truckload
ICP	Initial Collection Point
IRC	Initial Return Center
JIT	Juste à temps
KPI	Key Performance Indicator
LT	Lead Time
MARL	Multi-Agent Reinforcement Learning
MDP	Processus Décisionnel de Markov
MHT	Maximum Holding Time
MILP	Mixed Integer Linear Programming
ML	Machine Learning
MTO	Production sur Commande
MTS	Produire pour le Stock
OF	Order Fulfilled
OFA	Allocation des Items pour l'Exécution des Commandes
OFR	Taux d'Exécution des Commandes
OFS	Système d'Exécution des Commandes
OR	Operations Research
PDP	Plan Directeur de Production
POFT	Priority Order Fulfilled on Time

PPO	Proximal Policy Optimization
PSL	Priority Service Level
RL	Apprentissage par Renforcement
RLT	Replenishment Lead Time
SC	Chaîne d'Approvisionnement
SCM	Gestion de la Chaîne d'Approvisionnement
SKU	Unité de Gestion de Stock
SL	Service Level
SLA	Service Level Agreement
SNR	Signal to Noise Ratio
TSL	Target Service Level
VMI	Pilotage des Niveaux de Stock par les Consommations
VRP	Vehicle Routing Problem
WOFR	Weighted Order Fill Rate
WOFR-EC	Weighted Order Fill Rate Encouraging Consolidation
WOFR-OSP	Weighted Order Fill Rate with Overdraw Stock Penalty
WOLFR	Weighted Order Line Fill Rate

LISTE DES ANNEXES

CHAPITRE 1 INTRODUCTION

Face aux défis de marchés dynamiques qui évoluent sans cesse et où les demandes des consommateurs se transforment rapidement, la chaîne d'approvisionnement se révèle être un maillon crucial pour le succès entrepreneurial [1]. Elle forme un réseau complexe et interdépendant d'activités, englobant la planification, la production, la distribution et le service client. Face à ces dynamiques, les entreprises ont été contraintes d'accélérer leur modernisation technologique pour assurer la résilience et la pérennité de leurs chaînes d'approvisionnement [2]. Une étude de McKinsey & Company [3] révèle que les entreprises qui ont démontré la plus grande résilience durant la crise du COVID-19 (2020-2021) sont celles qui ont intégré de nouvelles technologies dans leur chaîne logistique, par leurs adaptation au changement, ces entreprises ont fait preuve de flexibilité structurale [4]. L'utilisation de ces nouvelles technologies, telles que l'intelligence artificielle, a la capacité de renforcer les solutions traditionnelles existantes grâce à leur capacité à gérer les données massives et ainsi améliorer les différents maillons de la chaîne logistique, surtout en matière de gestion de l'inventaire [4].

Amazon, le géant du commerce électronique, illustre parfaitement l'efficacité de l'intégration de l'intelligence artificielle (AI) et d'autres technologies novatrices dans la gestion de sa chaîne d'approvisionnement [5], offrant ainsi une livraison rapide et économique à des millions de clients. Sa capacité à consolider les marchandises, permettant aux clients de regrouper leurs achats selon leurs besoins, et son service Fulfillment by Amazon (FBA) contribuent à sa prospérité continue [6]. Toutefois, il est crucial de reconnaître que les innovations d'une entreprise d'une telle envergure ne sont pas toujours applicables à d'autres entreprises, qui opèrent à une échelle différente et adoptent souvent des stratégies technologiques distinctes. Malgré ces divergences, l'engagement à fournir un service client exceptionnel reste une priorité commune.

Dans ce contexte en mutation, trois composants clés se distinguent dans la gestion de la chaîne d'approvisionnement : le contrôle d'inventaire, l'allocation des items pour l'exécution des commandes et la consolidation de marchandise. Le contrôle d'inventaire implique principalement de gérer stratégiquement les niveaux de stock pour répondre à la demande tout en minimisant l'excès de stock et les coûts associés [7]. L'allocation des items pour l'exécution des commandes est un processus qui consiste à déterminer la meilleure attribution des stocks disponibles pour satisfaire les commandes clients en considérant les priorités relatives des commandes, et des niveaux d'inventaire limités [8]. Enfin, la consolidation de marchandises est la pratique visant à optimiser l'efficacité du transport et à réduire les coûts d'expédition

en combinant de petits envois en envois consolidés plus importants [9]. Cette thèse a pour objectif général la conception d'outils d'aide à la décision visant à améliorer l'efficacité opérationnelle de la chaîne d'approvisionnement dans un contexte de modernisation technologique.

L'état de l'art soulève divers problèmes et défis auxquels cette thèse tentera de répondre à travers trois objectifs spécifiques. Ces derniers représentent nos contributions et permettront d'atteindre notre objectif général.

La première contribution propose une approche novatrice de consolidation, exploitant les règles d'association pour comprendre les comportements des clients via des schémas de données, dans un cadre de données intermittentes et dans le but de réduire les coûts de transport.

La deuxième contribution incorpore la consolidation spatiale (issue d'une méthodologie améliorée de la la première contribution) à l'allocation des articles pour l'exécution des commandes (OFA) par une méthodologie multi-objectifs, qui vise simultanément à réduire les coûts d'expédition et à accroître le niveau de service client, grâce à l'intégration d'un modèle de programmation linéaire en nombres entiers mixtes.

La troisième contribution développe un système novateur de contrôle d'inventaire qui intègre l'apprentissage par renforcement multi-agents (MARL) avec des prévisions par quantiles et une programmation linéaire en nombres entiers mixtes. Ce système aborde directement les complexités des environnements de distribution réels : multiples SKUs aux profils de demande interdépendants, accords de niveau de service (SLA) hiérarchisés, demandes intermittentes et non-stationnaires, et délais d'approvisionnement dynamiques. L'architecture développée permet d'optimiser simultanément les coûts d'inventaire, la ponctualité des commandes et le respect des niveaux de service contractuels, sans nécessiter d'hypothèses simplificatrices sur les distributions de demande.

Pour atteindre l'objectif général de cette thèse, nous explorons les interconnexions complexes entre le contrôle d'inventaire, l'OFA et la consolidation de marchandises dans l'optimisation de la chaîne d'approvisionnement. Nos travaux visent à établir de nouveaux paradigmes à travers le prisme de la prise de décision basée sur les données, et à formuler des méthodologies rigoureuses grâce à l'utilisation des techniques de fouille de données, de recherche opérationnelle, et d'apprentissage par renforcement. Ces approches permettent d'extraire des informations actionnables à partir de volumes importants de données hétérogènes, facilitant ainsi une prise de décision optimisée dans des environnements opérationnels complexes.

La mise en œuvre de ces outils permettra aux entreprises de survivre et de prospérer sur des marchés compétitifs en améliorant l'efficacité de leur chaîne d'approvisionnement, tout en maintenant des niveaux de service élevés et en réalisant d'importantes économies de coûts.

Notre recherche s'appuie sur des données empiriques et des études de cas fournies par notre partenaire industriel, Logistik Unicorp, une entreprise textile du secteur de la distribution spécialisée dans la gestion d'uniformes via une plateforme de commerce électronique.

Chaque jour, Logistik reçoit des commandes de membres d'organisations clientes via une application web. La demande de chaque membre pour chaque article d'uniforme est enregistrée.

Cette collaboration industrielle garantit que nos méthodologies sont conçues pour répondre aux défis concrets des opérations de chaîne d'approvisionnement et peuvent être implémentées dans un contexte opérationnel réel.

La structure de cette thèse s'organise comme suit : le chapitre 2 présente une revue approfondie de la littérature, établissant le cadre théorique et contextualisant les problématiques abordées. Le chapitre 3 définit la problématique de recherche et le cadre méthodologique global. Les chapitres 4 à 6 exposent en détail les contributions scientifiques et méthodologiques de nos travaux. Le chapitre 7 propose une synthèse critique des résultats, tandis que le chapitre 8 présente les conclusions et perspectives futures.

CHAPITRE 2 REVUE DE LITTÉRATURE

2.1 Introduction

À la lumière de l'objectif global décrit dans le chapitre précédent, qui vise à développer un outil d'aide à la décision pour améliorer les opérations de la chaîne d'approvisionnement grâce à une approche basée sur les données, cette revue de littérature se penche sur les composants clés essentiels à ces opérations. Cette thèse, structurée sous forme d'articles, présente pour chaque contribution un état de l'art détaillé spécifique. Dans cette section, l'état de l'art se concentre principalement sur la problématique, afin de mieux contextualiser et situer le sujet dans son cadre scientifique.

La section 2.2 abordera en détail les subtilités de la consolidation de marchandises, en explorant ses définitions, sa classification, ainsi que les avantages et défis associés.

Ensuite, la section 2.3 s'attachera à l'aspect critique de l'allocation des items pour l'exécution des commandes, en élucidant son importance dans la détermination des moyens les plus efficaces pour satisfaire les commandes des clients en fonction de l'inventaire disponible et des considérations logistiques.

La section 2.4 déplacera ensuite l'accent sur le contrôle de l'inventaire, en fournissant un aperçu de son importance et des différentes techniques employées pour garantir une bonne disponibilité des produits au bon moment et en quantités correctes.

De plus, la section 2.5 introduira le concept de l'apprentissage par renforcement et son application potentielle dans le domaine de la logistique, offrant des perspectives sur la manière dont cette technologie émergente peut révolutionner les processus de prise de décision au sein de la gestion de la chaîne d'approvisionnement. Enfin, la section 2.6 aboutira à une conclusion critique, synthétisant les résultats de la revue de littérature et préparant le terrain pour les chapitres suivants.

2.2 Consolidation de marchandises

Dans la littérature, on trouve plusieurs définitions de la consolidation de marchandises, classées en trois catégories principales : la consolidation spatiale, la consolidation temporelle et la consolidation par produits. Des recherches telles que [10] et [11] décrivent la consolidation de marchandises comme le processus d'agrégation de multiples envois de petite taille pour former un envoi plus conséquent et plus économique. Cette stratégie peut être mise en œuvre

à différents niveaux de la chaîne d'approvisionnement, que ce soit en amont ou en aval, avec pour buts principaux la diminution des frais de transport [12], [13] et l'optimisation de la performance logistique [14]. La consolidation vise ainsi à une gestion optimisée des coûts de distribution, en alliant efficacité et efficience [14]-[18]. En regroupant et en combinant des produits à différents endroits et moments en un seul chargement ou tournée, la consolidation de marchandises permet de maximiser les bénéfices tout en minimisant les coûts associés au transport et à la gestion des stocks.

La littérature offre diverses taxinomies et stratégies de consolidation, dont la pertinence dépend du contexte d'application. Parmi celles-ci, la classification proposée par Brennan [9] est fréquemment reprise et se compose de :

La **consolidation temporelle**, qui consiste en l'agrégation de petites commandes au fil du temps afin d'équilibrer la qualité du service client, les coûts élevés d'inventaire et la réduction des coûts de livraison, est connue sous le nom de consolidation pure et a été décrite en détail par Çetinkaya et Lee [11]. La **consolidation spatiale** met l'accent sur le regroupement géographique des expéditions en tenant compte des points de départ et d'arrivée, afin de maximiser la quantité transportée par envoi. Enfin, la **consolidation par produits** vise à combiner plusieurs articles spécifiques afin d'augmenter les quantités livrées à chaque client. Chacune de ces stratégies est présentée plus en détails ci-dessous.

Higginson et Bookbinder [19] proposent différentes variantes de la **consolidation temporelle** qui se traduisent en diverses politiques. La politique basée sur le temps consiste à organiser l'envoi d'une cargaison consolidée à chaque période définie T, facilitant ainsi le traitement de l'ensemble des commandes en attente. La politique basée sur la quantité prévoit l'organisation d'un envoi consolidé dès qu'une quantité économique Q est atteinte. Quant à la consolidation hybride, elle représente la combinaison des deux approches et se manifeste naturellement dans la gestion quotidienne des opérations et la résolution de problèmes liés aux commandes urgentes. Elle est souvent intégrée dans les contrats dans le but d'assurer des livraisons rapides et une optimisation des chargements [19], [20]. Les exemples de l'industrie de l'ordinateur donnés par Çetinkaya [20] illustrent ces politiques selon le volume et la valeur des produits.

Il existe également une approche de consolidation intégrée, qui coordonne la gestion de l'inventaire avec la consolidation des expéditions pour équilibrer la livraison en temps voulu et les économies sur les coûts d'inventaire [20], le pilotage des niveaux de stock par les consommations communément appelé VMI est l'un des exemples les plus connues de l'approche intégrée.

La **consolidation spatiale** se réfère au processus de détermination des points de départ

et d'arrivée, de l'itinéraire, et du regroupement des petites commandes pour expédier en un grand envoi [21]. Cette stratégie vise à diminuer les coûts en combinant plusieurs petits envois au sein d'un même terminal de consolidation, analogue au problème de trouver le chemin le plus court de A à B. À l'opposé de Min [21], qui assimile la consolidation spatiale au problème de tournée de véhicules avec entrepôts multiples, Pooley et Stenger [14] soutiennent que la spécificité réside dans la façon dont le gestionnaire logistique orchestre l'agencement des envois individuels, mais aussi dans l'utilisation d'un aller simple. La combinaison sélective des expéditions peut réduire les coûts et améliorer le service. Les auteurs tels que Pooley et Stenger [14] et Min [21] s'accordent que le problème de consolidation spatiale inclut la tournée de véhicules à travers la consolidation de points plutôt qu'une tournée directe des dépôts aux clients.

Min [21] considère que la consolidation des transports fait partie intégrante de la consolidation spatiale. De manière plus large, la consolidation des transports et, de manière plus spécifique, la consolidation des expéditions sont perçues comme des stratégies cruciales pour améliorer l'efficacité logistique d'un système.

La consolidation des transports peut se manifester de différentes manières, comme le décrivent Pooley et Stenger [14] en s'appuyant sur la classification proposée par Sheffi [16] : La **consolidation indépendante** correspond au cas où l'expéditeur soumet de petits envois aux transporteurs qui évaluent chaque expédition de façon individuelle, tout en réalisant leurs propres regroupements pour une prestation économique. Dans la **consolidation de réseau**, l'expéditeur groupe de petites expéditions dans un seul véhicule jusqu'à un point de chargement, où elles sont ensuite acheminées à leur destination. Pour la **consolidation par tournées de véhicules**, l'expéditeur utilise une flotte de véhicules et sélectionne les itinéraires pour minimiser le nombre de véhicules et les kilomètres parcourus. Enfin, avec la **consolidation des expéditions**, l'expéditeur regroupe plusieurs petits envois en une charge complète, soumise à un transporteur de transport en lot complet qui effectue toutes les livraisons en un seul trajet.

La consolidation spatiale est fréquemment abordée sous l'angle de la conception de réseaux, de la planification du chargement et de l'optimisation des itinéraires des véhicules. Cette approche intègre divers facteurs, tels que l'emplacement des terminaux et des clients, ainsi que la séquence optimale de livraison des produits. L'inclusion de ces éléments ajoute à la complexité de sa mise en œuvre, mais est essentielle pour générer des avantages significatifs.

La littérature décrit plusieurs approches de la **consolidation de produits**, généralement définie comme l'agrégation de divers articles en une expédition unique afin d'augmenter le volume livré à chaque destinataire. Min [21] inclut dans cette définition la possibilité pour

plusieurs entrepôts de partager un canal de distribution commun, tout en répondant à des demandes variées en petites quantités. Hall [15] identifie, dans le cadre de la consolidation en terminal, le regroupement d'articles de provenances diverses en un point unique pour tri et réaffectation avant leur acheminement final. Svensson [22] décrit le *cross-docking* comme une variante, consistant à expédier immédiatement après réception et tri, afin de réduire les coûts d'inventaire et d'améliorer le service. D'autres travaux, tels que ceux de Bookbinder et Higginson [17], combinent la consolidation de produits avec la consolidation temporelle, prolongeant notamment le modèle stochastique de Gupta et Bagchi [13], qui suppose que la quantité de commandes accumulées suit une distribution gamma. Hanbazazah, Abril, Erkoc et al. [18] soulignent pour leur part la complexité de l'organisation des produits issus de sources multiples afin de garantir une livraison efficace à un client unique.

La consolidation de marchandises se distingue principalement par son potentiel de réduction significative des coûts associés aux inventaires et aux frais de transport. En consolidant les cargaisons, les sociétés limitent le besoin de maintenir de vastes stocks, générant ainsi d'importantes économies sur les coûts de stockage [17], [18]. Cette méthode engendre des économies d'échelle grâce à des tarifs avantageux pour les grandes expéditions, tout en encourageant une synergie entre les fournisseurs de services logistiques qui proposent une gestion intégrée des stocks et du transport [23], [24]. Par ailleurs, une coordination soignée entre les transporteurs améliore l'efficacité des itinéraires, réduisant de ce fait les temps de livraison par l'adoption de trajets plus directs et la limitation des manipulations [25]. L'optimisation des ressources et la rapidité du transport sont accentuées, grâce à une utilisation astucieuse des moyens de transport et des infrastructures de stockage, ce qui minimise les temps morts et maximise le rendement [26]. Une planification et une programmation rigoureuses exploitent au mieux la capacité disponible, éliminant le gaspillage et augmentant l'efficacité de la chaîne d'approvisionnement [25]. Ainsi, la consolidation de la marchandises se présente comme une stratégie à la fois économique et performante, cruciale pour stimuler les activités logistiques dans un marché concurrentiel. En outre, la consolidation émerge comme une alternative respectueuse de l'environnement, contribuant à la réduction de l'empreinte carbone grâce à la diminution des déplacements, ce qui entraîne une baisse de la consommation de carburant et des émissions [27], [28]. Elle aide également à alléger la congestion routière, avec moins de véhicules nécessaires au transport, rendant les routes et les zones urbaines moins saturées [29]. Ces bénéfices n'auraient pas émergé sans les modèles développés pour surmonter les défis logistiques complexes associés à la consolidation. Min et Cooper [30] proposent une analyse comparative qui classe ces modèles en deux grandes catégories : d'une part, une approche conceptuelle centrée sur la prise de décision, et d'autre part, un cadre mathématique empirique, élaboré à partir d'études de cas réels. Ces deux catégories se déclinent en trois

familles principales : (méta)heuristiques, stochastiques et optimisation. Malgré ces avancées, la consolidation reste confrontée à plusieurs enjeux majeurs : la complexité des décisions, qui doivent intégrer de multiples facteurs logistiques pour déterminer le moment et la méthode optimale de consolidation ; la difficulté algorithmique liée à la résolution de problèmes d'optimisation dynamiques et multi-objectifs ; et l'incertitude inhérente aux demandes ainsi que les perturbations possibles de la chaîne d'approvisionnement, qui introduisent une dimension stochastique supplémentaire. Ces éléments soulignent la nécessité d'adopter des approches innovantes pour exploiter pleinement le potentiel de la consolidation de marchandises.

2.3 Allocation des items pour l'exécution des commandes (OFA)

L'allocation des items pour l'exécution des commandes est un élément essentiel de l'efficacité opérationnelle des centres de distribution et des chaînes d'approvisionnement. Cette importance s'accroît avec la complexité croissante du commerce électronique et des réseaux de distribution mondiaux [31]. Ce processus d'allocation implique une sélection et une distribution stratégiques des commandes clients pour optimiser les opérations d'exécution, en tenant compte de variables telles que la disponibilité des stocks, la priorité des clients, la taille des commandes et les délais d'expédition [32]. Dans cette section, nous mettons en avant le rôle crucial de l'OFA au sein du processus de la chaîne d'approvisionnement, en le distinguant des étapes inter-connectées comme la planification de la préparation des commandes, la production sur commande (MTO), l'acceptation et la planification des commandes, ainsi que la gestion des commandes. À la différence de la planification de la préparation des commandes, qui vise une récupération efficace des produits [33], [34], et des phases axées sur la MTO ainsi que sur l'acceptation et la planification des commandes, qui déterminent quand et comment la production démarre selon les commandes clients [35], l'allocation des items pour l'exécution des commandes est déterminante dans la distribution stratégique des ressources pour le traitement des commandes.

Le principe fondamental de l'allocation des items pour l'exécution des commandes a été introduit par [36], qui a souligné l'importance de hiérarchiser les commandes clients pour répartir efficacement les stocks limités. Ces premiers travaux ont posé les bases des avancées ultérieures dans le domaine, soulignant la nécessité d'une approche systématique pour gérer les ruptures de stock, un défi courant dans les grands centres de distribution. Les ruptures de stock surviennent souvent lorsque la demande d'un article dépasse le stock disponible, nécessitant un équilibre entre le maintien de niveaux de stock adéquats et l'évitement de coûts de stockage prohibitifs. L'efficacité des stratégies d'exécution des commandes est souvent évaluée par le taux d'exécution des commandes (OFR), qui mesure la proportion de

commandes clients pouvant être entièrement satisfaites à partir du stock sans rencontrer de ruptures de stock [8], [37].

Dans le cadre de l'optimisation de l'OFA, les méthodologies traditionnelles telles que la règle du premier arrivé, premier servi (FCFS) ont été largement utilisées. Cette approche, qui consiste à honorer les commandes dans l'ordre où elles ont été reçues tant que le stock le permet, privilégie les commandes en fonction de leur date de création plutôt que de leur importance ou de leur urgence. Cependant, cette méthode ne tient pas compte des différents niveaux de priorité des commandes. Une autre stratégie consiste à attribuer un poids à chaque commande en fonction de facteurs tels que la date d'échéance et l'importance du client, assurant ainsi que les commandes prioritaires soient traitées en premier, contribuant à la livraison en temps voulu et à la satisfaction générale du client [38].

Meyr [38] explore les avantages de l'utilisation de la segmentation des clients et de la fonctionnalité "capable de promettre" (CTP) dans le processus d'exécution des commandes. Cette approche, particulièrement efficace dans les environnements à capacité limitée où la demande dépasse souvent les capacités de production, met en évidence le potentiel des stratégies centrées sur le client pour améliorer l'efficacité de l'exécution des commandes.

Seyedrezaei, Najafi, Aghajani et al. [39] ont développé les travaux de Rim et Park [8] en abordant des défis tels que la prévision de la demande des clients sur plusieurs périodes, les contraintes de capacité des entrepôts et la gestion des denrées périssables. Ils ont élaboré un modèle mathématique dynamique et un algorithme génétique (AG) pour améliorer le taux de remplissage des commandes (OFR), en tenant compte des besoins individuels des clients, de la demande probabiliste et de la dégradation des stocks.

Lečić-Cvetković, Atanasov, Babarogić et al. [40] ont introduit un système d'exécution des commandes "Make-to-Stock" pour une demande aléatoire et différentes catégories de clients, utilisant le disponible à la promesse (ATP) d'un plan directeur de production (PDP) pour allouer les stocks aux commandes et aux reliquats. Cette approche classe les clients selon les niveaux de service ciblés et intègre les reliquats pour améliorer le service global et respecter les accords contractuels.

Malgré ces contributions, Li, Li, Aneja et al. [41] ont noté le manque de modèles quantitatifs pour l'OFA. Ils ont proposé une méthode de recherche adaptative de grands voisinages pour résoudre simultanément les problèmes d'allocation et de routage des commandes dans un système d'exécution des commandes (OFS) à deux échelons. Contrairement aux méthodes d'allocation traditionnelles, cette approche alloue les stocks de plusieurs entrepôts aux commandes actives pour déterminer quel entrepôt devrait fournir chaque commande, en supposant que les stocks sont suffisants. Cette méthode offre de précieuses perspectives pour

optimiser l'OFA dans des systèmes de chaîne d'approvisionnement plus complexes.

2.4 Contrôle de l'inventaire

Qu'il s'agisse d'inventaire en stock ou en transit, son coût moyen aux États-Unis représente environ de 30 à 35 % de sa valeur. La réduction de ces stocks entraîne des économies considérables, améliorant directement la rentabilité de l'entreprise. Il est donc important d'utiliser des techniques adaptées pour une gestion efficace des stocks dans la chaîne logistique [42].

Le texte qui suit met en lumière, de manière concise, l'importance du contrôle de l'inventaire ainsi qu'un aperçu de ses techniques fondamentales.

Dans la gestion de la chaîne d'approvisionnement, le contrôle des stocks constitue un pilier essentiel pour concilier la satisfaction de la demande avec la minimisation des passifs liés au sur-stockage. Ross [43] met en évidence les défis inhérents à cette gestion, soulignant l'équilibre délicat entre l'utilité des stocks pour répondre à la demande en temps opportun et les coûts financiers liés à leur excédent. Plusieurs études [44], [45] insistent sur l'importance de stratégies de contrôle des stocks fondées sur les principes de gestion de la chaîne d'approvisionnement, et mettent en avant leur rôle déterminant dans l'optimisation des performances logistiques. Dans le prolongement de ces travaux, d'autres recherches [46] rappellent que le maintien d'une gestion efficace des stocks reste un facteur clé pour la performance globale des opérations commerciales. Chopra et Meindl [47] soulignent également l'importance d'une gestion précise des stocks.

De plus, l'équilibrage des niveaux de stock ne représente qu'un aspect du contrôle des stocks. La gestion des stocks englobe également la réduction des frais de stockage et de détention, ce qui influe directement sur la santé financière d'une entreprise [7]. Une gestion maîtrisée des stocks améliore par ailleurs la précision de l'exécution des commandes et la capacité de prévision de la demande, permettant aux organisations de s'adapter aux évolutions du marché [48].

Une gouvernance efficace des stocks est essentielle au succès d'une entreprise, à la satisfaction de sa clientèle et à son efficacité opérationnelle [42]. Dans un contexte de contrôle des stocks en constante évolution, les entreprises doivent adopter des approches stratégiques et fondées sur les données [49].

En matière de contrôle de l'inventaire, la capacité à répondre efficacement à la demande des clients, à réduire les coûts et à maximiser la rentabilité repose sur des décisions stratégiques relatives au réapprovisionnement et à l'optimisation des niveaux de stock. Diverses techniques de contrôle sont utilisées, chacune adaptée à des besoins spécifiques et à des modèles d'affaires

variés [7]. Parmi les facteurs déterminants figurent la prévision de la demande, les niveaux de stock de sécurité et la fiabilité des fournisseurs, qui influencent directement la performance du système de gestion des stocks [47].

Parmi les approches fondamentales de gestion des stocks se trouvent la Quantité de Commande Économique (Economic Order Quantity, EOQ), introduite en 1913 par Harris [50], qui vise à minimiser les coûts totaux de stock en optimisant la taille des commandes, en tenant compte de paramètres tels que la périssabilité et les remises. Des travaux récents [51] en ont examiné la flexibilité et l'application dans des secteurs allant de la vente au détail à la production, en soulignant son rôle dans l'amélioration de l'efficacité et la réduction des dépenses.

Une autre méthode classique toujours pertinente est le Juste-à-Temps (JIT), qui synchronise précisément l'approvisionnement des matières premières avec les besoins de production. Dans ce cadre, des recherches [52] ont abordé les défis liés à la détérioration des produits et à la demande incertaine, notamment dans des modèles de type marchand de journaux intégrant des livraisons multiples JIT. Ces travaux proposent une stratégie optimale de production et de réapprovisionnement combinant politiques de retour et de garantie, afin de réduire les déchets, améliorer l'efficacité et aligner la production sur la demande.

L'analyse ABC constitue une autre dimension du contrôle des stocks, offrant un cadre de classification permettant d'allouer stratégiquement les ressources aux articles critiques. Cette approche est particulièrement utile dans les secteurs à forte valeur ajoutée, tels que la pharmacie [53]. Dans le même esprit, l'application de cette méthode dans un contexte de restauration collective [54] a permis d'optimiser les niveaux de stock, de réduire les coûts de stockage et d'améliorer l'efficacité opérationnelle, aboutissant à un système de gestion des inventaires plus économiquement viable.

Plusieurs politiques et techniques de contrôle des stocks sont utilisées pour optimiser les niveaux d'inventaire, réduire les coûts et améliorer la qualité du service. La politique de commande jusqu'au niveau [55] réapprovisionne les stocks à un niveau prédéfini à intervalles réguliers, synchronisant le stock de sécurité avec les cycles de production pour générer des économies de coûts. La politique de point de commande [56] déclenche les réapprovisionnements dès qu'un seuil critique est atteint, assurant ainsi la continuité opérationnelle. L'inventaire de stock de sécurité [57] agit comme un tampon contre les incertitudes et garantit la disponibilité des produits lors des pics de demande, en particulier dans le commerce de détail.

Le système de révision continue [58] ajuste dynamiquement les stocks de sécurité en fonction de la demande en temps réel, réduisant ainsi les excédents et les coûts grâce à une planification stratégique. Le système de révision périodique [59] aligne les contrôles d'inventaire sur les

cycles de demande, optimisant à la fois les niveaux et les emplacements des stocks de sécurité. La politique de stock de base [60] utilise des approches intelligentes pour maintenir les stocks alignés sur les besoins du marché, renforçant l'agilité du système.

La prévision de la demande occupe également un rôle central. Des travaux [61] explorent des méthodes fondées sur les données, telles que la régression et la programmation linéaire, pour estimer les besoins en inventaire à partir de facteurs externes, soulignant l'importance stratégique d'une prévision précise.

Au-delà de ces approches spécifiques, une vision plus globale du contrôle des stocks est nécessaire. Une revue récente [46] analyse un large éventail de méthodes, des modèles analytiques aux approches non analytiques telles que la programmation dynamique et l'optimisation par simulation et met en lumière leurs atouts et leurs limites face à la complexité croissante des enjeux logistiques. Les auteurs insistent sur la nécessité d'approximation numériques efficaces pour traiter des problématiques multidimensionnelles. Dans le même esprit, Boute, Gijsbrechts, Van Jaarsveld et al. [62] soulignent que les hypothèses classiques sur la demande et les coûts sont souvent irréalistes, conduisant à des décisions sous-optimales.

2.5 Apprentissage par renforcement en logistique

L'apprentissage par renforcement (RL) incarne un paradigme transformateur dans le domaine de l'apprentissage automatique, un sous-domaine de l'intelligence artificielle, ancré dans le principe de l'apprentissage par interaction [63]. Cette méthode a trouvé une application solide dans divers domaines tels que la robotique, les jeux et, plus récemment, la logistique [64]. L'adoption de cette approche dans le domaine de la logistique a particulièrement suscité l'intérêt, en raison de son potentiel à résoudre des défis opérationnels complexes, visant à accroître la performance, la flexibilité et l'efficacité globale du système [65]. Cette section vise à explorer les fondements de l'apprentissage par renforcement et ses applications dans la logistique, en soulignant l'interaction dynamique entre la théorie et la pratique.

L'apprentissage par renforcement est un domaine captivant, fusionnant des perspectives issues de l'apprentissage automatique, de la psychologie et de la théorie de la décision. Au cœur du RL se trouve la prise de décision, avec un agent interagissant avec son environnement pour atteindre des objectifs spécifiques, en optimisant les récompenses totales sur une période donnée. Ce processus implique une séquence répétitive de réalisation d'actions, de réception de réponses, d'obtention de retours d'information et d'ajustement du comportement, saisissant les aspects fondamentaux de l'expérimentation, de l'essai et de l'erreur, et de l'acquisition de connaissances qui sont essentiels au RL [66].

Au cœur du RL se trouve le processus décisionnel de Markov (MDP), formulé mathématiquement par $MDP = (S, A, P_a, R_a)$, où S représente l'ensemble de tous les états possibles, A l'ensemble de toutes les actions possibles, $P_a(s, s') = Pr(s_{t+1} = s'|s_t = s, a_t = a)$ la probabilité de transition de l'état s à s' sous l'action a, et $R_a(s, s')$ la récompense reçue après être passé de s à s', étant donné l'action a. L'objectif de l'agent est d'identifier ou d'apprendre une politique $\pi : S \to A$ qui optimise le retour attendu, noté comme la somme des récompenses futures actualisées, $G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$, avec γ reflétant la préférence de l'agent pour les récompenses immédiates par rapport aux récompenses différées [67]. Comme le souligne [68], la particularité du RL réside dans sa faculté à élaborer des stratégies optimales par une interaction directe avec l'environnement, ce qui rend inutile l'élaboration préalable d'un modèle environnemental spécifique. Cette capacité s'avère particulièrement bénéfique dans des situations complexes ou changeantes où les approches traditionnelles peuvent s'avérer limitées. Au cœur du RL, l'enjeu majeur est de trouver le juste équilibre entre l'exploration, ou la recherche de nouvelles actions pour déterminer leurs résultats, avec l'exploitation, ou l'utilisation des connaissances acquises pour la prise de décision [69]. Maîtriser cet équilibre est essentiel pour améliorer le processus d'apprentissage dans le RL [66]. Dans cette lignée, la fusion de l'apprentissage par renforcement avec l'apprentissage profond a engendré l'apprentissage par renforcement profond (DRL). Cette approche tire parti des réseaux neuronaux profonds pour l'approximation des fonctions et la reconnaissance des schémas. Grâce à cette intégration, les agents peuvent traiter des données sensorielles de grande dimension et aborder des problèmes jusqu'alors irrésolus, signalant ainsi une avancée significative dans le domaine [68].

Le cadre de l'apprentissage par renforcement (RL), tel qu'illustré par Sutton et Barto [66], repose sur une modélisation mathématique de l'interaction entre un agent et son environnement, visant un apprentissage optimal par essais et erreurs.

L'agent est représenté par une politique, notée π, qui associe à chaque état s une distribution de probabilité sur les actions possibles a, soit $\pi(a|s)$. Cette fonction définit la stratégie décisionnelle de l'agent en réponse à son environnement. **L'environnement** est modélisé par un espace d'états S, un espace d'actions A, une fonction de récompense $R(s, a, s')$, ainsi qu'une dynamique de transition $P(s'|s, a)$ décrivant la probabilité de passer à l'état s' après avoir exécuté l'action a depuis l'état s.

Les états $s \in S$ représentent les différentes configurations possibles de l'environnement, servant de domaine aux fonctions de politique et de valeur. **Les actions** $a \in A$ sont les décisions que l'agent peut prendre dans chaque état, guidées par la politique $\pi(a|s)$. **Les récompenses**, fournies par la fonction $R(s, a, s')$, constituent un signal scalaire évaluant

l'efficacité immédiate d'une action à travers la transition d'un état à un autre.

La politique π définit la probabilité de choisir chaque action dans un état donné, jouant un rôle central dans le comportement de l'agent. **La fonction de valeur** $V^{\pi}(s)$ estime le retour escompté à partir de l'état s en suivant la politique π, offrant une mesure de la désirabilité de cet état. **La fonction de Q-valeur** $Q^{\pi}(s, a)$ évalue quant à elle le retour attendu en prenant l'action a dans l'état s, puis en poursuivant selon π, et elle est souvent utilisée pour raffiner la politique.

La dynamique de transition $P(s'|s, a)$ encode la nature stochastique de l'environnement, en indiquant la probabilité d'atteindre l'état s' après avoir pris l'action a dans s. **Le retour** G_t, défini comme $G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$, où γ est un facteur d'actualisation, représente la somme pondérée des récompenses futures à partir du temps t.

Un aspect fondamental de l'apprentissage est le **compromis entre exploration et exploitation**, incarné dans la politique π, qui doit équilibrer la découverte de nouvelles actions et l'exploitation des actions déjà identifiées comme bénéfiques. Le **processus d'apprentissage** consiste à mettre à jour itérativement la politique π ainsi que les fonctions V et Q à partir des expériences accumulées, à l'aide d'algorithmes comme l'apprentissage par différence temporelle (TD) ou le Q-learning. Enfin, la **fonction objectif** de l'apprentissage par renforcement est généralement la maximisation du retour espéré à long terme, en optimisant conjointement la politique et les fonctions de valeur associées.

Ce cadre structuré permet aux agents d'apprendre et d'adapter des comportements optimaux grâce à une interaction systématique avec leurs environnements.

Il existe différentes catégories d'algorithmes d'apprentissage par renforcement (RL). On y trouve ceux basés sur la valeur, ceux basés sur la politique, ceux en politique (On-policy), ceux hors politique (Off-policy) et ceux sans modèle (model-free), entre autres. Cependant, il convient de noter qu'un algorithme peut appartenir simultanément à plusieurs catégories [66].

Contrairement à l'apprentissage par renforcement Off-policy, illustré par les méthodes Q-learning et Deep Q-Networks (DQN), les agents peuvent apprendre à partir d'expériences passées collectées sous diverses politiques [70]. Cela améliore l'efficacité des données en réutilisant les ensembles de données existants et facilite le transfert de connaissances à travers différents domaines. Grâce à un tampon d'expériences passées maintenu pour des échantillonnages répétés, les agents peuvent continuer à interagir avec et à s'adapter à leur environnement sans écarter les interactions précédentes [71]. Dans les méthodes On-policy, telles que Actor-Critic et l'Optimisation de Politique Proximale (PPO), la politique est mise à jour

en fonction des trajectoires générées en adhérant à la politique actuelle [70]. Cette approche permet un apprentissage stable et efficace, particulièrement dans les scénarios où les compromis entre exploration et exploitation sont primordiaux, et vise à être atteint en utilisant des techniques en politique.

L'algorithme PPO, introduit par Schulman, Wolski, Dhariwal et al. [72], se distingue par sa stabilité et son efficacité dans l'optimisation des politiques. Il améliore le processus de mise à jour des politiques en introduisant un mécanisme d'écrêtage qui empêche les changements trop brusques, réduisant ainsi le risque de détériorations significatives de la performance pendant l'entraînement. Cette caractéristique est particulièrement avantageuse dans les environnements complexes et stochastiques tels que la gestion d'inventaire, où de légères modifications de politique peuvent avoir des répercussions importantes sur la performance du système.

Un autre algorithme model free et On-policy est l'apprentissage par imitation pondéré avec avantage monotone (MARWIL), spécifique aux tâches d'apprentissage par imitation. Il se concentre sur l'apprentissage à partir de démonstrations d'experts pour imiter le comportement souhaité [73]. Cependant, contrairement à PPO et Actor-Critic, qui apprennent directement de l'environnement, MARWIL est un algorithme hors ligne (Off-line), qui apprend à partir d'un ensemble fixe d'expériences (trajectoires collectées à partir d'interactions précédentes) sans interagir avec l'environnement pendant l'entraînement. L'algorithme vise à surmonter les limitations des approches traditionnelles d'apprentissage par imitation et à obtenir une meilleure généralisation et des performances dans des environnements difficiles [73]. Les modèles en ligne et hors ligne ont tous les deux leurs avantages et peuvent ainsi être exploités selon les besoins.

Pour améliorer la performance de certains algorithmes de RL selon leurs catégories, il existe des modèles de support tels que les modèles autorégressifs dynamiques. Ces modèles améliorent les modèles dynamiques traditionnels en générant chaque dimension de l'état suivant de manière séquentielle, capturant ainsi les dépendances entre les dimensions de l'état. Ils se caractérisent par leur capacité à modéliser avec précision des espaces d'état complexes [74]. Plus de détails sur RL seront fournis au chapitre 6.

Au cours du dernier siècle, l'intégration de la recherche opérationnelle et des ordinateurs a facilité le développement et l'application de modèles et d'algorithmes d'optimisation. Cela a fourni une approche scientifique et systématique pour résoudre les défis logistiques et de gestion de la chaîne d'approvisionnement (SCM) [75]. Avec les avancées en technologie informatique et en science des données, l'apprentissage par renforcement (RL) est devenu un outil populaire pour aborder les problèmes de prise de décision complexes, particulièrement ceux présentant de grands espaces d'états et de décisions ainsi que des incertitudes systémiques

[76]. En raison de la complexité des systèmes logistiques et de chaînes d'approvisionnement, le RL a gagné en popularité dans ce domaine, permettant une résolution de problèmes basée sur l'observation empirique et l'utilisation de données réelles.

Les recherches sur l'impact élargi du RL en logistique ont été approfondies grâce à des études fondamentales, comme celle de Murdivien et Um [77], qui ont appliqué le DRL pour résoudre le problème de l'empaquetage en 3D dans le secteur logistique. Cela a été réalisé en entraînant un agent au sein d'un environnement de moteur de jeu. Cette méthode a non seulement permis de visualiser le processus d'apprentissage, mais a également abouti à un modèle capable de relever efficacement les défis de l'empaquetage séquentiel en 3D en temps réel, révélant un potentiel considérable pour des applications logistiques complexes dans des environnements dynamiques. Ainsi, l'application du DRL a un impact significatif sur la logistique, en offrant des solutions à des problèmes complexes et à grande échelle, difficiles à résoudre pour les algorithmes traditionnels, et améliore par conséquent l'efficacité de tâches telles que le chargement et le déchargement.

En ce qui concerne le transport dans la chaîne d'approvisionnement, une approche basée sur le DRL a été développée pour aborder le problème complexe de la tournée de véhicules, à la fois dynamique et incertain. Elle s'appuie sur un modèle de réseau d'encodage-décodage graphique pour traiter des données variables et incertaines, comme le démontre Pan et Liu [78]. Cette méthode a optimisé les stratégies de tournée dans des contextes changeants.

Par ailleurs, des avancées notables ont été réalisées en intégrant l'apprentissage par renforcement multi-agents (MARL) à grande échelle, afin d'améliorer la collaboration entre les travailleurs robotiques et humains dans la logistique d'entrepôt. Cela a conduit à d'importantes améliorations en termes d'efficacité opérationnelle et de taux de prélèvement, comme l'indique [79]. Dans un domaine connexe, la fabrication en nuage, le DRL a été appliqué pour résoudre des problèmes complexes de composition de services, avec un modèle basé sur DRL utilisant un réseau Dueling Deep Q enrichi d'une relecture priorisée, proposé par Liang, Wen, Liu et al. [80]. Cette méthode vise à optimiser la composition des services en adaptant les décisions à l'environnement, afin de minimiser les coûts et maximiser l'efficacité, offrant une alternative robuste aux algorithmes méta-heuristiques traditionnels.

L'application du RL s'étend également à l'amélioration de l'efficacité de la préparation des commandes dans les entrepôts intelligents, une approche soutenue par Zhou, Lin et Cao [81]. Wang, Wang, Yang et al. [82] ont, quant à eux, introduit un cadre d'apprentissage par renforcement coopératif basé sur des groupes, ciblant les aspects dynamiques et spatiaux des tâches de livraison. Leur méthode, qui incorpore un mécanisme d'attention multi-niveaux et une stratégie basée sur des graphes, vise à améliorer l'efficacité et la coopération entre

les coursiers. Cela optimise à son tour l'efficacité de la livraison du dernier kilomètre et l'expérience client.

Dans le domaine spécifique de la gestion d'inventaire, l'apprentissage par renforcement offre des perspectives particulièrement prometteuses. Les approches traditionnelles de contrôle d'inventaire, comme les politiques de stock de base et les modèles de quantité économique de commande, reposent souvent sur des hypothèses simplificatrices concernant la distribution de la demande et les structures de coûts. Ces simplifications peuvent conduire à des décisions sous-optimales dans des environnements réels caractérisés par des demandes non-stationnaires, des contraintes de service multiples et des dynamiques complexes de la chaîne d'approvisionnement [83].

L'application du RL à la gestion d'inventaire permet de surmonter ces limitations en apprenant directement à partir des données et en s'adaptant aux complexités du monde réel. Plusieurs recherches récentes ont démontré la supériorité des approches basées sur le RL par rapport aux méthodes traditionnelles, particulièrement dans des contextes caractérisés par des dynamiques complexes. Parmi ceux-ci, on retrouve : **des demandes intermittentes ou hautement variables** [84], où les approches statistiques classiques peinent à capter la structure des séries temporelles ; **des accords de niveau de service (SLA) multiples et hiérarchisés** [85], qui requièrent une prise de décision différenciée selon les priorités client ; **des contraintes interdépendantes entre produits multiples** [86], qui introduisent des effets croisés difficiles à modéliser de façon analytique ; ainsi que **des délais d'approvisionnement dynamiques et incertains** [87], où l'adaptabilité des politiques apprises par RL s'avère particulièrement avantageuse.

L'intérêt du RL pour la gestion d'inventaire est d'autant plus marqué dans les environnements complexes caractérisés par des accords de niveau de service (SLA) hiérarchisés. Contrairement aux approches traditionnelles qui traitent souvent les objectifs de niveau de service de manière simplifiée, le RL permet de modéliser explicitement la nature asymétrique et non linéaire des implications de coûts associées aux excès d'inventaire et aux pénuries, particulièrement dans le contexte du respect des SLA. Cette capacité est cruciale pour les entreprises opérant avec des groupes de clients distincts, chacun ayant ses propres exigences contractuelles de service.

De plus, l'émergence de l'apprentissage par renforcement multi-agents (MARL) offre un cadre particulièrement adapté pour les systèmes d'inventaire complexes. Contrairement aux approches à agent unique qui centralisent la prise de décision, le MARL permet à plusieurs agents de contrôler différents aspects du système, décomposant ainsi naturellement le problème en sous-problèmes gérables [88]. Cette approche décentralisée améliore l'efficacité et la scalabilité des systèmes de gestion d'inventaire, permettant de traiter des problèmes de

grande dimension qui seraient autrement intractables.

Parmi les algorithmes de RL, l'Optimisation de Politique Proximale (PPO) s'est révélée particulièrement efficace pour les applications de gestion d'inventaire. Des études récentes ont démontré ses avantages par rapport à d'autres algorithmes comme A3C et DQN, notamment sa stabilité lors de l'entraînement, sa capacité à gérer des espaces d'action complexes, et sa convergence plus rapide vers des politiques optimisées [89], [90]. Ces caractéristiques font de PPO un choix privilégié pour développer des systèmes de contrôle d'inventaire adaptatifs capables de répondre efficacement aux fluctuations de demande, aux perturbations de la chaîne d'approvisionnement et aux autres éléments stochastiques inhérents aux systèmes d'inventaire.

Toutefois, les défis liés à la scalabilité des algorithmes de RL et aux ressources computationnelles conséquentes nécessaires pour le réglage des hyperparamètres des modèles de réseaux neuronaux constituent également des obstacles significatifs.

2.6 Synthèse

Cette revue de littérature a permis d'explorer les différentes facettes de la chaîne d'approvisionnement moderne, en mettant l'accent sur trois composantes fondamentales : la consolidation de marchandises, l'allocation des items pour l'exécution des commandes (OFA) et le contrôle d'inventaire. La littérature révèle que ces domaines, bien qu'étudiés séparément, présentent des interconnexions significatives et des opportunités d'optimisation intégrée encore peu explorées.

La consolidation de marchandises, qu'elle soit temporelle, spatiale ou par produits, offre des avantages considérables en termes de réduction des coûts de transport et d'amélioration de l'efficacité logistique. Cependant, sa mise en œuvre efficace reste complexe, particulièrement lorsque les demandes sont intermittentes ou que les contraintes opérationnelles sont multiples. Les modèles actuels, bien que sophistiqués, peinent souvent à capturer pleinement la dynamique stochastique et multi-objectif des environnements réels.

L'allocation des items pour l'exécution des commandes (OFA) constitue un maillon essentiel entre la gestion d'inventaire et la satisfaction client. Les approches traditionnelles comme le premier arrivé, premier servi (FCFS) se montrent insuffisantes face à la complexité des environnements modernes caractérisés par des priorités client variables et des contraintes de stock. Les modèles d'optimisation, notamment ceux intégrant des poids pour classer les niveaux de priorités des commandes en fonction de différents facteurs et utilisant la programmation linéaire en nombres entiers mixtes, offrent des perspectives prometteuses.

Le contrôle d'inventaire demeure un défi central dans les chaînes d'approvisionnement, où l'équilibre entre disponibilité des produits et minimisation des coûts de stockage est délicat à maintenir. Les techniques classiques comme la Quantité de Commande Économique (EOQ), le Juste-à-Temps (JIT) et les politiques de stock de base reposent sur des hypothèses souvent simplificatrices concernant la distribution de la demande et les structures de coûts, limitant leur efficacité dans des contextes réels caractérisés par l'incertitude et la complexité.

L'émergence de l'apprentissage par renforcement (RL) représente une avancée significative pour surmonter ces limitations. En permettant l'apprentissage de politiques optimales à travers l'interaction directe avec l'environnement, le RL offre une alternative aux approches analytiques traditionnelles qui nécessitent des modèles explicites et des hypothèses restrictives. Les algorithmes comme l'Optimisation de Politique Proximale (PPO) et l'apprentissage par renforcement multi-agents (MARL) se révèlent particulièrement adaptés aux défis de la gestion de chaîne d'approvisionnement moderne, offrant stabilité d'apprentissage, scalabilité et capacité d'adaptation aux environnements dynamiques.

La littérature récente suggère un potentiel considérable pour l'intégration de ces approches. Toutefois, plusieurs lacunes importantes subsistent :

1. Le manque de systèmes intégrant simultanément le contrôle d'inventaire et l'OFA de façon optimale
2. L'absence de modèles prenant adéquatement en compte les environnements multi-SLA où différents groupes de clients ont des exigences distinctes de niveau de service, avec une compréhension sophistiquée des impacts asymétriques et non linéaires du non-respect des niveaux de service contractuels
3. La rareté des modèles adaptés aux demandes intermittentes et non-stationnaires caractéristiques de nombreux environnements d'affaires
4. Le besoin d'architectures modulaires permettant l'application pratique dans des contextes organisationnels variés

Ces lacunes représentent des opportunités de recherche significatives que cette thèse vise à adresser. En particulier, le développement d'un système de contrôle d'inventaire basé sur l'apprentissage par renforcement multi-agents qui intègre des prévisions par quantiles et la programmation linéaire en nombres entiers mixtes constitue une approche novatrice pour répondre aux défis des environnements de distribution caractérisés par des SKUs multiples aux profils de demande interdépendants, des accords de niveau de service hiérarchisés avec des implications de coûts asymétriques, et des délais d'approvisionnement dynamiques.

