Titre: Title:	Analyse du niveau de granularité dans la conception d'un jumeau numérique de chaîne d'approvisionnement : cas de l'entreprise Kruger
Auteur: Author:	Achraf Challakhi
Date:	2025
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Challakhi, A. (2025). Analyse du niveau de granularité dans la conception d'un jumeau numérique de chaîne d'approvisionnement : cas de l'entreprise Kruger [Master's thesis, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/72001/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/72001/
Directeurs de recherche: Advisors:	Maha Ben Ali, & Jean-Marc Frayret
Programme: Program:	Maîtrise recherche en génie industriel

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Analyse du niveau de granularité dans la conception d'un jumeau numérique de chaîne d'approvisionnement : cas de l'entreprise Kruger

ACHRAF CHALLAKHI
Département de mathématiques et de génie induriel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie industriel

Décembre 2025

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Ce mémoire intitulé :

Analyse du niveau de granularité dans la conception d'un jumeau numérique de chaîne d'approvisionnement : cas de l'entreprise Kruger

présenté par **Achraf CHALLAKHI**
en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Camelia DADOUCHI, présidente
Maha BEN ALI, membre et directrice de recherche
Jean-Marc FRAYRET, membre et codirecteur de recherche
Raphaël OGER, membre

DÉDICACE

À mes parents, pour votre amour inconditionnel,
votre patience et vos encouragements constants. . . .

REMERCIEMENTS

Avant tout, je remercie Dieu pour la force et la patience qu'Il m'a accordées tout au long de ce parcours.

Je souhaite exprimer ma profonde gratitude à mes directeurs de recherche, Maha BEN ALI et Jean-Marc FRAYRET, pour m'avoir offert l'opportunité de réaliser cette maîtrise. Je les remercie sincèrement pour leur confiance, leur disponibilité, leur soutien constant, leurs corrections attentives et, surtout, leur rigueur scientifique, qui ont grandement enrichi ce travail.

Je remercie également les membres du jury, Camelia DADOUCHI et Rafaël OGER, d'avoir accepté d'évaluer ce mémoire et de contribuer à son amélioration par leurs remarques pertinentes.

Je tiens à exprimer ma reconnaissance la plus sincère à ma famille pour leur soutien moral, mental et financier, ainsi que pour leur amour inconditionnel, qui ont été une source essentielle de motivation et de persévérance.

Enfin, j'adresse mes remerciements les plus chaleureux à tous mes amis, à Montréal comme en Tunisie, pour leur soutien, leur bienveillance et leurs encouragements tout au long de ce parcours.

RÉSUMÉ

Dans un environnement industriel marqué par la complexification croissante des chaînes d'approvisionnement et la généralisation de la transformation numérique, la maîtrise de la performance logistique repose désormais sur la capacité à exploiter les données massives et à modéliser les processus opérationnels de manière réaliste. La simulation à base d'agents et les jumeaux numériques offrent dans ce contexte de nouvelles perspectives pour représenter les interactions au sein des systèmes logistiques et évaluer différents scénarios de gestion. Toutefois, la conception de tels modèles soulève un défi majeur : celui du choix du niveau de granularité approprié. Un modèle trop détaillé devient coûteux et lent à exécuter, tandis qu'un modèle trop simplifié perd en fiabilité et en pertinence décisionnelle.

Ce mémoire traite précisément de cette question du compromis entre précision et efficacité de simulation. L'objectif est d'étudier le niveau de granularité et les agrégations dans une simulation de chaîne d'approvisionnement, afin d'assurer, à la fois, une représentation fidèle des réalités opérationnelles et une performance computationnelle satisfaisante permettant de concilier complexité maîtrisée et validité opérationnelle.

Pour ce faire, une analyse comparative a été mise en œuvre, s'appuyant sur les données réelles d'un partenaire industriel. Trois niveaux de granularité ont été modélisés et simulés : une configuration de base sans agrégation, une agrégation managériale par analyse ABC croisée, et une agrégation statistique par clustering K-means. Deux méthodes de génération des variables aléatoires pour modéliser la demande (l'une ajustée par des lois probabilistes, l'autre rigide par une loi normale tronquée) ont également été testées. Face à la nature très volatile de la demande réelle, les trois configurations ont produit de bons résultats, chacune mettant en lumière un compromis différent. La configuration de base a fidèlement reproduit la moyenne (ratio de 0.966) et la volatilité inhérente aux données (écart-type de 0.74). La configuration ABC, via une approche managériale, a offert une stabilité remarquable (écart-type de 0.077) tout en conservant une bonne précision moyenne (ratio de 0.86). Enfin, la configuration K-means, via une approche statistique, a atteint une précision moyenne quasi parfaite (ratio de 1) tout en maîtrisant les extrêmes de la volatilité. L'étude a surtout démontré que l'efficacité d'une agrégation dépend de l'adéquation entre la logique de groupement et la méthode de génération des variables aléatoires (ajustée ou rigide).

Au-delà de son cadre spécifique, ce travail contribue à la réflexion sur l'adaptation du niveau de granularité dans les modèles de jumeaux numériques (Human actuated Digital Twins) des chaînes d'approvisionnement. Il met en avant l'importance d'un équilibre entre exhaustivité

et performance pour garantir la fiabilité des simulations et l'efficacité des décisions logistiques dans un environnement dynamique et incertain.

ABSTRACT

In an industrial environment characterized by the increasing complexity of supply chains and the widespread adoption of digital transformation, logistics performance management now relies on the ability to exploit large-scale data and to realistically model operational processes. In this context, agent-based simulation and digital twins offer new perspectives for representing interactions within logistics systems and for evaluating different management scenarios. However, the design of such models raises a major challenge: the selection of an appropriate level of granularity. A model that is too detailed becomes costly and slow to execute, whereas an overly simplified model loses reliability and decision-making relevance.

This thesis specifically addresses the issue of the trade-off between simulation accuracy and computational efficiency. The objective is to study the level of granularity and aggregation strategies in supply chain simulation, in order to ensure both a faithful representation of operational realities and satisfactory computational performance, thereby reconciling controlled complexity with operational validity.

To this end, a comparative analysis was conducted based on real data provided by an industrial partner. Three levels of granularity were modeled and simulated: a baseline configuration without aggregation, a managerial aggregation based on cross ABC analysis, and a statistical aggregation using K-means clustering. Two demand generation methods were also tested: a flexible approach based on probabilistic distributions and a rigid approach based on a truncated normal distribution. Given the highly volatile nature of real demand, all three configurations produced good results, each highlighting a different trade-off. The baseline configuration accurately reproduced both the mean (ratio of 0.966) and the inherent volatility of the data (standard deviation of 0.74). The ABC configuration, through a managerial approach, provided remarkable stability (standard deviation of 0.077) while maintaining good average accuracy (ratio of 0.86). Finally, the K-means configuration, through a statistical approach, achieved near-perfect average accuracy (ratio of 1.008) while effectively controlling extreme volatility. The study primarily demonstrated that the effectiveness of aggregation depends on the alignment between the grouping logic and the method used to generate random variables.

Beyond its specific scope, this work contributes to the broader reflection on adapting the level of granularity in supply chain digital twin models (Human-Actuated Digital Twins). It highlights the importance of balancing model completeness and performance to ensure simulation reliability and effective logistics decision-making in a dynamic and uncertain environment.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES SIGLES ET ABRÉVIATIONS

MBA	modèle à base d'agents
DT	jumeau numérique
OFR	Order Fulfillment Rate
TVL	Total des Volumes Livrés

LISTE DES ANNEXES

CHAPITRE 1 INTRODUCTION GENERALE

1.1 Contexte Général

Dans un contexte où les chaînes d'approvisionnement deviennent de plus en plus complexes et dynamiques, la digitalisation s'impose effectivement comme une solution clé. Selon une étude menée par PwC, 96% des responsables de la chaîne d'approvisionnement estiment que la transformation numérique a permis une amélioration significative de la visibilité sur les coûts et les processus logistiques, renforçant ainsi la réactivité et l'agilité organisationnelle [1]. Par ailleurs, une enquête rapportée par Syntactics Inc. révèle que 70% des entreprises considèrent leur chaîne d'approvisionnement comme très ou extrêmement complexe, ce qui justifie l'intégration croissante d'outils numériques [2]. Ces avancées technologiques constituent un outil de gestion essentiel pour assurer la prise de décision, tant stratégique qu'opérationnelle, grâce à leur capacité à manipuler et à prendre en compte un grand nombre de variables et de situations, telles que la fluctuation de la demande, la gestion des stocks, la logistique et le transport [3, 4].

De nombreuses recherches ont été menées sur ces technologies, notamment le data mining, le machine learning, ainsi que la modélisation et la simulation [5]. Parmi ces approches, la simulation à base d'agents s'est imposée comme un outil puissant pour modéliser les chaînes d'approvisionnement en capturant les interactions entre les différentes parties prenantes et en représentant les dynamiques émergentes du système à travers la notion d'agent [6].

Cette approche a conduit à l'émergence du concept de jumeau numérique, offrant la possibilité de contrôler les chaînes d'approvisionnement et de créer une réplique virtuelle du système réel afin de tester divers scénarios en temps réel [7, 8]. Cependant, cette avancée soulève de nombreux défis, notamment en matière de manipulation et d'exploitation des données, qui deviennent encore plus complexes lorsqu'elles sont appliquées à un système aussi dynamique que la chaîne d'approvisionnement [9, 10].

L'évolution vers des cyber-physical systems a permis d'intégrer les simulations en temps réel dans les environnements industriels, renforçant ainsi la flexibilité et l'efficacité des chaînes d'approvisionnement [11, 12]. Toutefois, cette transformation repose sur la capacité des entreprises à exploiter les données massives générées par ces systèmes. Le data mining joue un rôle fondamental en permettant d'analyser ces données, de les valoriser et d'extraire des connaissances exploitables pour affiner les modèles et améliorer la prise de décision [13].

L'interaction entre la simulation à base d'agents et le data mining a ainsi donné naissance au

concept d'agent mining, qui combine l'autonomie des agents dans l'acquisition d'informations avec les algorithmes d'exploitation de données afin d'optimiser les modèles simulés [14, 15]. Cette approche permet d'intégrer des processus d'apprentissage automatique et d'analyse des données dans les simulations, facilitant ainsi l'adaptation et l'optimisation des chaînes d'approvisionnement [16].

Généralement, un modèle trop détaillé peut devenir difficile à gérer, en raison de la complexité accrue des calculs et du temps de simulation nécessaire. À l'inverse, un modèle trop simplifié peut entraîner la perte d'informations cruciales, compromettant ainsi la qualité des décisions fondées sur les simulations. L'un des principaux défis réside donc dans la gestion de la granularité : comment trouver un équilibre entre un niveau de détail suffisant pour garantir la fidélité du modèle et une simplicité qui permet de maintenir son efficacité opérationnelle. Cette question devient encore plus complexe lorsqu'on intègre des technologies comme la simulation à base d'agents et les jumeaux numériques, qui nécessitent de traiter des volumes massifs et variés de données, tout en modélisant de manière précise la dynamique des interactions entre les différents acteurs de la chaîne d'approvisionnement.

Bien que ce travail s'inscrive dans le cadre du développement d'un jumeau numérique pour la chaîne d'approvisionnement du partenaire industriel, il convient d'en préciser le périmètre technique. Le jumeau numérique représente l'écosystème global de réplication et de décision ; toutefois, le cœur décisionnel et prédictif de ce dispositif repose sur un modèle de simulation. Ce mémoire se concentre spécifiquement sur le moteur de simulation à base d'agents, qui constitue la fondation technologique nécessaire à l'existence du jumeau numérique.

1.2 Problématique de recherche

La problématique centrale de ce projet est de garantir une granularité des modèles de jumeaux numériques, permettant d'atteindre un équilibre entre précision et performance, et ainsi obtenir des résultats fiables pour la gestion des chaînes d'approvisionnement, dans un environnement de plus en plus complexe et incertain.

Dans ce mémoire, nous étudierons la chaîne d'approvisionnement de notre partenaire industriel, Kruger, qui fait face à des défis croissants dans la gestion de ses opérations logistiques. En effet, Kruger est une entreprise familiale privée, active dans le secteur de la fabrication et du recyclage des produits de papier depuis 120 ans en Amérique du Nord. Consciente de l'importance de l'innovation technologique pour améliorer la gestion de son inventaire, l'entreprise a identifié la nécessité d'adopter des technologies avancées telles que les jumeaux numériques. Cependant, en raison de la grande variété de ses produits, du nombre élevé de

ses clients, et des nombreux centres de distribution répartis à travers l'Amérique du Nord, la modélisation d'une telle chaîne d'approvisionnement devient particulièrement complexe. La question de la granularité du modèle, ainsi que son impact sur la performance des simulations, est cruciale, car elle conditionne directement la fiabilité des décisions stratégiques et opérationnelles. Ce partenariat industriel nous permet non seulement de mieux comprendre les besoins réels du secteur, mais aussi de disposer des données nécessaires pour entamer cette recherche. Ainsi, nous pourrons adapter notre approche afin de répondre de manière spécifique et efficace aux enjeux de Kruger et contribuer à une prise de décision plus éclairée dans un environnement dynamique et en constante évolution.

1.3 Structure du rapport

Pour répondre à la problématique de recherche, ce mémoire est structuré en sept chapitres. Suite à ce chapitre qui représente l'introduction, une revue de la littérature est réalisée dans le deuxième chapitre afin d'établir le cadre théorique et positionner cette recherche dans le domaine de la gestion des chaînes d'approvisionnement. Dans le troisième chapitre, sont définis l'objectif principal et les sous-objectifs du projet, ainsi que l'approche méthodologique adoptée pour répondre à la problématique. Ensuite, le quatrième chapitre détaille la modélisation et l'implémentation, en présentant le modèle de simulation, la configuration de base développée, ainsi que sa validation. Le cinquième chapitre est consacré à l'expérimentation, où sont exposées les différentes configurations d'agrégation testées pour évaluer l'impact du niveau de granularité. Le sixième chapitre analyse les résultats obtenus, en proposant une interprétation approfondie des performances et des erreurs observées selon chaque configuration. Enfin, le dernier chapitre synthétise les contributions, discute les limites détectées et ouvre sur les perspectives d'amélioration et d'applications futures.

CHAPITRE 2 REVUE DE LITTÉRATURE

Ce chapitre vise à faire une synthèse des travaux antérieurs et des théories fondamentales liées à notre sujet. Une telle synthèse nous permettra d'identifier les lacunes et de bien situer l'étude dans un contexte scientifique plus large et de mettre l'accent sur les approches méthodologiques pertinentes pour résoudre notre problématique de recherche.