Les contributions proposées dans cette thèse s'inscrivent donc dans une perspective d'innovation à l'intersection de la recherche opérationnelle, de l'intelligence artificielle et de la gestion

de la chaîne d'approvisionnement, avec pour objectif de développer des outils d'aide à la décision adaptés aux réalités complexes et dynamiques des entreprises modernes.

CHAPITRE 3 DÉMARCHE ET ORGANISATION

3.1 Introduction

Dans un environnement économique où les marchés évoluent rapidement et les attentes des consommateurs se transforment continuellement, la chaîne d'approvisionnement constitue un élément stratégique déterminant pour le succès des entreprises. Les avancées technologiques, particulièrement dans le domaine de l'intelligence artificielle et de l'analyse de données, offrent aux organisations des opportunités inédites pour transformer leurs opérations logistiques. L'exploitation judicieuse des vastes quantités de données maintenant disponibles permet d'améliorer la prise de décision et d'optimiser les différents maillons de la chaîne d'approvisionnement.

Notre recherche s'inscrit dans ce contexte en examinant comment concevoir des outils décisionnels visant à améliorer l'efficacité opérationnelle de la chaîne d'approvisionnement dans un contexte de modernisation technologique. Cet objectif répond directement aux défis contemporains de la gestion logistique tout en exploitant les nouvelles possibilités offertes par les technologies avancées.

3.2 Objectifs de la recherche

Notre étude s'inscrit dans un contexte où les entreprises doivent gérer efficacement leurs chaînes d'approvisionnement face à des défis multiples et où la data-driven decision making représente un avantage compétitif majeur. Notre travail s'articule autour de trois piliers principaux : la consolidation des marchandises, l'allocation des articles pour l'exécution des commandes et le contrôle d'inventaires. Ces aspects, introduits dans les chapitres précédents, définissent le cadre de notre recherche, exposant les outils disponibles et les limites actuelles des connaissances dans ce domaine.

Face aux défis identifiés dans la littérature et à la question de recherche formulée, notre recherche poursuit un objectif général et trois objectifs spécifiques.

L'objectif général de cette thèse est de concevoir des outils décisionnels visant à améliorer l'efficacité opérationnelle de la chaîne d'approvisionnement dans un contexte de modernisation technologique.

Pour atteindre l'objectif général, nous avons défini trois objectifs spécifiques qui correspondent aux trois piliers de notre recherche.

Objectif spécifique 1 : Développer une méthodologie permettant d'identifier et d'exploiter les opportunités de consolidation géographique des expéditions à partir de l'analyse des comportements de commande.

- *OS1.1 — Sous-objectif 1.1 :* Extraire les motifs récurrents dans les historiques transactionnels.
- *OS1.2 — Sous-objectif 1.2 :* Traduire ces motifs en règles opérationnelles de consolidation exploitables par les gestionnaires logistiques.

Objectif spécifique 2 : Concevoir un modèle d'allocation optimale des commandes aux centres de distribution intégrant des objectifs multiples et des contraintes opérationnelles.

- *OS2.1 — Sous-objectif 2.1 :* Formaliser mathématiquement le problème décisionnel en capturant les équilibres entre efficacité logistique et qualité de service.
- *OS2.2 — Sous-objectif 2.2 :* Développer une approche de résolution permettant d'explorer le front de Pareto des solutions non dominées.

Objectif spécifique 3 : Élaborer un système de contrôle des stocks capable d'adapter dynamiquement ses politiques de réapprovisionnement en fonction de l'évolution de l'environnement.

- *OS3.1 — Sous-objectif 3.1 :* Concevoir un mécanisme d'apprentissage à partir des données opérationnelles permettant d'optimiser les décisions de réapprovisionnement.
- *OS3.2 — Sous-objectif 3.2 :* Développer une architecture capable d'intégrer simultanément les contraintes de capacité, les incertitudes de demande et les variations de délais.

Ces objectifs spécifiques structurent notre démarche de recherche et déterminent les contributions qui seront présentées dans les chapitres suivants. La méthodologie employée pour atteindre ces objectifs sera présentée dans la section 3.4, après une description du contexte industriel qui a informé notre recherche (section 3.3).

3.3 Contexte industriel

Le contexte industriel de cette thèse se concentre sur Logistik Unicorp, une entreprise spécialisée dans la gestion intégrale des programmes d'uniformes pour diverses organisations commerciales et agences gouvernementales, tant au niveau national qu'international. Logistik assure l'ensemble du processus, de la conception à la distribution des uniformes, en intégrant efficacement tous les aspects de la chaîne d'approvisionnement, incluant la conception, l'évaluation, la production, le contrôle qualité et la mise en place de plateformes web transactionnelles personnalisées pour ses clients.

La chaîne d'approvisionnement considérée présente une structure complexe typique des environnements de distribution multi-clients et multi-produits, illustrée dans la figure 3.1.

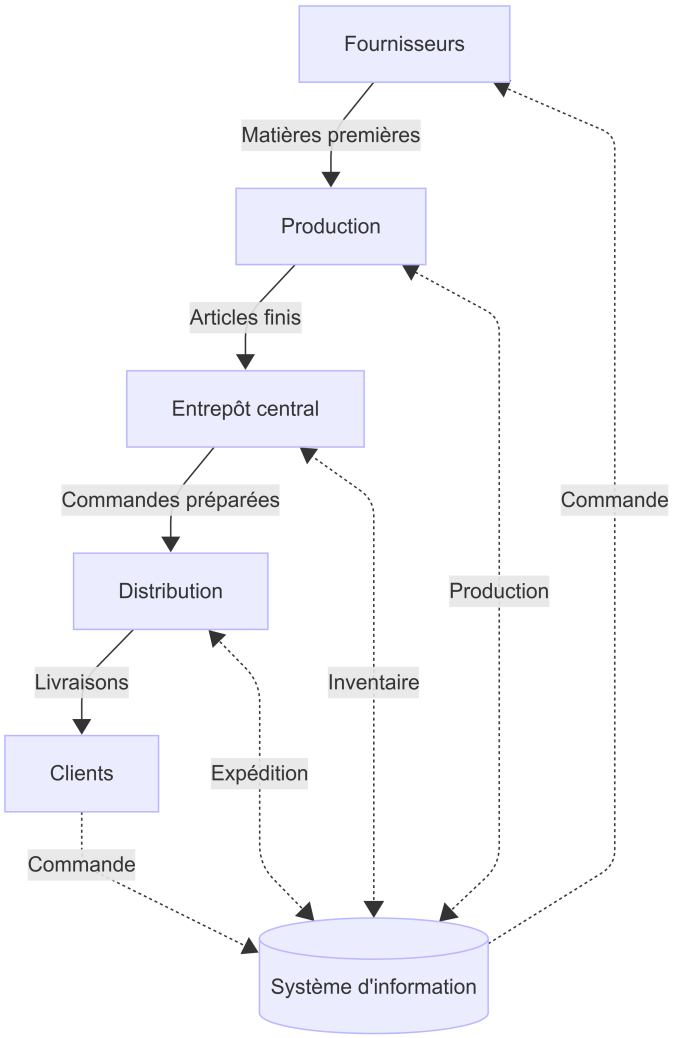

FIGURE 3.1 Représentation schématique de la chaîne d'approvisionnement considérée

Cette structure illustre les principaux flux de matières et d'informations dans la chaîne d'approvisionnement.

Les **fournisseurs** constituent le point de départ, représentant les multiples sources d'approvisionnement pour les matières premières (tissus, boutons, etc.) avec des délais variables. Ces matières premières sont ensuite acheminés vers la **production**, où s'effectue la transformation des matières premières en articles d'uniformes finis, suivant les spécifications des clients. Les produits finis sont alors dirigés vers l'**entrepôt central** qui assure le stockage et la gestion des inventaires pour des milliers d'items différents, organisés par clients et programmes. Le processus se poursuit avec la **distribution**, qui englobe la préparation des commandes,

l'allocation des articles et l'expédition vers les clients finaux. À l'extrémité de cette chaîne se trouvent les **clients**, organisations et leurs membres individuels qui commandent via la plateforme web. L'ensemble de ce flux est coordonné par le **système d'information** qui centralise les données et assure une visibilité continue sur l'ensemble des opérations de la chaîne d'approvisionnement.

Afin de satisfaire les demandes dans les délais et respecter les accords de niveau de service (SLAs), un stock adéquat est maintenu. Les demandes émanent directement des membres des organisations clientes via une application web, offrant ainsi une visibilité immédiate et précise sur les besoins réels, ce qui est crucial pour l'organisation et la planification des opérations.

Le système de contrôle des stocks est crucial à cause des coûts importants liés au stockage et de l'impact sur la réputation et les obligations contractuelles de l'entreprise. Le respect des SLAs est fondamental pour la satisfaction et la fidélité des clients. La stratégie de stock vise à garantir une disponibilité constante d'un certain pourcentage de la demande.

Pour prévenir les pertes de ventes dues aux ruptures de stock, l'entreprise adopte un système de commandes en attente, nécessitant un réapprovisionnement proactif, surtout que les délais des fournisseurs excèdent souvent ceux promis aux clients.

La gestion d'un large inventaire destiné à de nombreux articles et clients, chacun ayant ses propres priorités, délais et exigences en termes de niveau de service, représente un défi majeur. L'environnement opérationnel est marqué par des modèles de demande intermittents ; les requêtes pour les articles d'uniformes varient, incluant des commandes ponctuelles importantes, des motifs saisonniers et des tendances, ce qui complexifie la planification de l'inventaire et la gestion des stocks. Cette forte incertitude de la demande peut engendrer des inefficacités coûteuses en termes de gestion de stock. De plus, cette variabilité remet en question les hypothèses statistiques des modèles de contrôle des stocks traditionnels, conduisant à des performances sous-optimales.

3.4 Méthodologie

Cette thèse adopte une approche méthodologique intégrée et progressive pour développer des solutions d'optimisation de la chaîne d'approvisionnement. Notre démarche constitue un ensemble cohérent où chaque contribution s'appuie sur les précédentes et enrichit la suivante, formant ainsi un ensemble de solutions complémentaires.

Notre approche se structure en trois phases principales, chacune correspondant à un objectif spécifique et aboutissant à une contribution scientifique :

Cette progression méthodologique reflète une complexification graduelle des problèmes abor-

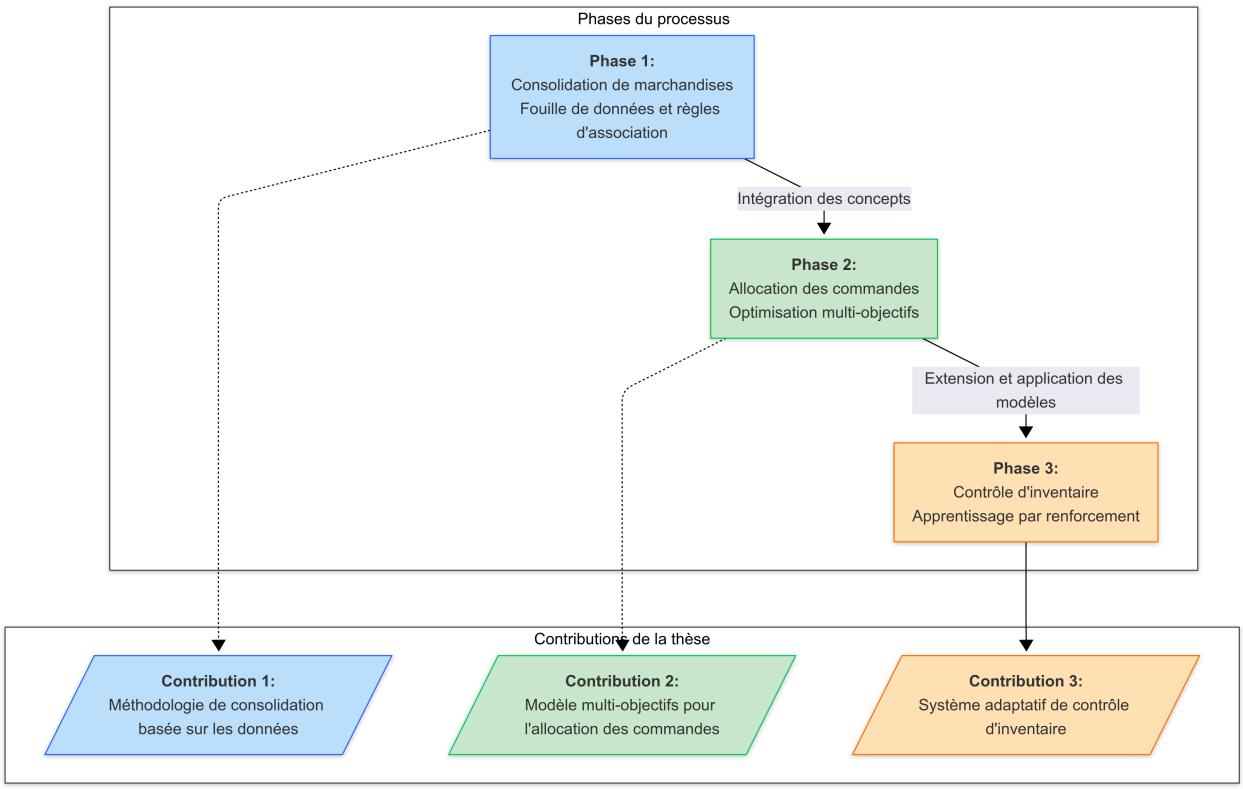

FIGURE 3.2 Progression de la méthodologie de recherche

dés. La **Phase 1** (Chapitre 4) consiste au développement d'une approche basée sur la fouille de données pour la consolidation de marchandises, créant une base pour l'optimisation spatiale. Dans la **Phase 2** (Chapitre 5), nous procédons à l'intégration des concepts et de la méthode de consolidation spatiale dans un cadre d'allocation des commandes multi-objectifs, établissant le lien entre consolidation et satisfaction client. Finalement, avec la **Phase 3** (Chapitre 6), nous étendons notre approche vers un système plus complet de contrôle d'inventaire qui incorpore à la fois l'allocation des commandes et la gestion des réapprovisionnements dans un contexte de demande complexe.

Chaque phase construit sur les connaissances et les outils développés dans les phases précédentes, formant ainsi une progression logique vers un système intégré d'optimisation de la chaîne d'approvisionnement (voir 3.2).

La méthodologie adoptée pour chacune des trois phases s'inspire du processus Cross-Industry Standard Process for Data Mining (CRISP-DM), adapté à la nature spécifique de nos travaux de recherche. Cette approche itérative, illustré dans la figure 3.3, comprend six étapes essentielles pour guider les projets de fouille de données et d'intelligence artificielle [91].

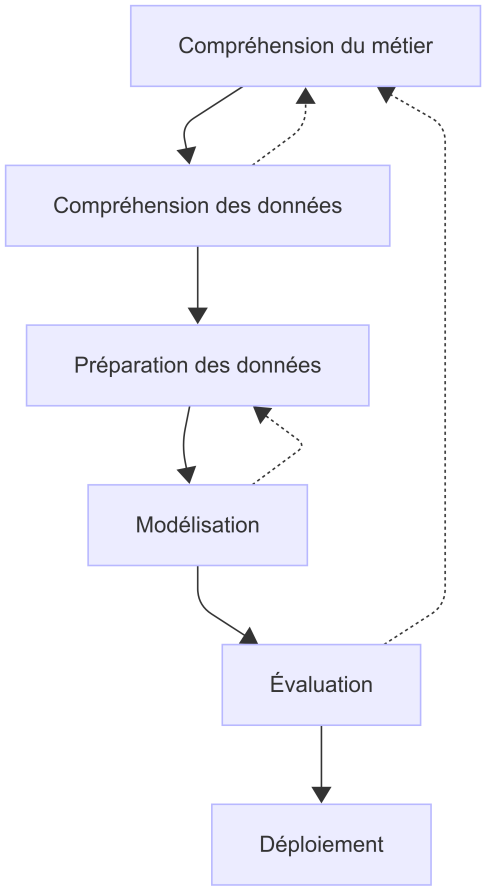

FIGURE 3.3 Méthodologie CRISP-DM

Pour chacune de nos contributions, nous avons adapté ce cadre méthodologique aux spécificités des problèmes abordés. La **compréhension du métier** constitue la première phase, où nous nous immergeons dans le contexte métier du cas d'étude et effectuons une revue de l'état de l'art pour comprendre précisément les exigences, les besoins et les avancées scientifiques pertinentes. L'objectif est de traduire la vision et les objectifs métier en une problématique de recherche claire, qui intègre la fouille de données, et de définir un plan d'action préliminaire. S'ensuit la **compréhension des données**, étape où nous entreprenons une exploration détaillée des données disponibles. Cela commence par la collecte des jeux de données nécessaires (historique des demandes clients, retours clients, contraintes logistiques, niveaux de stock) et se poursuit par une analyse approfondie pour identifier leurs caractéristiques, leurs limites et leur potentiel. Tout au long de ce processus, nous formulons des hypothèses sur les informations cachées qu'elles pourraient révéler. La **préparation des données** établit ensuite les fondations de notre analyse en sélectionnant les données pertinentes, les transformant et les nettoyant, afin de constituer un ensemble de données finalisé et prêt pour les étapes suivantes.

Après cette préparation, **la modélisation** nous permet de développer une méthodologie sur mesure qui intègre les techniques et les algorithmes les mieux adaptés à notre cas d'étude. Cette étape est de nature itérative et peut nécessiter un retour aux phases antérieures pour peaufiner notre approche. Avant la mise en production, l'**évaluation** s'avère cruciale pour valider la qualité et la pertinence des solutions obtenues à l'aide de données réelles et de simulations, afin de nous assurer que les modèles développés répondent efficacement aux objectifs métier et à l'objectif de la recherche. Cette évaluation comparative quantifie l'efficacité relative de nos approches par rapport à des méthodes de référence.

Finalement, le **déploiement** vise à intégrer les modèles développés dans les processus opérationnels de l'organisation, assurant ainsi leur utilité pratique dans le contexte métier spécifié.

Dans notre troisième contribution, qui mobilise l'apprentissage par renforcement, nous avons particulièrement adapté ce cadre pour intégrer des éléments spécifiques comme la simulation d'environnements, l'entraînement d'agents, et l'équilibre entre exploration et exploitation, tout en conservant la structure itérative fondamentale de la méthodologie.

Précision terminologique : Dans cette thèse, nous distinguons entre la validation de modèles (démontrer qu'un modèle de simulation représente fidèlement la réalité) et l'évaluation de performance (mesurer et comparer l'efficacité de solutions proposées). Nos contributions relèvent principalement de la seconde catégorie : nous évaluons quantitativement nos approches décisionnelles en les comparant à des alternatives sur des environnements de simulation ou des datasets contrôlés. Bien que le terme « validation expérimentale » apparaisse occasionnellement dans nos travaux par convention linguistique, il convient de l'interpréter au sens d'évaluation comparative de performances.

3.5 Conclusion

Ce chapitre a présenté l'approche méthodologique adoptée dans cette thèse pour développer des solutions d'optimisation de la chaîne d'approvisionnement. Nous avons détaillé les objectifs de recherche, le contexte industriel qui a motivé notre étude, ainsi que le cadre méthodologique qui structure notre démarche.

Les trois chapitres suivants exposent la mise en œuvre de cette approche à travers trois contributions complémentaires, formant une progression logique cohérente de complexité croissante.

Le chapitre 4 présente notre première contribution, qui propose une approche de consolidation exploitant les règles d'association pour analyser les comportements des clients à travers des motifs de données, dans le cadre de demandes intermittentes, avec pour objectif de diminuer les coûts de transport.

Le chapitre 5 expose la deuxième contribution, qui intègre la consolidation spatiale à l'allocation des articles pour l'exécution des commandes (OFA), grâce à une approche multi-objectifs. Celle-ci cherche à réduire simultanément les coûts d'expédition tout en augmentant le niveau de service client.

Le chapitre 6 présente la troisième contribution, qui développe un système de contrôle d'inventaire fondé sur l'apprentissage par renforcement. Cette approche intègre l'OFA dans un cadre plus large tout en démontrant que, bien que l'optimisation de l'allocation des commandes soit importante, ce sont les décisions de réapprovisionnement qui exercent l'influence la plus déterminante sur les performances globales de la chaîne d'approvisionnement.

Cette progression reflète une complexification méthodologique : nous passons d'une analyse focalisée sur l'optimisation du transport (fouille de données) à une intégration tactique transport-allocation (programmation mathématique), pour aboutir à un pilotage stratégique et dynamique des stocks (apprentissage par renforcement).

Chaque chapitre présentera non seulement la modélisation retenue, et aussi une analyse comparative rigoureuse de sa performance, démontrant la valeur ajoutée des approches basées sur les données par rapport aux pratiques conventionnelles.

CHAPITRE 4 ARTICLE 1: IMPROVEMENT OF FREIGHT CONSOLIDATION THROUGH A DATA MINING BASED METHODOLOGY

Zineb Aboutalib, Bruno Agard

Published in International Journal of Logistics Systems and Management

Submitted: 03-Feb-2022, Accepted: 05-Mar-2022, Published: 01-Oct-2024

Abstract *— Freight consolidation is a complex logistics practice supported by a broad spectrum of strategies and methods to improve supply chain cost-effectiveness. It consists of grouping products in a single batch to reduce distribution costs. Literature review revealed that Operational Research (OR) is typically used for freight consolidation, and their inputs are often aggregated over time. While necessary to accommodate computationally expensive OR algorithms, such data simplifications are responsible for losing valuable data patterns. Our contribution is a novel data mining methodology that uses Association Rules to leverage data patterns in the context of intermittent demand. Our approach is compared to a typical Operational Research approach from a literature case study. Simple to implement, our methodology gives good results and can flexibly accommodate and exploit data patterns while being able to scale to a much larger amount of data, making it a more suitable approach for the big data world.*

Keywords *— Transportation, Association rules, Cost reduction, Data mining, Freight consolidation, Intermittent data, Decision making, Logistics, Big data, Pattern extraction*

4.1 Introduction

Companies that face intense competition, open markets, and rapid evolution of information technologies have become more interested in improving their management of physical and information flows to maintain a satisfying level of service to their customers [92]. A quest for efficiency and gain maximisation is taking place to please and satisfy customers. Shorter planning cycles and accelerated deliveries at a lower cost are expected from companies to remain competitive. A trade-off between reactivity and delivery costs needs to be found. For that, a feasible solution to the problem can be identified as Freight consolidation. Freight

Published in International Journal of Logistics Systems and Management; Aboutalib, Zineb; Agard, Bruno. vol. 49, no. 2, pp. 255–273, 2024.

consolidation is an old concept that still arouses the interest of many researchers, leading to better supply chain management. For this reason, many operational research models were tailored and used to implement and improve this concept while maintaining an enhanced supply chain [93]. However, we sometimes forget that companies are inundated by an impressive amount of data, often very little of which is used. This can be an obstacle for some operational research models; in this case, Operation research (OR) models will experience difficulties solving problems because of many variables, computational time, and a massive amount of available data [94]. Not to mention that data can express a demand trend that is sometimes uncertain, unknown, and intermittent, thus complicating the resolution of problems and making chain management quite tricky. So, to get around these challenges, some assumptions are made by OR models to simplify the data and fit the model, but by doing so, valuable information may be lost. The literature for solving the freight consolidation problem has mainly used operational research. So the positioning of this article aims to propose a novel freight consolidation methodology that is data mining driven to help companies enhance the efficiency of their processes and allow them to take advantage of the massive amounts of data stored. The contribution of this paper is to propose a new methodology to consolidate and reduce costs for freight, based on a data mining approach in an intermittent demand context.

The remainder of this paper is organised as follows. Section 4.2 is a brief review of freight consolidation. The proposed methodology to approach freight consolidation is presented in section 4.3. The details of the experiments conducted are represented in section 4.4 from the previous literature [95], [96] and the results of our comparison are in section 4.5. This is followed by a conclusion in section 4.6.

4.2 State of the art

Freight consolidation is a common concept used to optimise a company's logistic parameters in order to reduce costs and to maintain a great service level. Many reviews of freight consolidation can be found in the literature. This section will detail an overview of its definitions, objectives and tools that will facilitate an understanding of our experiment. Freight consolidation is an old concept that is still relevant today and can be applied to all modes of transportation. As a result, over the years, different definitions of freight consolidation have emerged.

Freight consolidation was represented by Tyan, Wang, and Du [10] and Çetinkaya and Lee [11] as the joining of many small shipments. A more significant and more economical load of goods could be grouped in the same vehicle. Traditionally, classified as temporal, spatial, or

product-based [9], the consolidation objectives benefit from a price reduction associated with larger loading sizes, typical to the railways, road haulers, and airlines [12], [15]. Consolidation allows distribution costs to be managed efficiently and effectively [15]–[18], to increase earnings and reduce transportation and inventory losses by grouping and combining products at different locations and at other times in a single load or tour [13], [14], [17], [18]. In short, a compromise between the benefits and the losses of using consolidations should be examined [15]. The literature provides different classifications of freight consolidation problems that offer solutions to various contexts.

4.2.1 Freight consolidation classification and strategies

The literature presents several classifications concerning freight consolidation. Brennan's categorisation [9] considers three main groups: spatial, temporal, and product freight consolidation strategies.

4.2.1.1 Temporal consolidation.

Temporal consolidation aggregates small orders over time to ponder inventory costs against delivery costs while maintaining good customer service [15]. This could be called a pure consolidation policy when no coordination is needed [20]. According to Higginson and Bookbinder [19], there are two types of temporal consolidation policies. The application cases of every policy were given by Çetinkaya [20] using a computer industry example:

— A time-based policy consists of shipping a consolidated load every period T. It allows all pending orders to be released. According to Çetinkaya [20], it is for lower volume and higher value items.

— A quantity-based policy sends a consolidated load when an economic quantity of a shipment is available. Following the example of Çetinkaya [20], it is for higher volumes and lower value.

When both of the policies above are joined, we refer to a hybrid policy that emerges implicitly in the management of day-to-day operations for fast delivery and optimised loading [20].

4.2.1.2 Spatial consolidation.

Spatial consolidation defines the clustering of small orders that are shipped in one large shipment, the starting and arriving point, and the route [21]. Spatial consolidation studies converge on ascertaining the minimum cost correlated with coupling small loads using a consolidation terminal [21], [97], [98]. Min [21] estimates that spatial consolidation is similar

to that of a multi-warehouse vehicle tour problem. Pooley and Stenger [14] on the other hand, clarifies that a consolidation problem is different from a vehicle routing problem. First, consolidation can be selective and does not need to combine all products, lower costs, or enhance service. Second, vehicles are not necessarily making round trips, particularly when for-hire carriers are used.
Through a different perspective, Min [21] meets Pooley and Stenger [14] in explaining that the spatial consolidation problem involves vehicle touring using a consolidation point instead of a direct tour from deposits to the customer. In short, spatial consolidation takes into account several parameters and applying it enables a potentially significant profit.

4.2.1.3 Product consolidation.

According to Min [21], product consolidation groups different types of products into a single shipment. Gurnani [99] refers to it as different buyers sharing an agreement to combine one order with a simple supplier, to benefit from a quantity discount. In the same vein, product consolidation can also recognise that distribution centers might share the same distribution channel and that each customer could request different products in small quantities [9]. When a customer buys other products, consolidation may increase the quantity delivered to each customer by delivery and decrease the delivery costs, resulting in significant gains [99], [100].

4.2.2 Resolution models for different strategies.

Consolidation studies have employed many resolution models to provide technical solutions to logistics models. In the comparative analysis of different studies on consolidation and returns management, Min and Cooper [30] developed a table classifying the studies according to the consolidation strategy and the different categories of models chosen to address the issues. In the same study [30], the pattern categorisation of how the strategies mentioned above could be addressed were simplified into a conceptual model based on a logical flow for decision-making and a mathematical model based on real-world case studies. These two models are individually classified into three categories: Heuristic, Stochastic, and Optimisation. The idea that comes up most frequently in the articles is finding a quantity needed to group products that will be consolidated together and finding a temporal window for when customers will receive the orders.

— (meta)Heuristic: These are some technical solutions used to solve consolidation issues. As an extension of the study of Çetinkaya and Lee [11], Moon, Cha, and Lee [101] have developed four heuristic algorithms QSP-H, SP-H, SP-RAND, and QSP RAND,

focusing on the joint replenishment and consolidated freight delivery scheduling of a warehouse in a supply chain. In the same direction as using heuristics, Khouja, Michalewicz, and Satoskar [102] and Moon and Cha [103] have performed Genetic Algorithms (GA) for the joint replenishment problem. To address freight consolidation and containerisation, Hajiaghaei-Keshteli, J Afshari, and Nasiri [104] have used a hybridised metaheuristic, such as the Red Deer Algorithm (RDA) and Hybridised RDA & GA (RDGA). Qin, Zhang, Qi, *et al.* [105] addresses the same problem while using a memetic algorithm approach to provide a practical solution and to minimise costs.

— Stochastic: the literature provides many examples of using stochastic models to address freight consolidation problems. The stochastic clearing model is one of them. It is applied to solve consolidation problems in a just-in-time environment [106]. For example, Gupta and Bagchi [13] constructed a compensation model to calculate the minimum economic quantity at the consolidation center. Bookbinder and Higginson [17] used the stochastic clearing theory of Stidham Jr [106] to obtain a policy based on time and quantity. For the integrated inventory replenishment and delivery problem, with the general compound Poisson demand, Nguyen, Dessouky, and Toriello [23] present a stochastic model and approximation method.

— Optimisation: the literature presents a variety of solutions that use optimisation to solve the problem of freight consolidation [30], such as [107], which evaluates an inventory control policy at the distribution center and provides an accurate optimisation technique for identifying optimal lot sizes and shipment consolidation cycle times. Hanbazazah, Abril, Erkoc, *et al.* [18] proposed an optimisation model for a transportation problem with freight consolidation to obtain economies of scale, in addition to a three-stage solution algorithm to address the large size problems. The optimisation models are proposed so that cost can be minimised while considering capacity and time constraints. Results vary depending on the costs, the computational times, and whether the company implemented the solution.

Freight consolidation depends on the distribution of demand and its variation over time. The following subsection depicts how the relationship between freight consolidation and demand affects shipments.

4.2.3 Interdependency of consolidation and demand.

Knowing the demand pattern of costumers and having good visibility on inventory with accurate forecasts enables many companies to stay efficient and competitive [107]. Under a dynamic demand distribution with variations between periods, Nguyen, Dessouky, and Toriello [23] studied the effect of changing the demand distribution on the profits of consolidation for perishable products. They solved the optimisation problem using a dynamic stochastic programming approach. They noticed that, over time, consolidation of demand can generate enough volume to be economically profitable at the cheapest FTL rate. They show that the profits of consolidation vary, depending on the size of the demand. Consolidation works best when demand levels are moderate. In the same environment of stochastic demand as [23], Zikopoulos [108] presents a decision-making process that can achieve savings on transportation costs.

4.2.4 Synthesis

Freight consolidation is old, but is still a highly topical concept [10]. Over the years, the research literature has shown different definitions of freight consolidation, strategies, and classifications depending on the objectives pursued. The same is true for the operational research methods presented to implement it. To the best of our knowledge, very little data mining, such as [109], and machine learning, such as [110], techniques have been used to address freight consolidation. As the present context evolves to deal with massive data, OR models can not support it, or can only support it with difficulty and with significant computational time and higher costs. An opportunity lay the focus on implementing an Association Rule [AR] freight consolidation methodology and apply it to a mathematical model has arisen. This paper proposes such a methodology. To evaluate the methodology, it is compared to [95], [96], which aimed to solve the consolidation with an OR approach while using their data. The contribution of this research is thus to develop a data-driven approach that brings a new framework to freight consolidation and the supply chain.

4.3 Methodology

This section presents a new methodology for addressing freight consolidation. Our methodology is a data-driven approach that integrates customer data to extract useful information and patterns for consolidation to provide insights for a manager's decision-making. Our approach is different from the classical approaches because it applies freight consolidation concepts while using Association Rules (AR). Association Rules is a data mining technique

used for descriptions. The objective of AR is to discover patterns or co-occurrences in a set of data [111] It has been widely proposed for different areas such as market basket analysis for commercial planning, Medical diagnosis, Protein sequences, and CRM of credit card business, etc [112].

To comprehend how AR operates, let us take two item sets X and Y; according to Agrawal, Mannila, Srikant, *et al.* [113], AR are defined as {X}→{Y} which means if X then Y, where X is the antecedent and Y is the consequent. The most used criteria to evaluate the quality of an AR are the following:

— Support is defined as the ratio between the number of transactions containing both X and Y, and the total number of transactions [114],
$Support = P(X \cap Y) = \frac{number-of-transactions-containing-X and-Y}{total-number-of-transactions}$;

— Confidence is the probability that a transaction will contain Y knowing that X is already contained in the transaction [114],
$Confidence = \frac{P(X \cap Y)}{P(X)} = \frac{number-of-transactions-containing-X-and-Y}{total-number-of-transactions-containing-X}$;

— Lift is the ratio between the observed support and the one expected if X and Y were independent [114],
$Lift = \frac{P(X \cap Y)}{P(X) \times P(Y)}$.

Figure 4.1 below displays the three-stage process of the proposed AR freights consolidation methodology. In stage I, presented in the "Data preparation" section, a customer's data is verified and preprocessed. The preprocessed data is used as an input for the consolidation in stage II, which is presented in the "AR consolidation methodology" section. Stage III is described in the section "Integration of the consolidation results into the mathematical model," where the consolidation result is integrated into the mathematical model for the cost calculation.

4.3.1 Data preparation

As displayed in section 4.3, our method takes customer data information as input. Thus, we check whether the customer's data information is in the proper format. If not, we must first prepare the data to perform freight consolidation. This step is crucial for the algorithm used in our methodology in order to produce quality results. Data preprocessing is a step that takes our raw data and transforms it into a format that can be understood and analyzed by the programming tool and the data mining technique. The preprocessed transactional data includes the following: customer ID, FSA (geographical unit), transaction date, and quantity. Our AR consolidation methodology is then applied in stage II.

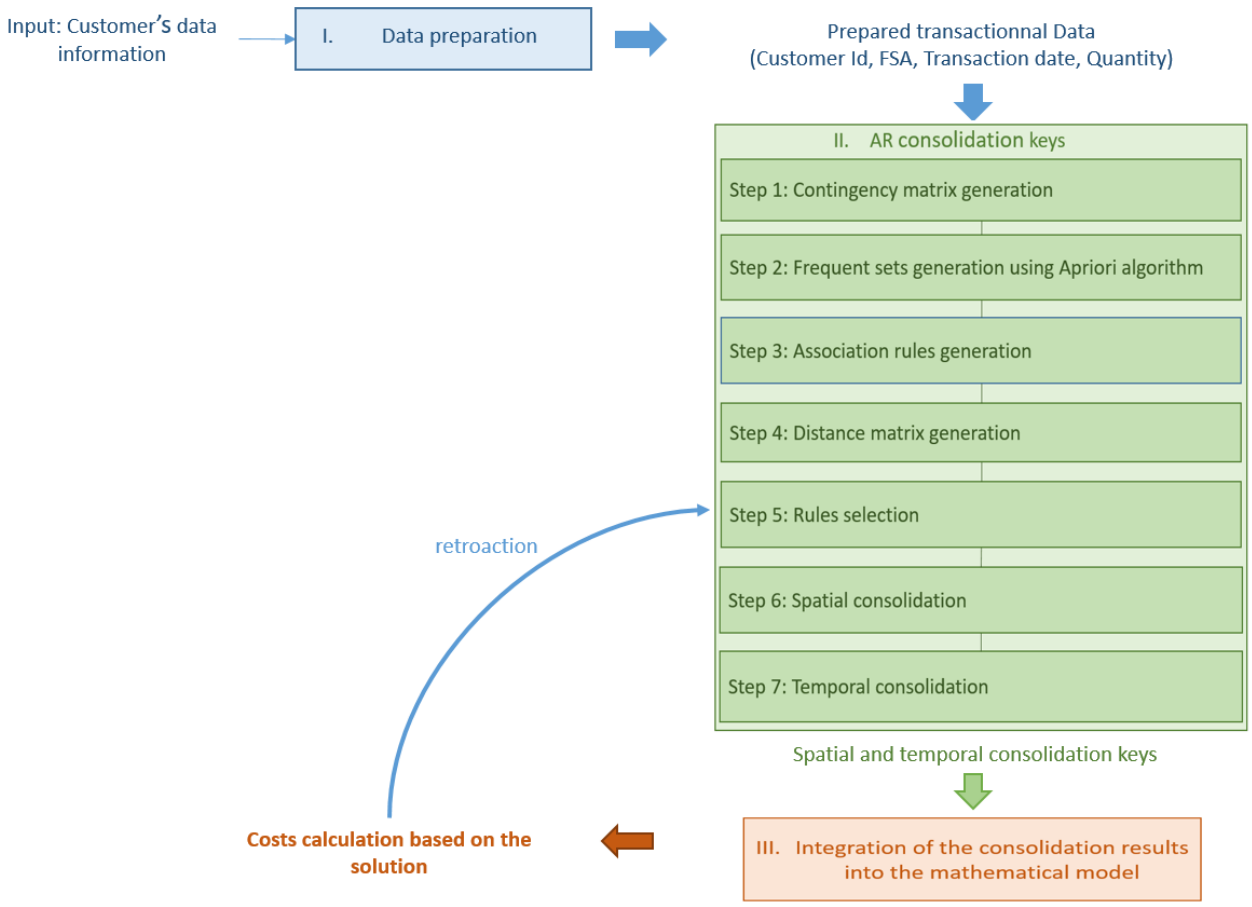

Figure 4.1 AR freight consolidation methodology

4.3.2 AR consolidation keys

The algorithm presented in this section explains the methodology and allows spatial and temporal consolidation using Association Rules (AR). In other words, this algorithm highlights our contribution, which consists of making consolidation by taking into account the demand patterns through the application of AR. Thus, it allows us to recognise the opportunities to consolidate customers' returns. We refer to the results of the applied algorithm as AR consolidation keys. The method takes transactional data as inputs. The following information is required:

— Customer ID (i.e., a unique identifier for the customer);
— FSA (i.e., the customer's geographical unit);
— Transaction date;
— Quantity.

The algorithm is divided into the following steps:

1. Contingency matrix generation: generate the Contingency Matrix from the transactional data, using the date as an index, the FSA as columns and the X as values (where X = 1 if quantity >0, and X = 0 otherwise);
2. Generation of frequent sets using Apriori algorithm: use Apriori algorithm on the Contingency Matrix, in order to generate frequent sets of items;
3. Generation of Association Rules: generate Association Rules from frequent sets;
4. Generation of distance matrix: generate a distance matrix from each FSA pair;
5. Selection of rules: Select the best Association Rules (based on AR scores, such as confidence and lift, and proximity threshold between FSA);
6. Spatial consolidation: Create the spatial consolidation keys based on the selected Association Rules;
7. Temporal Consolidation: Create the temporal consolidation keys based on the selected approach (e.g., delivery every five days, or delivery every Tuesday and Friday);

The resulting outputs are the spatial and temporal consolidation keys. The spacial consolidation keys tell the program how to group FSAs into larger clusters (LC). For instance, {FSA1: LC1, FSA2: LC1, FSA3: LC2, ...}. The temporal consolidation keys tell the program how to perform temporal consolidation; in other words, how to assign a delivery date to each transaction date. For instance, if the temporal consolidation strategy is to do deliveries each Tuesday and Friday, {2021-01-01: 2021-01-05, 2021-01-02: 2021-01-05, 2021-01-03: 2021-01-05, 2021-01-04: 2021-01-05, 2021-01-05: 2021-01-08, ...}.

4.3.3 Integration of the consolidation results into a mathematical model

This section integrates the results of the consolidations in the mathematical model, which depend on the case study used. The chosen mathematical model is a mixed-integer non-linear programming model depicted in the model development section of [95]. We have programmed the cost function using Python programming language based on the following costs: warehouse rental cost, warehouse start-up cost, inventory cost, and transportation cost.

Firstly, we created an Excel sheet to store all of the parameters from the article (a, s, b, w, r, h, etc.). The parameter values can then be easily read by python. The distance matrix can be calculated from the coordinates by using the python package "geopy."

Secondly, we developed a python function to translate the consolidation results of 4.3.2 into the appropriate decision variables. We build the matrix "Y" using our spacial consolidation keys: $Y_{i,j,p} = 1$ if customer j is allocated to ICP i (at period p); 0 otherwise. Similarly, we can

build the matrix "Z" using spacial consolidation keys: $Z_{i,p} = 1$, if an ICP is opened at site i (at period p); 0, otherwise. Then, for "X" (the consolidated volume of products returned from the ICPs to the CRC), the matrix can be calculated by multiplying the variable "T" with the *dot product* of the parameter "r" and the variable "Y."

Thirdly, we developed a cost function that takes the model parameters and decision variables as input and returns the costs as output. The function verifies that the constraints are satisfied, using assertions. After that, the renting costs, setup costs, inventory costs, handling costs, and transportation costs are calculated, using simple linear algebra with the package "numpy" and following the formula in [95] section 3.

Fourthly, the retroaction loop enables the generation of multiple model solutions by adjusting the rule selection to obtain the best results.

4.4 Experiment and contribution

To solve the reverse logistics problem for the spatial and temporal consolidation of returned products in a multiple time horizon, Min, Ko, and Ko [95] used a mixed-integer non-linear programming model as well as a genetic algorithm. Next, our methodology will be demonstrated and compared to [95] using the same data set, constraints and objective function. We chose this article, firstly because it deals with order returns, a subject that is highly topical and costly to industries, and secondly because all the input data of the case study are available in the article, thus allowing comparison and validation.

This section will experiment with the difference between a datamining-driven approach and an operational research approach, using a case study concerning spatial and temporal consolidation for returned products.

The problem in the case study will be detailed in 4.4.1.

4.4.1 Case study

A set of 30 customers are making product returns. The dataset shows their latitude and longitude and their average quantity of daily returns each year (see Table 2 in [95]).
Products must be returned to the centralised return center (CRC). The CRC is located at (x = 15, y = 20).
The company is investigating the possibility of renting Initial Collection Points (ICP), acting as relay points, where customers could return their products. There are ten potential ICPs. The dataset gives their latitude and longitude, as well as their fixed and variable costs. The

problem lies in choosing, for each year, which ICP is open and their Maximum Holding Time (MHT, i.e., the time window for the temporal consolidation). Table 4.1 gives the locations of ICPs and their associated costs.

Table 4.1 ICP locations and its fixed and variable costs.

Index	x	y	**Fixed cost ($)**	**Variable cost ($)**
1	43.97	49.89	300	1,000
2	9.34	21.67	300	1,000
3	41.23	30.25	300	500
4	5.04	58.97	300	500
5	24.79	19	300	1,500
6	16.18	20.66	300	1,500
7	30.18	45.3	300	1,500
8	42.8	12.89	300	1,000
9	6.94	33.58	300	500
10	54.71	57.06	300	500

Then, for each year, each customer is assigned an ICP. This can be done simply by assigning them to the nearest open ICP.

The following section, 4.4.2, will discuss an additional step.

4.4.2 Data preparation: disaggregation

As shown in section 4.3, our method takes transactional data as input. Thus, since this case study provides aggregated data (i.e., average quantity of returns), a disaggregation operation must be performed in order to prepare the data for our method. As such, the following assumptions are made about a customer's transactional data:

1. For each customer and for each year, 250 daily quantities will be generated.
2. Data is intermittent. The customers are very unlikely to make return trips every day. Our assumption is that they occur once or twice per week.
3. The proportion of days without returns is arbitrarily set at 70%.
4. When the quantity of returns is not null, it is a positive integer.
5. When the quantity of returns is not null, it follows a random Poisson distribution.
6. The average quantity of daily returns for each customer and for each year must match the data provided by the case study (see [95] table 2).

The following pseudo-code allows us to generate transactional data, from aggregated data while satisfying the above assumptions:

1. numberofzero = 70% * 250 (rounded to the nearest integer, if needed)
2. numberofnonzero = 250 - numberofzero
3. For each aggregated demand:

   — nonzeroaverage = averagedquantityofreturns * 250 / numberofnonzero;
   — nonzeros = Generate a series of numbers following the Poisson distribution, where the average is nonzeroaverage, and the length is numberofnonzero;
   — Make necessary adjustments to the series of nonzeros to ensure that its average is nonzeroaverage;
   — zeros = Generate a series of zeroes in which the length is numberofzero;
   — Concatenate the two series into one series, then shuffle it.

4.4.3 Problem translation

Our method (section 4.3) is designed to solve problems that are slightly different than the problem solved in the case study (section 4.4.1). Whereas our method is designed to form clusters of FSA, the case study needs to select which ICPs are going to be open. In order to solve the problem of the case study, we are going to make the following translations:

1. We consider that each potential ICP has its own FSA (e.g., ICPi is assigned to FSAi).
2. For each customer, we determine the nearest ICP, and then the customer inherits its FSA from the nearest ICP.
3. Then, we can apply our method to form clusters of FSA. (We have the choice to apply it once to all data, or to apply it for each year individually.).
4. For clusters containing a single FSA, the corresponding ICP must be open.
5. For clusters consolidating multiple FSAs, exactly one ICP must be open (and we choose the one that results in the lowest total cost).

4.4.4 Distance Normalisation

A problem with the coordinates (x & y) of the case study was encountered. By visualising them on the map and calculating the distance matrix between points, we noticed that all distances were enormous. We used the python package "geopy" in order to accurately calculate the distances in miles (one mile is equivalent to 1.60 kilometre). With the given coordinates, the minimum distance between a customer and an IRC was over 200 miles. However, the model constraints specify that the distance between a customer and its assigned IRC must not exceed 5 miles. To solve this issue, we normalised all distances by applying a factor of

1/340. To find this factor, we looked into the author's solution to find the customer that was the farthest from their allocated IRC. Customer 15 is located 1653 miles from their IRC. In order to make the author's solution feasible, we had to use a factor denominator of at least 331 and we used a round number, 340. After normalisation, the farthest distance in the author's solution is 4.9 units of distance.

Our methodology, its algorithms, and the mixed-integer programming model were developed and executed using Python. In terms of CPU time, the results were given in less than a second.

4.4.5 Models

Two Models were created using our methodology.

4.4.5.1 Model A

Based on the most reliable Association Rules score and distance threshold, the four best rules selected are the following: ({FSA1}→{FSA7} ; {FSA2}→{FSA5}; {FSA2}→{FSA6}; {FSA2}→{FSA9}). These rules are sets composed with two FSA in the form {X}→{Y} defined as the rule: if X then Y. According to customer density in each FSA and the number of returns, rules provide which returns will be grouped and allocated to the best IRC. In other words, we could say that customers living in FSA1 and FSA7 are associated and therefore the returns could be consolidated from both regions, giving the group {1+7}. Thus, FSAs were consolidated into these more significant groups: ({1+7}; {2+5+6+9}; {3}; {4}; {8}; {10}). To minimise costs, the following IRCs were selected: (1; 2; 3; 4; 8; 10). The customers are assigned to these IRCs based on proximity. So, IRC1 was chosen for group {1+7}, which means that all customers from FSA1 or FSA7 were allocated to IRC1. The customers from the group {2+5+6+9} are attributed to IRC2; all costumers from FSA3, FSA4, and FSA8 are assigned to IRC3, IRC4, and IRC8, respectively. Finally, all customers from FSA10 were attributed to IRC 1.

4.4.5.2 Model B

For this model, we used a larger distance threshold, allowing us to obtain fewer, more uniform groups. The following six best ARs were chosen: ({FSA1}→{FSA7}; {FSA1}→{FSA10}; {FSA2}→{FSA5}; {FSA2}→{FSA6}; {FSA4}→{FSA9}; {FSA3}→{FSA8}). The FSAs were therefore consolidated into these groups: ({1+7+10}; {2+5+6}; {4+9}; {3+8}). The following IRCs minimise the total costs: (1; 2; 4; 8). Customers were allocated to the IRC in

the same manner as the previous model. All customers from FSA 1, 7, and 10 were allocated to IRC 1, and so on.

4.4.5.3 Maximum holding time

It is important to note that in the above models, only the spacial consolidation was defined. In terms of temporal consolidation, the chosen method considers 7 scenarios by varying the Maximum Holding Time from 1 to 7 days.

4.5 Results

Our method allowed us to try two models (named A and B) that will be compared, in terms of cost, to [95]'s solution. Model costs will be presented for each scenario.

4.5.1 Model A

The following are the results from our first consolidation model.

Table 4.2 below displays the cost components for each period, and for each scenario. For each period, we have calculated the costs of the mathematical model using the consolidation results of model A using the AR methodology. Those results were calculated for each Holding time, which refers to the temporal consolidation of the returns.

We observe that with longer holding times come higher total costs. Results show that our methodology provides very similar results. Table 4.3 zooms the maximum holding time of 3 days in order to compare the results with [95].

For model A, we observe that our accumulated cost is 1747525 dollars for six open ICPs for all periods, against 1747162 dollars for [95] and 4 open ICPs for each period, which is 0.02% higher.