Dans ce contexte, plusieurs domaines ont adopté l'approche de l'agrégation et de la granularité pour analyser les systèmes complexes dans le but de les simplifier tout en maintenant la capacité à reproduire des comportements globaux pertinents. En effet, le défi central dans le choix du niveau de granularité, c'est-à-dire le degré de détail, est de maintenir la précision, la validité empirique, et le coût computationnel.

Plus précisément,nous allons explorer dans ce chapitre la littérature existante sur l'effet de la granularité et de l'agrégation appliquée dans les différents domaines et particulièrement la simulation à base d'agent des chaînes d'approvisionnement, et les différentes méthodologies employées pour déterminer le niveau optimal de granularité, notamment les techniques d'agrégation basées sur les données et l'utilisation de l'apprentissage machine afin d'atteindre un équilibre entre la complexité et la simplicité.

2.1 Notion de la granularité

Maier et al. [17] ont bien défini la granularité comme étant le niveau de détail utilisé afin de représenter les entités ou bien les processus d'un système modélisé. Selon les auteurs de cet article, la granularité peut être classifiée en:

- **Granularité fine:** Les entités sont modélisés individuellement avec un haut niveau de détail. Ce qui augmente la complexité du modèle, mais en contre partie, nous offre une grande précision.
- **Granularité grossière:** Les entités sont regroupées ou bien simplifiés, ce qui améliore la lisibilité globale du modèle et réduit la complexité du système modélisé mais aussi réduit le niveau de détail et la précision [17].

Cette notion de granularité est généralement reliée à plusieurs concepts qui peuvent décrire les activités qui influent sur la granularité du modèle (comme le clustering, l'agrégation, la décomposition, la granulation...), des concepts qui décrivent la granularité du modèle elle-même (comme la hiérarchie, la résolution, la complexité...), ainsi que des concepts qui décrivent les propriétés du modèle qui peuvent être influencées par la variation de la granularité (comme la

précision, la fidélité, la modularité...) Généralement, dans ce type de problème on est toujours à la recherche d'un trade-off Granularité-Agrégation. En effet, une granularité bien choisie implique souvent une forme d'agrégation où les entités n'influencent pas les dynamiques clés du système.

2.1.1 Définition de l'agrégation

Maier et al. [17] ont défini l'agrégation comme étant le processus de regrouper plusieurs éléments, entités ou agents individuels (granules) en un seul représentant ou groupe homogène, afin de représenter un système complexe d'une façon simplifiée tout en préservant sa dynamique ou son comportement global essentiels. Une telle technique aide à simplifier l'analyse, améliorer l'efficacité computationnelle et à se concentrer sur un aspect spécifique. De plus, Maier et al. [17] ont proposé deux dimensions de granularité:

- **Granularité Structurelle :** Ce niveau décrit à quel niveau les entité sont (dis)agrégées. Plus précisément, il décrit à quel point les élément sont (dés)agrégées par rapport au système réel. On parle dans ce cas là des niveaux de granularités qu'on en déjà parlé au paravent. Dans ce dimension, on parle aussi de la nature de la relation entre les éléments qui est dépendante du niveau d'agrégation qu'on a choisis.
- **Granularité Informationnelle :** Dans ce niveau, la granularité est liée au type et la quantité d'information qu'une granule peut apporter: Plus les éléments contiennent d'informations ou différents types d'information, plus la granularité informationnelle est fine et vis-versa. Tous ces aspect peuvent être interprétées à travers la résolution de l'analyse qui peut nous indiquer les propriété d'un modèle: la fidélité, la précision, la modularité...

2.1.2 Domaines d'application

Herbert [18] a abordé la notion de granularité et d'agrégation comme moyen de simplifier les systèmes complexes. Il souligne que ces systèmes sont généralement structurés de manière hiérarchique, avec des groupes imbriqués à différents niveaux de granularité. En biologie et dans les systèmes physiques, par exemple, on observe que les organismes vivants sont constitués d'organes, eux-mêmes formés de cellules, qui peuvent être décomposés en atomes, molécules et macromolécules, et ainsi de suite jusqu'aux planètes, au système solaire et aux galaxies. De même, dans les systèmes sociaux, les individus se regroupent pour former des familles, des communautés et des institutions. Simon évoque également les systèmes économiques, considérés comme des systèmes sociaux, où l'agrégation des individus en fonction de leurs comportements au sein des entreprises et des marchés de consommation permet

de simplifier l'analyse économique globale [18].

Même dans l'analyse par éléments finis, qui est une approche de discrétisation largement adoptée pour les modélisations mécaniques, la notion de granularité est essentielle. Elle désigne le niveau de détail du maillage, qui impacte directement la précision du modèle. Un maillage plus fin améliore la résolution, permettant de capturer des caractéristiques complexes et d'obtenir des résultats plus précis, surtout dans les zones critiques. Cependant, cette granularité accrue entraîne un coût computationnel plus élevé. Il existe donc un équilibre entre granularité et qualité : il faut maximiser la précision tout en optimisant la complexité des calculs [19].

Tous ces exemples mettent en évidence l'importance de l'agrégation pour étudier les interactions au sein des systèmes complexes.

2.2 Granularité et Agrégations dans la chaîne d'approvisionnement

La granularité et l'agrégation jouent un rôle crucial dans la gestion des chaînes d'approvisionnement en répondant à des besoins et défis opérationnels spécifiques. Dans ce contexte, la granularité désigne le niveau de détail avec lequel les éléments de la chaîne d'approvisionnement, tels que les produits, les données et les processus, sont gérés. Une granularité fine permet un suivi et une surveillance précis des unités, ce qui est indispensable dans des secteurs comme l'industrie alimentaire et pharmaceutique, où la traçabilité et le contrôle qualité sont primordiaux [20, 21]. En assurant une granularité fine, il devient possible d'identifier avec précision les articles contaminés ou défectueux, réduisant ainsi les risques de rappels coûteux et protégeant la réputation des entreprises [20].

De même, dans les chaînes d'approvisionnement modulaires, la granularité permet de décomposer les systèmes complexes en modules plus gérables, favorisant la flexibilité et facilitant la personnalisation des produits [22]. Cependant, une granularité élevée génère un volume important de données, nécessitant des technologies avancées pour leur gestion efficace, telles que les systèmes RFID, les ERP (Enterprise Resource Planning) et la blockchain [21].

En parallèle, l'agrégation simplifie les opérations de la chaîne d'approvisionnement en regroupant des unités plus petites dans des catégories ou processus plus larges. Cette approche réduit la complexité de gestion, améliore l'efficacité logistique et diminue les coûts opérationnels grâce aux économies d'échelle. Par exemple, les systèmes de suivi agrégés se concentrent sur les expéditions ou les palettes plutôt que sur les produits individuels, rationalisant ainsi la logistique et réduisant les besoins en traitement des données [21]. L'agrégation renforce également la flexibilité stratégique en offrant une vue d'ensemble macroéconomique,

essentielle pour la gestion des stocks, la prévision de la demande et la planification de la production. En regroupant les unités en catégories plus larges, cette approche atténue l'impact des problèmes spécifiques à certaines unités, renforçant ainsi la résilience et la souplesse de la chaîne d'approvisionnement face aux imprévus et aux perturbations [20, 22].

La relation entre granularité et agrégation exige une solution équilibrée et adaptée aux objectifs spécifiques de la chaîne d'approvisionnement. Alors que l'agrégation apporte rentabilité et flexibilité, particulièrement dans des environnements stables ou pour les opérations logistiques en vrac, la granularité garantit précision et traçabilité, indispensables aux secteurs à haut risque. Cet équilibre repose en grande partie sur l'utilisation d'outils numériques et d'analyses avancées, qui permettent aux entreprises de s'adapter aux conditions dynamiques du marché tout en optimisant leurs performances [21, 22].

En adoptant une gestion stratégique de ce compromis entre granularité et agrégation, les organisations peuvent accroître leur résilience, réduire leurs coûts et améliorer l'efficacité globale de leurs chaînes d'approvisionnement, leur offrant ainsi un avantage compétitif durable.

2.3 Granularité et Agrégation dans la Modélisation à base d'agents des Systèmes complexes

La granularité et l'agrégation dans les modèles à base d'agents MBAs sont essentielles pour réduire la complexité des systèmes à modéliser. Dans ce contexte, la granularité fait référence au niveau de détail auquel les agents et leurs interactions sont représentés, tandis que l'agrégation implique la simplification des modèles en regroupant les agents sur la base de dimensions temporelles ou spatiales pour améliorer l'efficacité du calcul et découvrir les comportements émergents. Franceschini et al. [23] mettent l'accent sur la notion d'abstraction adaptative, qui bascule dynamiquement entre les niveaux granulaires et agrégés, permettant aux simulations d'équilibrer la précision et la performance, en particulier dans des domaines tels que la modélisation du trafic où le regroupement spatial révèle des embouteillages émergents. En outre, l'agrégation de l'espace et du temps est mise en évidence en tant que dimensions critiques.Par exemple, Morvan et Kubera [24] soulignent l'importance de la cohérence temporelle entre les niveaux d'abstraction pour maintenir une dynamique de système précise, tandis que le modèle de ségrégation de Schelling démontre comment l'agrégation spatiale peut produire des modèles émergents à grande échelle à partir d'interactions locales. Ces méthodes ne simplifient pas seulement les systèmes complexes, mais donnent également un aperçu de l'interaction entre les comportements des agents individuels et les phénomènes au niveau du système, faisant le lien entre les actions au niveau micro et les résultats au niveau macro [23, 24].

2.4 L'exploitation du data mining et l'analyse des données dans la simulation à base d'agents des systèmes complexes

L'exploitation du data mining est devenue incontournable pour la compréhension et la modélisation des systèmes complexes. Cette exploitation est par ailleurs réciproque, créant une synergie particulièrement efficace. Baqueiro et al. [25] ont analysé l'interaction entre le data mining et les MBAs, explorant les apports mutuels de ces deux domaines. D'une part, la simulation à base d'agents peut générer un volume massif de données synthétiques, qui peuvent être exploitées par le data mining pour améliorer les modèles existants, valider des hypothèses sur le système étudié, ou encore prédire les impacts de différents scénarios. D'autre part, le data mining permet d'extraire, à partir de grandes quantités de données empiriques, des informations précieuses et des règles comportementales. Ces informations peuvent être utilisées pour définir et affiner les agents, ou encore pour segmenter la population d'agents grâce à des techniques telles que la classification ou le clustering. Cette interaction entre les systèmes multi-agents et le data mining a été formalisée dans de nombreux articles sous le concept d'Agent Mining. Ce dernier se concentre sur la synergie entre les agents et le data mining, en exploitant les atouts de chaque domaine pour résoudre des problèmes complexes dans divers champs d'application.Parmi ces articles, on peut citer Cao et al. [15], qui constitue une référence majeure dans ce domaine, en explorant de manière approfondie les fondements théoriques et les applications pratiques de cette interaction. Ce concept peut être utilisé dans plusieurs domaines, comme l'e-commerce pour la négociation et le trading, ou encore dans l'intelligence d'affaires, incluant les chaînes d'approvisionnement et la gestion des relations avec les clients. D'autre part, Bemthuis et al. [26] ont proposé une méthodologie structurée pour évaluer les modèles de simulation à base d'agents en utilisant le process mining basé sur le modèle Cross-Industry Standard Processing for Data Mining (CRISP-DM). Ce modèle est constitué de six phases majeures: Comprehension des affaires, Compréhension des données, Préparation des données, Modélisation, Evaluation, Déploiement.
L'intégration du process mining dans ces étapes a pour but d'analyser le comportement des agentsn et d'améliorer la validité du modèle ainsi que de fournir des informations plus approfondies pour la prise de décision.

2.5 Les jumeaux numériques et le rôle central de la simulation

Les jumeaux numériques (DTs)(Digital Twins en anglais) sont des représentations virtuelles dynamiques de systèmes physiques, connectées à ces derniers via des flux de données en temps réel. Leur objectif principal est d'améliorer la compréhension, le contrôle et l'optimisation des

systèmes qu'ils modélisent. Depuis leur émergence dans le domaine industriel, les DTs ont connu un développement rapide dans divers secteurs, allant de la fabrication intelligente à la logistique, en passant par la santé, les villes intelligentes et les systèmes cyber-physiques [27].

Au cœur de cette technologie se trouve la simulation, qui permet de modéliser et de prédire le comportement d'un système physique sous différents scénarios. Selon plusieurs travaux récents, dont celui de Sowe et al. [28], la simulation n'est pas seulement un complément des DTs, mais constitue leur fondement opérationnel. En effet, les DTs utilisent des moteurs de simulation pour évaluer les états futurs, anticiper des dysfonctionnements ou tester des décisions avant leur mise en œuvre dans le monde réel.

Parmi les différentes approches de simulation, la modélisation à base d'agents apparaît comme l'une des plus pertinentes et les plus utilisées dans le contexte des DTs. Cette approche permet de représenter des entités autonomes (agents), capables d'interagir entre elles et avec leur environnement, ce qui en fait un outil particulièrement adapté à la modélisation de systèmes complexes, distribués et dynamiques [29]. Sowe et al. [28] souligne que ce type de modélisation est essentiel pour capturer les comportements émergents issus de l'interaction entre les composants d'un système multi-agents, notamment dans des contextes tels que les réseaux logistiques, les systèmes urbains ou les infrastructures critiques.

L'intégration des MBAs dans les DTs permet non seulement une représentation plus fidèle des comportements individuels et collectifs, mais aussi une capacité accrue de simulation "what-if" et d'adaptation en temps réel. Cette synergie renforce ainsi le rôle stratégique des DTs dans la prise de décision et l'optimisation de systèmes complexes.

2.6 Revue Critique de la Littérature

La question de l'agrégation et de la granularité est un enjeu majeur dans plusieurs domaines, notamment dans l'optimisation des modèles et la prise de décision. Différents travaux ont proposé des approches systématiques permettant d'ajuster le niveau de granularité en fonction des besoins spécifiques. Par exemple, dans l'industrie manufacturière, l'étude [30] a introduit une approche d'agrégation multi-niveau basée sur la connaissance pour améliorer la prise de décision. Cette méthode exploite des connaissances manufacturières à chaque niveau afin de concevoir des critères de surveillance et des opérateurs d'agrégation, permettant ainsi une meilleure interprétation des données et une prise de décision optimisée. De même, dans le domaine de l'optimisation des séries temporelles, l'étude [31] a mis en évidence les limites des approches standard d'agrégation et proposé des méthodes alternatives pour équilibrer précision et coût computationnel. Cette recherche montre que l'agrégation peut

être systématisée en fonction des erreurs de sortie, permettant ainsi une optimisation plus fine du modèle.