Table A.1 represents the detailed information on the volume of consolidated products returned by customers at the different initial collection points selected by our solution model.

If we look at the number of ICP's used in total, we notice that for both of the results that were compared, six ICP's were opened.

4.5.2 Model B

Since we have detailed information for our model solution, we have noticed an ICP that was opened only for one customer, for customers 3 and 5. Therefore, we proposed a relaxed

Table 4.2 Total cost of model solution A with AR consolidation methodology

Period	Holding time	Renting	Setup	Inventory	Handling	Transportation	**Total**
	1	4500	1800	39400	39400	263200	348300
	2	4500	1800	59100	39400	258100	362900
	3	4500	1800	78800	39400	253300	377800
1	4	4500	1800	98500	39400	245650	389850
	5	4500	1800	118200	39400	240850	404750
	6	4500	1800	137900	39400	240850	424450
	7	4500	1800	157600	39400	240850	444150
	1	4500	0	58950	58950	388000	510400
	2	4500	0	88425	58950	375100	526975
	3	4500	0	117900	58950	366950	548300
2	4	4500	0	147375	58950	364400	575225
	5	4500	0	176850	58950	359650	599950
	6	4500	0	206325	58950	356250	626025
	7	4500	0	235800	58950	356250	655500
	1	4500	0	89825	89825	581400	765550
	2	4500	0	134738	89825	547450	776513
	3	4500	0	179650	89825	547450	821425
3	4	4500	0	224563	89825	544900	863788
	5	4500	0	269475	89825	544900	908700
	6	4500	0	314388	89825	543200	951913
	7	4500	0	359300	89825	543200	996825
	1	13500	1800	188175	188175	1232600	**1624250**
	2	13500	1800	282263	188175	1180650	**1666388**
	3	13500	1800	376350	188175	1167700	**1747525**
TOTAL	4	13500	1800	470438	188175	1154950	**1828863**
	5	13500	1800	564525	188175	1145400	**1913400**
	6	13500	1800	658613	188175	1140300	**2002388**
	7	13500	1800	752700	188175	1140300	**2096475**

Table 4.3 Total cost of model solution A for a maximum holding time of 3 days

	Time period			
	P1	P2	P3	*Accumulation*
Renting ICP	6300	4500	4500	15300
Inventory	78800	117900	179650	376350
Handling	39400	58950	89825	188175
Transportation	253300	366950	547450	1167700
Total	377800	548300	821425	**1747525**

second model solution B based on the vicinity, so that if a customer is located near the FSA, we would allocate them to the nearest ICP. Table 4.4 below provides the results of the total cost of model B. For a maximum holding time of 3 days, we observe a reduction of 2.42% for the accumulated cost for the maximum holding time compared to model A, and a decrease of 2.39% compared to the model in [95].

Table 4.4 Total cost of model solution B for a maximum holding time of 3 days

	Time period			
	P1	P2	P3	*Accumulation*
Renting ICP	4700	3500	3500	11700
Inventory	78800	117900	179650	376350
Handling	39400	58950	89825	188175
Transportation	236400	353700	538950	1129050
Total	359300	534050	811925	**1705275**

Table A.2 details the changes brought about by model B. We notice that customer 5 is no longer assigned to ICP 10, while customer 3 is no longer the only customer allocated to ICP 4. Model B has therefore allowed 2 centers to be closed (3 and 8) by dispatching customers to the nearest and least expensive centers, while respecting the distance constraints issued by Min, Ko, and Ko [95]. Unlike [95], the locations chosen for each customer by our model do not change over time, and the holding times for consolidation are not dynamic or heterogeneous, which can be penalising in some cases.

4.6 Conclusion

The contribution of this study is to offer a new methodology to consolidate and reduce costs for freight based on a data mining approach in an intermittent demand context. This paper has solved a consolidation problem of finding the number and location of ICP's involving spatial and temporal freight consolidation. To achieve this, we have used a data mining driven

approach that uses Association Rules for consolidation. Once the output of our AR freight consolidation methodology is available, we implement it in the mixed-integer programming model developed by Min, Ko, and Ko [95]. This allows us to compare our results with those from Min, Ko, and Ko [95] who opted for an operational research approach using a genetic algorithm and a mixed-integer non-linear programming model. The results obtained from the AR freight consolidation methodology were about the same as the approach in [95]. By selecting more Association Rules, we get a model with fewer and more uniform spatial groups, which have a lower total cost than the solution found by [95]. Our methodology takes advantage of raw data, whereas [95] is limited to aggregated data. This forces them to make the hypothesis that the customer's demand will be constant throughout the year, which does not match the reality of many companies, such data simplifications are responsible for the loss of lower-level patterns, such as demand intermittence and customer behaviors. Not to mention that it works both ways (it can handle returns and deliveries), and our fast algorithm can easily take into account massive data, contrary of the genetic algorithm in [95] that may not scale well with large data sets. Our method is fast even with large data sets and gives homogeneous holding times and ICP's, which is more tailored to some companies that can not handle the logistical complexity of constantly changing the ICP's number of locations and holding times. This logistical complexity requires a precise number of trucks and drivers if the company does not outsource deliveries. Not to mention the complexity that planners and dispatchers would have to deal with in modifying their routes.

For future work, we can make the following improvements:

— A selection algorithm that automates the choice of Association Rules

— A fast methodology that can propose a heterogeneous consolidation cycle when desired.

CHAPITRE 5 ARTICLE 2: A DATA-DRIVEN METHODOLOGY FOR ORDER FULFILLMENT ALLOCATION THROUGH SPATIAL CONSOLIDATION: BETTER DECISION-MAKING FOR CUSTOMER SATISFACTION

Zineb Aboutalib, Bruno Agard

Accepted in International Journal of Logistics Systems and Management

Submitted: 31-Dec-2024, Accepted: 04-Jul-2025

Abstract *— This paper introduces a novel multi-objective methodology for order fulfilment allocation and spatial consolidation in outbound logistics. At its core, this research presents a mathematical model that forms a holistic approach to managing these critical aspects, achieving a balance between customer satisfaction and cost reduction. The methodology employs a data-driven strategy, leveraging historical data to discover patterns in customer behaviour and demand intermittency. By incorporating a unique mixed-integer linear programming model that emphasises spatial consolidation, we seek to enhance two critical objectives in logistics operations: minimising shipping costs and maximising customer satisfaction, represented by a specific metric called the weighted order fulfilment rate. We demonstrate through simulation experiments that our approach significantly enhances both profitability and customer satisfaction. A comparative analysis further illustrates the superiority of this method in terms of these key metrics.*

Keywords *— Optimisation, Association Rules, Pattern extraction, Order fulfillment allocation, Data mining, Simulation, Logistics.*

5.1 Introduction

As global commerce expands, so does the competition among companies, particularly in shipping markets. This competitive landscape prompts organisations to seek innovative methods to maintain profitability and efficiency Nikolopoulos [93]. Freight consolidation, a strategy that merges various small shipments into one large shipment, has become vital in achieving cost savings and maintaining high service levels Hall [15], Tyan, Wang, and Du [10], and SteadieSeifi, Dellaert, Nuijten, *et al.* [115]. Despite its significance, decision-making for freight consolidation can be extremely challenging. Previous studies have often used oper-

ational research based on aggregated datasets, neglecting valuable information on customer behaviour and specific trends Ranyard, Fildes, and Hu [94].

This paper aims to overcome this limitation by introducing a data-driven methodology and a mathematical model that form a holistic approach to managing order fulfilment allocation and spatial consolidation in outbound logistics. Balance between customer satisfaction and cost reduction is achieved through a methodology for spatial consolidation that leverages non-aggregated data to unveil beneficial patterns, trends, and behaviours. The methodology can uncover previously inaccessible optimisation opportunities that aggregated data methods fail to provide.

Our research is guided by two crucial and often conflicting goals: to enhance the order fulfilment process and to reduce shipping costs simultaneously. Specifically, we target two key metrics: maximising the weighted order fill rate (WOFR), associated with customer satisfaction, and minimising the number of shipments in a route through freight consolidation to reduce shipping costs. The term 'weighted order fill rate' (WOFR) extends the concept of the customer order fill rate (OFR), a prevalent metric in the industry used to measure the ability to fulfil customer orders Song [116] and Iqbal, Malzahn, and Whitman [117]. Differing from the traditional order fill rate, WOFR assigns weights to orders based on their importance, incorporating factors such as sales amount, profit margin, or marketing strategy. This enhancement allows for more strategic alignment with organisational goals, as maximising WOFR ensures that more significant orders are satisfied more frequently, aligning supply chain processes with higher-priority customers Rim and Park [8].

Recognising the inherent conflict between these two objectives, we propose a compromise solution by maximising the difference between the weighted order fill rate (first objective) and a penalty proportional to the number of shipments (second objective). The balance between these objectives can be customised to reflect specific company priorities and the trade-off between customer satisfaction and shipping costs. Through extensive comparisons with existing methods, we demonstrate that our approach can result in remarkable gains in both objectives.

Our main contributions are twofold. First, we introduce a novel mixed-integer linear programming model that seeks spatial consolidation to optimise customer satisfaction and cost reduction. This approach specifically addresses the order fulfilment allocation problem, a key process in logistics operations, including distribution centres.

Second, we enhance and adapt a data-driven association rules-based methodology for freight consolidation by incorporating an automatic rule selection feature. This extension provides a more adaptable and autonomous approach to improve spatial consolidation and reduce costs

in the context of order allocation.

The paper is organised as follows: Section 5.2 offers a brief review of the relevant and state-of-the-art literature. The proposed methodology is detailed in Section 5.3, followed by experimental design and results in Sections 5.4 and 5.5. Finally, Section 5.6 concludes the paper.

5.2 State of the Art

In order fulfilment operations, such as those in distribution centres (DC), several decision-making processes are involved in fulfilling customer orders.

One of the key decision-making processes is the allocation of available stock to active orders, a subject discussed in detail in section 5.2.1. This process, typically referred to as order fulfilment allocation, order allocation, or order picking planning in the literature, involves selecting the optimal subset of customer orders to fulfil, considering factors like item availability, customer priority, order size, and shipping deadlines. It's a crucial aspect of order fulfilment efficiency and directly impacts customer satisfaction levels.

Downstream in the order fulfilment process lies another critical component: the vehicle routing problem (VRP). This classic optimisation problem, which is widely studied in logistics literature, involves efficiently allocating customer orders to delivery routes and determining the optimal sequence of stops to minimise the total cost or distance travelled.

A notable approach to improve the efficiency of order fulfilment operations involves integrating the order allocation process with the vehicle routing problem. The Inventory Routing Problem (IRP) is a well-known optimisation problem in the logistics literature that aims to determine the optimal way to allocate inventory to customer orders while simultaneously planning delivery routes. However, the feasibility of Inventory Routing depends on various factors, such as company size, the complexity, the nature of the supply chain network, and the customer demand patterns. Inventory Routing and its limitations will be discussed in section 5.2.2.

Beyond IRP, the literature lacks comprehensive approaches that effectively bridge order fulfilment allocation (OFA) and routing cost.

This paper introduces an innovative solution to fill this gap by leveraging freight consolidation in the context of order fulfilment. Freight consolidation, detailed in section 5.2.3, involves combining multiple smaller shipments into a single large shipment to reduce shipping costs and lead times. What sets our approach apart is the integration of freight consolidation with order fulfilment allocation through spatial consolidation. This unique combination allows for

simultaneous consideration of customer satisfaction and route costs, similar to the goals of IRP, but without IRP's inherent challenges.

5.2.1 Order fulfilment allocation

Order fulfilment allocation, also known as order allocation or order picking planning, is a critical decision-making process in order fulfilment operations such as those found in distribution centres (DCs). As defined by Rim and Park [8], this is a decision process by which a subset of orders is selected to be fulfilled. Many factors can be considered, including inventory availability, customer priority, order size, and shipping deadlines.

Traditional Approaches The problem of allocating limited stocks to customer orders was introduced by Guerrero and Kern [36], focusing on prioritising orders. Traditional approaches like the First-Come-First-Served (FCFS) rule and prioritising orders by due dates have been widely discussed Meyr [38].

Quantitative Approaches Quantitative models such as Mixed Integer Linear Programming (MILP) to maximise the Order Fill Rate (OFR) have been explored by Rim and Park [8]. Seyedrezaei, Najafi, Aghajani, *et al.* [39] extended this by considering challenges like perishable goods and warehouse capacity limitations, using a dynamic mathematical model and a genetic algorithm (GA) to optimise the OFR over selected timeframes. Lečić-Cvetković, Atanasov, Babarogić, *et al.* [40] proposed a Make-to-Stock system with targeted service levels.

In recent literature, Giat [118] tackled the allocation of scarce resources in scenarios where orders are continually made, but deliveries are periodic. They introduced the concept of the 'window fill rate'—a measure of the probability that a customer's wait time doesn't exceed their tolerance.

Further, Zhang, Bi, and Zhang [119] delved into the optimisation of the Logistics Service Supply Chain (LSSC) order allocation. Their focus was on enhancing LSSC efficiency and customer satisfaction. By combining K-Means clustering and QoS matching, they formulated an algorithm tailored to the logistics service integrator order allocation process.

Integrative Solutions Recent studies show a trend towards integrative approaches that jointly address multiple facets of the problem. Li, Li, Aneja, *et al.* [41] proposed an adaptive large neighbourhood search (ALSN) that address both order allocation and order routing problems.

Similarly, Jiang and Li [120] presented a comprehensive approach to the order fulfilment problem with time windows and synchronisation for online retailing. They consider the e-order fulfilment process, emphasising the integration of decisions regarding order allocation with other related operations such as order consolidation, synchronisation, and delivery. Such integrative solutions underline the growing need for holistic approaches in tackling the multifaceted challenges of order fulfilment. A growing body of research also highlights the need to move beyond single-objective optimisation, which often focuses on cost, to embrace multi-faceted goals that include customer satisfaction and environmental impact. For instance, in the context of e-grocery, Ekren, Perotti, Foresti, *et al.* [121] used simulation to evaluate various depot allocation policies against a triad of key performance indicators: fill rate, delivery cost, and carbon emissions. Their findings reveal that the optimal policy depends on the strategic weight a firm assigns to each objective, with hybrid policies considering both customer proximity and product availability showing great promise for balancing these competing goals. This underscores the importance of multi-objective frameworks that can be tailored to a company's specific strategic priorities, a central feature of the methodology proposed in this paper.

Gaps and Future Directions The limited availability of quantitative models has been noted by Li, Li, Aneja, *et al.* [41]. Additionally, a notable void in the literature is the general omission of vehicle routing costs in many models Rim and Park [8], Lečić-Cvetković, Atanasov, Babarogić, *et al.* [40], Seyedrezaei, Najafi, Aghajani, *et al.* [39], Giat [118], and Zhang, Bi, and Zhang [119]. Although Li, Li, Aneja, *et al.* [41] and Jiang and Li [120] incorporate route costs, their focus diverges from the traditional order fulfilment allocation paradigm. Rather than allocating limited inventory to orders to determine which subset of orders to fulfil, their models prioritise the selection of the most suitable warehouse (or fulfilment centre) to service each active order, operating under the assumption of an ever-abundant inventory.

This observation underscores the imperative for further inquiry into integrating route costs within the conventional order fulfilment allocation framework, particularly in scenarios with a single warehouse where inventory shortages might be a reality.

This disconnect is especially significant because the method of fulfilment itself has a profound behavioural impact on the end customer. For instance, recent empirical work by Wagner, Calvo, and Amorim [122] demonstrates that consolidating a multi-item e-commerce order into a single delivery measurably increases customer satisfaction and reduces product returns, even if the delivery takes slightly longer. This insight reinforces the value of integrating consolidation-aware thinking directly into the order allocation process, moving beyond simple

cost metrics to enhance the quality of the customer experience.

5.2.2 Inventory Routing and Its Challenges

The Inventory Routing Problem (IRP) integrates order allocation and vehicle routing to optimise both aspects simultaneously, aiming to minimise costs and enhance customer satisfaction Federgruen and Simchi-Levi [123], Campbell, Clarke, Kleywegt, *et al.* [124], Golden, Raghavan, and Wasil [125], and Cui, Long, Qi, *et al.* [126]. As logistics problems grow in complexity, the literature reflects a move towards sophisticated solution methods. For large-scale integrated problems like the IRP, researchers are developing advanced matheuristics that combine exact methods with heuristics to achieve high-quality solutions efficiently Skålnes, Vadseth, Andersson, *et al.* [127], Shaabani, Hoff, Hvattum, *et al.* [128], Najy, Archetti, and Diabat [129], and Charaf, Taş, Flapper, *et al.* [130]. Although beneficial, the implementation of IRP is limited by challenges such as computational complexity, change management, and unsuitability for companies heavily relying on third-party logistics providers (3PLs) Bertazzi, Bosco, and Laganà [131] and Barbosa-Povoa and Pinto [132].

5.2.2.1 Challenges of Inventory Routing

The Inventory Routing Problem (IRP) offers the potential to streamline order fulfilment by combining inventory allocation and routing. Still, the challenges associated with its implementation can make it a less suitable choice for some companies. The following subsections delve into these challenges.

Change Management Implementing an IRP can entail significant changes to an organisation's functions, processes, and systems, including the need to hire, train, and change management efforts. When a company is currently using a two-step approach to order fulfilment allocation and order routing, the integration of these processes can result in a loss of control and flexibility over the solutions and their impacts on customer service levels and route costs. Challenges can be further exacerbated by the conflicting objectives and incentives between the different processes and functions, which are often siloed in their decision-making. Initial investments and effective change management efforts are necessary to address these challenges [133], [132].

Computational Complexity Solving the IRP can be computationally intensive, making it challenging for companies that handle large-scale problems or have limited computational

resources. Even with advanced techniques to reduce computation time, the increased complexity may create barriers for users in configuring or managing the system. This complexity must be carefully weighed against the quality and efficiency of the solutions[133], [132].

Outsourced Deliveries For companies that rely heavily on third-party logistics providers (3PLs), the IRP might not be a suitable approach. Suppliers often choose to outsource deliveries to 3PLs for various reasons, such as achieving cost savings through economies of scale and reducing the administrative burden associated with managing drivers, vehicles, and routing processes. In such situations, the company does not have full control over the routing decisions and inventory management of the 3PL. Moreover, the company and the 3PL may have conflicting objectives, such as differing priorities for cost reduction and service quality. Additionally, the fees paid by the suppliers to the 3PLs are typically independent of the 3PLs' routing decisions. This disconnect between supplier objectives and 3PL routing decisions can hinder the effectiveness of the IRP for the supplier. Bertazzi, Bosco, and Laganà [131]

Summary of Challenges While the IRP has undeniable potential for enhancing efficiency and customer satisfaction, these challenges present significant barriers for many companies. The complexities of implementation, computational demands, incompatibility with outsourced deliveries, and potential organisational resistance underscore the need for more adaptable, less resource-intensive solutions. These challenges open a pathway for alternative approaches, such as the method proposed in this paper, to provide more tailored solutions that can better align with a company's unique requirements and constraints.

5.2.3 Freight consolidation

The increasing focus on sustainability and the need for efficient supply chain management has led to advancements in freight consolidation techniques recently. Freight consolidation, as described by Çetinkaya [20], is a strategic practice that allows managers to determine the appropriate volume or time to accumulate shipments before releasing them, thereby leveraging economies of scale to reduce shipping costs. This approach can be employed at multiple points within the logistics chain, ranging from supplier operations Lin, Hjelle, and Bergqvist [134] to last-mile urban delivery processes Haider, Hu, Charkhgard, *et al.* [135] and Lyons and McDonald [136]. Its primary goal is to efficiently manage distribution expenses by capitalising on cost reductions associated with larger shipment sizes and by combining products from various locations and time periods Bookbinder and Higginson [17], Hanbazazah, Abril, Erkoc, *et al.* [18], and Kumar and Anoop [137]. Foundational research

continues to refine core strategies; for instance, Yang, Lee, and Lee [138] uses a MILP model to compare fundamental time-based and quantity-based consolidation policies, concluding that time-based strategies excel in stable markets while quantity-based approaches are better suited to volatile demand.

Spatial consolidation, a method of grouping shipments by geographic locations, is particularly pertinent to this paper. By consolidating small shipments from nearby regions into larger loads, shipping costs are decreased, and delivery times are improved, additionally reducing the environmental impact Min [21]. The broader impact of freight consolidation is also gaining attention, with studies like Orhan, Goez, Guajardo, *et al.* [139] using vehicle routing models to assess how consolidation strategies affect the "livability" of small cities by reducing traffic and congestion.

Several methods for implementing spatial consolidation exist, such as Cross-docking [140], [141]; Milk run [23], [137]; Shuttle service [142].

Various models can be employed to solve spatial consolidation problems, such as Vehicle Routing Problem (VRP) [143], [135], [144], [145]; Inventory Routing Problem (IRP) [146], [147]; Location and allocation model [148]; Network Design Model [149]; Fuzzy model [28]; Multi-objective optimisation model [150], [144].

Numerous algorithms and techniques can be utilised to tackle spatial consolidation problems, encompassing Genetic algorithms [151], [152]; Ant colony optimisation [153]; Simulated annealing [154]; Tabu search [155]; Particle swarm optimisation (PSO) [156]; Memetic Algorithm [157], [158]; Lagrangian relaxation [159]; Adaptive large neighbourhood search (ALNS) [160], [161]; Reinforcement learning [162]. These models and algorithms can be employed individually or in conjunction to address spatial consolidation problems.

Notably, a data-driven approach was introduced by Aboutalib and Agard [163] that leverages Association Rules to exploit data patterns for freight consolidation in the context of intermittent demand.

A promising direction in this domain is the use of data mining to uncover non-obvious consolidation opportunities. Notably, a data-driven approach was introduced by Aboutalib and Agard [163] that leverages Association Rules to exploit latent patterns in customer demand. This line of research, focused on strategic pattern discovery, is rapidly evolving. Recent advancements, such as the 'predict-then-optimise' framework proposed by Cheng, Hijazi, Konak, *et al.* [164]—which uses spatio-temporal clustering and frequent itemset mining to identify promising consolidation points that are then fed into an optimisation model—now underscore the subsequent challenge: tightly coupling these data-mined insights with opera-

tional optimisation. They observe that to fully realise the value of historical patterns, they must be translated into feasible, constraint-respecting fulfilment plans.

This highlights a critical opportunity for a methodology focused on the integration of spatial consolidation with order fulfilment allocation, bridging strategic pattern discovery with prescriptive allocation decisions. Our current study is specifically designed to address this challenge. Building upon the foundation of Aboutalib and Agard [163], we introduce two key innovations to create this bridge: first, a novel, automated rule selection algorithm that refines the quality of the consolidation patterns, and second, our WOFR-EC model that integrates these patterns into a holistic order fulfilment allocation decision. This approach moves beyond simply identifying what *could* be consolidated, to prescribing what *should* be consolidated to balance customer satisfaction and route costs.

5.2.4 Synthesis

The landscape of order fulfilment operations presents complex decision-making processes involving order fulfilment allocation and vehicle routing, both of which are vital to customer satisfaction and operational efficiency. Research on vehicle routing (VRP) is extensive and well-established, whereas the study of order fulfilment allocation remains fragmented and inconclusive. Existing quantitative models for order allocation are limited and often overlook route costs.

The Inventory Routing Problem (IRP) has been recognised as a solution to optimise order allocation and vehicle routing simultaneously. However, its practical application is notably constrained, especially for small and medium-sized enterprises, due to its inherent complexity. This constraint presents a noticeable gap in available solutions to solve the OFA that consider both customer satisfaction and route costs. Within this context, the literature lacks alternative, applicable approaches that can fulfil the goals of IRP without facing its challenges.

This paper addresses these gaps by introducing a novel multi-objective methodology that seamlessly integrates the concepts of freight consolidation with order fulfilment allocation. Our contributions are twofold:

1. **WOFR Encouraging Consolidation (WOFR-EC) Model:** Unlike traditional approaches, the proposed WOFR-EC model bridges the gap between order allocation and route costs. Through a unique Mixed Integer Linear Programming (MILP) model, our solution maximises the Weighted Order Fill Rate (WOFR) (a metric for customer satisfaction) while minimising route stops. This serves as an attractive alternative to

IRP, specifically designed for companies unable or unwilling to implement more complex solutions.

2. **Enhanced Spatial Consolidation:**

   Building upon the Association Rules-based methodology introduced by Aboutalib and Agard [163], we extend this approach with a novel rule selection algorithm. This innovation facilitates automatic and deterministic rule selection, providing a more adaptable and efficient strategy for spatial consolidation within the context of order allocation. The enhanced method eradicates the need for manual intervention in rule selection, thereby streamlining the process and further reducing costs.

In conclusion, by integrating freight consolidation with order fulfilment allocation, our work creates a novel pathway that addresses the shortcomings of existing methods and offers a practical and versatile alternative. The paper's contributions not only bridge existing gaps in the literature but also open new avenues for future research and applications. The dual focus on enhancing customer satisfaction through maximising WOFR and reducing shipping costs through spatial consolidation represents a significant academic advancement in the fields of order fulfilment and logistics, particularly catering to the needs of small and medium-sized enterprises.

5.3 Methodology

This section presents a novel methodology for order fulfilment allocation, integrating freight consolidation. Our approach leverages customer data and historical order data to extract valuable insights and patterns for consolidation.

The methodology is divided into two parts. The first part, as detailed in 5.3.1, comprises three principal stages: Data Preparation, spatial Consolidation, and Order Fulfilment Optimisation. The second part, described in 5.3.2, addresses the validation process, encompassing data preparation, simulation experiments, and performance evaluation to benchmark the proposed methodology against other existing methods.

5.3.1 Proposed Methodology

Our proposed methodology starts with the Data Preparation stage, transforming raw data into valuable inputs. Subsequently, during the Spatial Consolidation stage, we derive consolidation keys, which are a mapping of orders to the customer's collection points. Collection Point (CP), also known as *drop-off locations* or *relay points*, is the designated delivery location for orders. Lastly, the Order Fulfilment Optimisation stage leverages a MILP model

to solve the order fulfilment allocation for the current planning horizon. For instance, if the company has a planning horizon of one day, these three steps will be executed each day to determine which orders will be fulfilled.

The sequential flow of the proposed methodology, as illustrated in Figure 5.1, is detailed in the subsequent sections.

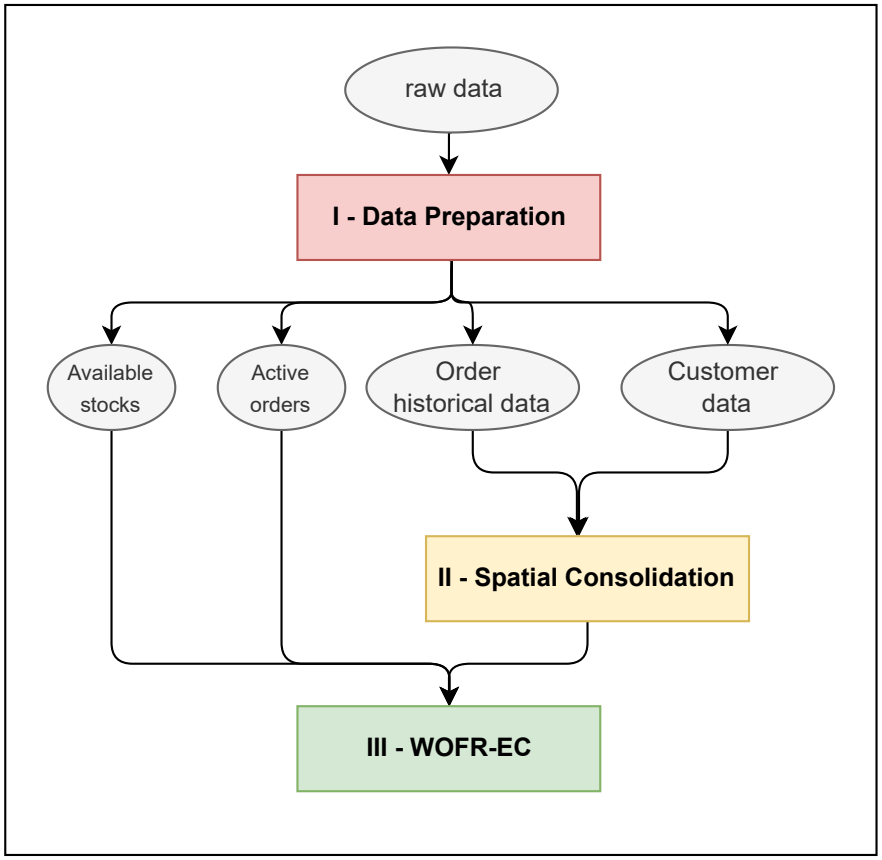

Figure 5.1 Proposed methodology

5.3.1.1 Phase 1 - Data preparation

The preliminary stage of our proposed methodology is data preparation. In this stage, we collect and pre-process customer and order data.

Inputs for spatial consolidation To utilise the spatial consolidation algorithms described in section 5.3.1.2, certain information is required. Specifically, the necessary attributes for order history data include the order ID, creation date, and customer ID (needed to link to the customer data). For customer data, the required attributes are the customer ID, customer

latitude and longitude, collection point ID, collection point latitude, and longitude.

Inputs for WOFR-EC To solve the MILP model, various inputs must be collected and prepared. For active order data, the attributes include the order ID, collection point ID, order weight (w_o), and order lines info such as item IDs and item quantities. The inventory data should specify the available quantities for each type of item. Delivery constraints should provide information on the maximum number of orders that can be shipped for the current planning timeframe. Additionally, spatial consolidation keys, representing the mapping between orders and their collection point, are obtained from the previous step (spatial consolidation).

Depending on the company's specific goals, additional information might be needed to calculate order weight (i.e., order's relative importance). Further details will be provided in the following section.

Order Weight Formula The weight assigned to each order (w_o) plays a crucial role in determining the optimal solution of the MILP model. This parameter gives the relative priority of each order, considering various factors such as the order's deadline, the importance of the customer, the quantity or value of the items, marketing strategy, profit margins, etc.

This section presents a general formula for calculating the order *weights* (w_o) based on two key factors: the number of days until the order deadline (dud_o) and the relative priority level of the customer (p_o). The factors are defined as follows:

> dud_o: days until deadline (≥ 0), which is calculated as the positive difference between the order's due date and the current date. If the deadline has already passed, the value is set to zero.
>
> p_o: relative priority level of the customer ($0 \leq p_o \leq 1$), which is determined by a predefined ranking system that considers various factors, such as the customer's previous order history, payment history, contractual obligations, and overall importance to the business.

The order weight can be calculated using the following formula, which ensures that orders with a higher customer priority and a closer deadline are assigned a higher weight :

$$w_o = (p_o + 1)/(dud_o + 1) \tag{5.1}$$

The example formula presented in this section can be easily modified to suit the specific needs and goals of the business or organisation.

5.3.1.2 Phase 2 - Spatial Consolidation

In the following sections, we explain how we use freight consolidation and how it contributes to the goal of reducing shipping costs.

Spatial Consolidation using Association Rules To reduce shipping costs through spatial consolidation, we propose a method that consolidates collection points (CPs) by merging them. By decreasing the number of CPs and reallocating customers to the nearest ones, this strategy proves valuable in areas with high CP density and low customer-to-CP ratios, thus maximising freight consolidation opportunities.

This task is executed using an algorithm developed by Aboutalib and Agard [163], employing a data mining technique called Association Rules (AR). By extracting and leveraging customer behaviour patterns from historical data, it selects CPs based on long-term consolidation opportunities to reduce costs. The spatial consolidation keys, outputted by the algorithm, guide the CP grouping. The process is depicted in Figure 5.2.

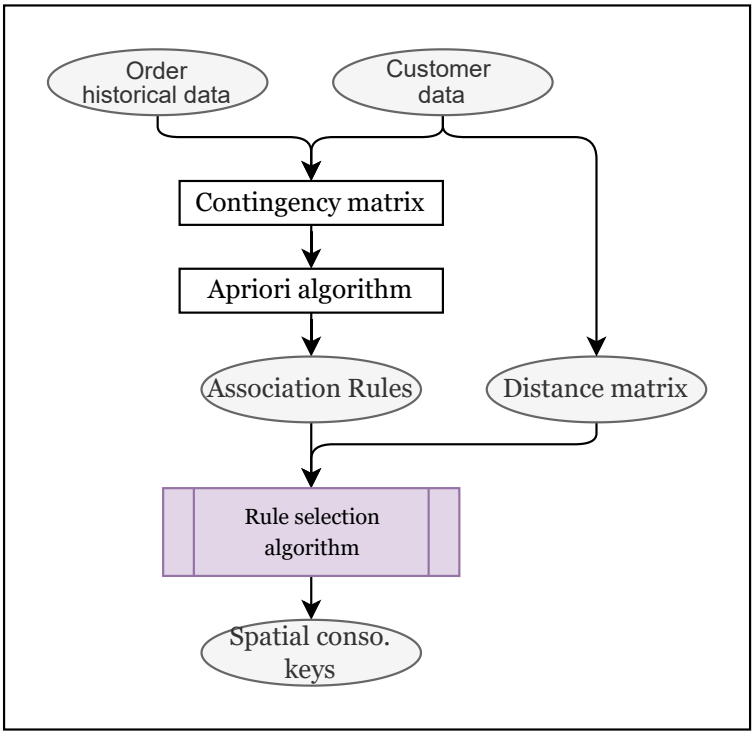

Figure 5.2 Spatial consolidation using Association Rules

However, the original algorithm requires manual intervention for rule selection. To address this limitation, we introduce an automatic rule selection methodology, detailed below.

Automatic AR Selection Our proposed algorithm automates rule selection, considering three criteria. To this end, we will define for each rule r, its metrics d_r, c_r, l_r, for the Distance, Confidence and Lift, respectively.

Distance criterion The distance criterion ensures that consolidated CPs are within reasonable proximity to the farthest customers. It prevents the consolidation of CPs that are too far apart and promotes the consolidation of nearby CPs.

Confidence criterion In the context of this algorithm, confidence can be understood as the likelihood of observing similar order patterns between customers' orders from different collection points (CPs). For example, when an order is made at CP-X, if it is common for an order to also be made at CP-Y, then it indicates strong confidence. Such patterns offer opportunities for spatial consolidations, allowing orders for CP-X and CP-Y to be delivered at the same collection point, thus reducing the number of shipments on a route.

More formally, in the domain of Association Rules (AR), confidence refers to the probability that a given rule's consequent (Y) holds true when its antecedent (X) is true Larose and Larose [114]. The antecedent is the condition of the rule, and the consequent is the outcome. In mathematical terms:
$Confidence = \frac{P(X \cap Y)}{P(X)}$
A high confidence value reveals a robust association between X and Y, making them suitable candidates for spatial consolidations, as long as other conditions are met.

Lift criterion In this context, the lift criterion measures how much more likely two collection points (CPs), such as CP-X and CP-Y, are to appear together in similar order patterns than would be expected by chance.

Formally, lift is a measure of the strength of association between the antecedent (X) and consequent (Y) of a rule. It is calculated by dividing the observed support (i.e., frequency) of both X and Y by the expected support if they were independent Larose and Larose [114]:
$Lift = \frac{P(X \cap Y)}{P(X) \times P(Y)}$
A lift score greater than 1 indicates a positive association, where a score of 1 means that there is no association, and a score that is less than 1 indicates a negative association Larose and Larose [114]. In the context of spatial consolidation, a high lift score means that the collection points are good candidates for consolidation, provided other conditions are also met.

Thresholds and grey areas To enable quantitative and automated rule selection, thresholds must be defined. Using a single threshold can result in unwanted outcomes, even when carefully configured. Instead, employing two thresholds for each criterion is preferred, as it defines three distinct ranges: a red area where rules are automatically rejected, a green area where rules are kept (and accepted if other criteria are not in the red area), and a grey area where no immediate rejection or acceptance occurs.

This dual-threshold strategy offers several advantages. First, it allows interaction between criteria, ensuring that when a score is in the grey zone for one criterion, other criteria can be used to make a decision, thereby reducing ambiguity. Second, it minimises sensitivity to threshold settings, meaning that small changes in the value of a threshold have a less pronounced effect on the results compared to a single-threshold system.

Selection Algorithm The rule selection algorithm facilitates automatic and deterministic selection of association rules by evaluating their quality metrics (confidence and lift) and distance to the farthest customer.

Seven key parameters must be defined:

1. D_0: This represents the maximum distance allowed between a consolidated collection point and its farthest allocated customer, defined as d_r. This strict upper limit is never broken, regardless of how good the other metrics might be between two collection points.
2. D_1: This distance represents a proximity threshold between two collection points. If two CPs are closer than this threshold, a "merge" is strongly recommended. In such cases of merging, customers are redirected to a consolidated collection point, which might be a short travel away from the original CP. Given the minimal distance, customers are likely to find this adjustment convenient for picking up their packages.
3. C_0: This is the minimum confidence required for an association rule to be considered. If the confidence falls below this value, the quality of the association rule is considered too low to justify merging the consolidation points, regardless of the other criteria.
4. C_1: This is a confidence value considered to be "very good," above which a merge is strongly recommended.
5. L_0: This is the minimum lift required for an association rule to be considered. If the lift falls below this value, the quality of the association rule is considered too low to justify merging the consolidation points, regardless of the other criteria.
6. L_1: This is a lift value considered to be "very good," above which a merge is strongly recommended.

7. Z: This threshold value (ranging from 0 to 1) is compared to the normalised score (also ranging from 0 to 1) to resolve ambiguity in the final grey zone.

The algorithm operates in three steps:

1. **Rejection Criteria:** Eliminate all rules that violate any of the rejection thresholds:
   If $(d_r \geq D_0)$ or $(c_r \leq C_0)$ or $(l_r \leq L_0)$ Then reject
2. **Acceptance Criteria:** Accept, from the remaining rules, those meeting a strong-recommendation threshold:
   Else If $(d_r \leq D_1)$ or $(c_r \geq C_1)$ or $(l_r \geq L_1)$ Then accept
3. **Score Calculation and Rule Selection:** For the remaining rules, calculate the normalised score and use it for selection:
   — Calculate:

$$NormalisedScore = \sqrt[3]{\frac{d_r - D_0}{D_1 - D_0} \cdot \frac{c_r - C_0}{C_1 - C_0} \cdot \frac{l_r - L_0}{L_1 - L_0}}$$

   — Select rules based on normalised score:
   If $NormalisedScore \geq Z$ Then Accept Else Reject

Steps (1) and (2) address most ambiguities by leveraging interactions between criteria. Step (3), 'Score Calculation and Rule Selection', evaluates rules in the grey intersection based on their position relative to both the rejection and strong-recommendation thresholds. Within this step, the normalised score is also matched against the seventh threshold (Z), finalising the rule selection.

Synthesis of Spatial Consolidation Phase In the Spatial Consolidation phase, the focus is on leveraging spatial consolidation to reduce shipping costs in areas with a high density of collection points (CPs). The methodology involves merging CPs and reallocating customers, aided by an algorithm that employs Association Rules (AR). This is further enhanced by an automatic rule selection strategy, considering criteria such as distance, confidence, and lift to guide CP grouping. The section introduces the rule selection algorithm, a systematic and deterministic approach that streamlines rule selection, ensuring a nuanced, configurable, and efficient decision-making process.

This methodology not only tailors traditional concepts from Data Mining to the context of freight consolidation and order fulfilment allocation, but also forges a new path that could be adopted in other domains. The spatial consolidation strategy represents a significant contribution to the field, confronting a multifaceted challenge in freight consolidation with a comprehensive and automatic method that recognises and quantifies patterns in customer

order data across different regions. Thus, this contribution also offers a valuable addition to the existing body of knowledge on automated decision-making and pattern recognition.

5.3.1.3 Phase 3 - Optimisation of order allocation fulfilment (WOFR-EC)

In order fulfilment operations such as those found in distribution centres, the recurrent decision process of selecting and delivering a subset of active orders is known as order allocation fulfilment. This is often constrained by limited inventory and delivery factors, leading to some orders being postponed.

Assigning weights (i.e., priority scores) to these orders can prioritise them during selection. To maximise the Weighted Order Fill Rate (WOFR), a problem that can be modelled and addressed through a Mixed Integer Linear Programming (MILP) model, as shown by Rim and Park [8].

Maximising total weight in this context becomes a significant indicator of customer satisfaction. Applying this MILP model, with carefully designed order weights (w_o), offers a path to enhance key performance indicators (KPIs) like service level and bolster overall customer satisfaction.

While the traditional WOFR approach solves order selection based on available inventory, delivery constraints, priorities, deadlines, and customer importance, it overlooks implications on delivery costs and downstream effects within the transport logistics chain.

In the literature, the well-known logistics problem that deals with deciding which orders to fulfil, when to deliver them, and which route to take to minimise shipping costs and meet customer demand is called the Inventory Routing Problem (IRP). The IRP is a complex problem that can help companies to improve customer service and reduce shipping costs simultaneously.

However, implementing an IRP solution typically involves a significant investment, including the cost of software, customisation, ongoing maintenance and support, as well as the labour needed during implementation, hiring and training staff, change management, and other related expenses. Depending on the size, complexity, and nature of a company's logistics operations, implementing an IRP solution may not be a feasible option. In such cases, a simpler and less intrusive approach, such as the one we propose in the next section, can still enable companies to improve customer service and reduce shipping costs.

Our proposed solution, the WOFR Encouraging Consolidation (WOFR-EC) model, offers an alternative to the complex IRP, aiming to reduce shipping costs through spatial freight consolidation. This approach serves as an attractive solution for companies that are unable,

unwilling, or unprepared to implement a more complex IRP solution.

Our model incorporates the concept of a collection point (CP), also known as *drop-off locations* or *relay points*, as the designated delivery location for orders. By consolidating orders delivered to the same CP, we enable a more efficient fulfilment process.

The crux of our strategy lies in minimising the CPs visited during order selection. We introduce a penalty term to the objective function, proportional to the number of visited collection points (V_c), tracked using constraint C5. This encourages geographical consolidation of orders, effectively reducing route stops, serving as a novel and effective method to reduce route costs. In doing so, we present a simpler, less intrusive approach that still enables companies to improve customer service and reduce shipping costs, particularly when an IRP solution may not be practical.

Below, we present the mathematical formulation of our proposed WOFR Encouraging Consolidation (WOFR-EC) model.

Indices

o: order index, where $o = 1, 2, \ldots, O$ for O total orders

i: item index, where $i = 1, 2, \ldots, I$ for I total items

c: collection point index, where $c = 1, 2, \ldots, C$ for C total collection points

Parameters

w_o : weight of order o, where $w_o \geq 0$

s_i : available stock of item i, where $s_i \geq 0$

$d_{o,i}$: demand for item i in order o, where $d_{o,i} \geq 0$

k : delivery capacity (maximum number of orders that can be fulfilled), where $k \geq 0$

$z_{o,c}$: 1 if order o is assigned to collection point c; 0 otherwise

ω : weight given to the desire to minimise the number of visited collection points, where $\omega \geq 0$

M : an arbitrarily large positive number (e.g. total number of orders O would be large enough)

Variables

F_o : 1 if order o is fulfilled; 0 otherwise

V_c : 1 if collection point c is visited; 0 otherwise

Objective function

$$max \left(\sum_{o} F_o w_o - \omega \sum_{c} V_c \right)$$

Subject to

$$\begin{array}{lrcll}
C1 & F_o & \in & \{0,1\} & \forall o \\
C2 & \sum_{o} F_o & \leq & k & \\
C3 & \sum_{o} F_o \cdot d_{o,i} & \leq & s_i & \forall i \\
C4 & V_c & \in & \{0,1\} & \forall o \\
C5 & M \cdot V_c & \geq & \sum_{o} z_{o,c} \cdot F_o & \forall c
\end{array}$$

The developed mixed-integer linear programming model is governed by the following constraints. Constraint (C1) ensures that the fulfilment of order o, represented by F_o, is binary. Constraint (C2) limits the maximum number of orders that can be delivered for the period, guaranteeing that the total fulfilled orders do not exceed capacity k. Constraint (C3) considers the stock constraints, ensuring that the total delivered quantity for an item i does not exceed the available stock s_i. Constraint (C4) specifies that the visitation status of the collection point c, denoted by V_c, is binary. Finally, constraint (C5) ensures that if at least one fulfilled order is assigned to collection point c, the collection point must be flagged by setting V_c to 1

With a ω value of 0, the model is equivalent to the WOFR maximisation, meaning that the total weight is optimised without any consideration for the number of visited collection points. However, when ω is set to a value larger than 0, the number of visited collection points is considered, which encourages consolidation. A larger value of ω indicates a greater willingness to trade off customer satisfaction for lower shipment costs.

This mathematical model forms a holistic approach to managing order fulfilment allocation and spatial consolidation in outbound logistics, achieving a balance between customer satisfaction and cost reduction. By carefully choosing order weights (w_o) and spatial consolidation weight (ω), we can achieve a balance between maximising customer satisfaction and minimising the number of deliveries. By enabling a balance between these often conflicting goals, our WOFR-EC model contributes a novel perspective to the literature.

5.3.2 Validation methodology

Our validation methodology is designed to evaluate the proposed methodology through a comparative study with other order allocation fulfilment methods. It consists of three main

stages: data preparation, simulation experiments, and performance evaluation, as shown in Figure 5.3 and detailed below.

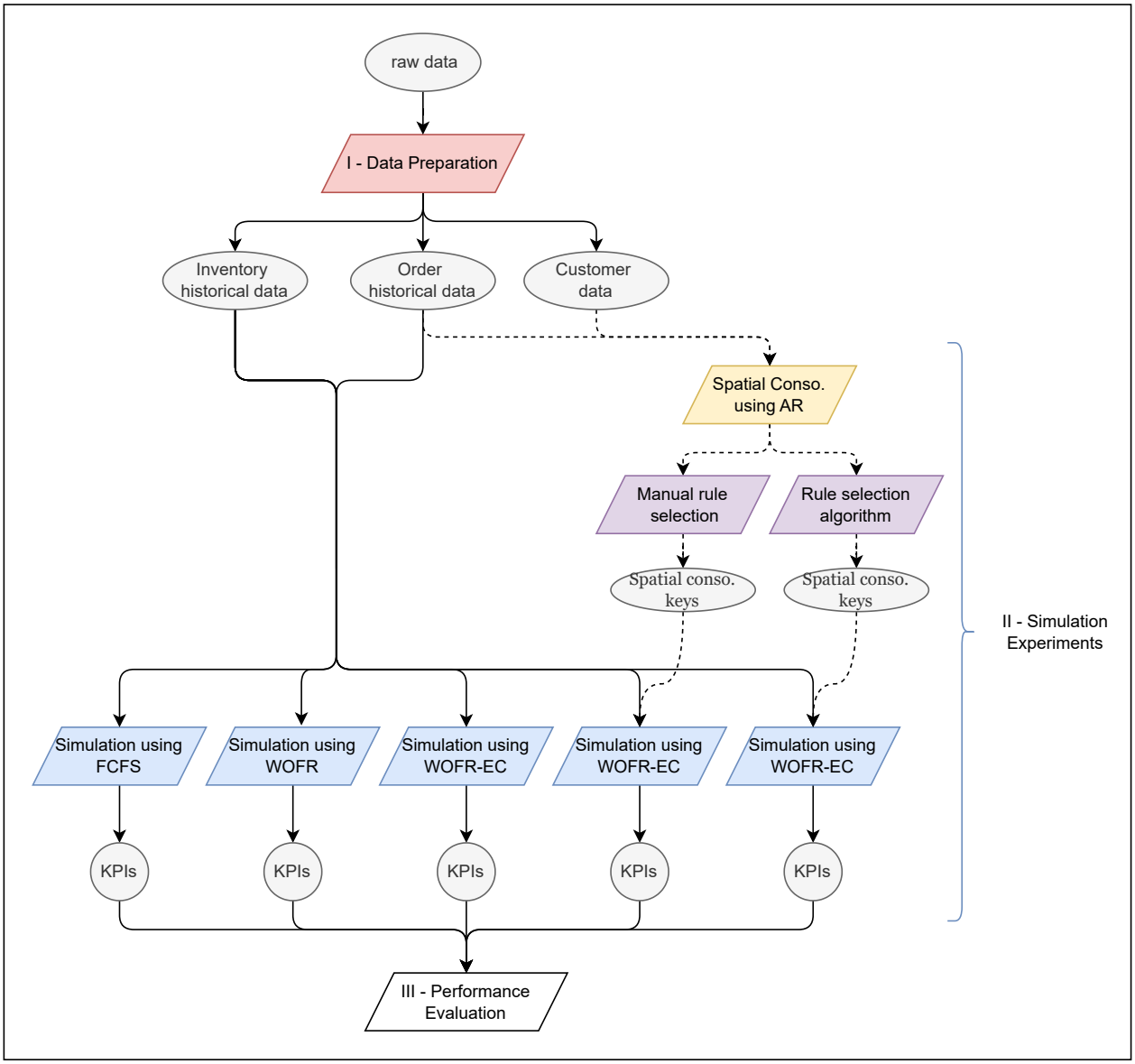

Figure 5.3 Validation Methodology

5.3.2.1 Data Preparation

The goal is to collect and prepare the necessary data to be used in the subsequent steps. The required datasets encompass past customer orders and inventory receptions, which should include the following attributes:

— **Historical Order Data**: Order ID, creation date, collection point ID, order lines info (item ID, and quantities), as well as all the necessary metadata to calculate order weights (w_o), such as customer importance or profit margin.