Malgré ces avancées dans d'autres domaines, la littérature présente une lacune significative concernant le compromis entre la granularité et l'agrégation dans le cadre de la simulation de MBAs. La plupart des travaux en simulation multi-agents privilégient une modélisation détaillée des entités sans adopter d'approches systématiques pour ajuster le niveau de granularité. Cette absence d'étude sur l'impact spécifique de la granularité et de l'agrégation sur les résultats de la simulation de MBAs est particulièrement problématique dans le contexte des DTs. En effet, la gestion d'un grand nombre d'agents dans ces systèmes peut considérablement alourdir les coûts computationnels et augmenter le temps d'exécution. De plus, bien que des études aient exploré la modélisation de la chaîne d'approvisionnement par des agents, un écart persiste en ce qui concerne l'étude spécifique de l'impact de la granularité sur les résultats de cette simulation dans le cadre des DTs.

Parallèlement, la littérature sur l'agent mining a mis en évidence la synergie entre les systèmes multi-agents et le data mining. Le data mining est utilisé pour analyser les données générées par les agents et affiner le modèle simulé. Cette interaction est particulièrement pertinente dans des contextes où les données sont massives et complexes. Ainsi, il apparaît nécessaire d'explorer des stratégies d'agrégation adaptées aux MBAs, en s'inspirant des approches utilisées dans d'autres domaines, tout en prenant en compte les spécificités des modèles basés sur les agents et l'apport de l'agent mining.

2.7 Conclusion

La revue de la littérature confirme l'importance de la granularité pour la modélisation et la simulation des chaînes d'approvisionnement. Elle met en lumière l'existence de méthodes d'agrégation efficaces dans divers domaines, mais révèle un gap de connaissance crucial concernant leur application systématique et leur impact sur les MBAs, spécifiquement les modèles et les DTs.

Ce projet s'inscrit dans une démarche visant à combler la lacune en explorant l'impact de la granularité sur les MBAs pour les chaînes d'approvisionnement.Le chapitre suivant sera consacré à la présentation détaillée de l'objectif de ce projet, des sous-objectifs spécifiques qui en découlent, ainsi que de la méthodologie qui sera appliquée pour y répondre.

CHAPITRE 3 OBJECTIFS DE RECHERCHE ET MÉTHODOLOGIE

3.1 Objectifs de recherche

L'objectif principal de ce projet est d'étudier le niveau de granularité des produits à inclure dans le modèle de simulation d'une chaîne d'approvisionnement, afin de garantir un compromis entre la fidélité et la simplicité du modèle. Il s'agit de déterminer le degré de détail nécessaire pour que le modèle soit à la fois représentatif des réalités opérationnelles tout en étant suffisamment simple pour garantir des performances de simulation efficaces et rapides. Ce compromis est crucial pour garantir l'efficacité du DT comme outil d'aide à la décision tout en minimisant les coûts computationnels et le temps de simulation. L'objectifs spécifiques de ce projet sont les suivants :

- S.O1 : Caractériser le système réel et établir une référence, soit un modèle de simulation de base au niveau de granularité le plus fin pour servir de modèle de référence validée par rapport aux données historiques.
- S.O2 : Proposer d'autres approches alternatives basées sur l'agrégation pour modéliser les produits dans le MBA.
- S.O3 : Etudier l'impact de l'agrégation. L'idée d'identifier comment comparer la performance des modèles agrégés à celle du modèle de base.
- S.O4 : Analyser les compromis entre la perte de précision (ou de fidélité) et le gain en efficacité (simplification).

3.2 Cas d'étude

Afin d'éviter toute confusion, il est important de distinguer le cadre global du projet de l'outil de modélisation utilisé. Le projet vise le déploiement d'un jumeau numérique. Ce jumeau nécessite un moteur de simulation capable de reproduire fidèlement la dynamique du système réel.

Dans ce contexte, la simulation à base d'agents a été choisie comme approche de modélisation. Elle constitue la couche logicielle du jumeau numérique, permettant de simuler les interactions entre les entités de la chaîne logistique. Ainsi, l'objet d'étude n'est pas l'architecture complète du jumeau, mais bien le modèle de simulation qui en constitue le socle indispensable.

Ce projet a été développé avec le secteur des tissus de Kruger, et plus précisément avec Kruger products, la subdivision manufacturière de l'entreprise.

Selon la classification des frameworks proposée dans l'étude [32] qui catégorise les jumeaux numériques en fonction de leur autonomie et de leur degré de liberté selon deux axes — l'aspect humain et l'aspect technique — nous visons à développer un jumeau numérique classé comme un "Human-Actuated Digital Twin" , soit le cadron supérieur gauche de la figure 3.1.

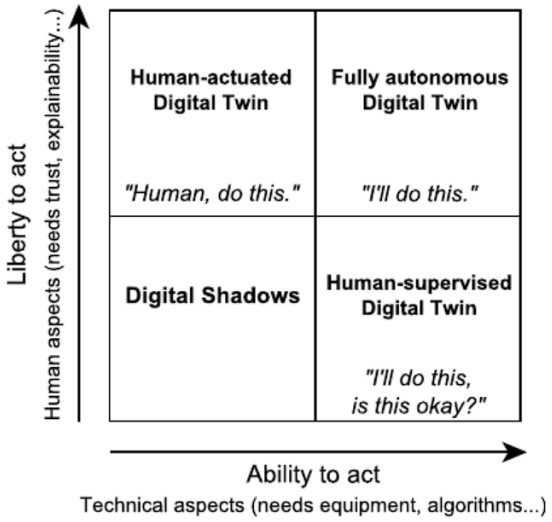


Figure 3.1 Niveaux d'autonomie des jumeaux numériques.

Cette catégorie se caractérise par une grande marge de liberté, mais sans capacité de contrôle autonome. Autrement dit, les jumeaux numériques de ce type formulent des recommandations, tandis que la prise de décision finale reste entre les mains des opérateurs humains. C'est précisément le cas du jumeau numérique visé, qui identifie les opportunités d'optimisation et propose des ajustements, laissant aux utilisateurs le soin d'effectuer les modifications ou le paramétrage nécessaires. Kruger possède une chaîne logistique complexe, s'étendant à travers l'Amérique du Nord. L'entreprise produit plus de 413 produits destinés à un portefeuille de plus de 566 clients. Cette chaîne d'approvisionnement comprend 4 usines et 16 centres de distribution.

Afin de simplifier la modélisation, nous avons restreint l'analyse aux usines et aux clients situés au Québec et en Ontario. Pour les opérations de transport, Kruger fait appel à un sous-traitant chargé d'assurer les livraisons.

En ce qui concerne la gestion des stocks, Kruger applique la politique DRP (Distribution Requirements Planning), un système de contrôle des stocks utilisé dans les environnements multi-niveaux de distribution [33]. Cette approche repose sur un modèle de gestion centralisé

et en flux poussé basé sur les prévisions de la demande des consommateurs finaux (dans la suite appelés clients), où la distribution des stocks est orchestrée depuis les centres de distribution (niveau inférieur du réseau) jusqu'aux usines (niveau supérieur).

Comme illustré dans la figure 3.2, la demande des produits au point final détermine la demande dans les centres de distribution intermédiaires, assurant ainsi une synchronisation entre les différents niveaux de la chaîne d'approvisionnement pour optimiser la gestion des flux et des stocks [34].