— **Inventory Historical Data**: Daily received quantities for each item, and the initial state of the inventory.

— **Delivery Constraints**: Maximum number of orders that can be shipped per day.

Examples of these data structures are provided in Appendix B.1 (refer to Tables B.1 to B.5).

5.3.2.2 Simulation Experiments

Simulation experiments will evaluate key performance metrics using different methods over a 30-day historical dataset. We will compare the following order fulfilment allocation approaches:

1. FCFS (Common approach; see Appendix B.2)
2. WOFR (State of the art; see Appendix B.3)
3. WOFR-EC (see section 5.3.1.3)
4. WOFR-EC, with Spatial Consolidation using AR (see section 5.3.1.2)
5. WOFR-EC, with Spatial Consolidation using AR and the rule selection algorithm (Proposed methodology, see section 5.3.1).

Algorithm 1 Simulation

t_0 = the first day of the simulation
Get the initial list of *open* orders (index "o") using Table B.1
Get the initial stocks (s_i) using Table B.3
for t in range(dates) **do**
    Add new orders (date created = t, from Table B.1)
    Prepare the matrix of demand ($d_{o,i}$) using Table B.2
    Add new stock (quantity received = t, from Table B.4)
    Calculate the updated weight (w_o) for all open orders (using the function)
    Get the value of k (date = t, from Table B.5)
    Apply order fulfilment allocation approach (the one we are evaluating)
    Remove fulfilled orders ($F_o = 1$, based on the solution)
    Store intermediate results for KPIs
end for
Evaluate KPIs

5.3.2.3 Performance Evaluation

In the final step, we compare our methodology's performance with that of existing methods in terms of key performance indicators (KPIs). The following KPIs were used for comparison:

1. the number of shipments in the routes during the simulation, which is equivalent to the total number of collection points (CPs) visited during the simulation. For each day, the number of unique collection points visited is counted, and these numbers are then summed to obtain the total (e.g., if CP1 and CP2 were visited on day 1, and CP2 and CP3 were visited on day 2, the total would be 4).
2. Number of Orders fulfilled during the simulation (OF).
3. Number of Orders fulfilled on Time (before or on their due date) during the simulation (OFT).
4. Number of high-Priority Orders fulfilled on Time during the simulation (POFT).
5. Service Level (SL): OFT/(OF + Number of Open Orders that are past their due date).
6. Priority Service Level (PSL): POFT/(Number of high-Priority Orders fulfilled during the simulation + Number of high-Priority Open Orders that are past their due date).

Note that an open order is an order that has been created but has not yet been fulfilled.

By conducting this structured validation methodology, we critically evaluate and compare the performance of the proposed methodology with existing methods, substantiating its efficiency and effectiveness.

5.4 Experiment and contribution

We have investigated the order fulfilment allocation problem within the context of a Canadian uniform distributor. Herein, the integration of various methodologies, including spatial consolidation and Weighted Order Fill Rate (WOFR), was assessed.

5.4.1 Case Study Context

The chosen case study revolves around a Canadian firm that manages the uniform program for several corporations and governmental bodies, encompassing design, manufacture, and quality assurance tasks. We focus on governmental clients availing a "clothing online" service, similar to e-commerce. The dataset contains orders delivered across Canada from a single warehouse. A major challenge currently faced by the company's distribution centre is optimising order fulfilment allocation, adhering to inventory constraints and delivery limitations

while ensuring optimal customer satisfaction. This relates to the larger research community by presenting a typical problem faced by many distribution centres.

This study will evaluate diverse strategies, leveraging a genuine dataset spanning thousands of items and orders. Simulations run over 30 days, as per the approach outlined in section 5.3.2.2. The order weights are recalculated daily during the simulation using a function that takes an order's relative due date and the customer's priority as parameters.

5.4.2 Benchmark: FCFS Approach

We begin by assessing the common First-Come-First-Served (FCFS) strategy. This method delivers orders straight to the recipient's address, devoid of spatial consolidation. The algorithm is presented in Appendix B.2.

As previously defined in section 5.3.2.3, the KPIs used in this study include the number of shipments in the routes; number of Orders fulfilled (OF); number of Orders fulfilled on Time (OFT); number of high-Priority Orders fulfilled on Time (POFT); Service Level (SL); and Priority Service Level (PSL). The findings are catalogued in Table 5.1.

Table 5.1 Results with FCFS Approach

Stops	**OF**	**OFT**	**POFT**	**SL**	**PSL**
6598	1719	1303	194	29%	28%

5.4.3 Exploring Spatial Consolidation

A basic form of spatial consolidation was introduced, where customers are allotted to nearby collection points. The ensuing KPIs, obtained via simulation, indicate a 43% reduction in the total stops. The outcomes are presented in Table 5.2.

Table 5.2 Results with FCFS and Consolidation

Stops	**OF**	**OFT**	**POFT**	**SL**	**PSL**
3731	1719	1303	194	29%	28%

5.4.4 Incorporating WOFR

Here, we address the order allocation problem from an Order Fill Rate (OFR) perspective. The WOFR strategy, presented in Appendix B.3, prioritises customer satisfaction, order ur-

gency, and deadlines. Simulation results, demonstrated in Table 5.3, showed a significant enhancement in order fulfilment. The models were solved using the Python package PuLP and the solver Cbc (Coin-or branch and cut), an open-source mixed integer linear programming solver written in C++. All solutions obtained were optimal.

Table 5.3 shows the results of the simulation.

Table 5.3 Results with WOFR

Stops	**OF**	**OFT**	**POFT**	**SL**	**PSL**
5715	3462	3027	325	62%	45%

5.4.5 Encouraging Consolidation with WOFR-EC

To further enhance efficiency via spatial consolidation, we applied the WOFR-EC model. The intention is to optimise customer satisfaction whilst cutting down shipping costs. Through the WOFR-EC approach, we discerned that by judiciously selecting the parameter ω (i.e., weight given to the desire to minimise the number of shipments), shipping costs could be minimised without substantially compromising service quality. Detailed results are displayed in Table 5.4, and a sensitivity analysis is provided in Figure 5.4.

With a value of 0, the penalty term becomes null, which makes the WOFR-EC equivalent to the WOFR model. The greater the value of ω, the more we are inclined to compromise customer satisfaction for lower shipping costs.

With a low ω value (between 1 and 6.5), we can improve the consolidation without affecting the KPIs. Indeed, the total number of visited collection points (CP) decreases by 1%, and the customers get the same service levels.

With higher ω values, we get some interesting trade-offs. For $\omega = 7$, we can decrease CP by 3%, and in exchange, we reduce the number of fulfilled orders (OF) and the service level (SL) by 2%. The service level of the priority orders (PSL) remains intact. For $\omega = 7.5$, we can decrease CP by 13%. In exchange, we reduce the number of fulfilled orders (OF) by 3%, and the service level (SL) by 8%. The service level of the priority orders (PSL) remains intact. For a value ω between 8 and 10, we can decrease CP by 21%. In exchange, we reduce the number of fulfilled orders (OF) by 3%, and the service level (SL) by 15%. The service level of the priority orders (PSL) remains intact.

In Figure 5.4, the sensitivity analysis illustrates the effect of varying omega values on multiple KPI ratios obtained with the WOFR-EC model. The X-axis represents omega values, and

Table 5.4 Results with WOFR-EC

ω	Stops	OF	OFT	POFT	SL	PSL
0	5715	3462	3027	325	62%	45%
1	5672	3468	3032	325	62%	45%
2	5672	3469	3032	325	62%	45%
3	5672	3468	3032	325	62%	45%
4	5668	3470	3032	325	62%	45%
5	5668	3470	3032	325	62%	45%
6	5669	3474	3034	324	62%	45%
6.5	5669	3474	3034	324	62%	45%
7	5547	3401	2963	324	61%	45%
7.5	4971	3342	2754	326	57%	45%
8	4498	3371	2593	327	53%	45%
9	4497	3372	2593	327	53%	45%
10	4497	3372	2593	327	53%	45%
15	4342	3336	2488	326	51%	45%
20	4228	3345	2425	329	50%	45%
25	4163	3337	2362	329	49%	45%
30	4147	3327	2341	327	49%	45%

the Y-axis displays the corresponding KPI ratios. The trends reveal the model's performance sensitivity to omega changes. Key observations include:

— For ω between 1 and 6.5, KPI ratios remain stable with a 1% increase.

— At ω of 7.5, KPI ratios rise between 5% (OFT/stops) and 15% (POFT/stop and PSL/stop).

— For ω between 8 and 10, ratios increase between 9% (OFT/stop) and 27% (POFT/stop and PSL/stop).

These insights help identify the optimal ω range for the desired model performance.

5.4.6 AR consolidation

Diving deeper into our dataset, we discerned patterns through the application of Association Rules (AR), revealing concurrent ordering tendencies among neighbouring clients. By targeting specific high-frequency order regions, better consolidation can be achieved.

To ensure the validity of our simulation results, we will be using historical orders that precede the 30-day simulation period outlined in section 5.3.2.2. This approach avoids the issue of using the same dataset for both training and testing.

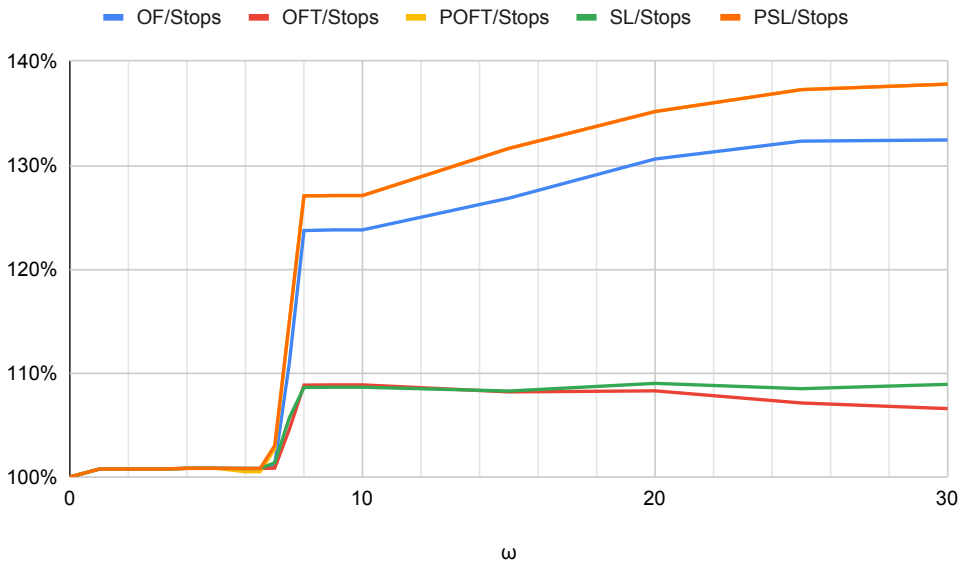

Figure 5.4 Sensitivity Analysis of KPI Ratios to ω using WOFR-EC Model

An example of the input data is available in Table 5.5.

5.4.6.1 ARs Generation

To prevent assigning customers to collection points that may be inconveniently distant, it's necessary to rely on a distance matrix. This matrix delineates, for every combination of collection points (CPs), the highest distance between a customer and the CP to which they would be reassigned. An illustrative instance can be found in Table 5.6, where the example demonstrates that merging CP1 with CP8 would be suitable from a distance perspective. This is because the farthest customer's travel distance to access their packages would be just 1 km. On the other hand, cases like CP1 and CP2 are unfavourable for consolidation, as this would necessitate the farthest customer travelling a distance of 177 km.

We can generate a contingency matrix from Table 5.5. This matrix takes the start date of the week as an index, the collection points (CPs) as columns, and employs boolean values to indicate the presence of placed orders. The resulting matrix resembles the one depicted in Table 5.7. Upon examining this table, we observe, for instance, that during the initial week (2021-02-08), customer orders were received from those allocated in CP1. However, during the same week, no orders were recorded from customers allocated in CP2 to CP8.

We utilise the *Apriori* algorithm, a popular method in Data Mining used to uncover frequent itemsets within a large dataset, to analyse the contingency matrix. Through the application

Table 5.5 Input data for AR consolidation (sample)

ORDER_ID	ENTRY_DATE	CUSTOMER	CP	LNG	LAT
O1	2021-03-15	C1	CP1	LNG1	LAT1
O2	2021-02-14	C2	CP1	LNG2	LAT2
O3	2021-03-17	C3	CP1	LNG3	LAT3
O4	2021-04-28	C4	CP1	LNG4	LAT4
O5	2021-04-06	C5	CP1	LNG5	LAT5
O6	2021-03-15	C6	CP2	LNG6	LAT6
O7	2021-03-15	C7	CP3	LNG7	LAT7
O8	2021-04-20	C7	CP3	LNG7	LAT7
O9	2021-03-24	C8	CP3	LNG8	LAT8
O10	2020-12-24	C8	CP3	LNG8	LAT8
O11	2021-04-13	C8	CP3	LNG8	LAT8
...	...	...	...	...	...

Table 5.6 Distance matrix (sample)

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	...
CP1	0	177	144	43	26	264	255	1	12	...
CP2	177	0	37	151	165	89	80	176	189	...
CP3	144	37	0	115	131	121	111	144	156	...
CP4	43	151	115	0	19	235	225	44	54	...
CP5	26	165	131	19	0	251	241	27	36	...
CP6	264	89	121	235	251	0	11	264	277	...
CP7	255	80	111	225	241	11	0	255	267	...
CP8	1	176	144	44	27	264	255	0	12	...
CP9	12	189	156	54	36	277	267	12	0	...
...	...	...	...	...	...	...	...	...	...	...

of this algorithm, we generate *frequent sets*, which denote combinations of collection points that exhibit recurring occurrences. These frequent sets can provide insights into common patterns and relationships within the data. A small sample of these sets is presented in Table 5.8 as an illustration. For instance, the table highlights that in 12.5% of the weeks, there are instances where both an order is placed by customers in CP1 and another by customers in CP19. This table provides the support (or frequency) for all collection points and combinations thereof, thereby allowing us to unveil valuable insights regarding collection points that share similar patterns and frequently place orders concurrently.

Subsequently, leveraging the *frequent sets*, we proceed to compute Association Rules (AR). Additionally, the distances for pairs of collection points (CPs), as depicted in Table 5.6, are

Table 5.7 Contingency matrix (sample)

	CP1	**CP2**	**CP3**	**CP4**	**CP5**	**CP6**	**CP7**	**CP8**	...
2021-02-08	1	0	0	0	0	0	0	0	...
2021-02-15	0	0	0	0	0	0	0	0	...
2021-02-22	0	0	0	1	0	1	0	0	...
2021-03-01	0	0	0	0	0	0	0	0	...
2021-03-08	0	0	0	0	0	1	0	0	...
2021-03-15	1	1	1	0	0	0	0	0	...
2021-03-22	0	0	1	0	0	1	0	0	...
2021-03-29	0	0	0	0	0	1	0	0	...
2021-04-05	1	0	0	0	0	1	0	0	...
2021-04-12	0	0	1	1	1	1	1	1	...
2021-04-19	0	0	1	0	0	1	0	0	...
2021-04-26	1	0	0	0	0	0	0	0	...
...	...	...	...	...	...	...	...	...	...

incorporated, as they play a pivotal role in rule selection. An illustrative instance of Association Rules, along with their corresponding metrics, is provided in Table 5.9. To comprehend the interpretation of the metrics, let's consider the first rule (topmost row) as an example. This rule pertains to the relationship between CP1 and CP19. The confidence of 100% indicates that whenever an order was placed in CP1, there was invariably a corresponding order placed in CP19 during the same week. This high confidence value signifies a strong association between the customer demand patterns from these two collection points. Moreover, the lift value of 2.91 offers valuable insight. It signifies that the occurrence of concurrent orders in CP1 and CP19 is three times more likely than if the orders were happening by chance alone. In other words, the lift value of 2.91 indicates that the likelihood of simultaneous orders in these two CPs is three times higher than what would be expected based solely on their individual order frequencies. This suggests a strong and significant relationship between CP1 and CP19, where orders in one CP are highly correlated with orders in the other CP, far beyond random chance.

5.4.6.2 ARs Sequential Selection

To identify the most impactful Association Rules (ARs) for consolidation, we implement a sequential selection approach based on various metrics, namely distance, support, confidence, and lift. The criteria for rule selection, as detailed in 5.3.1.2, are informed by prior literature and relevant statistical considerations Larose and Larose [114].

Initially, we filter out rules with low support (below 5%), effectively eliminating noise and

Table 5.8 Frequent sets (sample)

support	**itemsets**
0.125	{CP1, CP19}
0.125	{CP103, CP112}
0.1875	{CP103, CP19}
0.125	{CP407, CP103}
0.125	{CP486, CP103}
0.15625	{CP6, CP103}
0.1875	{CP122, CP19}
0.125	{CP126, CP72}
0.125	{CP226, CP271}
...	...

Table 5.9 Association Rules (sample)

antecedent	**consequent**	**support**	**confidence**	**lift**	**distance**
CP1	CP19	0.125	1	2.91	3.88
CP103	CP112	0.125	0.667	3.56	6.57
CP103	CP19	0.188	1	2.91	266.48
CP110	CP267	0.125	0.8	5.12	28.41
CP122	CP146	0.125	0.5	4	51.75
CP407	CP103	0.125	1	5.33	269.68
CP486	CP6	0.125	1	2.46	152.17
...	...	...	...	...	...

reducing the number of rules to about 8000. Next, we focus on rules with a distance value under 5 km, recognising that customer impact must be minimised. This further reduces the count to 72 rules.

Subsequently, we discard rules with a lift below 1.2, emphasising rules that represent a useful pattern rather than mere coincidence. All rules met this criterion, so none were discarded. We then scrutinise the rules based on confidence, eliminating those with a value under 0.7. This leaves 26 rules that signify a high probability of enhancing consolidation efficiency. Table 5.10 provides details of the selected rules.

5.4.6.3 Evaluation

The potential impact on freight consolidation of the selected ARs is assessed through simulations with the WOFR-EC model. This builds upon the methodologies explained in previous

Table 5.10 Selected Rules

antecedent	consequent	support	confidence	lift	distance
CP127	CP29	0.0625	1	2.67	2.62
CP150	CP549	0.0625	1	10.67	1.54
CP152	CP61	0.0625	1	16.00	4.09
CP162	CP19	0.0625	1	2.91	2.56
CP18	CP92	0.0625	1	4.57	3.45
CP263	CP19	0.0625	1	2.91	3.88
...	...	...	...	...	...

sections 5.4.4 and 5.4.5, adapting them to apply the 26 chosen rules.

Specific reallocations are implemented according to the selected rules, and results are summarised in Table 5.11.

Table 5.11 Results with WOFR-EC & AR consolidation

ω	Stops	OF	OFT	POFT	SL	PSL
6.5	5617	3474	3034	325	62%	45%
7	5501	3405	2967	325	61%	45%
7.5	4933	3346	2763	326	57%	45%
8	4460	3372	2602	330	54%	45%

Compared to the prior results of WOFR-EC without AR-based consolidation (Table 5.4), the total number of stops decreases by 1%, while the other KPIs remain unchanged (or were slightly improved).

5.4.7 AR consolidation using the rule selection algorithm

Building upon the previous section, the rule selection algorithm refines the rule selection process by facilitating the simultaneous evaluation of multiple criteria. This ensures a more comprehensive consideration of edge cases. For example, four rules were previously rejected for being slightly (within 1%) over the 5 km limit in distance, despite having excellent confidence and lift values (100% and >6); and two rules were rejected because the confidence value was slightly (within 3%) under the limit, although the distance and lift were both excellent (<2 km and >10).

In accordance with specific requirements and objectives, the following thresholds were established, resulting in the acceptance of 49 rules :

1. $D_0 = 6$: Maximum allowed distance between a consolidated collection point and its farthest allocated customer.
2. $D_1 = 3$: "Very close" distance between two collection points, under which a "merge" is strongly recommended.
3. $C_0 = 0.6$: Minimum allowed confidence for an AR.
4. $C_1 = 0.8$: "Very good" confidence, above which a "merge" is strongly recommended
5. $L_0 = 1.2$: Minimum allowed lift for an AR.
6. $L_1 = 2.0$: "Very good" lift, above which a "merge" is strongly recommended
7. $Z = 0.5$

This new approach adds flexibility to the selection process by defining green, red, and grey zones for different thresholds, enabling more nuanced decision-making. The rule selection algorithm makes defining and applying these thresholds more intuitive and manageable. This innovative method overcomes the limitations of the previous selection process, where each criterion was applied with a single, inflexible threshold, leading to a highly sensitive evaluation. The new approach enables more nuanced, accurate, and flexible rule selection.

5.4.7.1 Evaluation

As in the previous section, to evaluate the impact on freight consolidation, we will conduct simulations with the WOFR-EC model. We will repeat the same process as in previous sections 5.4.4, 5.4.5, and 5.4.6; however, we will apply the rules (that were selected using the rule selection algorithm) to the simulation dataset. Results are shown in Table 5.12.

Table 5.12 Results with WOFR-EC & AR consolidation & 7T

ω	**Stops**	**OF**	**OFT**	**POFT**	**SL**	**PSL**
6.5	5541	3474	3034	325	62%	45%
7	5433	3411	2975	325	61%	45%
7.5	4868	3355	2774	326	57%	45%
8	4407	3381	2619	330	54%	45%

Compared to the results with WOFR-EC (Table 5.4), the total number of stops decreased by 2%, while the other KPIs remain unchanged (or slightly improved). In other words, by applying the AR consolidation methodology combined with the rule selection algorithm, we can further reduce delivery costs.

5.5 Results

For ($\omega = 6.5$), Table 5.13 summarises the results obtained in the previous sections for WOFR-EC and its derivatives.

Table 5.13 Results Summary

	Stops	**OF**	**OFT**	**POFT**	**SL**	**PSL**
FCFS without consolidation	6598	1719	1303	194	29%	28%
FCFS with consolidation	3731	1719	1303	194	29%	28%
WOFR	5715	3462	3027	325	62%	45%
WOFR-EC	5669	3474	3034	324	62%	45%
WOFR-EC & AR	5617	3474	3034	325	62%	45%
WOFR-EC & AR & 7T	5541	3474	3034	325	62%	45%

We observe that 1) By introducing spatial consolidation, even in a straightforward form, we can significantly reduce the burden on shipping (43% reduction in the total number of visited CPs) 2) Compared to the FCFS algorithm, by using a MILP model, we can significantly improve customer service levels, but it comes naturally with an increased delivery cost. The stops/OF ratio, however, is considerably lower (1.65 vs 2.17), which means that the cost per order can decrease. 3) By replacing the WOFR model with our variation (WOFR-EC), we can encourage the optimiser to increase consolidation. Indeed, we can decrease the stops by 1% (without reducing the other KPIs). 4) By increasing the ω parameter value, we get even better trade-offs and ratios. At $\omega = 8$, for instance, the number of shipments is reduced by 21%, in exchange for only a 3% decrease in fulfilled orders. 5) By applying our Association Rules consolidation methodology, we can decrease the stops by another 1% (without reducing the other KPIs). 6) By applying our rule selection algorithm, we can decrease the stops by an additional 1% (without reducing the other KPIs).

The proposed methodologies lead to fulfilling more orders while simultaneously reducing the total number of stops, underscoring the value of this research in both academic and practical contexts.

5.6 Conclusion

This paper introduced a novel, data-driven methodology for order fulfilment allocation that strategically integrates spatial consolidation to balance the often-conflicting objectives of maximising customer satisfaction and minimising shipping costs. By moving beyond traditional aggregated approaches, our methodology leverages historical customer order data to

uncover latent consolidation opportunities, providing a more granular and effective decision-making framework. The simulation experiments, grounded in a real-world case study, confirm that our proposed approach significantly outperforms conventional methods like First-Come-First-Served (FCFS) and standard Weighted Order Fill Rate (WOFR) optimisation, leading to more fulfilled orders with fewer delivery stops.

This research offers two primary contributions to the logistics and supply chain management literature. First, we developed the WOFR-EC (Weighted Order Fill Rate - Encouraging Consolidation) model, a novel mixed-integer linear program that provides a practical and accessible alternative to the computationally intensive Inventory Routing Problem (IRP). By incorporating a penalty term proportional to the number of unique delivery locations, our model effectively bridges the gap between the order allocation decision and its impact on downstream routing costs, making it particularly valuable for firms that rely on 3PLs. A second key contribution is our automated data-driven consolidation framework. We enhanced an Association Rules (AR) based methodology with an automatic rule selection algorithm, which systematically selects high-value consolidation rules based on multiple criteria. This removes the need for manual intervention, creating a more robust and efficient process for leveraging customer demand patterns to reduce logistics costs.

Despite the promising results, we acknowledge several limitations that define the boundaries of our study. The proposed WOFR-EC model is deterministic and static, assuming all inputs are known and not accounting for real-world uncertainties like demand variability or transport delays. Additionally, the model uses the number of stops as a proxy for transportation costs, which, while effective for encouraging consolidation, does not capture the full complexity of routing costs. The current model also operates on a binary decision of full order fulfilment, not allowing for *partial deliveries*, which could be a practical strategy when inventory is scarce. Finally, the methodology was developed and tested within a single warehouse scope, and its application to more complex networks would require further adaptation.

These limitations offer several promising directions for future research. A primary avenue is to extend the WOFR-EC model into a stochastic or robust optimisation framework to handle uncertainty better. The model could also be enhanced by developing a more accurate cost function, perhaps through a two-stage approach where a VRP heuristic provides cost feedback. Future iterations should also consider allowing for partial deliveries and incorporating in-transit stock for more accurate available-to-promise calculations. Furthermore, while Association Rules proved effective, exploring advanced machine learning techniques like clustering or predictive modelling could yield even more powerful consolidation insights. Finally, the multi-objective nature of the model lends itself well to including sustainability

metrics, such as a third objective to minimise carbon emissions, creating a more holistic decision-making framework that aligns with modern supply chain challenges.

CHAPITRE 6 ARTICLE 3: RL-BASED INVENTORY CONTROL FOR NON-STATIONARY DEMAND AND DYNAMIC LEAD TIMES

Zineb Aboutalib, Bruno Agard

Submitted to Computers and Operations Research

Submitted: 15-Aug-2025, under review

Abstract *— This paper introduces a novel inventory control system that integrates multi-agent reinforcement learning (RL) with quantile forecasting and mixed-integer linear programming to address complex challenges faced by distribution-centric businesses. Inspired by collaboration with an industrial partner, the system directly tackles real-world complexities such as multiple SKUs with interdependent demand patterns, tiered service level agreements, non-stationary and intermittent demand profiles, and dynamic lead times. The system combines three key components: a multi-agent Proximal Policy Optimization framework for adaptively adjusting replenishment quantities; an order fulfillment allocation system to optimally distribute available inventory based on SLAs and due dates; and a comprehensive simulation framework that orchestrates these components through a realistic supply chain environment. This data-driven architecture enables customized objective functions that directly align with business goals, such as meeting SLAs, reducing order lateness, and minimizing inventory costs, without requiring simplified assumptions about demand distributions or cost structures. The RL-based system demonstrates a 9% normalized lift over an enhanced baseline, primarily achieved through a 16% reduction in inventory holding costs while maintaining service levels. The system shows particular strength in balancing competing objectives across different SLA tiers, with lower-priority groups seeing improvements of over 75% in some metrics. The work bridges a significant literature gap by tackling a combination of real-world challenges rarely addressed even in isolation within existing solutions. The system's modular architecture provides a foundation for both academic research and industrial applications, offering valuable insights for distribution-centric businesses facing similar complexities.*

Keywords *— Inventory control, Multi-Agent Reinforcement Learning (MARL), Proximal Policy Optimization (PPO), Mixed-Integer Linear Programming (MILP), Quantile Forecasting,*

6.1 Introduction

In today's rapidly evolving market, businesses in the distribution sector are under increasing pressure to make more informed, data-driven decisions in managing their supply chains. Inventory control and allocation decisions, which directly impact inventory holding costs and the ability to meet customer service level agreements (SLAs), are heavily tied to the complex, stochastic dynamics of the supply chain involving fluctuating demand and conflicting business objectives. Inventory control methods and heuristics often fail to address these challenges because they rely on simplified and inaccurate assumptions about demand distributions and cost functions [83].

Emerging machine learning technologies and computational capabilities present opportunities to improve and innovate inventory control systems. Reinforcement Learning (RL), a branch of artificial intelligence that has demonstrated effectiveness in addressing complex Markov Decision Processes (MDPs) representative of real-world systems, offers promising avenues for developing adaptive and intelligent inventory policies that can dynamically respond to changes and signals within the supply chain [85].

However, the application of RL in inventory management is not without its challenges: large observation and action spaces, slow learning speeds, the need for extensive data for model training, and substantial computational effort are significant hurdles that may impede their practical application [83], [165]. Due in part to these challenges, current state-of-the-art inventory control solutions often rely on oversimplified assumptions, overlooking crucial elements of real-world supply chain dynamics.

This research seeks to bridge some of these gaps by integrating more realistic considerations, such as 1) multiple stock-keeping units (SKUs) per customer orders, which introduces intricate interdependencies in item demand and order fulfillment; 2) dynamic lead times for both supplier and customer orders, reflecting the fluid nature of supply chain operations; 3) complex customer demand characterized by nonstationary, seasonal, intermittent, and erratic patterns that defy standard statistical assumptions; 4) tiered customer groups, each with its own set of service levels and lead time agreements, affecting the dynamics of inventory control and order fulfillment allocation; 5) service level objectives that recognize the asymmetric and nonlinear cost implications associated with inventory excesses and shortages, especially in the context of adhering to SLAs.

The data-driven, multi-agents RL-based inventory control system addresses these critical but often overlooked factors, covering inventory replenishments and order fulfillment allocation strategies. It is designed to optimize decisions based on multiple conflicting objectives: reduc-

ing inventory holding costs, minimizing customer order lateness, and honoring SLAs. These elements are essential for maintaining operational efficiency and a competitive edge in the fast-paced and increasingly demanding distribution sector.

This paper is organized as follows : Section 6.2 presents the problem statement, detailing the real-world industrial case that motivates this research. Section 6.3 reviews the state of the art in inventory control—including recent reinforcement learning applications—and supply chain simulation. Section 6.4 describes the architecture and the three components of the inventory control system. Section 6.5 presents the validation methodology, experimental results and analysis, demonstrating the effectiveness of the proposed approach. Finally, Section 6.6 concludes with a discussion of contributions and future research directions.

6.2 Problem Statement

This research is inspired by a real-world industrial case. The industrial partner specializes in providing comprehensive, end-to-end uniform management—from design to distribution—for a client base that encompasses a wide spectrum, from standard commercial accounts to high-stakes governmental bodies with mission-critical supply requirements.

Their operations involve processing user-submitted requests from members of client organisations via a web application, making each demand explicit and traceable through to delivery.

Meeting service level agreements (SLAs) is critical to ensure client satisfaction and long-term partnerships. To this end, the inventory strategy focuses on maintaining sufficient stock levels to fulfill a target percentage of demand at all times. The company operates under a backorder system to mitigate lost sales from stockouts, requiring a proactive replenishment approach, especially as supplier lead times often exceed the delivery lead time commitments made to clients.

Thousands of items and customers, each with distinct priorities, lead times, and target service level (TSL) requirements, complicates inventory control. The operational environment is characterized by intermittent and non-stationary demand patterns, including sporadic large orders, seasonal patterns, and trends that defy the simplified statistical assumptions of traditional inventory control models. Inventory control thus becomes a critical balancing act: maintaining target service levels, which are essential for client trust and business continuity, while minimizing the high inventory costs driven by this demand uncertainty.

This study focuses on two primary decision processes. The first is **order fulfillment allocation (OFA)**, a daily process that determines how available inventory is allocated to customer orders-considering inventory levels, customer service level agreements (SLAs), and

order due dates. The system supports various allocation strategies, from simple heuristics to mixed-integer linear programming optimizations, as detailed in Section 6.4.2. The second is the **inventory control policy (ICP)**, a periodic process that determines optimal replenishment quantities for each item to maintain target inventory levels. Operating on a periodic review basis, this process incorporates demand forecasting, as further described in Sections 6.4.3 and 6.5.1.

While considering this situation, several assumptions are made.

Supply assumptions include unlimited supplier capacity. Replenishment lead times (RLTs) are defined as the time from placing a purchase order to product availability in inventory. These lead times may vary for each purchase but are known at the time of ordering.

Demand is intermittent but exhibits predictable temporal patterns, including weekly fluctuations in overall order volumes, monthly seasonality for specific item-SLA combinations, and long-term annual trends modeled as growth or decline rates. Customer orders often include multiple stock-keeping units (SKUs), creating interdependencies in demand and fulfillment patterns. Once placed, customer orders are never canceled but may be delayed if inventory is insufficient.

Multiple customer SLA groups exist, each with specific target service levels (TSLs), which quantifies the desired probability of fulfilling customer demand without stockouts, and customer lead times (CLTs). This creates a hierarchy in which higher-tier SLA groups may receive priority in inventory allocation. Additionally, a portion of orders are classified as priority orders with stricter service level targets and shorter lead times than their standard SLA group requirements. Partial fulfillment policies, configurable per order, influence how available inventory is allocated when shortages occur.

The inventory control system follows a periodic review policy, with replenishment orders placed at fixed intervals. Customer orders are fully backordered in case of stockouts, meaning no lost sales occur. The system operates in discrete time, processing all daily events—order entry, fulfillment, and inventory receipt—in a specific sequence.

Constraints include a maximum inventory capacity across all items and a daily limit on the number of orders that can be processed and shipped. Additionally, order fulfillment and deliveries do not take place on weekends, reflecting operational limitations at the distribution center.

These assumptions collectively define the operational framework within which the inventory control system operates. While derived from the industrial partner's context, they represent common features of distribution-centric businesses, making the proposed approach applicable

to a wide range of inventory management challenges.

In this context, this research aims to optimize a global performance function that captures the trade-offs between three key metrics crucial for operational excellence: The first metric is **Inventory Costs** (Φ^{stock}), which aim to minimize inventory holding costs. The second is **Order Lateness** (Φ^{late}), targeting the reduction of customer orders delivered past their due dates, while accounting for both the quantity and duration of late items. The third is **Service Level Shortfall (SLS)** (Φ^{service}), which seeks to minimize the gap between the target service level (TSL) and the realized service level (RSL) for each customer group.

To formally define the objective function, we consider performance over discrete time intervals, or time-steps, indexed by t, which correspond to the periodic review cycle. Furthermore, since policies must be sensitive to customer prioritization, the metrics for lateness and service level are evaluated for each distinct Service Level Agreement (SLA) group, indexed by s.

The overall performance of an inventory policy for a given SLA group s at time-step t is thus evaluated using the following global objective function $\Phi_{s,t}$. This function aggregates the smoothed values of the three core metrics ($\bar{\Phi}$), where the bar notation indicates exponential smoothing over time to balance responsiveness with stability:

$$\Phi_{s,t} = \left(\bar{\Phi}_{t}^{\text{stock}} \cdot \beta_{\text{stock}} + \bar{\Phi}_{t,s}^{\text{late}} \cdot \beta_{\text{late}}\right) \cdot (\bar{\Phi}_{t,s}^{\text{service}})^{\beta_{\text{service}}} \tag{6.1}$$

where β_{stock}, β_{late}, and β_{service} are weighting factors representing the relative importance of each metric.

This formulation captures the multi-objective nature of inventory management: the linear combination of inventory and lateness costs reflects their direct trade-off, while the multiplicative relationship with the service level shortfall highlights the compounding business impact of failing to meet SLAs, which can damage customer trust and result in contract losses. The exponent β_{service} introduces a non-linear penalty, providing further control over how strongly service level shortfalls are weighted in relation to costs. The detailed mathematical definitions of each component metric (e.g., how Realized Service Level is calculated) and the temporal smoothing technique are provided in the experimental framework section (6.5.2.1).

6.3 State of the Art

6.3.1 Inventory Control

Inventory management has been a prominent topic in industrial engineering and operations research for over a century [166]. The core decisions center on when and how much to replenish to meet demand while optimizing costs and storage efficiency. However, the landscape of inventory control is undergoing a significant transformation, driven by challenges and opportunities in modern market. Traditional inventory management techniques, while foundational, often struggle to adapt to complex supply chain dynamics [46], as they rely on simplified assumptions about demand distributions and cost functions that rarely hold true in real-world scenarios, leading to suboptimal decision-making, increased costs, and challenges in meeting service level agreements (SLAs) [83].

Jackson, Tolujevs, and Kegenbekov [46] comprehensively reviewed inventory control models, focusing on evolving methodologies for deriving optimal control parameters. The review covers analytical models and non-analytical methods such as control theory, dynamic programming, simulation-based optimization, and metamodeling, discussing each in terms of application, advantages, and limitations.The study emphasizes the need for computationally efficient numerical approximations due to the complexity of business-driven inventory control issues.

The emergence of machine learning, particularly Reinforcement Learning (RL), has opened new possibilities in inventory control. RL's potential to learn and adapt to dynamic environments makes it promising for developing inventory policies that can respond in real-time to supply chain changes [85]. Integrating reinforcement learning techniques in inventory control systems presents a promising approach for optimizing supply chain performance, enabling adaptive inventory policies capable of responding dynamically to changes within the supply chain [167], [168].

Despite the promise of RL for inventory control, practical application has inherent complexities, including handling large observation and action spaces, slow learning pace, extensive data requirements, and computational demands [169]. These challenges are compounded by the need for extensive tuning of neural network hyperparameters, making implementation a subject of ongoing research [84], [169]. Current state-of-the-art RL models often oversimplify real-world supply chain dynamics, assuming stationarity of supplier lead times and customer demand, which does not reflect their fluctuating nature in practice [84]. Many studies focus on simplified models, such as single-SKU systems, which do not capture the complexity of real-world systems.

Decentralized, data-driven solutions Multi-Agent Reinforcement Learning (MARL), marks a significant advancement in managing inventory control problems within complex supply chain networks. By allowing each entity in the supply chain to be controlled by an independent agent, MARL naturally decomposes the problem into manageable sub-problems, enhancing the efficiency and scalability of inventory management systems [88].

Proximal Policy Optimization (PPO) is a particularly promising RL algorithm due to its stability, efficiency, and ability to handle complex scenarios. It excels in managing extensive action spaces while maintaining stability during training [72] - both crucial attributes for inventory management. PPO's adaptability is valuable in dynamic environments with fluctuating demand and supply chain disruptions [89]. Vanvuchelen, Gijsbrechts, and Boute [90] highlights PPO's stable behavior when adapting to complex inventory challenges, reducing the need for hyperparameter tuning compared to Asynchronous Advantage Actor–Critic (A3C) implementations. Kegenbekov and Jackson [85] leverages PPO to develop a deep reinforcement learning agent for synchronizing inbound and outbound flows in supply chains, demonstrating effectiveness in stochastic and nonstationary environments. Gokhale, Trasikar, Shah, *et al.* [89] shows that PPO discovers equivalent or superior policies in significantly fewer epochs than Advantage Actor–Critic (A2C) and Trust Region Policy Optimization (TRPO). Meisheri, Sultana, Baranwal, *et al.* [86] demonstrates that PPO outperforms Deep Q-Network (DQN) in both performance and convergence speed. Liu and Nishi [87] successfully applies PPO to determine ordering and lateral transshipment policies in a three-echelon supply chain, while [170] proposes Integrated Advantage Actor–Critic Proximal Policy Optimization (IACPPO), combining A2C and PPO to optimize warehouse inventory replenishment.

In summary, conventional inventory management approaches frequently fall short in adapting to the evolving and unpredictable characteristics of contemporary supply chains. The adoption of reinforcement learning introduces adaptive strategies that can respond in real-time to changes within the supply chain. Nonetheless, the practical implementation faces obstacles, including the requirement for substantial data, high computational resources, and the intricacy of real-world dynamics that existing models do not entirely address.

6.3.2 Supply Chain Simulation

In the rapidly evolving landscape of supply chain management, simulation techniques have become essential for developing agile, data-driven inventory management strategies, providing a robust framework for testing, analyzing, and optimizing complex supply chain dynamics. Researchers utilize simulation modeling to assess the impact of inventory management decisions on overall supply chain performance. Drakaki and Tzionas [171] provided a granular analysis

of multi-stage serial supply chains under various operating conditions, including disruptions, demonstrating the potential of simulation modeling to augment supply chain efficiency.

Simulation is also valuable for mitigating the bullwhip effect. Guo, Han, Wu, *et al.* [172] demonstrated the efficacy of digital twin technology in mitigating this effect within closed-loop supply chains through robust control strategies informed by simulations. The adoption of simulation technology for real-time supply chain modeling represents a significant advancement. By creating a digital replica of the physical supply chain, simulation enables enhanced decision-making and scenario analysis, bolstering resilience and responsiveness of operations [172], [173].

Simulation is particularly effective in addressing the challenges of aligning inventory with fluctuating demand. Pamulety and Pillai [174] advocates for the superiority of data-driven inventory policies over intuition-based decisions, highlighting the impact of ordering decisions on supply chain performance. Simulation enables examination of supply chain sensitivity to various parameters and development of robust policies under diverse scenarios. Perez, Hubbs, Li, *et al.* [175] conducted a comparative analysis of deterministic and stochastic modeling alongside reinforcement learning within a simulated environment, shedding light on varied impacts of these approaches on profit, service levels, and inventory policies.

Researchers leverage simulation for multi-objective optimization in supply chains. Azaron, Venkatadri, and Farhang Doost [176] used a multi-objective stochastic programming model within a simulated setting to maximize profit and minimize travel times for unsatisfied customers, demonstrating the potential of simulation in designing profitable and responsive supply chains when facing uncertainty. Recent advancements integrate simulation technology with real-time data for enhanced supply chain efficiency. Wu, Zhang, Li, *et al.* [177] proposed a scheme for real-time robust optimization of perishable supply chains based on digital twins, demonstrating the effectiveness of simulation in reducing inventory and production variability, even in complex environments. A powerful approach involves integrating stochastic simulation with optimization techniques, allowing for development of highly adaptable decision-making tools, especially for managing risks associated with supply chain inventory policies [178].

6.3.3 Synthesis and Research Gap

The state-of-the-art review reveals a clear evolution in inventory control, moving from traditional methods with static assumptions to more data-driven approaches that leverage the power of Reinforcement Learning and simulation technologies. While traditional methods struggle to handle the dynamic complexities of modern supply chains, RL offers a promising

framework capable of developing adaptive inventory policies.

However, bridging the gap between theoretical advancements and practical applicability in inventory control requires addressing key challenges. Many state-of-the-art RL-based approaches still rely on simplifications that do not hold in practice, such as stationary demand, simplified or absent SLA constraints, constant lead times, and single-SKU systems [84]. This disconnect between theoretical models and the multifaceted reality of supply chains limits the effectiveness of current RL-based solutions.

To bridge these gaps, this research introduces a novel inventory control system designed to directly address these challenges.

The approach combines several key components: a **multi-agent Proximal Policy Optimization (MARLPPO)** method to enhance learning efficiency and scalability in complex supply chain environments; a **comprehensive simulation framework** that captures the intricate dynamics of supply chains, including multiple SKUs, dynamic lead times, non-stationary demand patterns, and diverse SLA constraints; a **mixed integer linear programming (MILP) model** that optimizes daily order fulfillment allocation decisions while accounting for operational constraints; and **quantile forecasting**, which provides a more accurate representation of demand variability and enables agents to make better-informed decisions without relying on simplistic demand assumptions.

6.4 Proposed RL-Based Inventory System

6.4.1 System Architecture and Components

This section provides a detailed description of our primary scientific contribution: a novel, multi-component inventory control system designed to address complex, real-world supply chain challenges. We present the system's architecture, its core decision-making modules—including the Order Fulfillment Allocation (OFA) solver and the multi-agent Reinforcement Learning (MARL) framework—and the operational components required for its function, such as the simulation environment. The methodology used to validate this system is presented separately in Section 6.5.

The proposed system is structured around three core, synergistic modules that handle distinct aspects of inventory management

The architecture, illustrated in Figure 6.1, is built around two primary, recurring decision loops that interact with the central Supply-Chain Simulator. These loops operate on different timescales and serve distinct purposes.

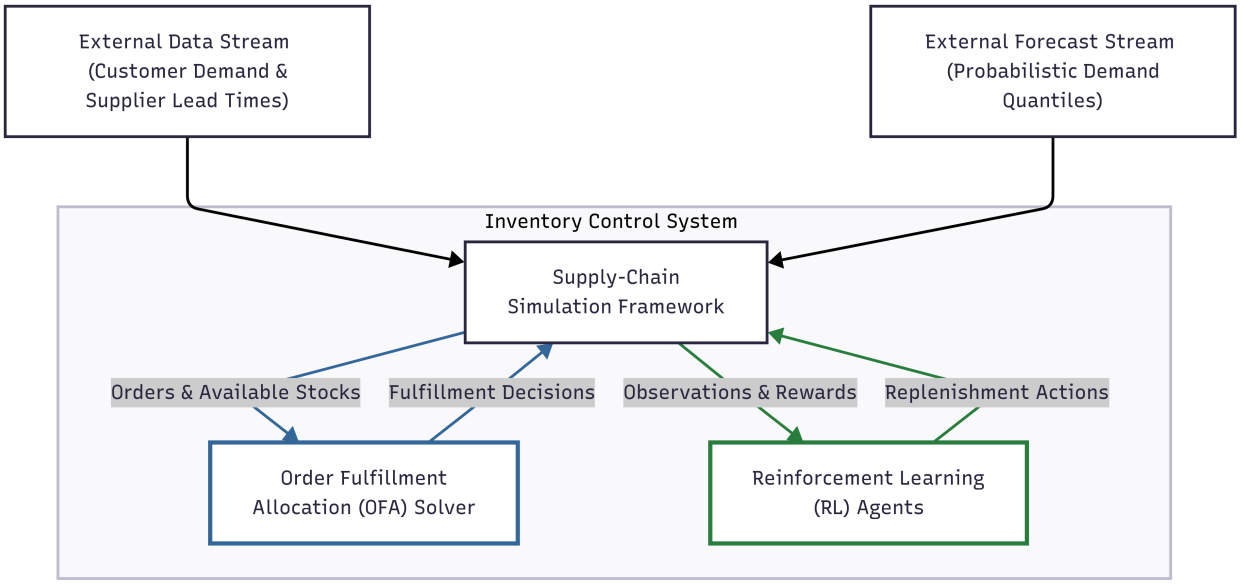

Figure 6.1 Overview of the system architecture and its primary information flows.

The first is the daily tactical process of Order Fulfillment Allocation (OFA), on the left side, and represented by the blue arrows in the figure. For this process, the simulator provides the OFA Solver with the necessary inputs: the current state of available on-hand inventory for each SKU, and the complete list of open customer orders with their respective attributes such as quantities, due dates, and SLAs. Based on this information, the solver computes an optimal allocation and returns its decision as an output to the simulator, namely the specific list of orders or order lines to be fulfilled and shipped for the current day.

The second loop, on the right side and highlighted in green, represents the periodic strategic process of inventory control, driven by the Reinforcement Learning (RL) agents. The inputs for this process consist of a comprehensive observation of the system state, which the simulator provides to the RL agents. This observation includes not only current inventory levels (on-hand and in-transit) and the status of orders, but also the probabilistic demand forecasts generated upstream. In response to this rich state representation, the RL agents produce an action as their output: the precise replenishment quantities of each SKU to order from suppliers. This decision is then passed back to the simulator to be integrated into the pending replenishment pipeline.

The Order Fulfillment Allocation (OFA) system (Section 6.4.2) optimizes daily tactical decisions about which customer orders to fulfill given current inventory constraints. This component implements multiple allocation strategies ranging from simple heuristics to Mixed Integer Linear Programming models. These models balance competing objectives such as

maximizing order fulfillment rates, prioritizing time-sensitive orders, and maintaining balanced service levels across customer groups.

The Reinforcement Learning framework (Section 6.4.3) makes periodic strategic inventory replenishment decisions. This component leverages Proximal Policy Optimization (PPO), and multi-agent reinforcement learning (MARL) that features specialized agents for each service level group plus a coordinating global agent, allowing for targeted optimization of performance across different customer priorities. The agents learn effective policies through continuous interaction with the simulated environment, adapting to demand fluctuations and supply chain dynamics without requiring explicit mathematical modeling of these complex relationships.

The Supply-Chain Simulation Framework (Section 6.4.4) integrates these components within a discrete-time Markov Decision Process. The simulation tracks the state of customer orders, inventory levels, and pending purchase orders. Key state transitions occur through OFA allocation decisions, stock replenishment actions, and stochastic events like incoming customer orders and supplier deliveries.

This architecture requires two critical data streams as inputs: (1) a source of customer demand and supplier lead times to drive the simulation, and (2) a stream of probabilistic demand forecasts to inform the RL agents' decisions. For the purpose of rigorously training and validating our system in a controlled and reproducible manner, we have developed a dedicated data generation environment. The methodology for this environment, which uses Monte Carlo simulation to produce synthetic data and quantile forecasts, is detailed in the validation section (Section 6.5.3). This approach allows us to test our system's performance across a wide range of scenarios, independent of the limitations of a single historical dataset.