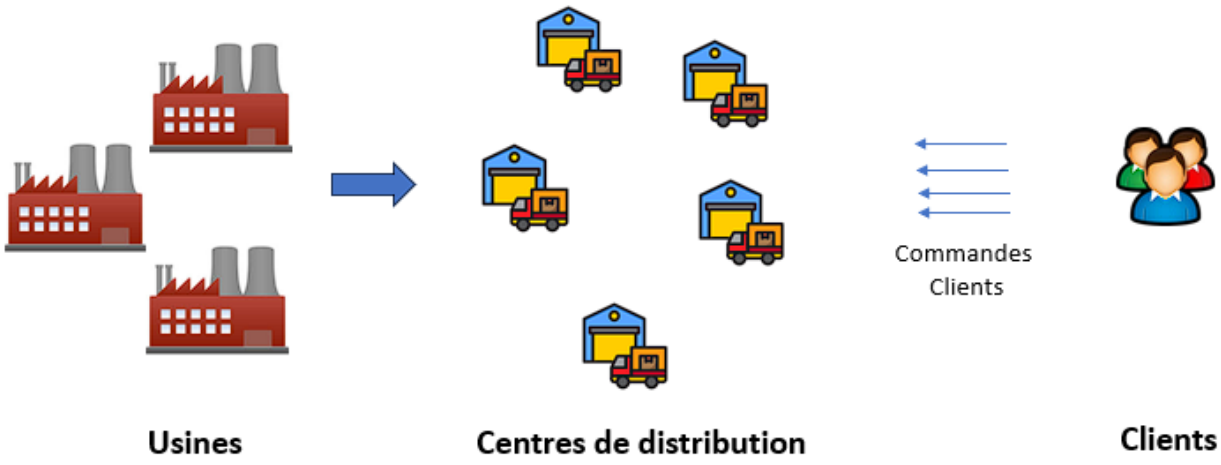


Figure 3.2 Figure simplificatrice du réseau étudié

3.3 Méthodologie de recherche

Après les étapes de définition de la problématique, d'analyse générale de la littérature, de formulation des objectifs spécifiques et de délimitation du cas d'étude, nous avons suivi les étapes méthodologiques illustrées dans la figure 3.3.

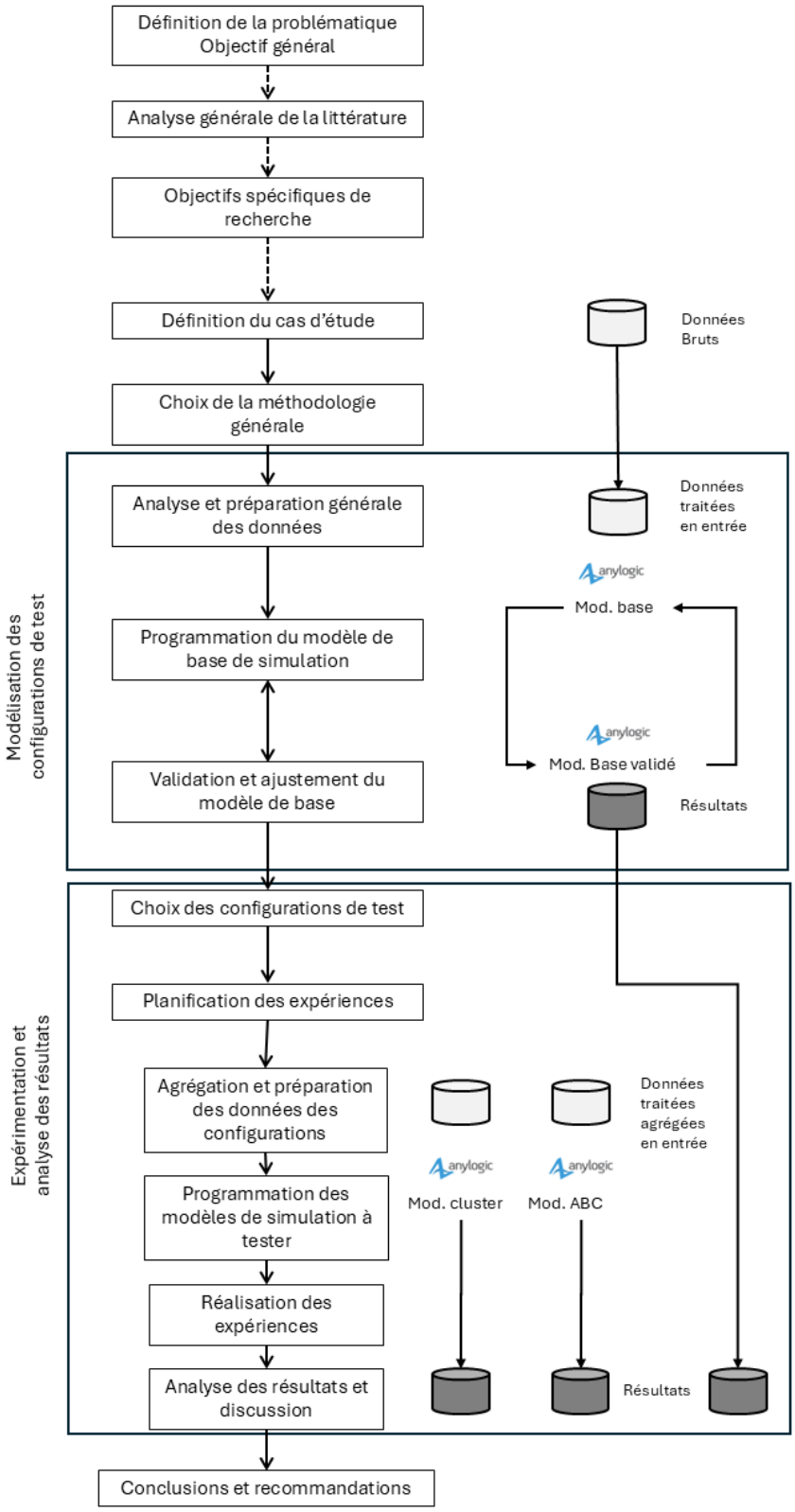


Figure 3.3 Schéma global de la méthodologie de modélisation et d'expérimentation

Nous avons proposé une méthodologie générale basée sur une approche expérimentale combinant analyse de données, modélisation, simulation et expérimentation. Elle s'articule autour de trois phases :

1. la modélisation des configurations de test,
2. l'expérimentation et l'analyse des résultats,
3. la formulation des conclusions et recommandations.

Chaque étape est interconnectée et suit un enchaînement logique permettant la validation progressive du modèle et la comparaison des configurations étudiées.

La phase de modélisation des configurations de test constitue le cœur du travail de développement du modèle de simulation. Elle comprend les étapes suivantes :

— Analyse et préparation générale des données : À partir des données collectées lors de l'étude de cas, un travail de nettoyage, de structuration et de validation est réalisé pour assurer la cohérence et la fiabilité du jeu de données afin de modéliser le modèle de base.

— Programmation du modèle de base de simulation : Sur la base des informations préparées, un modèle initial est développé dans le logiciel AnyLogic. Ce modèle de base représente le fonctionnement général du système étudié, sans intégrer encore les différentes configurations de test. La modélisation des agents, l'élaboration des hypothèses ainsi que la programmation de la logique du modèle de simulation ont été réalisées par mon collègue Enzo Domingos. Dans le cadre de ce mémoire, notre contribution dans la programmation du modèle de base s'est limitée à la modélisation du processus de traitement des données en entrée et modélisation des distributions pour la génération des variables aléatoires.

— Validation et ajustement du modèle de base : Le modèle est ensuite soumis à un processus itératif de test et d'ajustement. L'objectif est de vérifier la conformité entre les comportements simulés et les observations réelles issues des données. Une fois validé, le modèle sert de référence pour les phases d'expérimentation suivantes.