6.4.2 Order Fulfillment Allocation (OFA)

Order Fulfillment Allocation (OFA) is a critical process in supply chain management, particularly for distribution-centric businesses facing constrained inventory amid fluctuating demand. It can be defined as the strategic allocation of available stocks to fulfill customer orders while accounting for inventory levels, service level agreements (SLAs), and due dates. In this context, various allocation strategies can be employed, ranging from simple heuristics to optimization-based approaches.

The **first-come, first-served (FCFS)** strategy processes orders sequentially by their creation date, fulfilling each order if sufficient stock is available. This method prioritizes orders purely based on arrival time, without considering additional factors. A variant, the **FCFS**

by due date approach, instead prioritizes orders according to their due dates, aiming to reduce lateness by addressing more urgent orders first). An enhanced method, **FCFS by weight**, assigns a weight (w_o) to each order based on factors such as quantities, due dates, and SLAs. Orders are then processed in descending order of weight, allowing for more nuanced prioritization).

The **Weighted Order Fill Rate (WOFR)** strategy frames the allocation problem as a mixed-integer linear programming (MILP) model aimed at maximizing the sum of weights of fulfilled orders. As shown by [8], WOFR outperforms basic heuristics, particularly in high-volume scenarios. This method ensures feasibility (inventory and delivery constraints are respected) and optimality (weighted fulfillment is maximized). Building on this, the **Weighted Order Line Fill Rate (WOLFR)** strategy introduces partial deliveries, allowing dispatch of available items when complete fulfillment is not possible. To balance improved service levels with the risks of increased delivery frequency, the model incorporates penalties adjusted by coefficients α_{full} and α_{due}, which promote full deliveries while giving precedence to time-sensitive order lines.

Finally, the **WOLFR with Overdrawn Stock Penalty (WOLFR-OSP)** strategy introduces mechanisms to manage service level shortfalls (SLS) across SLA groups. This approach includes a reserved stock parameter $r_{i,s}$ for each item and SLA group, a variable $E_{i,s}$ to capture excess stock usage beyond reservation limits, and a penalty term weighted by α_{osp} to discourage imbalances. By adjusting $r_{i,s}$, the model allows direct control over how SLS is distributed between customer groups.

The effectiveness of these OFA strategies largely depends on the calculation of order weights (w_o), which in the implementation combine several factors. These include the **order quantity**, raised to a configurable power (α_{qty}); the **SLA importance**, which gives higher priority to orders from higher SLA groups (α_{sla}); and the **order urgency**, reflecting the elapsed time relative to the promised lead time and any days past due ($\alpha_{urgency}$).

The mathematical formulation and parameter details are provided in Appendix C.1.

While effectively allocating inventory can improve performance metrics, OFA's impact is inherently constrained by available stock. Without sufficient inventory, even the most advanced allocation strategy cannot meet service level targets. Conversely, with ample stock, the allocation process becomes trivial. Therefore, while OFA optimization is important, it addresses only one aspect of Supply Chain Management. The greater influence on key performance metrics comes from inventory control decisions—specifically, replenishment policies. The following Section 6.4.3 will introduce a reinforcement learning framework to optimize these more complex but more impactful decisions.

6.4.3 Reinforcement Learning Framework for Inventory Control

In Reinforcement Learning (RL), agents learn to make decisions through interactions with an environment, guided by the framework of a Markov Decision Process (MDP). Within this context, an agent observes the state of an inventory system (including inventory levels and demand forecasts), takes actions (stock replenishment decisions), and receives rewards based on service levels and inventory costs. Through these interactions, the agent develops a policy that maximizes cumulative rewards over time, effectively learning an optimal inventory control strategy.

The RL framework integrates with a supply chain simulator (described in Section 6.4.4), enabling agents to interact with a realistic representation of the supply chain. Policy Optimization (PPO) is selected as the reinforcement learning algorithm for its sample efficiency, stability, and adaptability to complex, stochastic environments such as supply chains. The implementation leverages RLlib for distributed training, PyTorch for neural network modeling, and Polars for high-performance dataframe operations. Further details on environment setup and training configurations are provided in Appendix C.5.

The framework employs a multi-agent approach with $S + 1$ agents, where S represents the number of customer SLA groups. Each of the S agents is responsible for a specific SLA group, tasked with: (1) determining the appropriate replenishment quantities, $Y_{i,s}$, for each item i to meet the service level targets of their assigned SLA-group s and (2) determining the amount of stock, $r_{i,s}$, to be reserved for their assigned group. In addition, a "global" agent oversees the entire system, determining the total replenishment quantity Y_i for each item i. To combine the decisions of individual SLA group agents and the global agent, a weighted average is employed:

$$\overline{Y_i} = \alpha_{global} \cdot Y_i + (1 - \alpha_{global}) \cdot \sum_{s=1}^{S} Y_{i,s} \tag{6.2}$$

where α_{global} represents the weight assigned to the global agent's decision. Each agent learns independently without sharing parameters with other agents, promoting exploration and specialization in managing inventory needs of assigned customer groups. The coordination mechanisms and competitive dynamics between agents are further detailed in Appendix C.5.

This multi-agent structure offers several advantages: each agent can be assigned a reward function specifically tailored to the performance metrics of its assigned customer group, the decentralized nature enhances scalability for complex supply chains with numerous SKUs and customer groups, and the multi-agent system can adapt more effectively to changes in

demand and supply chain dynamics.

The observation space ($S_{s,t}$) at time-step t for agent$_s$ encompasses relevant features that reflect the state of supply and demand, tailored to SLA-group s and its specific SLA requirements. It can be represented as a four-dimensional array where the dimensions correspond to items (i), customer SLA-groups (s), quantile positions (q), and observation measures (m). Key observation measures include supply observations (available inventory, pending inventory, reserved stock), demand observations (current demand, forecasted demand, net demand metrics), service level observations (target and realized service levels, service level shortfall), and probabilistic observations (probability metrics for reaching service level targets and satisfying demand without replenishment). These observation metrics ensure agents have a comprehensive view of current and future inventory and demand conditions. The complete mathematical formulations and implementation details of the observation space are provided in Appendix C.5.

The action space supports multiple implementations to provide flexibility in how agents make inventory decisions. The implemented action space types can be categorized into two main groups: discrete action spaces, which include mappings from discrete actions to fractions of maximum quantities, quantile positions, base stock multipliers, or target service level multipliers; and continuous action spaces, which include raw float values for actions, base stock multipliers, adjustments to target service levels, or quantile position selection.

For one implementation, a continuous action space is utilized where agents select optimal quantile positions from a pre-computed quantile forecast table for each item and customer group. This standardizes the action range (0 to 1) and creates a continuous relationship between similar decision points, facilitating gradient-based optimization and more efficient exploration. The quantile forecast approach addresses scalability challenges in inventory control by requiring fewer output nodes in the neural network, thereby reducing model complexity. The configuration parameters in `ActionSpaceConfigs` allow for fine-tuning of the action space behavior, including buffer ranges, weighting between due and total demand, and selection of appropriate net demand calculation methods. Detailed descriptions of each action space type and their mathematical formulations are provided in Appendix C.5.

The reward function $r_{s,t}$ for each agent s at time-step t incentivizes the minimization of the cumulative loss Φ (defined in sections 6.2 and 6.5.2.1). The reward is calculated as the difference between the current loss $\Phi_{s,t}$ obtained by an RL agent and the loss obtained using a baseline method $\Phi_{s,t}^{\text{baseline}}$, normalized to ensure consistency across different scenarios:

$$r_{s,t} = \frac{\Phi_{s,t}^{\text{baseline}} - \Phi_{s,t}}{\left(\sum_{t' \in T} \Phi_{s,t'}^{\text{baseline}}\right) \cdot S} \tag{6.3}$$

where:

— $\Phi_{s,t}$ is the loss incurred by the RL agent for SLA-group s at time step t

— $\Phi_{s,t}^{\text{baseline}}$ is the loss incurred by the baseline method (Weighted-Quantile-Forecast Multi-SLA Base Stock policy, defined in Section 6.5.1)

— T represents the set of all time-steps in the episode

— S represents the total number of SLA groups

This reward formulation centers rewards around zero, with positive values indicating performance better than the baseline and negative values indicating worse performance. The scaling ensures that if the total RL loss is X% higher than the baseline loss, the total reward is -X%, and if the total RL loss is X% smaller than the baseline loss, the total reward is X%.

For the RL agents to learn and operate effectively, the system requires a source of rich, varied data and probabilistic forecasts. The components presented in sections 6.5.3.1 and 6.5.3.2 fulfill this essential operational need within the system's architecture.

6.4.4 Supply-Chain Simulation Framework

The supply chain simulation framework serves as the central integration point for the inventory control system, effectively combining all other components within a unified operational environment. This framework implements a discrete-time Markov Decision Process (MDP) that models the complex dynamics of inventory management, order fulfillment, and stock replenishment.

Within this MDP framework, the environment's state encompasses customer orders, inventory levels, and pending purchase orders. This state evolves through three primary mechanisms: 1) order fulfillment decisions made by the OFA solver, 2) replenishment actions, and 3) stochastic events–including dynamic lead times and incoming customer orders generated from Monte Carlo simulations. This modular and structured approach enables both performance evaluation of traditional policies and effective training of reinforcement learning agents.

The simulation framework is built upon several key design principles that ensure both practical relevance and theoretical rigor : 1) The simulation advances in discrete daily time steps, with inventory control decisions made at configurable periodic review intervals. This structure mirrors real-world business cycles where, for example, replenishment decisions oc-

cur weekly while fulfillment operations run daily; 2) All system dynamics are represented through a comprehensive state representation, tracking inventory levels, order statuses, and replenishment pipelines. This state-centric approach facilitates both decision evaluation and reinforcement learning; 3) The framework explicitly separates OFA policies (for daily order fulfillment) from replenishment policies (for periodic inventory control), enabling independent optimization of each decision type; 4) The simulation incorporates practical business constraints such as weekday-only shipping, capacity limits, and service level agreements, ensuring that policies developed within the simulation remain applicable to real-world operations.

The simulation maintains a comprehensive state representation across four dimensions : The **inventory state** tracks available stock quantities for each item, including on-hand inventory and pending replenishments with their expected arrival dates. It also monitors reserved stock quantities allocated to specific SLA groups to ensure service level compliance. The **order state** captures the status of all customer order lines, including entry dates, due dates, fulfillment status, and priority weights, with orders transitioning through various states (active, fulfilled, late) as the simulation progresses. The **time state** manages the simulation's temporal progression, tracking the current day, review periods, and business calendar effects such as weekday/weekend distinctions that affect fulfillment activities. Finally, the **performance state** aggregates key performance metrics, including service levels, inventory costs, and order lateness, which are used for evaluation and reward calculation. This multidimensional representation enables the simulation to model the complex interdependencies between inventory levels, demand patterns, and fulfillment operations characteristic of real-world supply chains.

The simulation framework embeds two distinct decision-making processes, operating on different time scales : The first is **daily order fulfillment**, where, each business day, the OFA solver (Section 6.4.2) determines which customer orders to fulfill based on inventory availability, order priorities, and service level constraints. Fulfillment decisions are then executed within the simulation, updating inventory levels and order statuses. The second process is **periodic replenishment**, which takes place at each review period. Here, replenishment quantities for each item are determined either by applying through RL agent actions (Section 6.4.3) or base stock policies (Section 6.5.1). The resulting decisions feed into the replenishment pipeline for future inventory updates.

The simulation progresses through three nested cycles : The **daily cycle** processes daily events, including order activations, fulfillment operations, and stock receipts. On weekdays, it also executes the OFA solver and process order fulfillment. Additionally, the daily cycle updates order statuses, calculates priority weights, and records daily metrics. The **review cycle** occurs at predefined intervals, triggering a replenishment review. During this phase, the

simulation gathers state observations, evaluates performance metrics, executes replenishment decisions, and updates reservations across SLA groups. Finally, the **episode cycle** spans several review cycles, typically covering multiple weeks or months. Each episode provides a comprehensive evaluation of the inventory control policy's performance over an extended time horizon.

Agents receive a set of structured observations at each decision point : These include **item-level observations** such as stock levels, replenishment status, and demand forecasts for each item; **SLA-level observations** including service level performance, reserved stock status, and SLA-specific demand patterns; **temporal observations** in the form of time-series metrics that capture performance trends over the simulation period; and **performance observations** summarizing key metrics for inventory costs, service level adherence, and order lateness. For reinforcement learning, these observations are processed and structured into the agent observation space described in Section 6.4.3. The observation generation process includes aggregation, normalization, and filtering operations that transform raw simulation state into informative features for decision-making.

The simulation framework is designed with adaptability as a core principle, enabling its application across diverse distribution-centric business contexts. It provides **configurable parameters**, enabling the adjustment of business settings—such as review intervals, service level targets, and operational constraints—without altering the simulation logic. Its **modular components** allow individual elements like the OFA solver, forecasting methods, or replenishment policies to be independently replaced or extended. Moreover, **pluggable policies** can be integrated through defined interfaces, facilitating the testing of new decision-making strategies. Finally, the framework supports **diverse demand patterns**, including seasonality, trends, and intermittent demand, ensuring applicability across various industries.

This adaptability ensures that the simulation framework can be applied beyond the industrial partner's specific context, serving as a general platform for inventory control research and application in diverse distribution-centric businesses.

The supply chain simulation framework forms the operational core of the inventory control system, providing a realistic environment that integrates all other components. By maintaining a comprehensive state representation, incorporating practical business constraints, and separating different decision timeframes, the simulation enables the development and evaluation of inventory control policies that remain applicable to real-world distribution operations.

This design creates a natural division of labor within the system: the OFA solver focuses on short-term operational efficiency, finding optimal solutions to daily order allocation problems, while the RL framework concentrates on longer-term strategic objectives through periodic in-

ventory replenishment decisions. This separation of concerns allows each component to excel in its respective domain while working together seamlessly within the integrated simulation environment.

Through its principles of adaptability and extensibility, the framework transcends the specific context of the industrial partner, offering a general platform for inventory control research and application across diverse distribution-centric businesses dealing with complex multi-item, multi-SLA supply chains.

6.5 Validation Methodology and Experimental Analysis

This section details the empirical process used to validate our proposed RL-based inventory control system. The validation is structured to ensure a fair and comprehensive comparison against a strong heuristic. Our approach unfolds in four main parts. We first define the **heuristic benchmark policy**, derived from established base stock principles but enhanced to handle the complexities of our problem, such as multi-SLA environments. We then specify the **experimental framework**, which includes the performance metrics and the parameter tuning protocol applied to both the benchmark and our RL system to guarantee that each is evaluated at its optimal configuration. Next, we present the **results** of two sequential experiment stages: the optimization of the benchmark policy, followed by the training and evaluation of our proposed RL system. Finally, we conduct a detailed **comparative analysis** of the two optimized policies to quantify the performance gains.

6.5.1 The Heuristic Benchmark: Enhanced Base Stock Policies

To establish a credible point of comparison, we define a benchmark based on the widely recognized base stock inventory control policy. This policy is pivotal in supply chain management, aiming to maintain adequate stock to meet customer demand by replenishing inventory to a predetermined level at fixed intervals. While simple in principle, we enhance the classical model to create a robust heuristic capable of addressing the specific challenges of our study.

The Base Stock policy, a prominent inventory control strategy, stands out for its simplicity and reliability. In periodic review systems, inventory levels are evaluated at fixed intervals, and replenishment orders aim to restore inventory to predetermined "*base stock*" levels. These levels are calculated to ensure sufficient inventory to meet expected demand during the order cycle (the period until the next review plus the replenishment lead time).

6.5.1.1 Classical Base Stock Policy and Limitations

Traditional Base Stock policies primarily focus on minimizing the total costs associated with overstocking and understocking, guided by the statistical distribution of forecasted demand. A classical formulation can be expressed as:

$$B = \mu + Z\sigma \tag{6.4}$$

where μ represents mean demand, σ is the standard deviation of demand, and Z is a safety factor determined by cost trade-offs between stockouts and holding inventory.

However, this standard model presents several limitations when aligned with the study's objectives: First, it relies on a **linear cost assumption**, where overstocking and understocking costs are treated linearly. This approach conflicts with the goal of meeting predefined Target Service Levels (TSLs), requiring a shift from pure cost minimization to service level adherence. Second, the model is built on a **normality assumption** for demand, which often introduces inaccuracies, particularly in the tails of the distribution. In fact, these distribution tails are crucial when targeting high TSLs, and empirical evidence shows that assuming normality can impair the model's ability to meet these service levels. Third, it overlooks **time and item considerations**, as it generally assumes stationary demand in a single-SKU setting, limiting its relevance in multi-item environments with dynamic, time-varying demand patterns. Finally, the model lacks **backorder and due date considerations**. In backorder contexts with due dates, forecasting must focus exclusively on orders that are expected to be *both created and due within the same order cycle*, adding a layer of complexity that the classical model does not address. Indeed, orders created within the cycle but due afterward are largely irrelevant, as they will be addressed in the next order cycle. Conversely, orders due within the current period but created earlier are already included in the *observed demand* (i.e., backorders) and must not be double-counted in the *forecasted demand*.

6.5.1.2 Quantile-Forecast Base Stock Policy

To address the limitations of the classical model, **Quantile-Forecast Base Stock policy** is proposed. It explicitly aims for TSL adherence rather than cost minimization. The policy is defined as:

$$B_{t,i} = D_{t,i} + F_{t',i,q} \tag{6.5}$$

where:

— $B_{t,i}$ is the base stock level for item i at time t.

— $D_{t,i}$ is the existing demand for item i at time t that is due or will become due within the order cycle.

— $F_{t',i,q}$ is the forecasted demand at quantile position q, for item i, expected to be both created and due within the order cycle t'.

(For more information on quantile forecast, see Section 6.5.3.2.)

The *order cycle* t' is defined as the period starting at review period t and lasting for the duration of the review interval plus the item's replenishment lead time (RLT). For practical implementation, since the exact RLT for the next review period is uncertain, a reasonable upper bound, denoted as RLT_i^{max}, is considered.

The quantity to order for an item i at time t is calculated as:

$$Y_{t,i} = B_{t,i} - (S_{t,i}^{available} + S_{t',i}^{pending}) \tag{6.6}$$

where:

— $S_{t,i}^{available}$ is the quantity of item i available in inventory at time t.

— $S_{t',i}^{pending}$ is the quantity of item i ordered but not yet received, expected to become available within the order cycle t'.

This approach addresses the limitations of the conventional model by employing a non-parametric representation of forecasted demand across various quantiles, capturing the full spectrum of demand variability without specific distributional assumptions.

In environments with tiered customer Service Level Agreements (SLAs), additional complexity arises from the need to align inventory levels with diverse TSL requirements across customer segments. Three approaches are proposed next to address this challenge (6.5.1.3, 6.5.1.4, 6.5.1.5).

6.5.1.3 Grouped-Quantile-Forecast Multi-SLA Base Stock Policy

The **Grouped-Quantile-Forecast Multi-SLA Base Stock Policy** calculates base stock levels for each item and SLA group separately, using the forecasted demand at the quantile position equal to the group's TSL. It is formulated as:

$$B_{t,i,s}^{grouped} = D_{t,i,s} + F_{t',i,s,q=TSL_s} \tag{6.7}$$

where:

— $B^{grouped}_{t,i,s}$ is the base stock level for item i at time t required to achieve the TSL for group s.

— $D_{t,i,s}$ is the existing demand from customers in group s for item i at time t.

— $F_{t',i,s,q=TSL_s}$ is the forecasted demand from group s at quantile position $q = TSL_s$.

The order quantity is calculated by summing the base stock levels across all groups and adjusting for available and pending inventory:

$$Y_{t,i} = \sum_{s} B^{grouped}_{t,i,s} - (S^{available}_{t,i} + S^{pending}_{t',i}) \tag{6.8}$$

While straightforward, this approach can lead to overestimation when demand across groups is correlated. For high quantile positions ($q > 0.5$), the sum of individual quantiles typically exceeds the quantile of the aggregate demand, as demonstrated in the literature [179]. Despite this limitation, the policy provides a useful upper bound for inventory planning. For more details, see Appendix C.2.

6.5.1.4 Cumulative-Quantile-Forecast Multi-SLA Base Stock Policy

To address the overestimation issue of the grouped policy, the **Cumulative-Quantile-Forecast Multi-SLA Base Stock Policy** is introduced. This more sophisticated approach leverages the **hierarchical nature of order fulfillment**, where higher-SLA groups receive priority in inventory allocation.

The core insight is that to protect the service level of a specific group s, the inventory must be sufficient to cover not only the demand from group s but also the entire demand from all higher-priority groups that will be served before it. Therefore, instead of calculating requirements for each group in isolation, this policy calculates the necessary stock based on the cumulative demand of each SLA tier and all tiers above it.

The base stock level to protect all groups with an SLA equal to or higher than s' is formulated as:

$$B^{cumul}_{t,i,s'} = D^{cumul}_{t,i,s'} + F^{cumul}_{t',i,s',q=TSL_{s'}} \tag{6.9}$$

Here, $B^{cumul}_{t,i,s'}$ represents the base stock for item i required to protect all customer groups with a priority of s' or higher. This level is based on $D^{cumul}_{t,i,s'}$, the aggregated existing demand from these groups (i.e., $\sum_{s \geq s'} D_{t,i,s}$), and $F^{cumul}_{t',i,s',q=TSL_{s'}}$, the forecast of their aggregated future demand. The quantile q is set to the target service level $TSL_{s'}$ for each respective calculation,

ensuring the stock covers the cumulative demand up to the desired probability for that tier. (For more information on quantile forecast, see Section 6.5.3.2.)

The total order quantity is then determined by the **maximum** of these cumulative base stock levels calculated across all priority tiers:

$$Y'_{t,i} = \max_{s'}(B^{cumul}_{t,i,s'}) - (S^{available}_{t,i} + S^{pending}_{t',i}) \tag{6.10}$$

By taking the maximum, the policy ensures the inventory level is high enough to satisfy the most stringent cumulative requirement. Since the requirement for any lower-priority group already includes the full demand of all higher-priority groups, this single stock level is sufficient to meet all target service levels across the hierarchy.

This approach effectively avoids the statistical issue of summing quantiles found in the grouped policy. However, its accuracy hinges on a critical assumption: a perfectly hierarchical allocation. It presumes that higher-SLA groups are always served completely before any stock is allocated to lower-priority groups. If the actual order fulfillment process deviates from this strict priority, this policy may **underestimate** the required inventory. It therefore provides a reasonable **lower bound** for inventory planning, complementing the upper bound from the Grouped-Quantile policy. For detailed implementation, see Appendix C.2.

6.5.1.5 Weighted-Quantile-Forecast Multi-SLA Base Stock Policy

To balance the overestimation of the grouped approach and the underestimation of the cumulative approach, a **Weighted-Quantile-Forecast Multi-SLA Base Stock Policy** is proposed as a weighted average of the two:

$$\bar{B}_{t,i} = B^{cumul}_{t,i} \cdot \alpha_{cumul} + B^{grouped}_{t,i} \cdot (1 - \alpha_{cumul}) \tag{6.11}$$

where:

— $\bar{B}_{t,i}$ is the weighted average base stock level for item i at time t.

— $B^{cumul}_{t,i}$ is the base stock level calculated using the cumulative approach.

— $B^{grouped}_{t,i}$ is the base stock level calculated using the grouped approach.

— α_{cumul} is the weight parameter controlling the balance between the two approaches.

This hybrid policy provides a more accurate base stock estimation by combining the upper and lower bounds derived from the two approaches. The weight parameter allows for adjustment based on the specific characteristics of the demand environment. For the complete

implementation details of all three policies, refer to C.2.

6.5.2 Experimental Framework

A fair comparison requires a clearly defined experimental framework, including robust performance metrics and a systematic tuning protocol to ensure both the benchmark and the RL system are operating at their best. The following sections describe this framework.

6.5.2.1 Performance Metrics

This section details the performance metrics, their mathematical formulations, and the aggregation methods used across time periods and service level groups.

The metrics are defined along three dimensions: the **time-step** (t), which represents the interval between inventory reviews and captures the periodic nature of replenishment decisions; the **service level agreement (SLA) group** (s), which distinguishes customer segments with different service level requirements; and the **day** (d), referring to individual days within a time-step, allowing for fine-grained metrics calculations.

Performance is measured at the SLA group level rather than individual item or customer level because many items and customers generate orders too infrequently to provide stable metrics. This aggregation approach ensures robust measurements while maintaining the ability to differentiate between customer segments with varying service requirements.

This research aims to optimize a global performance function (Equation 6.1) that captures the trade-offs between three key metrics:

The first metric, **Inventory Costs** ($\Phi_t^{\mathbf{stock}}$), represents the inventory holding cost for time-step t, calculated as the cumulative inventory level across all items and days:

$$\Phi_t^{\text{stock}} = \sum_i \sum_{d \in t} s_{i,d} \tag{6.12}$$

where $s_{i,d}$ is the available quantity of item i at the beginning of day d.

The second metric, **Order Lateness** ($\Phi_{t,s}^{\mathbf{late}}$), quantifies the lateness penalty for time-step t and SLA-group s, considering both quantity and days late:

$$\Phi_{t,s}^{\text{late}} = \sum_{o \in O_s} \sum_i d_{o,i} \cdot \sum_{d \in t} \text{late}_{o,i,d} \tag{6.13}$$

where:

— $d_{o,i}$ is the demanded quantity of item i in order o

— $\text{late}_{o,i,d} \in \{0, 1\}$ indicates whether order-line (o, i) is unfulfilled and past its due date on day d

— O_s represents the set of orders from customers in SLA-group s

The third metric, **Service Level Shortfall ($\Phi^{\text{service}}_{t,s}$)** measures the gap between target and realized service levels:

$$\Phi^{\text{service}}_{t,s} = \max\left(\frac{1 - \text{RSL}_{s,t}}{1 - \text{TSL}_s}, 1\right) \tag{6.14}$$

where TSL_s is the target service level for SLA-group s (e.g., 0.95 for 95%), and $\text{RSL}_{s,t}$ is the realized service level:

$$\text{RSL}_{s,t} = \frac{\sum_{o \in O_s} \sum_i d_{o,i} \cdot X^{\text{ontime}}_{o,i,t}}{\sum_{o \in O_s} \sum_i d_{o,i} \cdot (X^{\text{ontime}}_{o,i,t} + X^{\text{late}}_{o,i,t} + \text{late}_{o,i,t})} \tag{6.15}$$

where:

— $X^{\text{ontime}}_{o,i,t} \in \{0, 1\}$ indicates whether order-line (o, i) was fulfilled on time within time-step t

— $X^{\text{late}}_{o,i,t} \in \{0, 1\}$ indicates whether order-line (o, i) was fulfilled late within time-step t

— $\text{late}_{o,i,t} \in \{0, 1\}$ indicates whether order-line (o, i) remains unfulfilled and late at the end of time-step t

The Service Level Shortfall metric is designed with three important properties : First, it is **SLA-aware**, meaning each group's performance is evaluated relative to its specific target. Second, it is **asymmetrical**, penalizing RSL underperformance while providing no additional benefit for overperformance. Third, it is **non-linear**, as the penalty increases disproportionately the further performance falls below the target.

To balance sensitivity to recent events with stability in metrics, exponential smoothing is applied across time-steps:

$$\bar{\Phi}^k_{t,s} = \sum_{t' \leq t} \Phi^k_{t',s} \cdot w^k_{t',t} \tag{6.16}$$

where:

— $k \in \{\text{stock}, \text{late}, \text{service}\}$ represents the metric type

— $w^k_{t',t} = \frac{(1-\beta_k)^{t-t'}}{\sum_{t'' \leq t}(1-\beta_k)^{t-t''}}$ is the normalized weight for time-step t'

— β_k is the smoothing factor for metric k, controlling the rate at which older observations lose influence

This approach ensures that metrics reflect recent performance trends without becoming overly volatile or stable.

These three smoothed metrics are aggregated into the global performance function, $\Phi_{s,t}$ (Equation 6.1), which provides a holistic measure of operational efficiency by balancing cost-control with service-level adherence.

This global performance function serves a critical dual role throughout our validation process, and it is important to distinguish its two purposes: 1) *As an objective function:* It forms the basis of the reward signal used to train the RL agents (as defined in Section 6.4.3). In this capacity, it is the internal compass that guides the agent's learning process toward desirable operational outcomes. 2) *As a comparison metric:* It is used as the final, external measure to objectively evaluate and compare the aggregate performance of the fully-optimized baseline policy against the fully-trained RL system.

For reinforcement learning, this performance function is transformed into reward signals by comparing an agent's performance against a baseline policy:

$$r_{s,t} = \frac{\Phi_{s,t}^{\text{baseline}} - \Phi_{s,t}}{\sum_{t' \in T} \Phi_{s,t'}^{\text{baseline}} \cdot S} \tag{6.17}$$

where:

— $\Phi_{s,t}$ is the loss incurred by the RL agent for SLA-group s at time step t

— $\Phi_{s,t}^{\text{baseline}}$ is the loss from a baseline policy (typically our enhanced base stock policy)

— T represents all time-steps in the episode

— S is the number of SLA groups

This normalization ensures that rewards are centered around zero, with positive values indicating improvement over the baseline and negative values indicating worse performance. The scaling ensures that if the total loss is X% higher than the baseline, the reward is approximately -X%, creating a proportional incentive system.

6.5.2.2 Parameter Tuning Protocol

We used a structured parameterization approach that separates tunable hyperparameters from fixed operational constraints. This separation enables systematic optimization while maintaining consistent business rules across experiment runs.

Alpha parameters represent tunable hyperparameters that control various aspects of inventory policies, order fulfillment strategies, and reinforcement learning configuration. These parameters are subject to optimization through the experimental process. **Beta parameters** represent fixed operational constraints and business requirements that remain constant across experiment runs. These parameters define fundamental aspects such as review inter-

vals, operational constraints, and reward calculations.

The hyperparameter optimization methodology follows a structured approach to efficiently explore the parameter space while maintaining experimental rigor. Systematic parameter sweeps were conducted to identify optimal parameter configurations for both baseline policies and reinforcement learning approaches. Throughout all experiment runs, the beta parameters remained fixed, as they represent business constraints and operational settings that do not change.

The optimization process was organized into two stages : 1) For the **non-RL alpha parameters**, preliminary sweeps were performed across the entire parameter space, followed by a final sweep refined to promising regions. Optimal configurations were then selected and fixed for subsequent RL experiment runs. These optimal parameter values also served as baselines for calculating relative rewards (see Equation 6.17) later; 2) For the **RL-specific alpha parameters**, an initial broad sweep was conducted across the parameter space, then the search was narrowed to focus on the most promising parameter combinations, leading to the identification of the final optimal parameter values.

This staged approach allowed isolation of the effects of each parameter group and ensured fair comparisons between the baseline and RL strategies.

6.5.3 Data Environment for Training and Validation

A robust validation of data-driven decision-support systems requires an environment that can generate consistent, reproducible, and complex scenarios. To this end, we developed a data generation pipeline based on Monte Carlo methods. This pipeline serves the dual purpose of creating vast datasets for training the RL agents and generating independent test sets for evaluating both the benchmark and RL policies. The following sections describe the two main components of this environment.

6.5.3.1 Monte Carlo Synthetic Data Simulation

The Monte Carlo Synthetic Data Simulators generates realistic supply chain data by modeling the stochastic processes that characterize real-world inventory systems. This simulation framework provides the foundation for both training reinforcement learning agents and evaluating their performance under diverse conditions. It is crucial to understand that this component serves a dual purpose: it is 1) an *operational requirement* of our system for generating training data for RL, and 2) a tool for creating independent test datasets for the *validation protocol* described in Section 6.5.4.

The data generation system simulates key aspects of supply chain operations through three components : First, **order line generation** replicates customer ordering patterns with realistic temporal and quantitative characteristics. Second, **replenishment lead time generation** captures variability in supplier delivery times. Finally, **quantile forecast generation** produces probabilistic forecasts derived from a large number of simulated demand datasets.

The **order line simulator** creates realistic demand patterns by combining multiple influencing factors : It models **customer-item relationships**, determining which customers order which items and at what frequency. It also accounts for **temporal patterns**, modeling weekly ordering cycles, weekday distributions, and seasonal variations. **Priority rules** are implemented to handle prioritized orders with adjusted service level requirements. The simulator generates order quantities using probabilistic **quantity distributions** that respect minimum quantity constraints. Finally, **long-term trends** are integrated through configurable growth or decline rates to capture evolving demand patterns.

The resulting order lines include key attributes such as customer ID, item ID, quantity, entry date, due date, and service level agreement ID.

The **replenishment lead time simulator** generates realistic supplier delivery timeframes by creating item-specific lead times that fluctuate within configurable minimum and maximum bounds. It models temporal variations to reflect real-world fluctuations in supplier performance and ensures alignment between lead time data and other components of the simulation.

This simulation framework enables the evaluation of inventory control policies under controlled yet realistic conditions, without being constrained by limitations of real-world historical data. Further technical details, including probability distributions and parameter settings, are provided in Appendix C.3.

6.5.3.2 Synthetic Quantile Forecast Simulator

Quantile forecasting plays a critical role in our inventory control system (described in Section 6.4.3), and in our Base Stock policies (described in Sections 6.5.1.2, 6.5.1.4, 6.5.1.5) by providing a comprehensive representation of demand variability. This approach is particularly valuable in supply chain operations, where understanding the full range of potential demand outcomes is essential for optimal decision-making under uncertainty.

A quantile forecasting approach provides robust probabilistic predictions without relying on restrictive distributional assumptions. Unlike traditional forecasting methods that focus primarily on the central tendency of demand, quantile forecasts capture the entire distribution,

including the tails that are crucial for meeting high service level agreements. This approach occupies a strategic middle ground between simplified distribution-based models and complex raw-data RL approaches : Compared to traditional forecasting models that assume specific distributions, quantile forecasts better represent the tails of demand distributions, which are essential for high SLA adherence. Compared to other RL-based inventory control approaches that incorporate extensive historical data directly into the observation space, our quantile forecast-based method provides a more structured and condensed representation of uncertainty, improving both scalability and signal-to-noise ratio.

The implementation utilizes Monte Carlo simulation to generate synthetic quantile forecasts through the following process : (1) Generate multiple synthetic demand datasets using the Monte Carlo simulation framework described in section 6.5.3.1; (2) Aggregate demand measures for each dataset according to relevant dimensions (item, SLA group, time horizon); (3) Calculate quantiles across all samples to capture the probabilistic distribution of future demand; (4) Produce both standard and cumulative SLA versions of the forecasts, where the latter includes quantities from higher SLA levels.

For each forecast, two distinct measures are calculated: demand created within the horizon, and demand both created and due within the horizon. This distinction enables agents to distinguish between overall future demand and urgent demand requiring immediate fulfillment.

The resulting quantile forecasts serve multiple purposes in the framework : 1) they form a critical component of the observation space for RL agents, 2) they enable the action space formulation based on quantile position selection, and 3) they provide the foundation for the baseline base stock policies. Further implementation details and mathematical formulations of the quantile forecast generation process are provided in Appendix C.4.

6.5.4 Experimental Dataset

The validation experiments described below were conducted using a synthetic dataset specifically generated to reflect the key characteristics and complexities of the target operational environment, including non-stationary demand and dynamic lead times.

This section presents the characteristics of the synthetic data used in the experiments. The data was generated using Monte Carlo techniques based on input parameters derived from the industrial partner's operational data, appropriately transformed to maintain statistical properties.

To facilitate efficient development experimentation, a small but representative subset of the available products and customers was considered, comprising 5 items from the thousands

available and 34 customers from the whole customer base. This simplification enables faster computation during development while preserving most of the essential characteristics of the inventory management problem.

Figure 6.2 shows the weekly distribution of orders across days of the week, reflecting typical business operational patterns with reduced weekend activity.

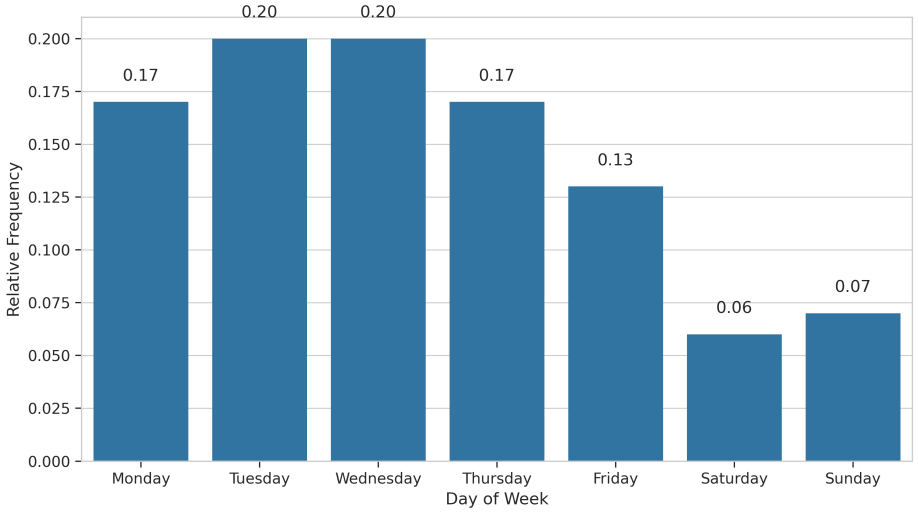

Figure 6.2 Weekly ordering pattern showing day-of-week distribution. Business days (Monday-Friday) show higher ordering activity compared to weekends.

The input parameters include seasonal factors that vary by month for each item-SLA combination. Figure 6.3 illustrates these seasonal patterns.

The generated order lines exhibit distinct ordering patterns by SLA group and item. Figure 6.4 shows the distribution of order quantities across different SLA groups.

Lead time variability represents a key operational consideration in inventory management. Figure 6.5 shows the distribution of lead times across different items.

The synthetic dataset exhibits several key characteristics. First, **order patterns** vary in both frequency and quantity across items and SLA groups, with notable variability in the time between consecutive orders. Second, **demand distributions** display multi-modal behaviors that significantly deviate from normality, with distinct patterns depending on the item and the SLA group. Third, **temporal variations** are present, including monthly seasonality, weekly cycles, and long-term annual trends affecting selected items. Finally, **supplier lead times** fluctuate by item and over time.

These characteristics reflect the operational reality faced by the industrial partner, where inventory policies must address demand heterogeneity and lead time variability while ensuring

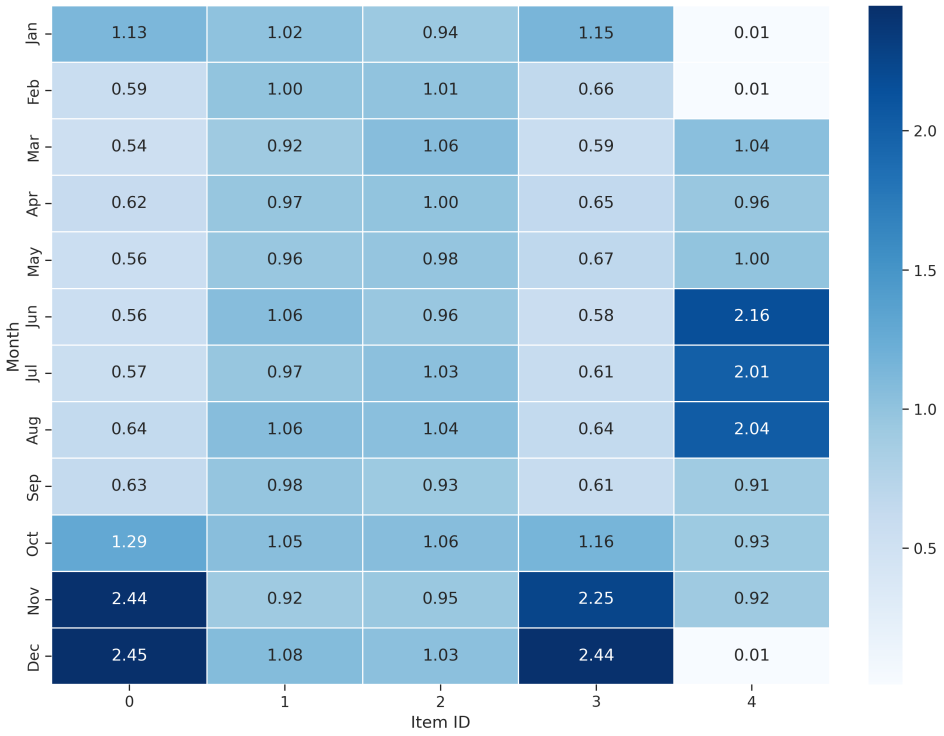

Figure 6.3 Seasonality factors by month and item. Each cell represents the demand multiplier applied in that month for that specific item, showing how demand fluctuates throughout the year.

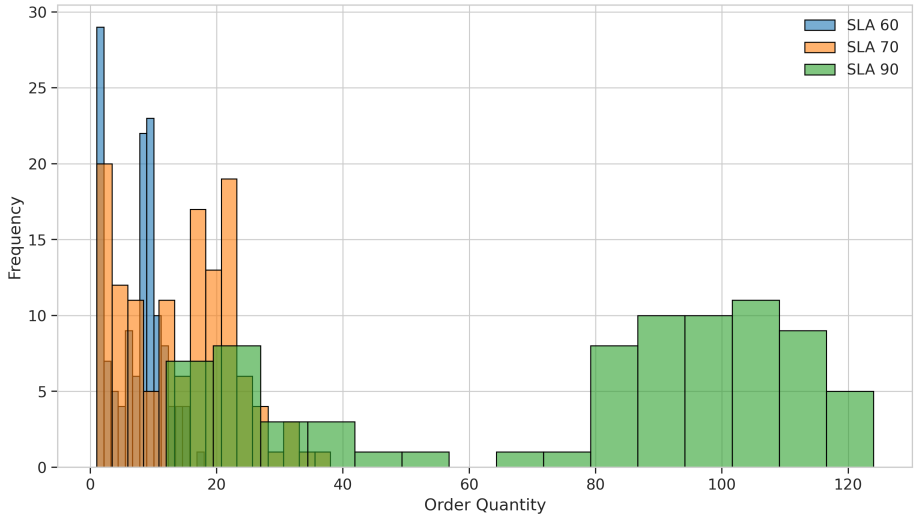

Figure 6.4 Distribution of order quantities by SLA group. Each SLA group demonstrates different ordering behaviors, with higher SLA groups (e.g., SLA 90) showing different quantity distributions compared to lower SLA groups.

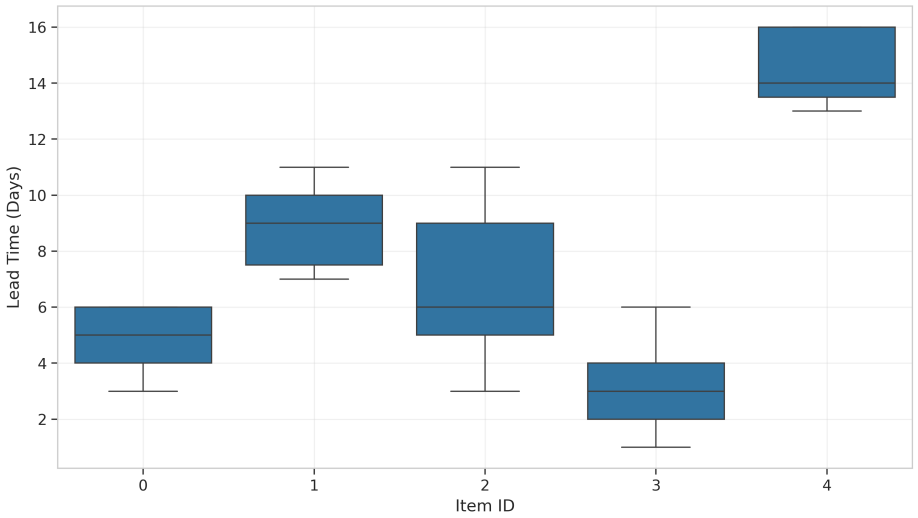

Figure 6.5 Box plots showing lead time distributions for each item. The range within each box plot illustrates the variability in supplier delivery times that inventory policies must accommodate.

service level adherence. The synthetic dataset preserves these essential features while allowing for controlled experimentation.

6.5.5 Stage 1: Benchmark Policy Optimization

In the first Stage of our validation, we apply the tuning protocol (Section 6.5.2.2) to the heuristic benchmark policies (Section 6.5.1) to identify their optimal configuration.

This section presents the experimental methodology and analysis of results prior to implementing Reinforcement Learning (RL). A comprehensive parameter sweep (7,750+ experiment runs with distinct parameter combinations) was conducted to understand the relative importance of different component parameters in the system, particularly comparing the impact of inventory control decisions versus order fulfillment allocation strategies.

The experimental approach consisted of systematically varying parameters related to both inventory control and Order Fulfillment Allocation (OFA). Each experiment run was replicated across 6 independent dataset samples, each running for 25 time-steps (with each step representing one week), to ensure statistical robustness.

The analysis involves two main categories of hyperparameters:

- Base Stock parameters:
  - Base stock multiplier ($\alpha_{\text{multiplier}}$): discrete values {0.95, 0.99, 1.0, 1.01, 1.05}
  - Cumulative weight ($\alpha_{\text{weight\_cumul}}$): uniform distribution [0.0, 1.0] with step 0.2

— Order Fulfillment Allocation (OFA) parameters:

  — Solver type: seven variants (FCFS, FCFS-BD, FCFS-BW, WOFR, WOLFR, WOFR-OSP, WOLFR-OSP)

  — Due date weight ($\alpha_{\text{weight\_due}}$): uniform distribution [0.25, 0.75] with step 0.05

  — SLA weight (α_{sla}): uniform distribution [0.5, 1.5] with step 0.1

  — Urgency sensitivity (α_{urgency}): uniform distribution [0.5, 1.5] with step 0.1

Several auxiliary parameters were fixed during experiment runs based on preliminary optimization. These include OFA overdraw penalty, OFA order weight settings, and various reserve adjustment factors. Additionally, the system operates under fixed beta parameters that define the problem configuration, such as review interval (7 days), training duration (25 weeks), and reward weighting factors.

The experimental campaign generated results from over 7,750 distinct parameter combinations. The raw results were processed through four sequential steps : (1) Sample Averaging : computing the mean of each loss metric across samples for each SLA group; (2) Global Loss Calculation : combining averaged individual losses with their respective weights; (3) Normalized Regret Computation : calculating the normalized regret for each SLA group; and (4) Total Regret : computing the total regret as the sum across all SLA groups.

Figure 6.6 presents the distribution of total regret across all experiment runs, with total regrets ranging from 1.24 to 2.40.

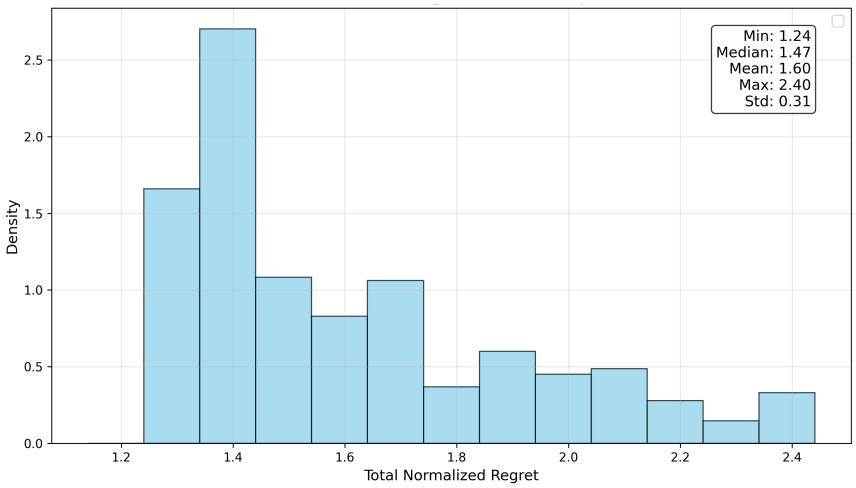

Figure 6.6 Distribution of total regret across all experiment runs.

Note that while each individual regret_s has a minimum value of zero (achieved by the best configuration for that SLA group), the total regret will typically be greater than zero, as it is unlikely for any single configuration to achieve the minimum loss simultaneously for all SLA groups.

Figure 6.7 show the distribution of total regret across combinations of Base Stock parameters ($multiplier \in \{0.95, 0.99, 1.0\}$ and $weight\_cumul \in \{0.6, 0.8, 1.0\}$).

Three critical insights emerge from this analysis:

First, Base Stock parameters dominate system performance. The top five combinations ($regret < 1.6$) show distinct, non-overlapping distributions, indicating that Base Stock parameters account for most of the variance in regret, while other parameters have minimal impact. Indeed, with optimal Base Stock settings ($multiplier = 0.99$, $weight\_cumul = 0.8$), the remaining variation in regret is merely $\pm 1\%$.

Second, the impact of Base Stock parameters exhibits steep slopes around the optimal values, with surprisingly high sensitivity in both directions. Increasing inventory above optimal levels (via a 1% rise in multiplier) raises regret from 1.24 to 1.39 (+12%). The effect is even more pronounced when decreasing inventory below optimal levels. This occurs either through reducing the multiplier or increasing *weight_cumul* (which increases reliance on the underestimating cumulative method). Such reductions trigger severe penalties from service level shortfalls that far outweigh any inventory cost savings, reflecting the business reality where compromised customer service can erode trust and jeopardize contracts.