La phase d'expérimentation et d'analyse des résultats vise à évaluer l'impact des différentes configurations sur les performances du système simulé. Elle s'appuie sur le modèle de base validé et comprend les étapes suivantes :

— Choix des configurations de test : Plusieurs configurations sont définies afin d'explorer différentes approches de modélisation, notamment selon le niveau d'agrégation des données.

— Agrégation et préparation des données des configurations : Les données nécessaires à chaque configuration sont préparées selon les niveaux d'agrégation choisis (par exemple, approche par clusters ou classification ABC).
— Planification des expériences : Les paramètres de simulation, les scénarios à tester et les indicateurs de performance sont spécifiés.
— Programmation des modèles de simulation à tester : À partir du modèle de base, plusieurs variantes sont implémentées dans AnyLogic, notamment le modèle cluster et le modèle ABC.
— Réalisation des expériences : Les simulations sont exécutées pour chaque configuration afin de recueillir les résultats de performance du système.
— Analyse des résultats et discussion: Les résultats obtenus sont comparés entre les différentes configurations pour évaluer l'effet du niveau d'agrégation sur la précision, la performance et la complexité de la simulation.

Et pour finir, les conclusions tirées des expérimentations sont analysés afin de formuler des recommandations méthodologiques et opérationnelles. Ces conclusions permettent d'identifier les configurations les plus adaptées selon les objectifs du projet et ouvrent la voie à des perspectives d'amélioration du modèle de simulation.

3.4 Méthodologie de validation

L'étape de validation est une étape cruciale de la phase de modélisation des configurations de tests. Dans ce qui suit, nous présentons les indicateurs clés utilisés pour la validation.

Total des Volumes Livrés (TVL)

Le Total des Volumes Livrés (TVL) représente la somme des quantités de produits effectivement livrés aux clients durant la période considérée, parmi l'ensemble des volumes générés lors de la simulation. Cet indicateur est particulièrement pertinent dans l'évaluation des performances logistiques d'une chaîne d'approvisionnement, car il mesure la capacité du système à satisfaire la demande réelle à partir des volumes produits ou disponibles. Dans le cadre de la simulation, le TVL correspond uniquement aux volumes dont la livraison a pu être assurée conformément aux prévisions du modèle, après prise en compte des éventuelles contraintes de stock.L'analyse de cet indicateur permet ainsi de comparer différents scénarios ou configurations.

Order Fulfillment Rate (OFR)

Le "Order Fulfillment Rate (OFR)" est un indicateur clé de performance pour mesurer l'efficacité du système de gestion des commandes. Il est calculé comme suit :

$$\text{OFR} = \frac{\text{Nombre de commandes satisfaites à temps}}{\text{Nombre total de commandes traitées}} \times 100$$

Pour calculer l'OFR dans la simulation, nous utilisons les données suivantes :

- Temps de Livraison Réel : est le temps réel pris pour livrer une commande.
- Expected Time Delivery (ETD) : est calculé en fonction de la distance entre le centre de distribution le plus proche et le client, ainsi que des délais d'entreposage et administratifs.
- Le temps réel de livraison est comparé à l'ETD. Si le temps réel est inférieur ou égal à l'ETD, la commande est considérée comme satisfaisante à temps.

3.5 Conclusion

Dans ce chapitre, nous avons détaillé les objectifs et les étapes méthodologiques adoptées pour les atteindre. Le chapitre suivant sera consacré à l'analyse et à la préparation des données, ainsi qu'à la modélisation et à la validation du modèle de simulation de base.

CHAPITRE 4 MODÉLISATION ET VALIDATION

Dans ce chapitre, nous détaillons les étapes de la phase de modélisation de la configuration de test initiale, qui sert de modèle de base. Nous commençons par l'analyse et la préparation générale des données, suivies de la définition du modèle de simulation détaillé. Enfin, nous procédons à la validation de ce modèle en le testant avec des données générées par deux différentes méthodes de génération de variables aléatoires.

4.1 Analyse et préparation générale des données

4.1.1 Description des données brutes

Le partenaire industriel Kruger nous a donné accès à une base de données contenant 168975 lignes, chacune représentant une transaction de livraison vers un client. Cette base a été générée par le système SAP, utilisé par Kruger pour gérer l'ensemble de sa chaîne logistique.

Les données fournies sont particulièrement exploitables car elles ne contiennent aucune valeur manquante, ce qui simplifie considérablement la phase de nettoyage, en évitant notamment les étapes d'imputation.

La base de données se compose des colonnes suivantes (voir le Tableau 4.1):

- Colonne 1 - Plant : désigne le point de départ des livraisons. Il peut s'agir soit d'un centre de distribution, soit d'un site de production (manufacturier). Dans le cadre de ce projet, nous nous concentrons exclusivement sur les centres de distribution.
- Colonne 2 - Sold-To Party / Ship-To Party : ces deux champs définissent la destination de la commande. Sold-To Party correspond à l'entité qui a passé la commande (ex. : Costco Inc.), tandis que Ship-To Party fait référence au magasin physique ayant reçu la livraison (ex. : magasin Costco à Montréal).
- Colonne 3 - Material : identifie les produits commandés par les clients. Chaque ligne représente un produit spécifique dans une livraison.
- Colonne 4 - ShipDate : indique la date d'expédition de la commande, telle qu'enregistrée dans le système.
- Colonne 5 - Shipment : représente l'identifiant unique attribué à chaque livraison. Cet identifiant permet de regrouper plusieurs lignes correspondant à un même envoi multi-produits.

— Colonne 6 - DelvQtCas : désigne la quantité livrée exprimée en Cases. Le Case est une unité personnalisée utilisée par Kruger, permettant de standardiser les volumes livrés selon la quantité de papier contenue dans chaque lot.

Tableau 4.1 Récapitulatif des colonnes de la base de données

Nom de la colonne	Description
`Plant`	Point d'origine de la livraison. Peut être un centre de distribution ou un site de production. Seuls les centres de distribution sont pris en compte dans cette étude.
`Sold-To Party`	Entité ayant passé la commande.
`Ship-To Party`	Magasin ou point de réception de la commande.
`Material`	Code produit commandé.
`ShipDate`	Date d'expédition.
`Shipment`	Identifiant unique de la livraison. Permet de regrouper plusieurs lignes associées à un même envoi.
`DelvQtCas`	Quantité livrée en Cases, unité standardisée par Kruger.

4.1.2 Préparation et transformation des données

Afin de constituer le socle de notre modèle, une étape initiale consiste à modéliser les données d'entrée pour refléter la réalité stochastique de la chaîne d'approvisionnement. Plus précisément, nous visons à modéliser les volumes commandés pour chaque couple unique « Ship-To Party » (client) et « Material » (produit). Ces volumes doivent être représentés par des distributions de variables aléatoires au sein de la simulation pour capturer la variabilité de la demande.

Pour ce faire, nous avons regroupé les données par couple client-produit à partir des colonnes critiques de notre base de données (`Ship-To Party`, `Material` et `DelvQtCs`). Nous avons ensuite appliqué l'algorithme de l'Écart Interquartile (IQR) pour identifier et traiter les valeurs aberrantes (outliers) de la variable `DelvQtCs`.

Cette démarche se justifie par la nécessité de garantir un ajustement statistique robuste et représentatif. En effet, la présence de valeurs extrêmes, dues à des événements exceptionnels non récurrents, peut biaiser significativement les paramètres des lois de probabilité.