Third, under insufficient inventory conditions (e.g., $multiplier = 0.95$, $weight\_cumul = 1.0$), substantially larger variations in the regret distribution as observed. This scenario, though undesirable for practical operations, reveals an important dynamic: when inventory is critically low, the OFA component becomes paramount. While simple allocation algorithms suffice with ample stock, constrained conditions demand more advanced prioritization to minimize stockout impacts on service levels and lateness.

After fixing Base Stock parameters to their optimal values ($multiplier = 0.99$, $weight\_cumul = 0.8$), the sensitivity of the remaining OFA parameters was analyzed. As anticipated from earlier findings, the impact of these parameters is relatively minor, with standardized coefficients orders of magnitude smaller than those of Base Stock parameters.

To handle the mixed nature of parameters (continuous and categorical), Ridge Regression with One-Hot Encoding was employed. Figure 6.8 presents the standardized coefficients from the Ridge Regression analysis.

Among OFA parameters, solver choice exhibits the strongest influence, though still mod-

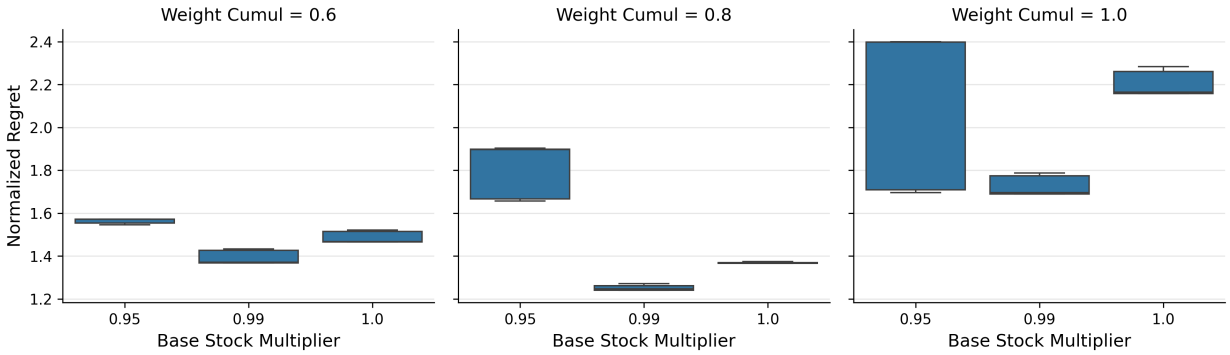

Figure 6.7 Box plots showing total regret distributions for different Base Stock parameter combinations. Non-overlapping distributions for top performers highlight the parameters' dominant impact on system performance.

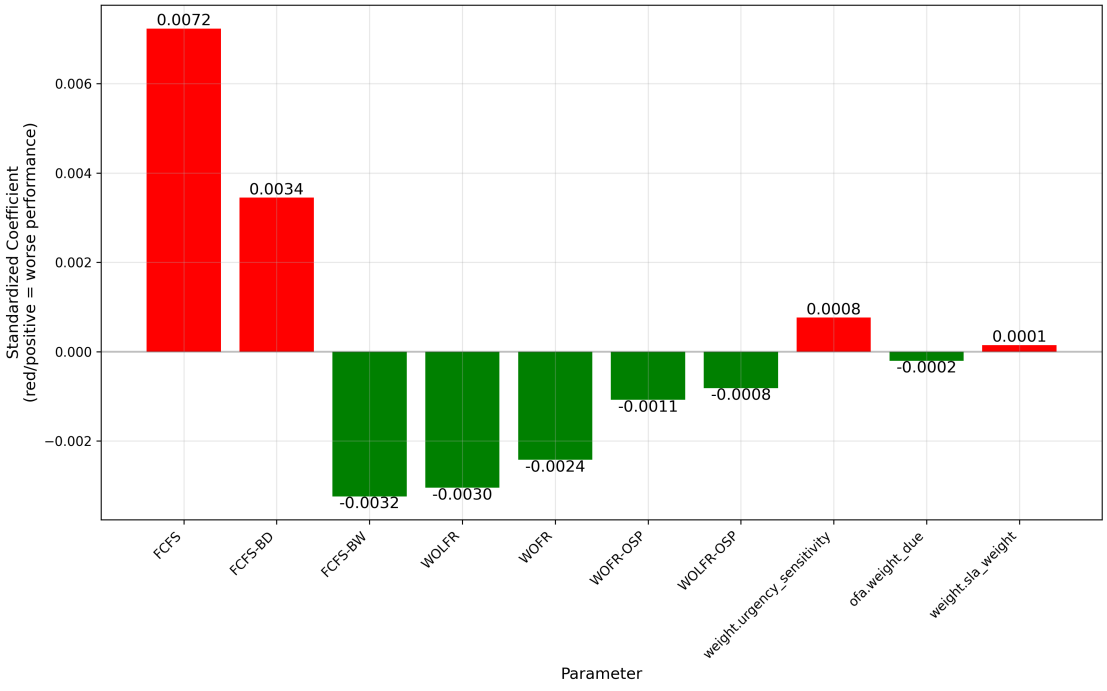

Figure 6.8 Standardized coefficients for OFA parameters using Ridge Regression with One-Hot Encoding. While solver choice shows the strongest effect among OFA parameters, all coefficients remain below 0.01, indicating minimal impact on overall performance.

est in absolute terms. The naive FCFS (First-Come-First-Served) solver shows the highest correlation with regret (coefficient +0.0072), indicating that this approach consistently underperforms compared to more advanced strategies. FCFS-BD (By Due Date) similarly shows a positive correlation (+0.0034), suggesting that due date prioritization alone is insufficient for optimal allocation.

Interestingly, FCFS-BW (By Weight) emerges as the most effective variant (coefficient -0.0032), aligning with the best-performing parameter combinations. This success could be attributed to its order weighting scheme specifically designed for optimizing allocation decisions. Apart from the 2 worst solvers, all other show minimal and statistically insignificant performance differences.

The pre-RL experiment runs reveals three overarching insights:

First, the system is highly sensitive to inventory control decisions. Minor deviations from the optimal base stock multiplier can produce disproportionately large variations in cost and service outcomes. Consequently, the correctness of replenishment policies is of paramount importance, overshadowing other factors like the OFA solver choice, especially when inventories are sufficient.

Second, when inventory replenishment policies are near-optimal, variations in OFA policies yield only marginal gains or losses. However, if inventory levels are insufficient, effective OFA becomes more relevant and can help mitigate some of the penalties from low service levels and lateness. While inventory replenishment parameters dominates system performance (76% of variance explained), order fulfillment parameters becomes critical precisely when inventory planning fails. This creates a decision hierarchy paradox : optimal performance requires near-perfect inventory control decisions (sensitivity $< 1\%$), which can make OFA parameters largely irrelevant, while suboptimal inventory turns OFA into damage control, where allocation logic explains over 25% of performance variance.

Third, having identified the best combination of base stock and OFA parameters, they are fixed as the *baseline configuration*. This configuration serves two purposes : it represents a **strong heuristic baseline** against which more advanced methods, including the upcoming RL approach, can be compared, and it ensures that subsequent experiment runs (e.g., with RL) focus on replenishment decisions that truly matter, without confounding suboptimal parameter selections.

6.5.6 Stage 2: RL System Evaluation and Comparative Analysis

In this second experiment Stage, we evaluate our proposed RL system, as described in Section 6.4. The optimized heuristic identified in Stage 1 serves two distinct and crucial roles in this evaluation. First, it acts as the primary *performance benchmark* against which the final RL policy is compared. Second, its loss values are used to construct the normalized *reward signal* for the RL agents, as formulated in Equation (6.16). This technique of relative reward shaping provides a dynamic and relevant learning target.

Prior to the main experiment runs, extensive preliminary sweeps (approximately 250 runs in total) were conducted to explore parameter importance and refine the hyperparameter distributions. These preliminary investigations helped identify promising regions of the parameter space and establish appropriate ranges for the final sweep. Several auxiliary parameters were fixed during the main experiment runs based on these preliminary optimizations.

For the main experimental campaign, a comprehensive hyperparameter sweep was conducted, testing 98 distinct parameter configurations across several categories: training dynamics (total training steps and steps per iteration), multi-agent architecture parameters (governing agent interactions and balance between global and local agent influence), action space design (different action space types, including both discrete and continuous formulations), and PPO-specific parameters (key algorithmic settings such as learning rate and discount factor).

Each experiment run was structured around multiple training iterations. Within each iteration, agents completed training episodes using a rotating set of 35 unique dataset samples, followed by 15 evaluation episodes using a separate test set. At each time step within every episode, agent-specific rewards were computed as the normalized lift compared to the baseline heuristic method.

For all experiment runs, hyperparameter sweeps were conducted using Weights & Biases (W&B) for experiment tracking and efficient parallelization across multiple machines. Bayesian optimization guided the exploration of the parameter space, automatically suggesting promising configurations based on previous results. This approach helped identify high-performing parameter combinations more efficiently than grid or random search would allow.

Figure 6.9 illustrates the progression of mean rewards across the 98 hyperparameter configurations, revealing distinct phases in the optimization process:

The initial exploration phase, covering runs 1 to 20, showed high variability but rapid improvement. The first two runs performed poorly, with rewards around -0.2, followed by high oscillations (standard deviation of 0.1) between -0.21 and +0.19. The mean reward hovered near zero but showed a slight upward trend, and by run 3, the system had already surpassed

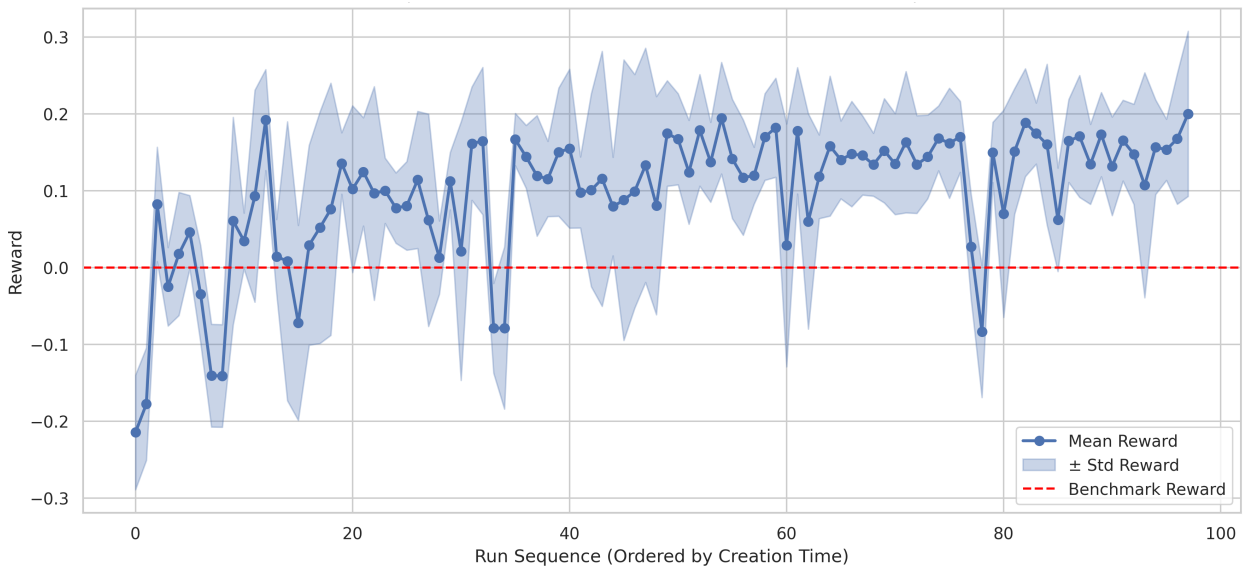

Figure 6.9 Hyperparameter Sweep Evolution

the baseline with an 8% improvement.

The rapid convergence to effective configurations in the final sweep was enabled by insights gained from preliminary runs. Early experiment runs refined parameter distributions and highlighted promising regions of the parameter space, allowing the final sweep to concentrate on the most impactful combinations. As a result, baseline-beating performance was achieved in just three runs, thanks to this initial groundwork.

In the refinement phase, spanning runs 20 to 40, performance stabilized while continuing to improve. Oscillations were reduced (standard deviation of 0.07), the mean reward increased to +0.09, and the maximum reward reached +0.17. All but two runs exceeded the baseline, and a 19% improvement was achieved by run 13.

The convergence phase, covering runs 40 to 98, exhibited robust and stable performance. Oscillations further decreased (standard deviation of 0.05), the mean reward stabilized at +0.13, and the maximum reward peaked at +0.20. Only one run failed to outperform the baseline.

Overall, the sweep demonstrated both computational efficiency and strong performance distribution. In terms of efficiency, the average run duration was 12 hours (ranging from 1 to 48 hours), with a total compute cost of approximately 1,200 CPU-hours, parallelized across three virtual machines (20 vCPUs each at $0.73/hour). The total wall time was about 50 hours, resulting in a total cost of roughly $110. From a performance perspective, the 90th percentile of runs achieved a 17% improvement, the median improvement was 12%, and even

the 10th percentile performed slightly above the baseline. This robust performance across different parameter combinations, coupled with the efficient convergence to high-performing configurations, suggests that the RL architecture is inherently stable and effective.

To understand the relative importance of different parameters and their impact on model performance, multiple linear regression analysis were conducted using normalized rewards as the dependent variable. The normalization was performed relative to the best run's performance and the standard deviation across all runs:

$$\text{Normalized Reward} = \frac{\text{Reward} - \text{Reward}_{\text{best}}}{\text{Reward}_{\text{std}}} \tag{6.18}$$

Figure 6.10 illustrates the explanatory power (R^2) of different parameter groups, revealing a clear hierarchy in parameter importance.

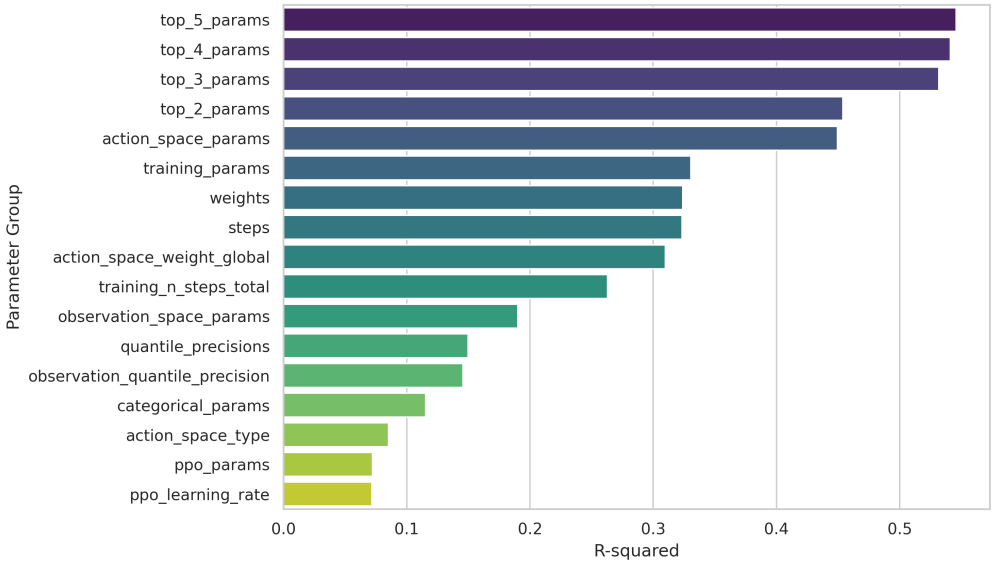

Figure 6.10 Model Explanatory Power by Parameter Group

The analysis demonstrates that the top three parameters (action.weight_global, training_n_steps_total, and action.space_type) collectively account for 53% of performance variance. This represents a significant increase over using just the top two parameters ($R^2 = 0.45$) or the single most influential parameter ($R^2 = 0.31$). PPO-specific parameters showed minimal impact, suggesting robustness to algorithm-specific settings when using reasonable initial values.

Figure 6.11 provides a detailed examination of the three key parameters, each revealing distinct patterns of influence:

Action Weight Global ($R^2 = 0.31$) showed the strongest individual influence, with a pro-

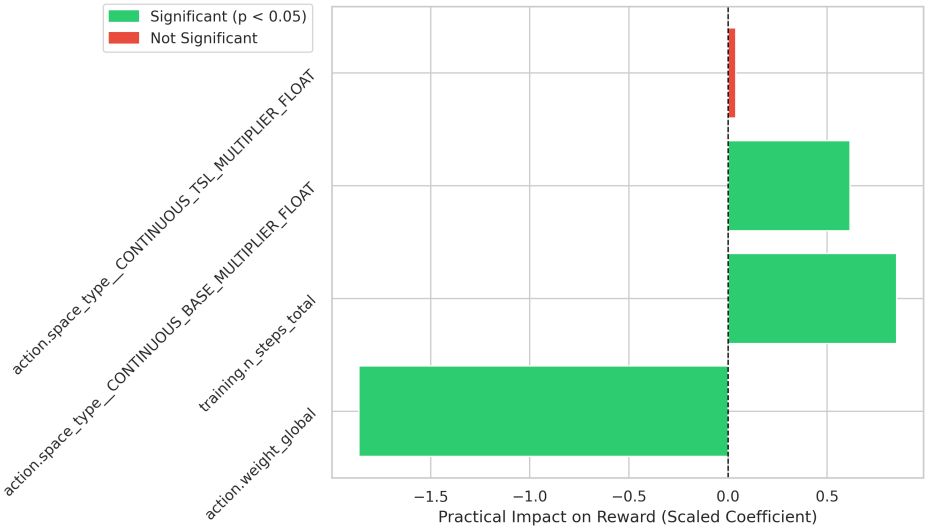

Figure 6.11 Parameter Sensitivity: Top 3 Parameters

nounced negative correlation with performance. The optimal configuration allocates only 20% of decision weight to the global agent, with the remaining 80% distributed among SLA-specific agents (approximately 27% each). The worst-performing runs used significantly higher global weights (up to 0.94), suggesting that excessive centralization of decision-making hinders performance. This finding indicates that while global coordination matters, specialized agents should retain primary control over their respective SLA groups.

Training Steps Total ($R^2 = 0.26$) demonstrated a strong positive correlation with performance, with higher step counts consistently yielding better results. The data shows a clear pattern: maximum steps (32,000) dominated the top performances, appearing in 7 of the top 8 runs; minimum steps (5,000) characterized poor performance, present in 6 of the bottom 8 runs; and the relationship appears monotonic, suggesting potential benefits from even longer training periods.

The CONTINUOUS_BASE_MULTIPLIER_FLOAT Action Space Type (see Appendix C.5) ($R^2 = 0.09$), demonstrated consistent association with superior performance. All top 8 runs employed this configuration, though its presence alone didn't guarantee success, as evidenced by its use in some poorly performing runs.

Figures 6.12 and 6.13 provide additional insights into specific parameter categories.

Among training parameters, n_steps_per_iteration emerged as particularly noteworthy, showing a significant negative correlation with performance. Smaller iteration sizes, which enable more frequent training and evaluation cycles, consistently outperformed larger ones. The optimal configuration used the minimum value (128 steps per iteration), while the poor-

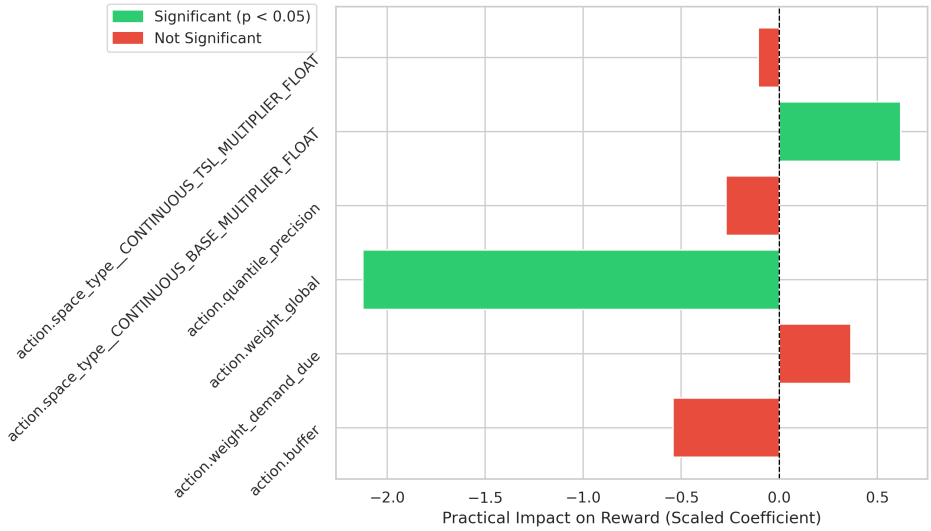

Figure 6.12 Parameter Sensitivity: Action Space Parameters

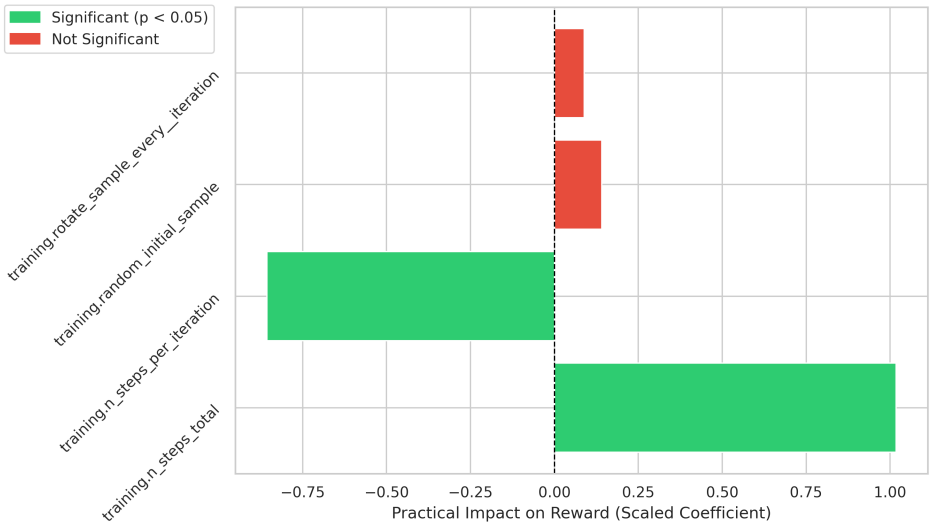

Figure 6.13 Parameter Sensitivity: Training Parameters

est performance occurred with the maximum value (2,048 steps). This suggests that in this environment, which is computationally more demanding than typical RL benchmarks like Atari games, more frequent feedback loops provide substantial benefits despite the classic exploration-exploitation trade-off.

Analysis of the best-performing model's learning progression reveals distinctive phases in the training process and provides insights into the RL system's adaptation capabilities. Figure 6.14 illustrates the evolution of both training and evaluation performance across iterations, with shaded regions representing one standard deviation around the mean reward.

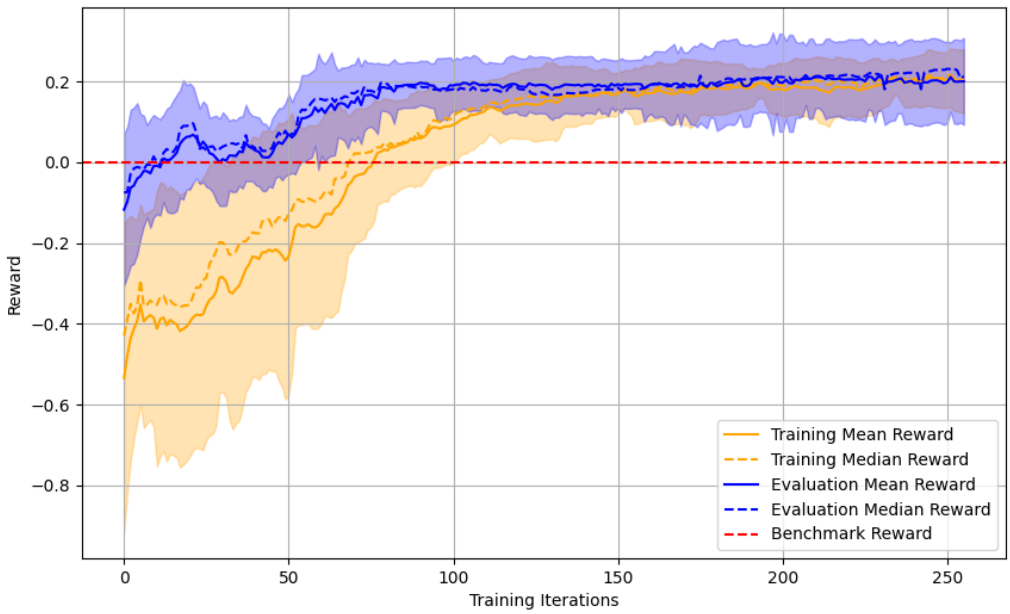

Figure 6.14 Learning Curve of Best Model

The learning curve reveals a characteristic pattern of RL training in complex environments. Initially, the agents perform worse than the baseline methods, reflecting the exploration phase where the model is learning the dynamics of the environment. However, as training progresses, the performance steadily improves, eventually surpassing the baseline consistently. This progression demonstrates the model's ability to learn effective strategies through experience, despite the challenging multi-objective nature of the task.

6.5.7 Comparative Performance Analysis

Having identified the optimal configuration for both the heuristic benchmark and the RL system, we now conduct a detailed head-to-head comparison to assess the benefits of our proposed approach.

Detailed analysis of the best-performing model reveals nuanced performance across different metrics and SLA groups.

Figure 6.15 shows the evolution of losses over time for each SLA group.

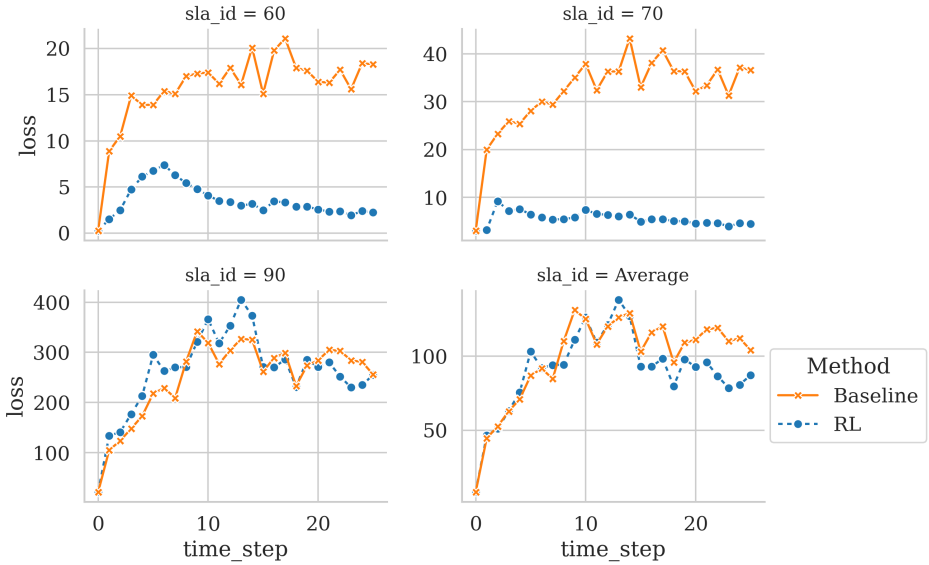

Figure 6.15 Loss Evolution Over Time by SLA Group

The RL approach demonstrates time-varying effectiveness across service level tiers. The system shows strongest initial gains in lower SLA groups (60, 70), where the performance advantage over the baseline continuously expands. The high-priority group (SLA 90) exhibits a more complex learning curve, starting below baseline performance before achieving competitive results. This pattern suggests that the RL approach requires more time to optimize high-priority service levels, but ultimately converges to superior performance across all groups.

The distribution of performance across samples (Figure 6.16) highlights the model's robustness:

For SLA groups 60 and 70, the RL method demonstrates remarkably stable performance with minimal variability, consistently achieving lower losses than the baseline. Group 90 shows higher variability (coefficient of variation ≈ 0.3) in both methods, with the RL approach showing slightly higher mean losses but comparable overall distributions.

Looking at the Average across groups, despite significant variability ($CV \approx 20\%$ for both methods) and distribution overlap, the RL approach maintains approximately 10% lower losses across all quartiles compared to the baseline. This indicates that the RL approach

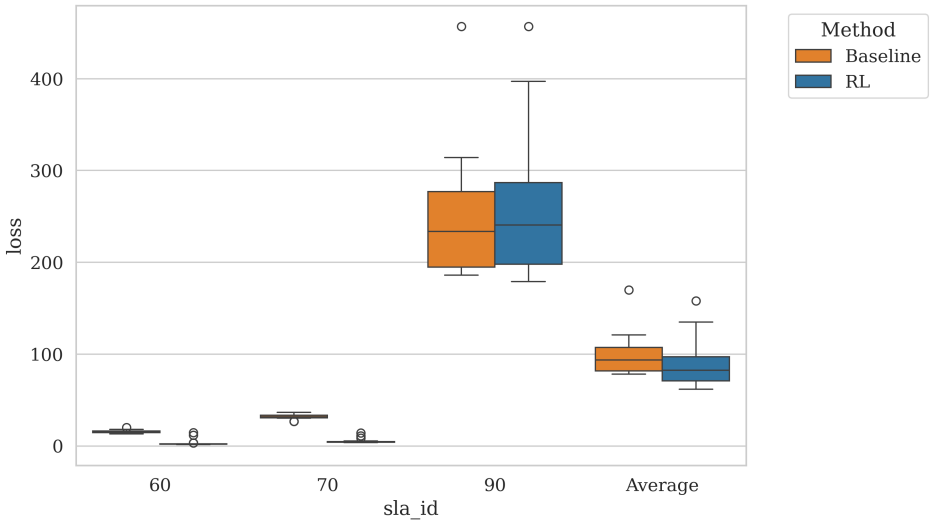

Figure 6.16 Loss Distribution by SLA Group

delivers robust performance improvement across various operational scenarios.

Figures 6.17 and 6.18 provide normalized views of improvements over the baseline.

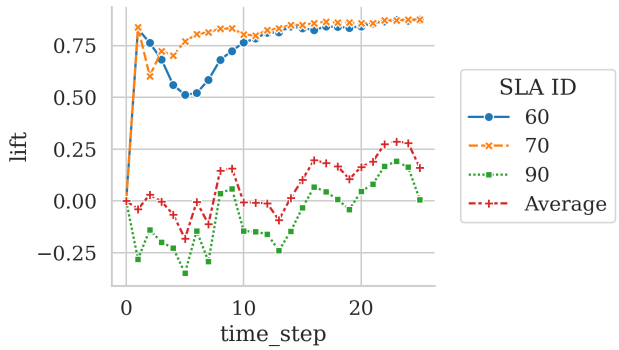

Figure 6.17 Lift Over Time by SLA Group

The lift metrics reveal exceptional performance for groups 60 and 70, achieving improvements of over 75% compared to the baseline. Group 90 shows more modest results, starting with a negative lift around -25% but improving over time to achieve slightly negative to neutral lift. The average lift across all groups, while not a simple average of individual group lifts due to the weighting of absolute losses, shows steady improvement throughout the time horizon, ranging from ±15% in the first half to consistently positive values exceeding 25% in the latter half.

The lift distribution reveals robust performance across samples, with the 25th percentile at 0, median at 0.14, and 75th percentile at 0.21, indicating that 75% of samples showed improve-

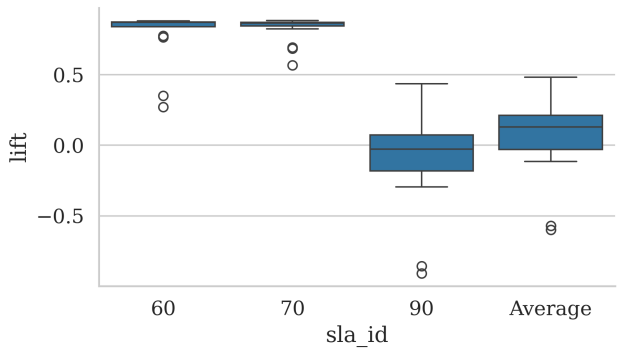

Figure 6.18 Lift Distribution by SLA Group

ment over the baseline. While two edge cases with lifts of -0.6 skew the mean downward, the median lift of 0.14 provides a representative measure of typical performance.

The model achieved significant improvements in key operational metrics, including a 16% reduction in inventory holding costs, which accounted for 98% of the combined loss (excluding SLS); a marginal increase in lateness from an already low (good) baseline, remaining well within acceptable operational limits; and a minimal increase in service level shortfall ($<$ 0.5%), maintaining near-perfect service levels. Increases in lateness and service level shortfall were reasonable, given their relatively small contribution to total costs. The trade-off between inventory reduction and service levels was particularly favorable, as the slight increase in lateness and service level shortfall occurred at already low baseline levels.

This performance profile demonstrates the model's ability to achieve substantial inventory cost reductions while maintaining high service levels. The 16% reduction in inventory costs represents significant potential savings, achieved without meaningfully compromising service quality or timeliness.

6.6 Conclusion

This research introduces a novel inventory control system that successfully bridges the gap between theoretical reinforcement learning advancements and practical supply chain challenges. By integrating multi-agent reinforcement learning, quantile forecasting, and mixed-integer linear programming, the approach addresses critical real-world complexities often overlooked in traditional and even state-of-the-art solutions.

The work represents a deliberate departure from the conventional approach of incremental improvements to existing solutions. Instead, it directly addresses previously unexplored complexities in inventory control that reflect actual operational challenges: multiple SKUs with

interdependent demand patterns, tiered service level agreements with distinct requirements, non-stationary and intermittent demand profiles, and dynamic lead times. While this combination of elements makes the problem substantially more challenging, it reflects the actual conditions faced by many distribution-centric businesses.

This methodological choice has produced several significant contributions. First, the research demonstrates that reinforcement learning can effectively handle these complex, real-world inventory control scenarios. The system's ability to achieve a 9% normalized lift over an already optimized baseline, primarily through a 16% reduction in inventory holding costs, validates the practical viability of RL-based approaches in distribution-centric businesses.

Second, the multi-agent architecture proves particularly effective at balancing competing objectives across different service level tiers. The decentralized approach, with specialized agents for each SLA group plus a global coordinator, provides a natural framework for handling the inherent tensions between inventory costs and service levels. Third, the integration of quantile forecasting with RL represents a novel approach to handling demand uncertainty, offering agents a richer representation of potential future scenarios while maintaining computational efficiency.

Rather than comparing the system against methods that operate under fundamentally different assumptions, a robust baseline was developed that incorporates these real-world complexities. This baseline itself represents an advancement over traditional approaches, incorporating quantile forecasting instead of normal distribution assumptions, explicit handling of multi-SLA constraints, and optimization for business-specific objective functions. The RL system's significant improvements over this enhanced baseline demonstrate its practical value.

A distinguishing feature of the approach is its architectural flexibility and modular design. Each component—from the observation space construction to the reward function definition—can be adapted to specific operational contexts without requiring fundamental changes. This adaptability extends to the objective function itself, allowing arbitrary cost structures and performance metrics—a significant advantage over common methods that often embed simplified assumptions in their mathematical foundations.

The system demonstrates robust performance across various scenarios, maintaining approximately 10% lower losses across all quartiles compared to the baseline. This reliability makes it particularly valuable for businesses facing inventory management challenges or operating under specialized constraints.

The comprehensive implementation details and code in the appendix create a flexible frame-

work for both academic exploration and industrial adaptation. While the system's effectiveness is demonstrated with specific components (e.g., synthetic quantile forecasting, particular reward structures), each element can be modified or replaced to suit different operational contexts.

Several methodological considerations warrant discussion for practical implementation. The cross-validation methodology used in model selection could be enhanced by incorporating a separate final evaluation phase with completely new dataset samples, followed by an independent test phase. This would provide more rigorous validation of the system's generalization capabilities.

Additionally, while the synthetic forecasting approach enables controlled experimentation and theoretical validation, real-world implementation would require integration with operational forecasting methods. However, the system's modular architecture allows for such adaptations without compromising its fundamental effectiveness.

This work opens several promising avenues for future research : The integration of curriculum learning techniques and reward shaping could accelerate training and improve performance. Investigating advanced architectural elements such as attention networks or transformers could enhance the system's ability to capture temporal dependencies in demand patterns. While the current implementation demonstrates strong performance on small problem sizes, systematic investigation of scaling behavior with larger datasets (thousands of SKUs and customers) would provide valuable insights for enterprise applications. This includes analyzing both computational requirements and opportunities for performance optimization through GPU acceleration and parallel environment sampling. Further research should examine the system's sensitivity to variations in demand patterns and forecast accuracy, operational constraints such as lead times and service level requirements, cost structures, including ordering costs and profit considerations, and implementation parameters and reward formulations.

In conclusion, this research demonstrates that reinforcement learning, when properly adapted to the complexities of real-world supply chains, can deliver significant practical benefits in inventory control. By explicitly tackling real-world complexities rather than abstracting them away, a foundation has been established for future research while providing distribution-centric businesses with tools that match their operational reality. While the results show substantial improvements over traditional methods, they also suggest considerable potential for further advances, particularly in contexts with complex constraints and competing objectives.

CHAPITRE 7 DISCUSSION GÉNÉRALE

7.1 Introduction

La présente discussion générale vise à synthétiser et à analyser de manière critique les contributions fondamentales de cette thèse, intitulée "*Conception d'outils d'aide à la décision visant à améliorer l'efficacité opérationnelle de la chaîne d'approvisionnement dans un contexte de modernisation technologique*". Face à l'accroissement des dynamiques de marché, aux attentes fluctuantes des consommateurs et à l'explosion des données, la problématique centrale de cette thèse a exploré l'impact transformateur des technologies d'intelligence artificielle sur la gestion de la chaîne d'approvisionnement, avec pour objectif de renforcer sa résilience opérationnelle, son efficacité logistique et la satisfaction client.

Pour décomposer cette problématique multidimensionnelle, notre recherche a été structurée autour de trois sous-problèmes distincts mais interconnectés, chacun ciblant des leviers d'optimisation spécifiques au sein de la chaîne d'approvisionnement. Ces axes ont été délibérément choisis pour couvrir des domaines critiques de la gestion logistique : la consolidation des marchandises, l'optimisation de l'allocation des articles pour l'exécution des commandes (OFA) et le contrôle de l'inventaire.

Les trois chapitres précédents ont détaillé les méthodologies développées et les résultats obtenus dans l'exploration de ces sous-problèmes, apportant des réponses concrètes à la question fondamentale de la thèse et mettant en lumière nos contributions scientifiques. Ce chapitre propose une réflexion approfondie sur ces contributions, en soulignant leur interdépendance, leur progression méthodologique et leur impact global sur l'optimisation de la chaîne d'approvisionnement à l'ère de la modernisation technologique. Nous procéderons à une synthèse de nos apports, suivie d'une analyse critique de nos choix méthodologiques et de nos résultats, avant de conclure sur les perspectives de recherche futures qu'ouvre ce travail.

7.2 Synthèse et complémentarité des contributions

7.2.1 Progression méthodologique et intégration des contributions

Notre recherche se caractérise par une progression méthodologique délibérée et cohérente, qui s'étend de l'analyse exploratoire par fouille de données à des approches d'apprentissage automatique, tout en s'ancrant dans les principes fondamentaux de la recherche opérationnelle. Cette trajectoire n'est pas une simple succession chronologique, mais représente une

évolution conceptuelle où chaque contribution s'appuie sur les précédentes, enrichissant la portée et la profondeur de notre cadre d'optimisation.

La première contribution s'est focalisée sur la consolidation des marchandises, en introduisant une méthodologie novatrice basée sur les règles d'association. Cette approche exploite les patterns cachés dans les données de transaction non agrégées pour optimiser la consolidation spatiale et temporelle. À la différence des méthodes traditionnelles de recherche opérationnelle, qui nécessitent souvent une agrégation des données et peuvent ainsi masquer des informations cruciales sur les comportements des clients et les patterns de demande intermittente, notre méthode préserve ces subtilités. Ce travail a posé les jalons méthodologiques des contributions subséquentes en démontrant l'efficacité des approches axées sur les données dans un contexte logistique complexe.

La deuxième contribution a étendu cette perspective en intégrant la consolidation spatiale dans un cadre plus large d'allocation des articles pour l'exécution des commandes (OFA). Le modèle WOFR-EC (Weighted Order Fill Rate Encouraging Consolidation) que nous avons développé représente une avancée significative dans la conciliation de deux objectifs souvent perçus comme contradictoires : la minimisation des coûts d'expédition et l'optimisation de la satisfaction client. Cette contribution incorpore et améliore la méthodologie de consolidation spatiale proposée dans la première, illustrant ainsi la complémentarité et la continuité de notre démarche de recherche.

La troisième contribution, plus ambitieuse et complexe, a abordé le contrôle d'inventaire à travers le développement d'un système basé sur l'apprentissage par renforcement (RL). Ce système intègre l'OFA dans un cadre plus holistique, soulignant que si l'optimisation de l'allocation des commandes est importante, ce sont les décisions de réapprovisionnement qui exercent l'influence la plus déterminante sur les performances globales de la chaîne d'approvisionnement. Cette approche représente une progression naturelle vers une vision plus intégrée et dynamique de la gestion de la chaîne d'approvisionnement, capable de gérer des environnements complexes et stochastiques.

7.2.2 Complémentarité des approches méthodologiques

La complémentarité de nos contributions se manifeste non seulement par leur progression thématique, mais également par la diversité de leurs approches méthodologiques. Chaque contribution a délibérément employé une méthodologie distincte, choisie pour sa pertinence et son efficacité face au sous-problème traité.

Pour la consolidation de marchandises, nous avons privilégié les règles d'association, une

technique de fouille de données particulièrement adaptée à l'identification de patterns non triviaux dans des données transactionnelles volumineuses. Cette approche exploratoire a permis de révéler des opportunités de consolidation qui seraient restées inaccessibles par des méthodes analytiques traditionnelles, souvent limitées par des hypothèses simplificatrices..

Pour l'OFA, nous avons développé un modèle de programmation linéaire en nombres entiers mixtes (MILP). Ce choix méthodologique est justifié par la nature du problème d'allocation, qui requiert une prise de décision déterministe et optimale, tout en respectant des contraintes strictes d'inventaire et de niveau de service client. Le MILP a offert une solution robuste et transparente à ce problème bien défini.

Pour le contrôle d'inventaire, nous avons eu recours à l'apprentissage par renforcement (RL), reconnaissant la nature intrinsèquement dynamique, stochastique et complexe de ce problème. Cette approche a permis d'apprendre des politiques adaptatives qui peuvent réagir aux conditions évolutives du marché, à la variabilité de la demande et aux exigences changeantes des clients, sans nécessiter une modélisation explicite et exhaustive de ces dynamiques.

Cette diversité méthodologique n'est pas fortuite ; elle reflète notre conviction qu'il n'existe pas d'approche universelle pour l'optimisation de la chaîne d'approvisionnement. Au contraire, nous avons adopté une vision pragmatique et éclectique, sélectionnant pour chaque sous-problème la méthodologie la plus appropriée à sa structure inhérente, à ses objectifs spécifiques et à la nature des données disponibles. Cette flexibilité méthodologique est un gage de robustesse et d'applicabilité de nos solutions aux réalités complexes des environnements industriels.

7.3 Analyse critique des contributions

7.3.1 Première contribution : Consolidation de marchandises par fouille de données

Notre première contribution a introduit une approche novatrice pour la consolidation des marchandises, s'appuyant sur les règles d'association pour identifier des opportunités de groupement spatial et temporel directement à partir des données transactionnelles.

En comparaison avec les résultats d'études de cas de la littérature, notamment ceux de [95] et [96], nos résultats ont démontré une capacité à générer des solutions de coûts similaires, tout en offrant des avantages qualitatifs distincts en termes de flexibilité et d'adaptabilité.

La caractéristique distinctive de cette contribution réside dans sa capacité à traiter directement des données brutes ainsi qu'à détecter puis exploiter des patterns qui demeureraient

inaccessibles aux approches traditionnelles.

Les méthodes conventionnelles de recherche opérationnelle exigent souvent une agrégation préalable des données pour des raisons de tractabilité mathématique. Cette agrégation, bien que nécessaire pour certains modèles, entraîne une perte inévitable d'informations granulaires sur les comportements spécifiques des clients et les dynamiques spatio-temporelles des commandes.

Notre approche, en opérant directement sur les données transactionnelles non agrégées, préserve ces subtilités et permet d'identifier des opportunités de consolidation autrement inexploitées, offrant ainsi une perspective plus fine et plus riche de l'optimisation logistique.

L'efficacité de cette approche repose toutefois sur la présence de patterns spatio-temporels exploitables dans les données. Dans des contextes de faible densité géographique ou de commandes temporellement dispersées, les opportunités de consolidation seraient naturellement limitées. Cette dépendance contextuelle suggère que l'application de notre méthodologie gagnerait à être précédée d'une analyse préliminaire de la structure des données pour évaluer son potentiel d'amélioration. Une caractérisation formelle des conditions minimales requises (densité géographique seuil, fréquence de commandes, variabilité temporelle) permettrait aux praticiens d'identifier a priori les situations où notre approche apporte une valeur ajoutée significative.

La consolidation spatiale modifie fondamentalement le paradigme de livraison en exigeant des clients qu'ils se déplacent vers des points de collecte. Ce transfert d'effort constitue une limitation significative de notre approche, à la fois pratique et méthodologique. D'un point de vue pratique, cette modification pourrait affecter l'adoption du système et la satisfaction client. D'un point de vue méthodologique, notre modélisation ne capture pas explicitement ce coût client : une livraison à domicile n'impose aucun déplacement au client (le coût de transport est entièrement supporté par le fournisseur), alors qu'un point de collecte génère un coût additionnel en temps, distance et potentiellement en frais monétaires (carburant, stationnement). Notre fonction objectif optimise les coûts logistiques du fournisseur mais ignore cette asymétrie dans la distribution des coûts totaux, risquant ainsi de surestimer l'acceptabilité réelle des solutions proposées.

Une extension méthodologique consisterait à intégrer une pénalité représentant ce coût client, potentiellement modélisée comme une fonction de la distance entre le domicile du client et le point de collecte assigné. Cependant, la quantification de ce coût soulève des défis : comment valoriser monétairement le temps client ? Comment capturer l'hétérogénéité des préférences individuelles (certains clients privilégiant la commodité, d'autres acceptant un déplacement pour un rabais) ? Des approches issues de l'économie des transports (valeur du

temps de déplacement) ou de la théorie du choix discret (modèles logit mixtes pour capturer l'hétérogénéité) pourraient fournir des cadres méthodologiques adaptés.

Du point de vue opérationnel, une stratégie d'atténuation consisterait à redistribuer une partie des économies logistiques réalisées sous forme d'incitations tarifaires ou d'autres avantages (flexibilité horaire étendue, services additionnels aux points de collecte), favorisant ainsi l'acceptabilité du modèle. L'efficacité de telles mesures reste toutefois à valider empiriquement, potentiellement via des enquêtes de préférences révélées ou des expérimentations contrôlées mesurant la sensibilité des clients à différents niveaux d'incitation et configurations de service.

Concernant la sélection des règles d'association, notre seconde contribution a significativement amélioré ce processus en introduisant un algorithme automatisé qui remplace la sélection manuelle initiale. Ce développement représente une évolution naturelle et positive de notre approche, renforçant son autonomie et sa scalabilité. La paramétrisation de cet algorithme ouvre des perspectives intéressantes pour de futures recherches, notamment l'intégration potentielle avec des techniques d'apprentissage par renforcement pour une optimisation adaptative des paramètres en fonction des résultats observés, permettant une adaptation dynamique aux conditions changeantes.

7.3.2 Deuxième contribution : Allocation des articles pour l'exécution des commandes

Notre deuxième contribution a introduit le modèle WOFR-EC (Weighted Order Fill Rate Encouraging Consolidation), qui intègre la consolidation spatiale dans l'optimisation de l'allocation des articles pour l'exécution des commandes. Les résultats empiriques ont démontré qu'avec un paramétrage adéquat, notre modèle peut réduire significativement le nombre d'arrêts dans les itinéraires de livraison (jusqu'à 21%) tout en maintenant des niveaux de service client élevés.

La force de cette contribution réside dans sa capacité à établir un équilibre explicite et quantifiable entre deux objectifs souvent perçus comme antagonistes : la minimisation des coûts logistiques et l'optimisation du taux de remplissage des commandes (WOFR), un indicateur clé du niveau de service. En introduisant un mécanisme de pénalisation proportionnel au nombre de points de collecte visités, notre modèle encourage naturellement la consolidation spatiale sans imposer une structure rigide qui compromettrait l'adaptabilité aux besoins spécifiques des clients. Cette approche offre une flexibilité stratégique aux décideurs, leur permettant d'ajuster le compromis en fonction des priorités commerciales.

Le paramètre de pondération (ω) introduit une dimension de calibration dans le modèle.

Bien que nos expérimentations aient démontré une certaine robustesse face à des variations modérées de ce paramètre, sa détermination optimale nécessite une phase d'exploration empirique. Cette calibration, moins coûteuse computationnellement que l'entraînement d'un système d'apprentissage par renforcement, pourrait être automatisée dans de futures implémentations pour renforcer l'autonomie du système. Des approches d'optimisation bayésienne ou de recherche en grille adaptative permettraient d'identifier efficacement les valeurs optimales contextuelles.

L'analyse des résultats empiriques révèle un continuum de compromis entre coût et service, permettant aux décideurs de sélectionner le point d'équilibre le plus adapté à leur contexte. Nos expérimentations ont mis en évidence deux zones particulièrement intéressantes sur cette courbe d'efficience. La première permet une réduction des coûts sans aucune dégradation mesurable de la qualité de service. La seconde offre une réduction substantielle des coûts moyennant une diminution marginale du niveau de service, générant ainsi un rapport favorable entre gain économique et impact opérationnel.

Cette asymétrie favorable dans le compromis coût-service s'explique par l'existence d'opportunités de consolidation spatio-temporelle qui n'auraient pas été exploitées par des approches conventionnelles. Le modèle WOFR standard, par exemple, considère uniquement l'urgence et l'importance des commandes, ignorant leur proximité géographique dans les décisions d'allocation. En conséquence, il pourrait planifier des livraisons distinctes pour des commandes géographiquement proches mais d'urgences différentes. Le modèle WOFR-EC, en revanche, intègre cette dimension spatiale, identifiant des opportunités de groupement qui réduisent les coûts sans compromettre significativement les délais critiques.

Un aspect particulièrement intéressant de notre modèle WOFR-EC est sa capacité à réaliser implicitement de la consolidation temporelle. La dynamique d'optimisation du modèle peut naturellement conduire à une forme de consolidation temporelle émergente : lorsqu'une commande non urgente est située dans une région sans autres commandes actives, le modèle sera incité à retarder son traitement jusqu'à l'arrivée potentielle d'autres commandes dans la même région. Cette temporisation permet de distribuer la pénalité de visite du point de collecte sur plusieurs commandes, optimisant ainsi le rapport coût-service.

Le modèle WOFR-EC utilise toutefois le nombre d'arrêts comme proxy du coût logistique, découplant les décisions d'allocation et de routage. Ce choix méthodologique offre un compromis entre tractabilité et précision : intégrer un module VRP complet dans notre MILP aurait rendu le problème difficilement résoluble à l'échelle industrielle, mais suppose implicitement une relation monotone entre le nombre d'arrêts et la distance parcourue. Dans certaines configurations géographiques, cette hypothèse peut ne pas être respectée. L'écart

entre notre proxy et le coût réel dépend fortement de la topologie du réseau de distribution. Une validation empirique post-optimisation comparant les tournées réelles aux prédictions du modèle quantifierait l'ampleur de cette approximation. Une extension naturelle de notre travail consisterait à coupler le modèle WOFR-EC avec un module de résolution VRP, soit de manière séquentielle (allocation puis routage), soit de manière intégrée dans une approche de type branch-and-price.

Une deuxième limitation méthodologique concerne la modélisation de la capacité logistique via le paramètre k (nombre maximum de commandes traitables par centre de distribution). Cette contrainte de cardinalité suppose implicitement que le goulot d'étranglement se situe au niveau de la capacité de traitement opérationnel (préparation, picking, packing) plutôt qu'au niveau de la capacité volumétrique des véhicules de transport. Cette simplification présente deux risques d'inefficacité : (1) si les commandes sont volumétriquement petites, satisfaire k commandes pourrait laisser une capacité de chargement sous-utilisée dans les camions, générant des coûts de transport unitaires élevés ; (2) inversement, si les commandes sont volumineuses, k commandes pourraient excéder la capacité physique réelle des véhicules, rendant la solution infaisable en pratique. Sans analyse préalable de la corrélation entre nombre de commandes et du volume total occupé, ce proxy reste approximatif. Une extension naturelle consisterait à remplacer ou compléter la contrainte de cardinalité par une contrainte de capacité volumétrique multidimensionnelle ($\sum_i v_i \cdot q_i \leq V_{\max}$), intégrant explicitement les dimensions volume et poids, assurant ainsi une meilleure utilisation du volume de chargement et une faisabilité opérationnelle accrue.

La synergie avec notre première contribution mérite d'être soulignée. Le modèle WOFR-EC peut opérer de deux manières : (1) sans consolidation préalable, en pénalisant simplement le nombre de points de collecte pour encourager naturellement le regroupement spatial ; ou (2) en exploitant les opportunités de consolidation identifiées par les règles d'association (Contribution 1). Nos résultats empiriques distinguent clairement ces deux sources d'amélioration : l'apport de la consolidation par règles d'association est mesurable indépendamment de l'effet du mécanisme de pénalisation du WOFR-EC. Cette modularité confirme que le WOFR-EC constitue une contribution autonome qui peut bénéficier, mais ne dépend pas, de la méthodologie de consolidation développée dans notre première contribution. L'intégration des deux approches produit néanmoins des résultats synergiques supérieurs à l'application isolée de chacune, démontrant la cohérence de notre cadre méthodologique unifié.

7.3.3 Troisième contribution : Contrôle d'inventaire par apprentissage par renforcement

Notre troisième contribution a développé un système de contrôle d'inventaire basé sur l'apprentissage par renforcement (RL), capable de gérer efficacement des environnements complexes caractérisés par des demandes intermittentes, des SKUs multiples et des accords de niveau de service (SLA) hiérarchisés.

Notre système a démontré une amélioration significative par rapport à notre heuristique de référence, avec une réduction de 16% des coûts de détention d'inventaire tout en maintenant des niveaux de service élevés. L'heuristique utilisée comme référence n'est pas une approche standard de la littérature, mais une méthode avancée que nous avons spécifiquement développée et optimisée pour ce cas d'étude, intégrant déjà des éléments novateurs comme la prévision par quantiles et la gestion explicite des contraintes multi-SLA. Cette référence exigeante démontre la viabilité du RL dans des contextes complexes, mais limite la comparabilité directe avec les méthodes standard de la littérature. L'amélioration normalisée de 9% obtenue valide l'approche RL, bien qu'une évaluation complémentaire face à des benchmarks établis—politiques (s, S), algorithmes génétiques, ou autres méthodes de la littérature—aurait renforcé le positionnement de notre contribution dans le paysage méthodologique existant et facilité l'appréciation de sa performance relative.

Le choix d'adresser directement des complexités jusqu'alors peu explorées dans le contrôle d'inventaire reflète les défis opérationnels réels : SKUs multiples avec des patterns de demande interdépendants, accords de niveau de service hiérarchisés, profils de demande non-stationnaires et intermittents, et délais d'approvisionnement dynamiques. Cette combinaison d'éléments rend le problème substantiellement plus complexe, mais reflète les conditions réelles auxquelles sont confrontées de nombreuses entreprises de distribution.

Ce choix méthodologique a produit plusieurs contributions significatives. Premièrement, nous démontrons que l'apprentissage par renforcement peut efficacement gérer ces scénarios complexes de contrôle d'inventaire. La capacité du système à réaliser une amélioration normalisée de 9% par rapport à une référence déjà optimisée, principalement grâce à une réduction de 16% des coûts de détention d'inventaire, valide la viabilité pratique des approches basées sur l'apprentissage par renforcement dans les entreprises de distribution. Deuxièmement, notre architecture multi-agents s'avère particulièrement efficace pour équilibrer les objectifs concurrents entre différents niveaux de SLA, les groupes de priorité inférieure bénéficiant d'améliorations dépassant 75% pour certaines métriques, tout en maintenant la performance pour les catégories prioritaires.

Plusieurs facteurs expliquent la performance de notre approche.

L'**adaptabilité dynamique** représente un avantage majeur, car notre système RL peut ajuster ses décisions en fonction des performances récentes et des signaux de demande, contrairement aux heuristiques qui utilisent des paramètres statiques. Cette méthodologie se distingue également par son **exploitation complète des distributions de prévision** : plutôt que de se limiter à un seul quantile ou à une combinaison fixe de quantiles, notre système RL peut intégrer l'ensemble de la distribution de la demande prévue, permettant des décisions plus nuancées. Le traitement des **observations complexes** représente un autre atout considérable, notre cadre RL pouvant traiter des états détaillés au niveau des articles, capturant des informations granulaires sur les caractéristiques de la demande de chaque SKU. Enfin, l'**équilibrage multi-objectifs** s'avère un avantage déterminant, l'apprentissage par renforcement gérant naturellement les compromis entre objectifs concurrents, équilibrant efficacement les niveaux de service, les retards et les coûts d'inventaire.

Il est important de clarifier que la contribution principale ne réside pas dans le modèle entraîné spécifique, mais dans la méthodologie, l'architecture flexible et l'implémentation développée qui peuvent être adaptées à diverses situations.

Ce caractère modulaire et adaptable constitue l'une des forces de cette contribution. Dans un nouveau contexte d'application, les données d'entrée, les paramètres de récompense, et d'autres éléments peuvent être configurés pour refléter les spécificités du domaine concerné. Cette modularité permet également de modifier, d'ajouter ou de supprimer certains composants selon les besoins spécifiques de l'implémentation.

Toutefois, cette flexibilité soulève des questions pratiques de maintenance : dans quelles circonstances un réentraînement est-il nécessaire, et comment minimiser le coût de ces adaptations ?

Typologie des changements et stratégies d'adaptation. Trois types de modifications impliquent des stratégies distinctes de maintenance.

Les *changements architecturaux* (ajout/suppression de SKUs ou de caractéristiques modifiant la dimensionnalité des entrées/sorties) nécessitent un réentraînement car l'architecture neuronale ou le signal d'apprentissage changent fondamentalement. Cependant, ce réentraînement peut exploiter les connaissances antérieures via *transfer learning* et *warm start*. La littérature en continual reinforcement learning propose plusieurs techniques : réutilisation des poids pré-entraînés plutôt qu'initialisation aléatoire (réduisant le temps d'entraînement de 60 à 80% [180]), *progressive networks* ajoutant de nouvelles colonnes tout en gelant les précédentes [181], ou *value function initialization* exploitant les Q-values de tâches antérieures

[182]. Une extension consisterait à évaluer empiriquement ces techniques dans le contexte du contrôle d'inventaire multi-produits.

Le *drift temporel* (évolution graduelle de la demande, changements comportementaux, modifications opérationnelles) ne requiert pas de réentraînement complet mais un *fine-tuning* ou *continual learning* adaptant progressivement les poids aux nouvelles distributions [183], [184]. Des techniques récentes comme l'Adaptive Memory Realignment (AMR) réalignent sélectivement le buffer de replay quand un drift est détecté [185], évitant le coût du réentraînement intégral.

L'*optimisation des hyperparamètres* (learning rate, architecture, batch size, facteur d'actualisation) représente un cas particulier. Bien que ces paramètres ne soient pas appris par le réseau, leur optimisation via sweeps d'expériences peut s'avérer coûteuse car certains hyperparamètres nécessitent un réentraînement complet pour chaque évaluation. Néanmoins, plusieurs stratégies atténuent ce coût : exploration initiale sur datasets réduits, application des techniques de warm start mentionnées précédemment, et fréquence d'optimisation espacée (annuelle ou semestrielle suffit généralement pour des systèmes stables).

Stratégies de déclenchement en production. La littérature MLOps identifie trois approches principales pour déterminer quand intervenir [186]-[189]. Le *monitoring de performance* déclenche une intervention lorsque des métriques clés (taux de rupture, coûts d'inventaire, respect des SLA) se dégradent au-delà d'un seuil prédéfini (ex. : augmentation >X% sur fenêtre glissante de Y semaines). La *détection de drift* utilise des tests statistiques (Kolmogorov-Smirnov, Population Stability Index) pour identifier les changements distributionnels avant que la performance ne dégrade visiblement [190], permettant une intervention anticipative. La *réévaluation périodique* effectue des fine-tunings préventifs selon une fréquence déterminée par backtesting : diviser les données historiques en périodes "passé" et "futur", mesurer le temps moyen jusqu'à dégradation, puis établir une fenêtre de maintenance égale à 50% de ce temps.

Extensions de recherche. Une caractérisation systématique enrichirait ces lignes directrices : (1) définir des métriques et seuils de déclenchement spécifiques au contrôle d'inventaire ; (2) évaluer empiriquement l'impact du drift temporel avec et sans interventions périodiques, quantifiant le *coût d'opportunité de l'inaction* ; (3) comparer différentes techniques d'adaptation (warm start, progressive networks, AMR, fine-tuning) selon les types de changements ; (4) développer des méthodes de détection de drift adaptées aux séries temporelles intermittentes ; (5) caractériser les conditions (variabilité de demande, complexité contraintes) sous lesquelles notre architecture RL justifie l'investissement en maintenance comparativement à des heuristiques adaptatives plus simples.

Au-delà de la maintenance adaptative, d'autres dimensions critiques méritent attention.

Tests de robustesse et généralisabilité. Bien que notre validation empirique démontre l'efficacité sur le cas d'étude présenté, elle porte sur un ensemble limité de configurations. Des expérimentations sur des jeux de données diversifiés renforceraient la confiance dans sa généricité et identifieraient les conditions limites de son applicabilité.

Scalabilité et complexité algorithmique. La validation porte sur un système de 5 produits et 34 clients, générant un espace d'états et d'actions de dimension modeste. Le passage à l'échelle industrielle — typiquement des milliers de SKUs distribués à des milliers de clients — pose un défi de scalabilité computationnelle majeur. La dimensionnalité des espaces d'observation ($O(n \times m)$ où n = produits, m = clients) et d'action croît linéairement avec le nombre d'entités, mais la complexité d'apprentissage des réseaux de neurones profonds croît de manière sur-linéaire. Les temps d'entraînement deviendraient prohibitifs sans architecture spécialisée.

Plusieurs stratégies méthodologiques pourraient soutenir la mise à l'échelle du système.

Premièrement, les architectures multi-agents permettent de décomposer le problème en sous-groupes d'agents coordonnés : la littérature récente montre que les décompositions hiérarchiques [191], [192] et les méthodes de factorisation de valeurs comme VDN [193] et QMIX [194] réduisent la complexité des calculs, tandis que le partage de paramètres sélectif améliore l'efficacité d'apprentissage [195].

Deuxièmement, les techniques de réduction de dimensionnalité notamment via autoencodeurs variationnels (VAE) pour compresser l'espace d'états [196], [197] atténuent la croissance de l'espace état-action.

Troisièmement, le transfer learning accélère l'apprentissage en transférant les politiques apprises sur certains produits vers de nouveaux SKUs [198].

Enfin, des approches de clustering dynamique de clients ou produits peuvent segmenter l'espace de décision pour supporter l'augemntation de la taille du problème.

Ces stratégies représentent des pistes prometteuses mais introduisent chacune des complexités additionnelles : les architectures multi-agents nécessitent des mécanismes de coordination sophistiqués, le transfer learning suppose des similarités structurelles entre produits, et le clustering dynamique peut créer des artefacts si les regroupements sont mal calibrés. Une évaluation empirique comparative de ces approches sur des instances de taille croissante permettrait d'identifier les architectures les plus adaptées à différentes tailles de problèmes.

Effets systémiques non anticipés. Notre fonction de récompense optimise un objectif local (minimiser ruptures et coûts de stockage pour notre entreprise). Deux effets systémiques pour-

raient émerger. Premièrement, une politique réactive aux fluctuations de demande pourrait amplifier la variabilité des commandes transmises en amont (effet coup de fouet), phénomène non capturé dans notre modélisation car nous n'observons pas les conséquences en cascade sur la chaîne d'approvisionnement. Une extension intégrerait une pénalité sur la variance des ordres dans la fonction de récompense, bien que la calibration de cette pénalité nécessiterait une compréhension fine des dynamiques de la chaîne d'approvisionnement complète.

Deuxièmement, notre modèle suppose des délais d'approvisionnement indépendants de la taille des commandes, alors que dans la réalité, commander massivement pourrait saturer le fournisseur et allonger les délais (boucle de rétroaction). Modéliser cette relation dynamique complexifierait le modèle mais améliorerait son réalisme. Une approche pourrait consister à intégrer un modèle de simulation du comportement fournisseur dans la boucle d'entraînement, permettant à l'agent d'apprendre à anticiper ces effets de congestion.

Opacité décisionnelle et barrières à l'adoption. L'opacité décisionnelle des réseaux de neurones représente une barrière significative à l'adoption industrielle [199], [200]. Un gestionnaire expérimenté peut légitimement hésiter à suivre des recommandations qu'il ne peut interpréter, particulièrement lorsqu'elles contredisent son expertise [201]. Cette limitation n'est pas purement technique mais organisationnelle : elle touche à la confiance, à la responsabilité en cas d'erreur, et à la capacité d'apprentissage humain du système [202]. Les approches XAI (eXplainable AI) [203] telles que SHAP [204], LIME [201], ou les mécanismes d'attention [205], [206] offrent des pistes pour révéler les patterns appris. par exemple, identifier que l'agent commande davantage quand les délais récents augmentent ou quand certains SKUs montrent des corrélations de demande, bien que leur applicabilité au RL pour inventaires reste à démontrer empiriquement. Une voie complémentaire consisterait à développer des interfaces homme-machine permettant aux gestionnaires de challenger et d'ajuster les recommandations du système [207], [208], créant ainsi une boucle de feedback qui renforce progressivement la confiance tout en capturant l'expertise humaine dans le processus décisionnel.

Limites épistémiques fondamentales : démarrage à froid et événements extrêmes.

Deux limitations structurelles méritent clarification car elles révèlent les frontières de ce que peut accomplir l'apprentissage statistique.

Le problème du **démarrage à froid**, pour déployer un système d'apprentissage sur de nouveaux produits avec historique limité, est souvent soulevé comme objection aux approches RL pour le contrôle des stocks. Cependant, ce problème se situe en réalité à l'interface entre deux modules distincts dans notre architecture : le système de prévision de la demande et l'agent de contrôle des stocks. Notre contribution suppose l'existence d'un module de prévi-

sion fournissant des distributions de demande future sous forme de quantiles ($q_{0.1}$, $q_{0.5}$, $q_{0.9}$). Ce module de prévision, qui n'était pas l'objet de notre recherche, constitue le lieu où le problème du démarrage à froid se pose véritablement. Notre agent RL, lui, consomme ces prévisions comme données d'entrée et peut fonctionner dès lors qu'un module de prévision, même rudimentaire, lui fournit des estimations. La question pertinente devient alors : quelle est la sensibilité de la politique apprise à la qualité des prévisions ?

La question de la réaction face aux **événements extrêmes** (pandémies, guerres, effondrements de chaînes d'approvisionnement et catastrophes naturelles majeures) touche à une limite épistémique fondamentale de toute approche d'apprentissage statistique.

Face à ces situations, il convient de distinguer deux types d'incertitude selon la distinction classique formalisée par Knight [209].

Le *risque mesurable* (*risk* chez Knight) désigne l'incertitude quantifiable pour laquelle des probabilités peuvent être assignées ou estimées. Notre système gère cette forme d'incertitude via l'apprentissage sur des distributions de probabilité : l'agent RL apprend à naviguer dans un espace de scénarios possibles reflétant la volatilité historique observée, modélisant ainsi la variabilité normale de la demande et des délais d'approvisionnement.

L'*incertitude vraie* (*uncertainty* chez Knight), également appelée incertitude radicale, désigne une incertitude fondamentalement non quantifiable pour laquelle aucune probabilité ne peut être assignée. Les ruptures de régime (changements fondamentaux des conditions opérationnelles sortant de toute distribution historique) relèvent de cette catégorie et ne peuvent être anticipées par construction.

Un système entraîné sur des données pré-COVID ne pouvait pas « apprendre » à gérer une pandémie globale car cet événement était hors de sa distribution d'entraînement. Aucune quantité d'entraînement sur données historiques ne peut préparer un système à des événements radicalement nouveaux dont la nature même est imprévisible a priori.

7.4 Perspectives d'avenir

Les travaux présentés dans cette thèse ouvrent de nombreuses perspectives pour des recherches futures, s'inscrivant dans une démarche d'approfondissement et d'élargissement des solutions d'optimisation de la chaîne d'approvisionnement. Trois directions nous semblent particulièrement prometteuses et méritent une exploration approfondie.

7.4.1 Approche plus intégrée de l'optimisation de la chaîne d'approvisionnement

Au-delà des extensions spécifiques à chaque méthodologie, une direction future particulièrement prometteuse est le développement d'une approche plus intégrée de l'optimisation de la chaîne d'approvisionnement. Bien que nos contributions actuelles abordent des aspects distincts mais interdépendants de la chaîne logistique (consolidation, OFA, contrôle d'inventaire), une approche véritablement holistique nécessiterait une intégration plus complète et plus dynamique de ces différentes composantes.

Cette intégration pourrait prendre la forme d'un système décisionnel hiérarchique, où les décisions stratégiques de contrôle d'inventaire guideraient les décisions tactiques d'allocation des commandes, qui à leur tour informeraient les décisions opérationnelles de consolidation et de routage. Une telle architecture multi-niveaux permettrait de capturer les interdépendances complexes entre les différentes décisions logistiques et d'optimiser la performance globale de la chaîne d'approvisionnement, plutôt que ses composantes individuelles. Alternativement, elle pourrait s'orienter vers un système décisionnel plus distribué, basé sur des agents autonomes mais coordonnés, chacun responsable d'une composante spécifique de la chaîne d'approvisionnement, mais communiquant et collaborant pour atteindre des objectifs globaux.

Une extension naturelle de cette intégration serait de confier au système d'apprentissage par renforcement un rôle de pilotage paramétrique. Au lieu de se limiter aux décisions de réapprovisionnement, l'agent pourrait ajuster dynamiquement les hyperparamètres des modèles opérationnels (par exemple, les poids d'urgence dans le modèle WOFR-EC).

La mise en œuvre de cette intégration dynamique pourrait s'effectuer soit par l'extension de l'espace d'action des agents existants, soit par l'ajout d'agents superviseurs dans l'architecture multi-agents.

Pour déterminer l'architecture la plus performante sans a priori, une approche méthodologique rigoureuse consisterait à implémenter plusieurs candidats architecturaux paramétrés par des hyperparamètres de structure (de type énumération pour la sélection d'architecture ou facteurs de pondération pour des approches hybrides). L'intégration de ces hyperparamètres architecturaux dans le processus d'optimisation (Sweep) permettrait alors d'identifier empiriquement quelle configuration — centralisée, hiérarchique ou compétitive — offre la meilleure performance selon les contextes spécifiques.

7.4.2 Exploration de nouvelles technologies et paradigmes

L'évolution rapide des technologies et des paradigmes informatiques ouvre des perspectives fascinantes pour l'avenir de l'optimisation de la chaîne d'approvisionnement, offrant des opportunités d'enrichir et d'étendre les méthodologies développées dans cette thèse. L'émergence de technologies comme l'apprentissage fédéré, qui permet l'entraînement de modèles d'apprentissage automatique sur des données distribuées sans les centraliser, pourrait faciliter la collaboration entre différents acteurs de la chaîne logistique tout en préservant la confidentialité des données, un enjeu majeur dans les écosystèmes complexes.

De même, les avancées dans les domaines de l'Internet des Objets (IoT) et de l'analyse en temps réel pourraient enrichir considérablement les données disponibles pour l'optimisation logistique, permettant des décisions encore plus précises et réactives. La visibilité accrue offerte par ces technologies pourrait notamment renforcer la capacité des systèmes d'apprentissage par renforcement à développer des politiques adaptatives face aux fluctuations dynamiques de la demande et des conditions opérationnelles, en fournissant des observations plus riches et plus opportunes.

Une direction particulièrement prometteuse concerne l'intégration du principe du "human-in-the-loop" dans les systèmes d'optimisation logistique. Inspiré du concept de Jidoka (automatisation avec une touche humaine) issu du système de production Toyota, cette approche reconnaît que certains processus décisionnels bénéficient d'une supervision humaine, même dans un environnement hautement automatisé. Le Jidoka permet aux machines ou aux opérateurs de détecter des problèmes et d'arrêter la production pour prévenir les défauts, intégrant des fonctions de supervision plutôt que simplement d'exécution.

Dans le contexte de nos contributions, l'application du principe "human-in-the-loop" pourrait prendre la forme d'interfaces intelligentes qui facilitent la collaboration entre les systèmes automatisés d'optimisation et les décideurs humains. Ces interfaces permettraient aux experts humains de valider les décisions critiques, d'apporter des ajustements basés sur des connaissances contextuelles non modélisées (par exemple, des événements imprévus, des relations clients spécifiques), et d'identifier des opportunités d'amélioration continue des algorithmes.

Les récentes avancées dans les modèles de langage (LLM) et les systèmes RAG (Retrieval-Augmented Generation) offrent des possibilités particulièrement intéressantes pour de telles interfaces. Ces technologies pourraient servir d'intermédiaires naturels entre les humains et les systèmes automatisés, traduisant les décisions algorithmiques complexes en explications compréhensibles et transformant les instructions en langage naturel en actions systémiques précises. Cette symbiose entre l'intelligence humaine et artificielle pourrait significativement

améliorer l'acceptabilité, l'efficacité et la robustesse des systèmes d'optimisation logistique avancés, en combinant la puissance de calcul avec le jugement expert.

Ces technologies émergentes, combinées avec des approches avancées d'apprentissage automatique et d'optimisation comme celles développées dans cette thèse, pourraient conduire à une transformation profonde des chaînes d'approvisionnement, les rendant plus résilientes, efficientes et centrées sur le client. Nos contributions actuelles, en démontrant la valeur des approches basées sur les données pour l'optimisation logistique, posent les jalons pour cette évolution future.

7.4.3 Intégration d'objectifs environnementaux

La troisième direction concerne l'incorporation d'objectifs environnementaux dans le cadre d'optimisation. Dans un contexte où la durabilité devient une préoccupation centrale pour les entreprises et leurs parties prenantes, l'extension de nos modèles pour intégrer explicitement l'empreinte carbone, la consommation d'énergie, ou d'autres métriques environnementales représenterait une avancée significative. Cela inclurait l'optimisation des itinéraires pour minimiser les émissions, la promotion de modes de transport plus écologiques, et la gestion des retours de produits pour réduire les déchets. Cette perspective multi-objectifs, incluant les dimensions économique et environnementale, est essentielle pour développer des chaînes d'approvisionnement véritablement durables et responsables.

7.4.4 Stratégies de résilience face aux limites épistémiques

Face aux limites structurelles identifiées dans la précédente section, notamment le problème du démarrage à froid et l'enjeu de la robustesse aux événements extrêmes, plusieurs stratégies opérationnelles permettraient de renforcer la viabilité du système.

7.4.4.1 Approches pour le démarrage à froid

La littérature sur le forecasting propose plusieurs réponses au problème du démarrage à froid : analogie avec des produits similaires, modèles hiérarchiques exploitant les similarités entre familles de produits, incorporation du jugement d'experts, ou approches de machine learning utilisant des attributs produits (prix, catégorie, saisonnalité du marché) pour transférer des connaissances d'un contexte à un autre.

Notre validation expérimentale utilisait des prévisions synthétiques de qualité contrôlée. Une extension naturelle consisterait à conduire une analyse de sensibilité systématique où l'agent entraîné serait testé avec des prévisions délibérément dégradées (biais additionnel, variance

amplifiée, retard temporel) pour cartographier sa robustesse. Autrement dit, si les prévisions sont biaisées ou très incertaines (large intervalle de quantiles), la politique RL se dégrade-t-elle graduellement ou catastrophiquement ?

Pour les déploiements industriels, une approche prudente consisterait à initialiser l'agent via imitation learning : plutôt que de démarrer l'apprentissage par exploration aléatoire, l'agent imiterait d'abord une politique heuristique éprouvée (par exemple, une politique (s, S) calibrée par expertise métier), puis raffinerait progressivement cette politique par apprentissage par renforcement au fur et à mesure que des données réelles s'accumulent. Cette stratégie minimise les risques durant la phase initiale tout en permettant l'amélioration continue par apprentissage adaptatif.

7.4.4.2 Postures face aux chocs systémiques

Face à la limite épistémique des événements radicalement nouveaux, deux postures sont possibles.

La première, que nous qualifions de **robustesse algorithmique**, consiste à incorporer des stress tests durant l'entraînement, exposant l'agent à des scénarios extrêmes synthétiques : demandes multipliées par un facteur 10, délais triplés, ruptures d'approvisionnement prolongées sur plusieurs périodes consécutives, pics saisonniers artificiellement amplifiés. Les techniques de résilience (entraînement adversarial, domain randomization, stress testing avec scénarios extrêmes synthétiques) peuvent élargir la couverture de l'incertitude modélisée, exposant l'agent à des situations plus diverses durant l'entraînement. Cette approche améliore la robustesse marginale en élargissant l'enveloppe des situations couvertes, mais ne peut, par définition, couvrir l'espace infini des catastrophes imaginables. Elle présente de plus le risque de sur-régularisation : un agent trop conservateur face à des scénarios extrêmes peut sous-performer dans les conditions normales d'opération.

La seconde posture, que nous recommandons et que nous qualifions de **flexibilité architecturale**, repose sur la modularité du système. Nos trois contributions sont conçues comme modules indépendants et couplés faiblement :

— **Module 1 (Consolidation spatiale)** : peut être désactivé temporairement en période de crise pour privilégier la réactivité.
— **Module 2 (Allocation WOFR-EC)** : ses paramètres de pondération peuvent être ajustés manuellement pour privilégier la fiabilité ou la diversification.
— **Module 3 (Contrôle des stocks RL)** : peut être contraint par des règles de sécurité (stocks minimums obligatoires calculés par des méthodes conservatrices, limites sur la

taille des commandes, interdiction temporaire de certaines actions jugées trop risquées).

Cette architecture modulaire offre une résilience supérieure aux systèmes monolithiques : lors d'une perturbation majeure, les gestionnaires conservent la capacité d'intervention manuelle sur chaque module, désactivant temporairement certaines optimisations automatiques au profit de règles conservatrices basées sur l'expertise humaine, tout en maintenant les autres composants opérationnels.

La robustesse face à l'imprévu radical n'est pas algorithmique (ce qui serait épistémologiquement impossible) mais organisationnelle : elle réside dans la capacité des décideurs humains à reconfigurer le système selon les circonstances émergentes, capacité facilitée par une architecture transparente, modulaire, et documentée.

En conclusion, cette thèse propose une lecture renouvelée de l'optimisation de la chaîne d'approvisionnement à l'ère numérique. En développant des méthodologies exploitant de manière ciblée les données pour soutenir la décision logistique, nous avons illustré comment les technologies avancées peuvent contribuer à mieux appréhender des problématiques opérationnelles complexes, en en faisant émerger de nouvelles opportunités de création de valeur. La posture pragmatique adoptée dans ce travail reconnaît la diversité des défis logistiques contemporains et met l'accent sur l'efficacité opérationnelle, tout en conservant une exigence de rigueur analytique.

CHAPITRE 8 CONCLUSION

Cette thèse s'est attachée à la conception d'outils décisionnels visant à améliorer l'efficacité opérationnelle de la chaîne d'approvisionnement dans un contexte de modernisation technologique. Face à des marchés caractérisés par une volatilité croissante et des attentes clients en constante évolution, nous avons exploré comment les technologies avancées d'analyse de données et d'intelligence artificielle peuvent transformer les processus décisionnels logistiques.

8.1 Synthèse des contributions

Notre recherche s'est articulée autour de trois axes complémentaires, chacun ciblant un maillon stratégique de la chaîne d'approvisionnement moderne.

La première contribution a introduit une méthodologie novatrice de consolidation de marchandises fondée sur l'exploitation des règles d'association. Cette approche se distingue fondamentalement des méthodes traditionnelles de recherche opérationnelle en travaillant directement sur les données transactionnelles non agrégées, permettant ainsi de détecter des opportunités de consolidation spatio-temporelle qui demeureraient invisibles avec les approches conventionnelles. Les expérimentations menées sur le cas d'étude de [95], [96] ont démontré l'efficacité de cette méthode, capable de générer des solutions de qualité tout en offrant une adaptabilité supérieure face aux environnements caractérisés par des données massives et des patterns de demande complexes.

La deuxième contribution a élargi cette perspective en proposant un modèle d'allocation des articles pour l'exécution des commandes intégrant explicitement la consolidation spatiale. Le modèle WOFR-EC (Weighted Order Fill Rate Encouraging Consolidation) développé réconcilie deux objectifs souvent perçus comme antagonistes : la minimisation des coûts d'expédition et l'adhérence aux niveaux de service visés. L'introduction d'un mécanisme de pénalisation proportionnel au nombre de points de collecte visités a permis de réduire significativement les coûts logistiques (jusqu'à 21% de réduction des arrêts) tout en maintenant des niveaux de service élevés. Cette contribution a également révélé la capacité du modèle à réaliser implicitement une forme de consolidation temporelle, démontrant des synergies entre les dimensions spatiale et temporelle de l'allocation des articles pour l'exécution des commandes.

La troisième contribution, plus ambitieuse dans sa portée, a développé un système de contrôle d'inventaire fondé sur l'apprentissage par renforcement multi-agents. Cette approche adresse directement des complexités souvent négligées dans les modèles traditionnels : SKUs mul-

tiples avec patterns de demande interdépendants, accords de niveau de service hiérarchisés, demandes non-stationnaires et intermittentes, et délais d'approvisionnement dynamiques. Notre système a démontré une amélioration significative par rapport à une heuristique sophistiquée de référence, réduisant les coûts de détention d'inventaire de 16% tout en maintenant les niveaux de service contractuels. Particulièrement remarquable est sa capacité à équilibrer les performances entre différents groupes de SLA, illustrant l'efficacité de l'apprentissage par renforcement pour la gestion de compromis complexes et multidimensionnels.

8.2 Originalité et portée scientifique

L'originalité de cette thèse réside dans son approche délibérément pragmatique et sa progression méthodologique cohérente. Plutôt que de chercher une solution universelle, nous avons adopté une vision différenciée de l'optimisation logistique, sélectionnant pour chaque problématique la méthodologie la plus adaptée à sa structure inhérente et à ses enjeux spécifiques. Cette perspective reconnaît implicitement la diversité des défis logistiques et propose un cadre méthodologique flexible qui peut être adapté à différents contextes opérationnels.

Un aspect particulièrement novateur de nos travaux est leur capacité à dépasser la vision traditionnelle des compromis logistiques. Là où les approches conventionnelles présupposent souvent un antagonisme inévitable entre efficience économique et qualité de service, nos méthodologies révèlent des zones d'optimisation synergique où ces deux dimensions peuvent être améliorées simultanément. Cette redéfinition du paysage d'optimisation logistique ouvre des perspectives prometteuses pour les praticiens et les chercheurs du domaine.

Notre recherche se démarque également par son engagement à aborder directement les complexités du monde réel plutôt que de les abstraire à travers des simplifications excessives. Cette approche, particulièrement évidente dans notre troisième contribution, contraste avec la tendance dominante dans la littérature à privilégier l'élégance mathématique au détriment de la pertinence opérationnelle. En démontrant la viabilité et l'efficacité de méthodes avancées comme l'apprentissage par renforcement dans des environnements complexes et dynamiques, nous contribuons à combler le fossé persistant entre théorie et pratique dans l'optimisation des chaînes d'approvisionnement.

La portée scientifique de cette thèse s'étend au-delà du domaine spécifique de la logistique. Notre approche basée sur les données et notre utilisation intégrée de techniques issues de la fouille de données, de la recherche opérationnelle et de l'apprentissage automatique illustrent le potentiel des méthodologies hybrides pour aborder des problèmes complexes d'optimisation. Cette perspective interdisciplinaire enrichit non seulement la littérature en gestion de la

chaîne d'approvisionnement, mais contribue également aux domaines connexes de la science des données appliquée et de l'optimisation multi-objectifs.

8.3 Implications pratiques

Sur le plan pratique, cette thèse offre aux gestionnaires de chaînes d'approvisionnement un cadre décisionnel renouvelé, adapté aux défis contemporains de leur domaine. Nos contributions présentent plusieurs avantages tangibles pour les organisations :

Premièrement, elles permettent une exploitation beaucoup plus efficace des données disponibles. Dans une ère où les entreprises accumulent des volumes massifs de données transactionnelles mais peinent souvent à en extraire des informations exploitables, nos méthodologies offrent des mécanismes concrets pour transformer ces données en avantages opérationnels. La méthodologie de consolidation basée sur les règles d'association, par exemple, peut révéler des opportunités d'optimisation logistique invisibles aux approches traditionnelles d'analyse.

Deuxièmement, elles proposent des outils flexibles et adaptables qui peuvent être calibrés pour refléter les priorités spécifiques de chaque organisation. Le modèle WOFR-EC, avec son paramètre de pondération ajustable, permet aux décideurs d'explorer différents équilibres entre coût et service, sélectionnant celui qui correspond le mieux à leur positionnement stratégique et aux attentes de leurs clients. Cette personnalisation représente un avantage considérable par rapport aux approches standardisées qui dominent souvent le paysage des solutions logistiques.

Troisièmement, elles offrent des capacités d'adaptation dynamique essentielles dans un environnement commercial caractérisé par l'incertitude et le changement rapide. Notre système de contrôle d'inventaire basé sur l'apprentissage par renforcement, en particulier, illustre comment les décisions logistiques peuvent être continuellement optimisées en fonction des conditions changeantes, sans nécessiter les recalibrations manuelles fréquentes qu'exigent les approches traditionnelles. Cette réactivité accrue peut constituer un avantage concurrentiel décisif dans des marchés volatils.

La mise en œuvre de ces méthodologies dans des contextes industriels réels nécessite bien sûr une approche progressive et réfléchie, comme nous l'avons discuté dans le chapitre précédent. Néanmoins, le potentiel de transformation qu'elles représentent justifie l'investissement initial requis, particulièrement pour les organisations opérant dans des environnements complexes où les approches conventionnelles atteignent leurs limites.

8.4 Mot final

Ce travail contribue à enrichir la littérature en démontrant qu'une approche hybride, articulant exploration de données, optimisation multi-objectifs et apprentissage adaptatif, permet de dépasser les dichotomies classiques entre efficience économique et qualité de service. Il ouvre ainsi la voie à des chaînes d'approvisionnement plus intelligentes, plus réactives et durablement créatrices de valeur.

Dans un monde où l'agilité et l'adaptation sont devenues des impératifs stratégiques, nos contributions offrent aux organisations des outils concrets pour naviguer efficacement dans la complexité. Elles démontrent qu'il est possible de dépasser les compromis traditionnels entre coût et service pour créer des chaînes d'approvisionnement qui sont à la fois plus efficientes économiquement et plus réactives aux besoins des clients.

Au-delà des méthodologies spécifiques que nous avons développées, cette thèse plaide pour une évolution plus fondamentale dans notre conception de l'optimisation logistique – une évolution qui embrasse la complexité plutôt que de la simplifier à outrance, qui exploite les données plutôt que de les agréger prématurément, et qui adapte les approches méthodologiques aux problèmes plutôt que l'inverse. C'est dans cette perspective renouvelée, à l'intersection de la rigueur analytique et de la pertinence opérationnelle, que réside peut-être la contribution la plus durable de ce travail.

RÉFÉRENCES

[1] M. Nakasumi, "Information Sharing for Supply Chain Management Based on Block Chain Technology", in *2017 IEEE 19th Conference on Business Informatics (CBI)*, IEEE, t. 1, 2017, p. 140-149.

[2] K. Oettmeier et E. Hofmann, "Impact of additive manufacturing technology adoption on supply chain management processes and components", *Journal of Manufacturing Technology Management*, t. 27, nº 7, p. 944-968, 2016.

[3] McKinsey & Company, *How COVID-19 has pushed companies over the technology tipping point and transformed business forever*, `https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/how-covid-19-has-pushed-companies-over-the-technology-tipping-point-and-transformed-business-forever`, Accessed: 2023-06-10, 2020.

[4] M. Ben-Daya, E. Hassini et Z. Bahroun, "Internet of things and supply chain management: a literature review", *International Journal of Production Research*, t. 57, nº 15-16, p. 4719-4742, 2019.

[5] B. Marr, *Artificial intelligence in practice: how 50 successful companies used AI and machine learning to solve problems*. John Wiley & Sons, 2019.

[6] G. Lai, H. Liu, W. Xiao et X. Zhao, "'Fulfilled by Amazon': A Strategic Perspective of Competition at the E-commerce Platform", *Manufacturing & Service Operations Management*, t. 24, nº 3, p. 1406-1420, 2022.

[7] S. Axsäter, *Inventory control*. Springer, 2015, t. 225, p. 2.

[8] S.-C. Rim et I.-S. Park, "Order picking plan to maximize the order fill rate", *Computers & Industrial Engineering*, t. 55, nº 3, p. 557-566, 2008.

[9] J. J. Brennan, "Models and analysis of temporal consolidation.", thèse de doct., University of California at Los Angeles, 1982.

[10] J. C. Tyan, F.-K. Wang et T. C. Du, "An evaluation of freight consolidation policies in global third party logistics", *Omega*, t. 31, nº 1, p. 55-62, 2003.

[11] S. Çetinkaya et C.-Y. Lee, "Optimal outbound dispatch policies: Modeling inventory and cargo capacity", *Naval Research Logistics (NRL)*, t. 49, nº 6, p. 531-556, 2002.

[12] G. C. Jackson, "Evaluating order consolidation strategies using simulation", *Journal of Business Logistics*, t. 2, nº 2, p. 110-138, 1981.

[13] Y. P. Gupta et P. K. Bagchi, “Inbound freight consolidation under just-in-time procurement”, *Journal of Business Logistics*, t. 8, n° 2, p. 74-94, 1987.

[14] J. Pooley et A. J. Stenger, “Modeling and evaluating shipment consolidation in a logistics system”, *Journal of Business Logistics*, t. 13, n° 2, p. 153-174, 1992.

[15] R. W. Hall, “Consolidation strategy: inventory, vehicles and terminals”, *Journal of Business Logistics*, t. 8, n° 2, p. 57-73, 1987.

[16] Y. Sheffi, “Carrier Shipper Interactions in the Transportation Market: An Analytical Framework”, *Journal of Business Logistics*, t. 7, n° 1, p. 1-27, 1986.

[17] J. H. Bookbinder et J. K. Higginson, “Probabilistic modeling of freight consolidation by private carriage”, *Transportation Research Part E: Logistics and Transportation Review*, t. 38, n° 5, p. 305-318, 2002.

[18] A. S. Hanbazazah, L. Abril, M. Erkoc et N. Shaikh, “Freight consolidation with divisible shipments, delivery time windows, and piecewise transportation costs”, *European Journal of Operational Research*, t. 276, n° 1, p. 187-201, 2019.

[19] J. Higginson et J. H. Bookbinder, “Policy recommendations for a shipment-consolidation program”, *Journal of Business Logistics*, t. 15, n° 1, p. 87-112, 1994.

[20] S. Çetinkaya, “Coordination of inventory and shipment consolidation decisions: A review of premises, models, and justification”, *Applications of supply chain management and e-commerce research*, t. 92, p. 3-51, 2005.

[21] H. Min, “The vehicle routing problem with product/spatial consolidation and backhauling”, thèse de doct., The Ohio State University, 1987.

[22] G. Svensson, “A conceptual framework for the analysis of vulnerability in supply chains”, *International Journal of Physical Distribution & Logistics Management*, t. 30, n° 9, p. 731-750, 2000.

[23] C. Nguyen, M. Dessouky et A. Toriello, “Consolidation strategies for the delivery of perishable products”, *Transportation Research Part E: Logistics and Transportation Review*, t. 69, p. 108-121, 2014.

[24] G. Y. Ke et J. H. Bookbinder, “Coordinating the discount policies for retailer, wholesaler, and less-than-truckload carrier under price-sensitive demand: A tri-level optimization approach”, *International Journal of Production Economics*, t. 196, p. 82-100, 2018.

[25] G. Zhou, Y. Van Hui et L. Liang, "Strategic alliance in freight consolidation", *Transportation Research Part E: Logistics and Transportation Review*, t. 47, n° 1, p. 18-29, 2011.

[26] T. G. Crainic, M. Hewitt, M. Toulouse et D. M. Vu, "Scheduled service network design with resource acquisition and management", *EURO Journal on Transportation and Logistics*, t. 7, n° 3, p. 277-309, 2018.

[27] J. Holguín-Veras et I. Sánchez-Díaz, "Freight demand management and the potential of receiver-led consolidation programs", *Transportation Research Part A: Policy and Practice*, t. 84, p. 109-130, 2016.

[28] K. Aljohani et R. G. Thompson, "A multi-criteria spatial evaluation framework to optimise the siting of freight consolidation facilities in inner-city areas", *Transportation Research Part A: Policy and Practice*, t. 138, p. 51-69, 2020.

[29] E. Taniguchi, R. Dupas, J.-C. Deschamps et A. G. Qureshi, "Concepts of an integrated platform for innovative city logistics with urban consolidation centers and transshipment points", *City logistics 3: Towards sustainable and liveable cities*, p. 129-146, 2018.

[30] H. Min et M. Cooper, "A comparative review of analytical studies on freight consolidation and backhauling", *Logistics and Transportation Review*, t. 26, n° 2, p. 149-169, 1990.

[31] G. Marchet, M. Melacini et S. Perotti, "Investigating order picking system adoption: a case-study-based approach", *International Journal of Logistics Research and Applications*, t. 18, n° 1, p. 82-98, 2015.

[32] J. Gu, M. Goetschalckx et L. F. McGinnis, "Research on warehouse operation: A comprehensive review", *European Journal of Operational Research*, t. 177, n° 1, p. 1-21, 2007.

[33] C. G. Petersen et G. R. Aase, "A comparison of picking, storage, and routing policies in manual order picking", *International Journal of Production Economics*, t. 92, p. 11-19, 2004. DOI : 10.1016/J.IJPE.2003.09.006.

[34] T. Van Gils, K. Ramaekers, A. Caris et R. B. De Koster, "Designing efficient order picking systems by combining planning problems: State-of-the-art classification and review", *European Journal of Operational Research*, t. 267, n° 1, p. 1-15, 2018.

[35] W. O. Rom et S. A. Slotnick, "Order acceptance using genetic algorithms", *Computers & Operations Research*, t. 36, n° 6, p. 1758-1767, 2009.

[36] H. H. Guerrero et G. M. Kern, “How to more effectively accept and refuse orders”, *Production and Inventory Management Journal*, t. 29, n° 4, p. 59, 1988.

[37] C. Larsen et A. Thorstenson, “The order and volume fill rates in inventory control systems”, *International Journal of Production Economics*, t. 147, p. 13-19, 2014.

[38] H. Meyr, “Customer segmentation, allocation planning and order promising in make-to-stock production”, *OR spectrum*, t. 31, n° 1, p. 229, 2009.

[39] M. Seyedrezaei, S. Najafi, A. Aghajani et H. Bagherzadeh Valami, “Designing a genetic algorithm to optimize fulfilled orders in order picking planning problem with probabilistic demand”, *International Journal of Research in Industrial Engineering*, t. 1, n° 2, p. 40-57, 2012.

[40] D. Lečić-Cvetković, N. Atanasov, S. Babarogić et al., “An algorithm for customer order fulfillment in a make-to-stock manufacturing system”, *International Journal of Computers, Communications & Control*, t. 5, n° 5, p. 783-791, 2010.

[41] X. Li, J. Li, Y. P. Aneja, Z. Guo et P. Tian, “Integrated order allocation and order routing problem for e-order fulfillment”, *IISE Transactions*, t. 51, n° 10, p. 1128-1150, 2019.

[42] F. R. Jacobs, R. B. Chase et R. Lummus, *Operations and supply chain management*. New York : McGraw-Hill, 2014, p. 533-535.

[43] D. F. Ross, “Managing Supply Chain Inventories”, in *Distribution Planning and Control: Managing In The Era Of Supply Chain Management*, New York : Springer, 2015, p. 245-296.

[44] Q. Zhang, Y. Chi, Y. Liu et Q.-q. Shi, “Enterprise Inventory Control Based on Supply Chain Management”, *Advanced Materials Research*, t. 472-475, p. 3273-3276, 2012.

[45] X. Zhang, “Inventory Control of Supply Chain Environment”, *Advanced Materials Research*, t. 971-973, p. 2346-2349, 2014.

[46] I. Jackson, J. Tolujevs et Z. Kegenbekov, “Review of inventory control models: a classification based on methods of obtaining optimal control parameters”, *Transport and Telecommunication*, t. 21, n° 3, p. 191-202, 2020.

[47] S. Chopra et P. Meindl, “Strategy, planning, and operation”, *Supply Chain Management*, t. 15, n° 5, p. 71-85, 2001.

[48] G. Buxey, “Inventory control systems: theory and practice”, *International Journal of Information and Operations Management Education*, t. 1, p. 158-170, 2006.

[49] V. Tom Jose, J. Akhilesh et M. Sijo, “Analysis of inventory control techniques: A comparative study”, *International Journal of Scientific and Research Publications*, t. 3, p. 1-6, 2013.

[50] F. W. Harris, “How many parts to make at once”, t. 10, n° 2, p. 135-136, 1913.

[51] R. Khanna, K. Patel et P. N. Rajnikant, “Review Presentation of Different Economic Order Quantity (EOQ) Models and Their Application.”, *Journal of Advanced Zoology*, t. 44, 2023.

[52] A. H. M. Mashud, H.-M. Wee, C.-V. Huang et J.-Z. Wu, “Optimal replenishment policy for deteriorating products in a newsboy problem with multiple just-in-time deliveries”, *Mathematics*, t. 8, n° 11, p. 1981, 2020.

[53] E. Mfizi, F. Niragire, T. Bizimana et M. F. Mukanyangezi, “Analysis of pharmaceutical inventory management based on ABC-VEN analysis in Rwanda: a case study of Nyamagabe district”, *Journal of Pharmaceutical Policy and Practice*, t. 16, n° 1, 30, 2023.

[54] T. F. Abdelmaguid, M. M. Dessouky et F. Ordóñez, “Heuristic approaches for the inventory-routing problem with backlogging”, *Computers & Industrial Engineering*, t. 56, n° 4, p. 1519-1534, 2009.

[55] Z. Bahroun et N. Belgacem, “Determination of dynamic safety stocks for cyclic production schedules”, *Operations Management Research*, t. 12, p. 62-93, 2019.

[56] C. A. Serna-Urán, M. D. Arango-Serna, J. A. Zapata-Cortés et C. G. Gómez-Marín, “An agent-based memetic algorithm for solving three-level freight distribution problems”, *Exploring Intelligent Decision Support Systems: Current State and New Trends*, p. 111-131, 2018.

[57] T. Abayomi et A. Adeyemi, “Dynamics of Inventory Cost Optimization–A Review of Theory and Evidence”, *Dynamics*, t. 5, n° 22, 2014.

[58] J. J. Kanet, M. F. Gorman et M. Stößlein, “Dynamic planned safety stocks in supply networks”, *International Journal of Production Research*, t. 48, n° 22, p. 6859-6880, 2010.

[59] B. Amirjabbari et N. Bhuiyan, “Determining supply chain safety stock level and location”, *Journal of Industrial Engineering and Management (JIEM)*, t. 7, n° 1, p. 42-71, 2014.

[60] B. Desmet, E.-H. Aghezzaf et H. Vanmaele, “Intelligent techniques for safety stock optimization in networked manufacturing systems”, in *Artificial intelligence techniques for networked manufacturing enterprises management*, Springer, 2010, p. 367-421.

[61] A.-L. Beutel et S. Minner, “Safety stock planning under causal demand forecasting”, *International Journal of Production Economics*, t. 140, n° 2, p. 637-645, 2012.

[62] R. N. Boute, J. Gijsbrechts, W. Van Jaarsveld et N. Vanvuchelen, “Deep reinforcement learning for inventory control: A roadmap”, *European Journal of Operational Research*, t. 298, n° 2, p. 401-412, 2022.

[63] M. Naeem, S. T. H. Rizvi et A. Coronato, “A gentle introduction to reinforcement learning and its application in different fields”, *IEEE Access*, t. 8, p. 209 320-209 344, 2020.

[64] K. Arulkumaran, M. P. Deisenroth, M. Brundage et A. A. Bharath, “Deep reinforcement learning: A brief survey”, *IEEE Signal Processing Magazine*, t. 34, n° 6, p. 26-38, 2017.

[65] Y. Yan, A. H. Chow, C. P. Ho, Y.-H. Kuo, Q. Wu et C. Ying, “Reinforcement learning for logistics and supply chain management: Methodologies, state of the art, and future opportunities”, *Transportation Research Part E: Logistics and Transportation Review*, t. 162, p. 102 712, 2022.

[66] R. S. Sutton et A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018, p. 1-18.

[67] F. Garcia et E. Rachelson, “Markov decision processes”, *Markov Decision Processes in Artificial Intelligence*, p. 1-38, 2013.

[68] V. Mnih, K. Kavukcuoglu, D. Silver et al., “Human-level control through deep reinforcement learning”, *Nature*, t. 518, n° 7540, p. 529-533, 2015.

[69] I. Giannoccaro et P. Pontrandolfo, “Inventory management in supply chains: a reinforcement learning approach”, *International Journal of Production Economics*, t. 78, n° 2, p. 153-161, 2002.

[70] B. Rolf, I. Jackson, M. Müller, S. Lang, T. Reggelin et D. Ivanov, “A review on reinforcement learning algorithms and applications in supply chain management”, *International Journal of Production Research*, t. 61, n° 20, p. 7151-7179, 2023.

[71] R. F. Prudencio, M. R. Maximo et E. L. Colombini, “A survey on offline reinforcement learning: Taxonomy, review, and open problems”, *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[72] J. Schulman, F. Wolski, P. Dhariwal, A. Radford et O. Klimov, “Proximal policy optimization algorithms”, *arXiv preprint arXiv:1707.06347*, 2017.

[73] Q. Wang, J. Xiong, L. Han, H. Liu, T. Zhang et al., "Exponentially weighted imitation learning for batched historical data", *Advances in Neural Information Processing Systems*, t. 31, 2018.

[74] M. R. Zhang, T. L. Paine, O. Nachum et al., "Autoregressive dynamics models for offline policy evaluation and optimization", *arXiv preprint arXiv:2104.13877*, 2021.

[75] S. A. Priya, V. Maheswari et V. Balaji, "Operations research for supply chain management–an overview of issue and contributions", in *Journal of Physics: Conference Series*, IOP Publishing, t. 1964, 2021, p. 022 010.

[76] A. Gosavi, "Reinforcement learning: A tutorial survey and recent advances", *INFORMS Journal on Computing*, t. 21, n° 2, p. 178-192, 2009.

[77] S. A. Murdivien et J. Um, "BoxStacker: Deep Reinforcement Learning for 3D Bin Packing Problem in Virtual Environment of Logistics Systems", *Sensors*, t. 23, n° 15, p. 6928, 2023.

[78] W. Pan et S. Q. Liu, "Deep reinforcement learning for the dynamic and uncertain vehicle routing problem", *Applied Intelligence*, t. 53, n° 1, p. 405-422, 2023.

[79] A. Krnjaic, R. D. Steleac, J. D. Thomas et al., "Scalable multi-agent reinforcement learning for warehouse logistics with robotic and human co-workers", *arXiv preprint arXiv:2212.11498*, 2022.

[80] H. Liang, X. Wen, Y. Liu, H. Zhang, L. Zhang et L. Wang, "Logistics-involved QoS-aware service composition in cloud manufacturing with deep reinforcement learning", *Robotics and Computer-Integrated Manufacturing*, t. 67, p. 101 991, 2021.

[81] L. Zhou, C. Lin et Z. Cao, "Reinforcement-Learning-Based Adaptive Iterated Local Search Approach to Integrated Order Batching and Job Assignment Problems in a Smart Warehouse: Driving Growth for Logistics and Retail", *IEEE Robotics & Automation Magazine*, 2023.

[82] H. Wang, S. Wang, Y. Yang et D. Zhang, "GCRL: Efficient Delivery Area Assignment for Last-mile Logistics with Group-based Cooperative Reinforcement Learning", in *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, IEEE, 2023, p. 3522-3534.

[83] R. Boute, J. Gijsbrechts, W. Jaarsveld et N. Vanvuchelen, "Deep reinforcement learning for inventory control: A roadmap", *European Journal of Operational Research*, t. 298, p. 401-412, 2021.

[84] J. Gijsbrechts, R. N. Boute, J. A. Van Mieghem et D. Zhang, “Can deep reinforcement learning improve inventory management? Performance on lost sales, dual-sourcing, and multi-echelon problems”, *Manufacturing & Service Operations Management*, t. 24, n° 3, p. 1349-1368, 2022.

[85] Z. Kegenbekov et I. Jackson, “Adaptive supply chain: Demand–supply synchronization using deep reinforcement learning”, *Algorithms*, t. 14, n° 8, p. 240, 2021.

[86] H. Meisheri, N. N. Sultana, M. Baranwal et al., “Scalable multi-product inventory control with lead time constraints using reinforcement learning”, *Neural Computing and Applications*, t. 34, n° 3, p. 1735-1757, 2022.

[87] Z. Liu et T. Nishi, “Inventory Control with Lateral Transshipment Using Proximal Policy Optimization”, in *2023 5th International Conference on Data-driven Optimization of Complex Systems (DOCS)*, IEEE, 2023, p. 1-6.

[88] M. Mousa, D. van de Berg, N. Kotecha, E. A. del Rio-Chanona et M. Mowbray, “An analysis of multi-agent reinforcement learning for decentralized inventory control systems”, *arXiv preprint arXiv:2307.11432*, 2023.

[89] A. Gokhale, C. Trasikar, A. Shah, A. Hegde et S. R. Naik, “A Reinforcement Learning Approach to Inventory Management”, in *Advances in Artificial Intelligence and Data Engineering: Select Proceedings of AIDE 2019*, Springer, 2021, p. 281-297.

[90] N. Vanvuchelen, J. Gijsbrechts et R. Boute, “Use of proximal policy optimization for the joint replenishment problem”, *Computers in Industry*, t. 119, p. 103 239, 2020.

[91] U. Shafique et H. Qaiser, “A comparative study of data mining process models (KDD, CRISP-DM and SEMMA)”, *International Journal of Innovation and Scientific Research*, t. 12, n° 1, p. 217-222, 2014.

[92] A. Gunasekaran, K.-h. Lai et T. E. Cheng, “Responsive supply chain: a competitive strategy in a networked economy”, *Omega*, t. 36, n° 4, p. 549-564, 2008.

[93] K. Nikolopoulos, “We need to talk about intermittent demand forecasting”, *European Journal of Operational Research*, t. 291, n° 2, p. 549-559, 2021.

[94] J. C. Ranyard, R. Fildes et T.-I. Hu, “Reassessing the scope of OR practice: The influences of problem structuring methods and the analytics movement”, *European Journal of Operational Research*, t. 245, n° 1, p. 1-13, 2015.

[95] H. Min, C. S. Ko et H. J. Ko, “The spatial and temporal consolidation of returned products in a closed-loop supply chain network”, *Computers & Industrial Engineering*, t. 51, n° 2, p. 309-320, 2006.

[96] G. Kannan, P. S. Kumar et S. Narendran, “Comments on the erratum to “The spatial and temporal consolidation of returned products in a closed-loop supply chain network”[Computers & Industrial Engineering, 51(2):309–320,(2006)]”, *Computers & Industrial Engineering*, t. 54, n° 4, p. 1087-1090, 2008.

[97] Y. Zhang, L. Sun, X. Hu et C. Zhao, “Order consolidation for the last-mile split delivery in online retailing”, *Transportation Research Part E: Logistics and Transportation Review*, t. 122, p. 309-327, 2019.

[98] Y. Zhang, W.-H. Lin, M. Huang et X. Hu, “Multi-warehouse package consolidation for split orders in online retailing”, *European Journal of Operational Research*, t. 289, n° 3, p. 1040-1055, 2021.

[99] H. Gurnani, “A study of quantity discount pricing models with different ordering structures: Order coordination, order consolidation, and multi-tier ordering hierarchy”, *International Journal of Production Economics*, t. 72, n° 3, p. 203-225, 2001.

[100] M. A. Ulku, “Analysis of shipment consolidation in the logistics supply chain”, thèse de doct., University of Waterloo, 2009.

[101] I. K. Moon, B. C. Cha et C. U. Lee, “The joint replenishment and freight consolidation of a warehouse in a supply chain”, *International Journal of Production Economics*, t. 133, n° 1, p. 344-350, 2011.

[102] M. Khouja, Z. Michalewicz et S. S. Satoskar, “A comparison between genetic algorithms and the RAND method for solving the joint replenishment problem”, *Production Planning & Control*, t. 11, n° 6, p. 556-564, 2000.

[103] I. K. Moon et B. Cha, “The joint replenishment problem with resource restriction”, *European Journal of Operational Research*, t. 173, n° 1, p. 190-198, 2006.

[104] M. Hajiaghaei-Keshteli, A. J Afshari et E. Nasiri, “Addressing the freight consolidation and containerization problem by recent and hybridized meta-heuristic algorithms”, *International Journal of Engineering*, t. 30, n° 3, p. 403-410, 2017.

[105] H. Qin, Z. Zhang, Z. Qi et A. Lim, “The freight consolidation and containerization problem”, *European Journal of Operational Research*, t. 234, n° 1, p. 37-48, 2014.

[106] S. Stidham Jr, “Stochastic clearing systems”, *Stochastic Processes and their Applications*, t. 2, n° 1, p. 85-113, 1974.

[107] İ. Çapar, “Joint shipment consolidation and inventory decisions in a two-stage distribution system”, *Computers & Industrial Engineering*, t. 66, n° 4, p. 1025-1035, 2013.

[108] C. Zikopoulos, “Determination of freight rates under stochastic demand and freight consolidation savings”, *International Journal of Production Research*, t. 57, n° 17, p. 5556-5573, 2019.

[109] Z. Aboutalib et B. Agard, “Improvement of freight consolidation with a data mining technique”, in *Proceedings of the Eighth International Conference on Information Systems, Logistics and Supply Chain (ILS Conference), Austin, TX, USA*, 2020, p. 22-24.

[110] A. Assemi, “A Random Forest-based Model for Anticipatory Freight Selection and Long-haul Consolidation”, thèse de doct., State University of New York at Binghamton, 2019.

[111] R. J. Bayardo Jr, “Efficiently mining long patterns from databases”, in *Proceedings of the 1998 ACM SIGMOD international conference on Management of data*, 1998, p. 85-93.

[112] A. Rajak et M. K. Gupta, “Association rule mining: applications in various areas”, in *Proceedings of international conference on data management, Ghaziabad, India*, 2008, p. 3-7.

[113] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, A. I. Verkamo et al., “Fast discovery of association rules.”, *Advances in knowledge discovery and data mining*, t. 12, n° 1, p. 307-328, 1996.

[114] D. T. Larose et C. D. Larose, *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons, 2014, t. 4, chap. Association rules.

[115] M. SteadieSeifi, N. P. Dellaert, W. Nuijten, T. Van Woensel et R. Raoufi, “Multimodal freight transportation planning: A literature review”, *European Journal of Operational Research*, t. 233, n° 1, p. 1-15, 2014.

[116] J.-S. Song, “On the order fill rate in a multi-item, base-stock inventory system”, *Operations research*, t. 46, n° 6, p. 831-845, 1998.

[117] Q. Iqbal, D. E. Malzahn et L. E. Whitman, “Selecting a multicriteria inventory classification model to improve customer order fill rate”, *Advances in Decision Sciences*, t. 2017, p. 1-11, 2017.

[118] Y. Giat, “Allocation of scarce resources in a network with periodic deliveries and customer tolerable wait”, *Computers & Industrial Engineering*, t. 159, p. 107 462, 2021.

[119] S. Zhang, C. Bi et M. Zhang, “Logistics service supply chain order allocation mixed K-Means and Qos matching”, *Procedia Computer Science*, t. 188, p. 121-129, 2021.

[120] D. Jiang et X. Li, “Order fulfilment problem with time windows and synchronisation arising in the online retailing”, *International Journal of Production Research*, t. 59, n° 4, p. 1187-1215, 2021.

[121] B. Y. Ekren, S. Perotti, L. Foresti et L. Prataviera, “Enhancing e-grocery order fulfillment: improving product availability, cost, and emissions in last-mile delivery”, *Electronic Commerce Research*, t. 25, p. 1-39, 2024.

[122] G. Wagner, F. Calvo et P. Amorim, “Better together! The consumer implications of delivery consolidation”, *Manufacturing & Service Operations Management*, t. 25, n° 3, p. 903-920, 2023.

[123] A. Federgruen et D. Simchi-Levi, “Analysis of vehicle routing and inventory-routing problems”, *Handbooks in operations research and management science*, t. 8, p. 297-373, 1995.

[124] A. Campbell, L. Clarke, A. Kleywegt et M. Savelsbergh, “The Inventory Routing Problem”, in *Fleet Management and Logistics*. Boston, MA : Springer US, 1998, p. 95-113.

[125] B. L. Golden, S. Raghavan et E. A. Wasil, *The vehicle routing problem: latest advances and new challenges*. Springer Science & Business Media, 2008, t. 43.

[126] Z. Cui, D. Z. Long, J. Qi et L. Zhang, “The inventory routing problem under uncertainty”, *Operations Research*, t. 71, n° 1, p. 378-395, 2023.

[127] J. Skålnes, S. T. Vadseth, H. Andersson et M. Stålhane, “A branch-and-cut embedded matheuristic for the inventory routing problem”, *Computers & Operations Research*, t. 159, p. 106 353, 2023.

[128] H. Shaabani, A. Hoff, L. M. Hvattum et G. Laporte, “A matheuristic for the multi-product maritime inventory routing problem”, *Computers & Operations Research*, t. 154, p. 106 214, 2023.

[129] W. Najy, C. Archetti et A. Diabat, “A column generation-based matheuristic for an inventory-routing problem with driver-route consistency”, *European Journal of Operational Research*, t. 324, n° 2, p. 382-397, 2025.

[130] S. Charaf, D. Taş, S. D. P. Flapper et T. Van Woensel, “A matheuristic for the two-echelon inventory-routing problem”, *Computers & Operations Research*, t. 171, p. 106 778, 2024.

[131] L. Bertazzi, A. Bosco et D. Laganà, “Managing stochastic demand in an inventory routing problem with transportation procurement”, *Omega*, t. 56, p. 112-121, 2015.

[132] A. P. Barbosa-Povoa et J. M. Pinto, “Process supply chains: Perspectives from academia and industry”, *Computers & Chemical Engineering*, t. 132, p. 106 606, 2020.

[133] A. P. Barbosa-Póvoa et J. M. Pinto, “Challenges and perspectives of process systems engineering in supply chain management”, *Computer Aided Chemical Engineering*, t. 44, p. 87-96, 2018.

[134] N. Lin, H. M. Hjelle et R. Bergqvist, “The impact of an upstream buyer consolidation and downstream intermodal rail-based solution on logistics cost in the China-Europe container trades”, *Case Studies on Transport Policy*, t. 8, n° 3, p. 1073-1086, 2020.

[135] Z. Haider, Y. Hu, H. Charkhgard, D. Himmelgreen et C. Kwon, “Creating grocery delivery hubs for food deserts at local convenience stores via spatial and temporal consolidation”, *Socio-Economic Planning Sciences*, t. 82, p. 101 301, 2022.

[136] T. Lyons et N. C. McDonald, “Last-mile strategies for urban freight delivery: a systematic review”, *Transportation Research Record*, t. 2677, n° 1, p. 1141-1156, 2023.

[137] M. Kumar et K. Anoop, “An innovative modelling framework for freight consolidation in transportation planning”, *IIMB Management Review*, t. 37, n° 1, p. 100 550, 2025.

[138] C. Yang, Y. Lee et C. Lee, “Data-Driven Order Consolidation with Vehicle Routing Optimization”, *Sustainability*, t. 17, n° 3, p. 848, 2025.

[139] C. C. Orhan, J. C. Goez, M. Guajardo, O. Osicka et S. W. Wallace, “Assessing macro effects of freight consolidation on the livability of small cities using vehicle routing as micro models: The case of Bergen, Norway”, *Transportation research part E: logistics and transportation review*, t. 185, p. 103 521, 2024.

[140] H. K. E. Abad, B. Vahdani, M. Sharifi et F. Etebari, “A bi-objective model for pickup and delivery pollution-routing problem with integration and consolidation shipments in cross-docking system”, *Journal of Cleaner Production*, t. 193, p. 784-801, 2018.

[141] V. Ghomi, D. Gligor, S. Shokoohyar, R. Alikhani et F. Ghazi Nezami, “An optimization model for collaborative logistics among carriers in vehicle routing problems with cross-docking”, *The International Journal of Logistics Management*, t. 34, n° 6, p. 1700-1735, 2023.

[142] T. G. Crainic, P. Dell"Olmo, N. Ricciardi et A. Sgalambro, “Modeling dry-port-based freight distribution planning”, *Transportation Research Part C: Emerging Technologies*, t. 55, p. 518-534, 2015.

[143] Y. Ancele, M. H. Hà, C. Lersteau, D. B. Matellini et T. T. Nguyen, “Toward a more flexible VRP with pickup and delivery allowing consolidations”, *Transportation Research Part C: Emerging Technologies*, t. 128, p. 103 077, 2021.

[144] Ö. Büyükdeveci, S. Özpeynirci et Ö. Özpeynirci, "Multi-objective shipment consolidation and dispatching problem", *Computers & Operations Research*, t. 169, p. 106 728, 2024.

[145] M. K. Zuhanda, S. A. R. S. Hasibuan, Y. Y. Napitupulu et al., "An exact and metaheuristic optimization framework for solving vehicle routing problems with shipment consolidation using population-based and swarm intelligence", *Decision Analytics Journal*, t. 13, p. 100 517, 2024.

[146] Y.-B. Park, J.-S. Yoo et H.-S. Park, "A genetic algorithm for the vendor-managed inventory routing problem with lost sales", *Expert Systems with Applications*, t. 53, p. 149-159, 2016.

[147] D. R. Sonntag, A. H. Schrotenboer et G. P. Kiesmüller, "Stochastic inventory routing with time-based shipment consolidation", *European Journal of Operational Research*, t. 306, n° 3, p. 1186-1201, 2023.

[148] M. Janjevic, P. Lebeau, A. B. Ndiaye, C. Macharis, J. Van Mierlo et A. Nsamzinshuti, "Strategic scenarios for sustainable urban distribution in the Brussels-capital region using urban consolidation centres", *Transportation Research Procedia*, t. 12, p. 598-612, 2016.

[149] T. G. Crainic, M. Hewitt, M. Toulouse et D. M. Vu, "Service network design with resource constraints", *Transportation Science*, t. 50, n° 4, p. 1380-1393, 2016.

[150] E. B. Tirkolaee, A. Goli, A. Faridnia, M. Soltani et G.-W. Weber, "Multi-objective optimization for the reliable pollution-routing problem with cross-dock selection using Pareto-based algorithms", *Journal of Cleaner Production*, t. 276, p. 122 927, 2020.

[151] K. Aljohani et R. G. Thompson, "Profitability of freight consolidation facilities: A detailed cost analysis based on theoretical modelling", *Research in Transportation Economics*, t. 90, p. 101 122, 2021, ISSN : 0739-8859.

[152] B. Lv, B. Yang et E. P. Chew, "Optimization of multistage timeliness transit consolidation problem using adaptive-weighted genetic algorithm", *Annals of Operations Research*, t. 346, n° 2, p. 1345-1376, 2025.

[153] R. Musa, J.-P. Arnaout et H. Jung, "Ant colony optimization algorithm to solve for the transportation problem of cross-docking network", *Computers & Industrial Engineering*, t. 59, n° 1, p. 85-92, 2010.

[154] S. D. Hosseini, M. A. Shirazi et B. Karimi, "Cross-docking and milk run logistics in a consolidation network: A hybrid of harmony search and simulated annealing approach", *Journal of Manufacturing Systems*, t. 33, n° 4, p. 567-577, 2014.

[155] O. Guemri, P. Nduwayo, R. Todosijević, S. Hanafi et F. Glover, "Probabilistic tabu search for the cross-docking assignment problem", *European Journal of Operational Research*, t. 277, n° 3, p. 875-885, 2019.

[156] S. R. P. Purba, H. S. Amarilies, N. L. Rachmawati, A. Redi et al., "Implementation of particle swarm optimization algorithm in cross-docking distribution problem", *Acta Informatica Malaysia*, t. 5, p. 16-20, 2021.

[157] M. A. Dulebenets, "An Adaptive Polyploid Memetic Algorithm for scheduling trucks at a cross-docking terminal", *Information Sciences*, t. 565, p. 390-421, 2021, ISSN : 0020-0255.

[158] C. Snoussi, A. El Fallahi et S. Hicham, "A Memetic Algorithm to Solve the Two-Echelon Collaborative Multi-Centre Multi-Periodic Vehicle Routing Problem with Specific Constraints.", *International Journal of Advanced Computer Science & Applications*, t. 14, n° 12, 2023. DOI : 10.14569/ijacsa.2023.0141296.

[159] M. Gaudioso, M. F. Monaco et M. Sammarra, "A Lagrangian heuristics for the truck scheduling problem in multi-door, multi-product Cross-Docking with constant processing time", *Omega*, t. 101, p. 102 255, 2021, ISSN : 0305-0483.

[160] C. Friedrich et R. Elbert, "Adaptive large neighborhood search for vehicle routing problems with transshipment facilities arising in city logistics", *Computers & Operations Research*, t. 137, p. 105 491, 2022.

[161] S. Jia, L. Deng, Q. Zhao et Y. Chen, "An adaptive large neighborhood search heuristic for multi-commodity two-echelon vehicle routing problem with satellite synchronization.", *Journal of Industrial & Management Optimization*, t. 19, n° 2, 2023.

[162] A. Shahmardan et M. S. Sajadieh, "Truck scheduling in a multi-door cross-docking center with partial unloading – Reinforcement learning-based simulated annealing approaches", *Computers & Industrial Engineering*, t. 139, p. 106 134, 2020, ISSN : 0360-8352.

[163] Z. Aboutalib et B. Agard, "Improvement of freight consolidation through a data mining-based methodology", *International Journal of Logistics Systems and Management*, t. 49, n° 2, p. 255-273, 2024.

[164] S. Cheng, A. Hijazi, J. Konak, A. Erera et P. Van Hentenryck, "SPOT: Spatio-Temporal Pattern Mining and Optimization for Load Consolidation in Freight Transportation Networks", *arXiv preprint arXiv:2504.09680*, 2025.

[165] T. Katanyukul et E. K. P. Chong, “Intelligent Inventory Control via Ruminative Reinforcement Learning”, *Journal of Applied Mathematics*, t. 2014, 238357:1-238357:8, 2014.

[166] M. A. Bushuev, A. Guiffrida, M. Jaber et M. Khan, “A review of inventory lot sizing review papers”, *Management Research Review*, t. 38, n° 3, p. 283-298, 2015.

[167] Z. Khanidahaj, “Deep Learning and Reinforcement Learning for Inventory Control”, thèse de doct., Ecole Polytechnique, Montreal (Canada), 2018.

[168] J. W. Chong, W. Kim et J. Hong, “Optimization of apparel supply chain using deep reinforcement learning”, *IEEE Access*, t. 10, p. 100 367-100 375, 2022.

[169] Y. Zhou, L. Jia, Y. Zhao et Z. Zhan, “Optimal dispatch of an integrated energy system based on deep reinforcement learning considering new energy uncertainty”, in *2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS)*, IEEE, 2023, p. 804-809.

[170] R. Tian, M. Lu, H. Wang, B. Wang et Q. Tang, “IACPPO: A deep reinforcement learning-based model for warehouse inventory replenishment”, *Computers & Industrial Engineering*, t. 187, p. 109 829, 2024.

[171] M. Drakaki et P. Tzionas, “A Colored Petri Net-based modeling method for supply chain inventory management”, *Simulation*, t. 98, n° 3, p. 257-271, 2022.

[172] J. Guo, K. Han, H. Wu et al., “Cmt: Convolutional neural networks meet vision transformers”, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, p. 12 175-12 185.

[173] K. T. Park, Y. H. Son et S. D. Noh, “The architectural framework of a cyber physical logistics system for digital-twin-based supply chain control”, *International Journal of Production Research*, t. 59, n° 19, p. 5721-5742, 2021.

[174] T. C. Pamulety et V. M. Pillai, “Impact of Ordering Decisions on Performance of a Supply Chain–An Experimental and Simulation Study.”, *Journal of Mines, Metals & Fuels*, t. 70, 2022.

[175] H. D. Perez, C. D. Hubbs, C. Li et I. E. Grossmann, “Algorithmic approaches to inventory management optimization”, *Processes*, t. 9, n° 1, p. 102, 2021.

[176] A. Azaron, U. Venkatadri et A. Farhang Doost, “Designing profitable and responsive supply chains under uncertainty”, *International Journal of Production Research*, t. 59, n° 1, p. 213-225, 2021.

[177] Y. Wu, J. Zhang, Q. Li et H. Tan, “Research on Real-Time Robust Optimization of Perishable Supply-Chain Systems Based on Digital Twins”, *Sensors*, t. 23, n° 4, p. 1850, 2023.

[178] B. Mielczarek, M. Gora et A. Dobrowolska, “Simulation–Based Optimisation Model as an Element of a Digital Twin Concept for Supply Chain Inventory Control”, in *International Conference on Computational Science*, Springer, 2023, p. 513-527.

[179] J. Baruník et T. Kley, “Quantile coherency: A general measure for dependence between cyclical economic variables”, *The Econometrics Journal*, t. 22, n° 2, p. 131-152, 2019.

[180] J. Ash et R. P. Adams, “On warm-starting neural network training”, *Advances in neural information processing systems*, t. 33, p. 3884-3894, 2020.

[181] A. A. Rusu, N. C. Rabinowitz, G. Desjardins et al., “Progressive neural networks”, *arXiv preprint arXiv:1606.04671*, 2016.

[182] S. Mehimeh, “Value Function Initialization for Knowledge Transfer and Jump-start in Deep Reinforcement Learning”, *arXiv preprint arXiv:2508.09277*, 2025.

[183] K. Khetarpal, M. Riemer, I. Rish et D. Precup, “Towards continual reinforcement learning: A review and perspectives”, *Journal of Artificial Intelligence Research*, t. 75, p. 1401-1476, 2022.

[184] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan et S. Wermter, “Continual lifelong learning with neural networks: A review”, *Neural networks*, t. 113, p. 54-71, 2019.

[185] A. Ashrafee, J. Kozal, M. Wozniak et B. Krawczyk, “Holistic Continual Learning under Concept Drift with Adaptive Memory Realignment”, *arXiv preprint arXiv:2507.02310*, 2025.

[186] R. Florence, S. Leo, S. Kyle, C. Mark et M. Thomas, “When to retrain a machine learning model”, *arXiv preprint arXiv:2505.14903*, 2025.

[187] A. Mahadevan et M. Mathioudakis, “Cost-aware retraining for machine learning”, *Knowledge-Based Systems*, t. 293, p. 111 610, 2024.

[188] A. Mahadevan et M. Mathioudakis, “Cost-effective retraining of machine learning models”, *arXiv preprint arXiv:2310.04216*, 2023.

[189] Y. Wu, E. Dobriban et S. Davidson, “Deltagrad: Rapid retraining of machine learning models”, in *International Conference on Machine Learning*, PMLR, 2020, p. 10 355-10 366.

[190] F. E. Casado, D. Lema, M. F. Criado, R. Iglesias, C. V. Regueiro et S. Barro, "Concept drift detection and adaptation for federated and continual learning", *Multimedia Tools and Applications*, t. 81, nº 3, p. 3397-3419, 2022.

[191] S. Ahilan et P. Dayan, "Feudal multi-agent hierarchies for cooperative reinforcement learning", *arXiv preprint arXiv:1901.08492*, 2019.

[192] S. Pateria, B. Subagdja, A.-h. Tan et C. Quek, "Hierarchical reinforcement learning: A comprehensive survey", *ACM Computing Surveys (CSUR)*, t. 54, nº 5, p. 1-35, 2021.

[193] P. Sunehag, G. Lever, A. Gruslys et al., "Value-decomposition networks for cooperative multi-agent learning", *arXiv preprint arXiv:1706.05296*, 2017.

[194] T. Rashid, M. Samvelyan, C. S. De Witt, G. Farquhar, J. Foerster et S. Whiteson, "Monotonic value function factorisation for deep multi-agent reinforcement learning", *Journal of Machine Learning Research*, t. 21, nº 178, p. 1-51, 2020.

[195] F. Christianos, G. Papoudakis, M. A. Rahman et S. V. Albrecht, "Scaling multi-agent reinforcement learning with selective parameter sharing", in *International Conference on Machine Learning*, PMLR, 2021, p. 1989-1998.

[196] D. Ha et J. Schmidhuber, "World models", *arXiv preprint arXiv:1803.10122*, t. 2, nº 3, 2018.

[197] Y. Uehara et S. Matumae, "Dimensionality reduction methods using VAE for deep reinforcement learning of autonomous driving", in *2023 Eleventh International Symposium on Computing and Networking Workshops (CANDARW)*, IEEE, 2023, p. 338-342.

[198] F. Zhuang, Z. Qi, K. Duan et al., "A comprehensive survey on transfer learning", *Proceedings of the IEEE*, t. 109, nº 1, p. 43-76, 2020.

[199] Z. C. Lipton, "The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery.", *Queue*, t. 16, nº 3, p. 31-57, 2018.

[200] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead", *Nature machine intelligence*, t. 1, nº 5, p. 206-215, 2019.

[201] M. T. Ribeiro, S. Singh et C. Guestrin, "" Why should i trust you?" Explaining the predictions of any classifier", in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, p. 1135-1144.

[202] A. Adadi et M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)", *IEEE access*, t. 6, p. 52 138-52 160, 2018.

[203] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser et al., “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI”, *Information fusion*, t. 58, p. 82-115, 2020.

[204] S. M. Lundberg et S.-I. Lee, “A unified approach to interpreting model predictions”, *Advances in neural information processing systems*, t. 30, 2017.

[205] A. Mott, D. Zoran, M. Chrzanowski, D. Wierstra et D. Jimenez Rezende, “Towards interpretable reinforcement learning using attention augmented agents”, *Advances in neural information processing systems*, t. 32, 2019.

[206] P. Madumal, T. Miller, L. Sonenberg et F. Vetere, “Explainable reinforcement learning through a causal lens”, in *Proceedings of the AAAI conference on artificial intelligence*, t. 34, 2020, p. 2493-2500.

[207] S. Amershi, M. Cakmak, W. B. Knox et T. Kulesza, “Power to the people: The role of humans in interactive machine learning”, *AI magazine*, t. 35, n° 4, p. 105-120, 2014.

[208] A. Holzinger, “Interactive machine learning for health informatics: when do we need the human-in-the-loop?”, *Brain informatics*, t. 3, n° 2, p. 119-131, 2016.

[209] F. H. Knight, “Risk, uncertainty and profit”, *Hart, Schaffner and Marx*, 1921.

ANNEXE A ANNEXES RELATIVES AU CHAPITRE 4

TABLEAU A.1 Details of model solution A

Time period = 1							
The total number of opened initial collection points	6						
Initial collection points :	1	2	3	4	8	10	
Maximum holding time in days	1	1	1	1	1	1	
Customers allocated to the initial collection point	7,8,14,16,17,20	1,2,4,9,10,13,15,18,19, 24,25,26,27,28,29,30	11,23	3	6,12,21,22	5	
Consolidated shipping volume of returned products	474	834	51	102	96	19	1576
Time period = 2							
The total number of opened initial collection points	6						
Initial collection points :	1	2	3	4	8	10	
Maximum holding time in days	1	1	1	1	1	1	
Customers allocated to the initial collection point	7,8,14,16,17,20	1,2,4,9,10,13,15,18,19, 24,25,26,27,28,29,30	11,23	3	6,12,21,22	5	
Consolidated shipping volume of returned products	434	1452	51	68	258	95	2358
Time period = 3							
The total number of opened initial collection points	6						
Initial collection points :	1	2	3	4	8	10	
Maximum holding time in days	1	1	1	1	1	1	
Customers allocated to the initial collection point	7,8,14,16,17,20	1,2,4,9,10,13,15,18,19, 24,25,26,27,28,29,30	11,23	3	6,12,21,22	5	
Consolidated shipping volume of returned products	394	2309	51	34	520	285	3593

TABLEAU A.2 Details of model solution B

Time period = 1					
The total number of opened initial collection points	4				
Initial collection points :	1	2	4	8	
Maximum holding time in days	1	1	1	1	
Customers allocated to the initial collection point	5,7,8,14,16,17,20	1,2,4,10,13,15,19,24 ,25,26,27,28,29	3,9,18,30	6,11,12,21,22,23	
Consolidated shipping volume of returned products	493	700	236	147	1576
Time period = 2					
The total number of opened initial collection points	4				
Initial collection points :	1	2	4	8	
Maximum holding time in days	1	1	1	1	
Customers allocated to the initial collection point	5,7,8,14,16,17,20	1,2,4,10,13,15,19,24 ,25,26,27,28,29	3,9,18,30	6,11,12,21,22,23	
Consolidated shipping volume of returned products	529	1068	452	309	2358
Time period = 3					
The total number of opened initial collection points	4				
Initial collection points :	1	2	4	8	
Maximum holding time in days	1	1	1	1	
Customers allocated to the initial collection point	5,7,8,14,16,17,20	1,2,4,10,13,15,19,24 ,25,26,27,28,29	3,9,18,30	6,11,12,21,22,23	
Consolidated shipping volume of returned products	679	1885	458	571	3593

ANNEXE B ANNEXES RELATIVES AU CHAPITRE 5

B.1 Examples of Input Data Structures

Tableau B.1 Example of order data.

Order ID	w_o	**Creation date**	**CP**
0001	0.3	2022-01-01	X
0002	0.7	2022-01-08	Y
...	...	...	...

Tableau B.2 Example of order lines data.

Order ID	**Item ID**	**Quantity**
001	111	10
001	222	10
002	111	10
002	333	10
...	...	...

Tableau B.3 Example of inventory data.

Item ID	**Quantity in stock**	**As of date**
111	70	2021-01-01
222	21	2021-01-01
333	3	2021-01-01
...	...	...

B.2 First-Come-First-Served (FCFS) Algorithm

A commonly used approach for the order fulfillment allocation problem is the First-Come-First-Served (FCFS) algorithm, also known as First-In-First-Out (FIFO). A naive implementation involves checking orders in ascending order based on their due date and fulfilling them if sufficient stock is available. This approach prioritizes orders solely on their due date without considering priority level. A more effective alternative calculates and assigns a weight

Tableau B.4 Example of inventory reception data.

Item ID	Quantity received	Date of inventory reception
111	500	2021-12-01
111	500	2022-01-05
222	100	2021-12-01
222	100	2022-01-05
333	100	2022-01-05
...	...	...

Tableau B.5 Example of data in delivery constraints data (k).

Date	k
2021-01-01	100
2021-01-02	100
2021-01-03	100
2021-01-04	0
2021-01-05	0
2021-01-06	150
2021-01-07	150
...	...

(w_o) to each order based on factors such as due date and customer importance, then checks orders in descending order based on weight, fulfilling those with sufficient stock available. This approach ensures higher priority orders are fulfilled first, improving overall customer satisfaction.

B.3 Maximum Weighted Order Fill Rate (WOFR) Model

The WOFR model uses mixed-integer linear programming to optimize order fulfillment decisions. The mathematical formulation includes indices for orders and items, parameters for order weights, available stock, demand quantities, and delivery capacity, and binary variables indicating whether orders are fulfilled.

Indices :

o : order index, where $o = 1, 2, \ldots, O$ for O total orders

i : item index, where $i = 1, 2, \ldots, I$ for I total items

Parameters :

w_o : weight of order o, where $w_o \geq 0$

Algorithm 2 FCFS

```
Get the current stock for all items
Get the current list of open orders
Get the current capacity (i.e., max number of orders that can be fulfilled)
Sort active orders by their weight (descending)
for each order in the sorted list do
    if the order can be fulfilled with the stocks available then
        Fulfil this order
        Subtract the used quantity from the item stocks
        if the capacity is reached then
            Stop
        end if
    end if
end for
```

s_i : available stock of item i, where $s_i \geq 0$

$d_{o,i}$: demand for item i in order o, where $d_{o,i} \geq 0$

k : capacity (max. number of orders that can be fulfilled), where $k \geq 0$

Variables :

F_o : 1 if order o is fulfilled ; 0 otherwise

Objective function :

$$max \sum_o F_o w_o$$

Subject to :

$$\begin{array}{lrcll} C1 & F_o & \in & \{0,1\} & \forall o \\ C2 & \sum_o F_o & \leq & k & \\ C3 & \sum_o F_o \cdot d_{o,i} & \leq & s_i & \forall i \end{array}$$

The model is governed by the following constraints : Constraint (C1) ensures that the fulfillment of order o, represented by F_o, is binary. Constraint (C2) limits the maximum number of orders that can be delivered for the period, guaranteeing that the total fulfilled orders do not exceed capacity k. Constraint (C3) considers the stock constraints, ensuring that the total delivered quantity for an item i does not exceed the available stock s_i.

ANNEXE C ANNEXES RELATIVES AU CHAPITRE 6

This appendix provides a detailed technical account of the system proposed.

Each logical component of the system is presented in a self-contained block that consolidates its conceptual description and mathematical formulation.The components are presented in a logical sequence that builds from core operational models to the intelligent framework that governs them.

C.1 Order Fulfillment Allocation

Component : Order Fulfillment Allocation (OFA)

Allocation Models and Logic

This component determines how to allocate available on-hand inventory to pending customer orders.

FCFS (First-Come, First-Served)

A simple heuristic that processes orders sequentially based on a sorting criterion. Variants include **FCFS by Due Date** and **FCFS by Weight**. The latter prioritizes orders based on a calculated weight reflecting factors like quantity, urgency, and SLA (as defined in the Weight Formulation section), not on a direct measure of business value.

WOFR (Weighted Order Fill Rate)

A Mixed-Integer Linear Programming (MILP) model that maximizes the weighted sum of *fully* fulfilled orders.

Indices : $o \in O$ (orders), $i \in I$ (items).

Parameters : w_o (weight), s_i (stock), $d_{o,i}$ (demand), k (delivery capacity).

Variable : $X_o \in \{0, 1\}$.

Objective : $\max \sum_o w_o X_o$

Subject to :

$$\sum_o X_o \leq k \qquad \text{(Delivery capacity)}$$

$$\sum_{o:i \in o} X_o \cdot d_{o,i} \leq s_i, \qquad \forall i \in I \quad \text{(Inventory constraint)}$$

WOLFR (Weighted Order Line Fill Rate)

An extension of WOFR permitting partial fulfillment for greater flexibility.

Parameters : Includes $w_{o,i}$ (line weight), p_o (partial fulfillment flag), α_{full} (full order bonus), α_{due} (lateness penalty).

Variables : $X_{o,i}$ (line fulfilled), F_o (fully fulfilled), P_o (partially fulfilled).

Objective : $\max \sum_{o,i} w_{o,i} \left(X_{o,i} + F_o \cdot \alpha_{full} - (1 - X_{o,i}) \cdot due_{o,i} \cdot \alpha_{due}\right)$

WOLFR-OSP (Overdrawn Stock Penalty)

A WOLFR variant that penalizes consuming inventory beyond quantities reserved for specific Service Level Agreement (SLA) groups.

Variable : $E_{i,s}$ (excess stock for item i, group s).

Objective : WOLFR Objective $- \alpha_{osp} \sum_{i,s} E_{i,s}$

Constraint : $\sum_{o \in s}(X_{o,i} \cdot d_{o,i}) - r_{i,s} \leq E_{i,s}, \quad \forall i, s$, where $r_{i,s}$ is reserved stock.

Weight Formulation

Order weights are calculated via `OrderWeightConfigs`. The formula for an order line is :

$$w_{o,i} = (d_{o,i})^{\alpha_{qty}} \cdot \left(\frac{1}{\max(\Delta t, 0) + 1}\right)^{\alpha_{urgency}} \cdot \left(\frac{1}{1 - TSL_{o,i}}\right)^{\alpha_{sla}}$$

where Δt is urgency, $d_{o,i}$ is quantity, and $TSL_{o,i}$ is target service level.

C.2 Baseline Inventory Policies

Component : Baseline Inventory Policies

Configuration Parameters

Policies are governed by `BaseStockPolicyConfigs`, including : `weight_cumul`, `multiplier`, `multiplier_for_cumulative`, `multiplier_for_grouped`, and `weight_demand_due_...` parameters.

Base Stock Policy Models

These methods implement non-RL reference policies for calculating target inventory levels.

Grouped-Quantile-Forecast Calculates a target stock level for each item-SLA pair independently.

Cumulative-Quantile-Forecast Aggregates demand hierarchically across SLA groups.

Weighted-Quantile-Forecast Combines the two approaches via a weighted average for a robust compromise.

Theoretical Analysis of Quantile Aggregation

A key property is that for high quantiles ($q > 0.5$), $\sum Q_q(X_i) \geq Q_q(\sum X_i)$. The *Grouped* policy is thus prone to overstocking (lacks risk pooling), while the *Cumulative* policy may understock if allocation is imperfect. The *Weighted* policy navigates this trade-off.

Implementation Notes and Parameter Tuning

Based on empirical testing, the following guidelines are recommended :

— `weight_cumul` : Start with a value of 0.2-0.3 to balance the two approaches.

— **Multipliers** : The default of 1.0 is a neutral starting point ; adjust based on risk tolerance.

— `weight_demand_due` : Higher values (closer to 1.0) prioritize short-term demand but may increase overall inventory.

C.3 Monte Carlo Simulation for Data Generation

Component : Monte Carlo Simulation for Data Generation

Input Configuration

The `OrderLinesSimulator` is configured via several tables defining the supply chain structure : `sla_groups`, `customers`, `items`, `customer_item_patterns`, `seasonality_patterns`, and `week_day_patterns`.

Order Line Generation Process

The simulator generates realistic order data via a multistep process :

1. **Relationship Definition** : Establishes all valid customer-item combinations.
2. **Temporal Duplication** : Replicates data across specified weeks and simulation samples.
3. **Ordering Cycle Application** : Filters records based on customer-specific ordering frequencies.
4. **Seasonality Integration** : Applies monthly seasonal factors.

5. **Probabilistic Selection** : Determines which items are ordered based on frequency and seasonality.
6. **Priority Determination** : Assigns priority status to a portion of orders.
7. **Date Calculation** : Determines entry and due dates based on lead times.
8. **Quantity Generation** : Creates order quantities using the stochastic models below.

Stochastic Modeling

Order Quantity Generation Quantities follow a shifted Poisson distribution. The mean incorporates a temporal trend.

$$Q = \text{Poisson}(\max(0, \mu - L)) + L$$
$$\mu_t = \mu_0 \times (1 + r)^{t/365}$$

where Q is quantity, L is a lower bound, μ_t is the mean at time t, and r is the annual trend rate.

Replenishment Lead Time Generation Supplier lead times are modeled with a uniform distribution.

$$LT_{i,w,s} = LT_i^{\min} + U(0,1) \times (LT_i^{\max} - LT_i^{\min})$$

C.4 Quantile Forecasting

Component : Quantile Forecasting

Advantages of Simulation-Based Quantile Forecasts

This approach is used for its significant advantages :

— **Distribution-Free :** Avoids restrictive assumptions about demand distributions.

— **Evaluation Independence :** Enables policy evaluation independent of specific forecasting model bias.

— **Data Abundance :** Generates ample data for robust training and testing.

— **Sensitivity Analysis :** Facilitates controlled experiments on the impact of forecast errors.

Forecast Generation Process

The `QuantileForecastSimulator` transforms simulated demand into probabilistic forecasts :

1. **Data Aggregation** : For each future period, aggregates simulated demand quantities (both created, and created-and-due).
2. **Quantile Computation** : Computes specified quantiles across all simulation samples.
3. **SLA Cumulation** : Generates a second version of forecasts with quantities cumulative across hierarchical SLA levels.

Mathematical Formulation

For an item i, horizon h, SLA group s, and quantile q, the forecast is :

$$F_{t,h,i,s,q} = \text{Quantile}_q \left(\sum_{j=1}^{N} D^{(j)}_{t,h,i,s} \right)$$

where $D^{(j)}$ is the simulated demand in sample j. The cumulative version aggregates demand over groups $\geq s'$ before computing the quantile.

C.5 Reinforcement Learning Framework

Component : Reinforcement Learning Framework

Environment Integration and Training Configuration

The framework uses **RLlib (Ray)** with a **PyTorch** backend, and **Weights & Biases (W&B)** for experiment tracking. The environment is a `MultiAgentEnv` subclass managing a hierarchy of **Steps** (decision points), **Episodes** (full simulations), **Iterations** (training milestones), and **Runs** (full experiments).

Observation Space Design

— **Dimensional Structure :** While conceptually a 4D tensor (item, SLA, quantile, measure), the observation space is flattened into a 1D vector for computational efficiency with RLlib's default MLP networks.

— **Cumulative Demand Observations :** The state representation is enriched with cumulative demand metrics (e.g., $D_{t,i,s'}$, demand for item i from groups

$\geq s'$; and $F_{t',i,s',q}$, the corresponding forecast). This is critical, as it allows agents to reason about stock competition from higher-priority groups.

— **Strategic Information Hiding :** The design allows for hiding information from lower-priority SLA groups from an agent responsible for a higher-SLA group. Since high-SLA groups have priority access to inventory, the state of lower-SLA groups has minimal impact on their optimal policy, and hiding this information reduces noise.

— **Dynamic Bounds :** Observation space bounds are calculated dynamically by analyzing a baseline simulation run, ensuring proper normalization.

Action Space Design

A flexible range of action spaces allows for testing different control strategies.

— **Discrete Spaces :** Map an integer action to a decision. For example, `DISCRETE_MAP_TO_QUANTILE` maps action $a \in [0, n-1]$ to a quantile position.

— **Continuous Spaces :** Map a float action to a decision. For example, `CONTINUOUS_BASE_MULTIPLIER_FLOAT` uses the action as a multiplier on a baseline stock quantity, allowing for fine-grained adjustments.

A processing pipeline translates these abstract agent outputs into concrete inventory orders.

Reward Engineering and Shaping

Crafting an effective reward signal is paramount for successful training.

— **Reward Function :** The core reward is a weighted sum of penalties for inventory holding costs, order lateness, and service level misses, controlled by parameters in `RewardConfigs`.

— **Reward Normalization :** To create a stable learning signal, the raw reward is normalized against the performance of a strong, pre-tuned baseline policy (the Weighted-Quantile-Forecast policy). This centers the reward around zero and scales it, effectively tasking the agent with learning a policy that consistently improves upon the strong heuristic.

— **Temporal Smoothing :** Reward components incorporate exponential smoothing to balance the impact of recent and historical performance, providing a more stable (less noisy) learning signal to the agent.