4.2 Programmation du modèle de base de simulation (Configuration 1)

Le modèle de simulation repose sur une architecture hiérarchique, avec les agents principaux suivants :

— Agent Client : Cet agent représente un client qui passe une commande pour un produit spécifique.
— Agent Manufacturier : Cet agent est responsable de la production des produits. Les manufacturiers reçoivent des commandes des centres de distribution et expédient les produits aux centres.
— Agent Camion : Cet agent représente un camion de livraison qui assure le transport des produits depuis les centres de distribution jusqu'aux clients finaux et depuis les manufacturiers jusqu'aux centres de distribution.
— Agent Commande : Cet agent représente une commande passée par un client pour un produit donné. Chaque commande contient des informations sur la quantité commandée et est associée à un centre de distribution.
— Agent Centre de Distribution : Chaque centre de distribution stocke une quantité de produits et reçoit des commandes des clients. Les centres de distribution sont connectés entre eux, permettant aux clients de se fournir à partir du centre le plus proche disponible.

4.2.1 Modélisation des agents

Il est à noter que la modélisation détaillée des agents sortait du périmètre de ce projet. Cette étape a constitué un travail parallèle, réalisé dans le cadre du doctorat de mon collègue Enzo Domingos [35]. Mon projet s'est donc concentré sur l'utilisation de ce modèle d'agents comme fondation pour l'expérimentation, en se focalisant spécifiquement sur l'impact de l'agrégation des données et de la granularité.

Agent Client

L'agent Client est responsable de la demande dans le système. Chaque client effectue les actions suivantes :

— Placer des Commandes : Les clients passent des commandes pour des produits spécifiques. Les quantités commandées sont modélisées à l'aide des distributions paramétriques ou empiriques (selon les méthodes).
— Choisir le Centre de Distribution : Chaque client cherche à commander auprès du centre de distribution le plus proche.
— Comportement Constant : Le comportement des clients reste constant sur la période

de simulation, ce qui signifie qu'il n'y a ni saisonnalité, ni tendance dans la demande.

Agent Manufacturier

L'agent Manufacturier est responsable de la production des produits et de leur expédition aux centres de distribution. Le processus inclut les étapes suivantes :

— Expédition : Une fois les commandes reçues, les manufacturiers expédient les produits aux centres de distribution pour assurer l'approvisionnement des stocks.
— Stockage : Chaque manufacturier dispose de stocks suffisants pour répondre aux demandes sans risque de rupture.

Agent Camion

L'agent Camion est responsable de la livraison des produits des centres de distribution aux clients. Ses caractéristiques sont les suivantes :

— Disponibilité : Les camions sont disponibles en nombre suffisant pour couvrir les besoins en livraison. Ils sont gérés en fonction des demandes des clients.
— Temps de Livraison : Le temps de livraison dépend de la distance entre le centre de distribution et le client. Les conditions de transport (trafic, capacité, etc.) ne sont pas pris en considération.

Agent Commande

L'agent Commande représente une commande passée par un client. Chaque commande contient :

— Quantité Commandée : La quantité de produits demandée par le client.
— Centre de Distribution : Le centre de distribution à partir duquel la commande sera traitée.
— Délai de Livraison : Le délai de livraison est calculé en prenant en compte la distance entre le centre de distribution et le client, ainsi que les délais d'entreposage et administratifs.

Agent Centre de Distribution

L'agent Centre de Distribution est responsable de la gestion des stocks et de la réception des commandes des clients. Ses caractéristiques sont les suivantes :

- Stockage des Produits : Chaque centre de distribution stocke une quantité de produits afin de satisfaire les commandes des clients.
- Réception et traitement des commandes : Il reçoit les commandes des clients et gère l'expédition des produits. En cas de rupture de stock, il peut rediriger la demande vers le centre de distribution le plus proche.
- Prévision de la demande (Logique DRP) : L'agent analyse les données historiques et les tendances de consommation pour établir des prévisions de demande. Cette étape est cruciale pour anticiper les besoins futurs du réseau.
- Passage des commandes : En se basant sur ses prévisions et l'état actuel de ses stocks, le centre de distribution émet des commandes de réapprovisionnement aux manufacturiers afin de garantir une synchronisation optimale des flux dans la chaîne d'approvisionnement.

4.2.2 Hypothèses du Modèle

Les principales hypothèses du modèle sont les suivantes :

- H1 : La fréquence des commandes des clients suit une loi exponentielle de paramètre λ (taux d'arrivée), calculé à partir de la base de données des transactions de livraison.
- H2 : Les clients commandent uniquement auprès du centre de distribution le plus proche. Si la quantité demandée n'est pas disponible, ils passent commande au centre suivant le plus proche, et ainsi de suite jusqu'à ce que la commande soit entièrement satisfaite ou qu'il n'y ait plus de stocks disponibles.
- H3 : Le comportement des clients est constant pendant toute la période de simulation, sans saisonnalité ni tendance. Cela implique que les paramètres des distributions des commandes ne changent pas au cours du temps.
- H4 : Les clients ne sont pas influencés par des promotions ou des variations de prix, ce qui simplifie l'analyse du modèle.
- H5 : La taille des commandes des clients est le sujet de test des deux méthodes . Dans chaque méthode , nous modélisons la taille des commandes en utilisant différentes distributions.
- H6 : Le taux de satisfaction des commandes à temps (Order Fulfillment Rate, OFR) est utilisé comme indicateur de performance du système. Le OFR est calculé comme

le rapport entre le nombre de commandes livrées à temps et le nombre total de commandes traitées. Dans le cas réel, ce taux se situe généralement entre 95% et 99%.

— H7 : L'expected time delivery (ETD) est calculé en prenant en compte la distance entre le client et le centre de distribution le plus proche, ainsi que les délais d'entreposage et les délais administratifs de traitement des commandes.

4.2.3 Génération des Prévisions de Demande

Comme mentionné dans la section méthodologie, l'entreprise partenaire applique une politique de planification des ressources de distribution (DRP) pour la gestion de ses stocks. Ainsi, nous avons utilisé la méthode de la moyenne mobile avec une période de deux semaines pour générer les prévisions de la demande au niveau des centres de distribution.

La moyenne mobile est un outil d'analyse et de prévision utilisé pour traiter les séries temporelles. Elle repose sur le calcul d'une moyenne glissante qui permet de lisser les fluctuations aléatoires et de dégager la tendance générale des données. Le choix de la période joue un rôle déterminant dans l'efficacité de la méthode : une période courte rend la moyenne plus réactive aux variations récentes, mais avec une plus grande sensibilité au bruit ; inversement, une période plus longue offre une courbe plus lisse et stable, mais peut retarder la détection des changements structurels. En matière de prévision, la moyenne mobile est utilisée pour extrapoler les tendances observées, en prolongeant la dynamique passée dans le futur [36].

4.2.4 Paramètres et conditions de simulation du modèle de base

Afin de garantir la validité des résultats, le modèle de simulation a été configuré pour refléter les conditions opérationnelles du partenaire industriel tout en assurant une stabilité statistique. Les paramètres de simulation retenus sont les suivants :

— **Horizon de simulation :** La simulation est exécutée sur une période de 730 jours, ce qui correspond à la durée totale des données historiques de livraison disponibles.

— **Phase de préchauffage (Warm-up) :** Un inventaire initial nul au début de la simulation peut biaiser les résultats. Pour atteindre un état stable, une période de préchauffage de 28 jours a été instaurée. Ce délai représente environ trois cycles de réapprovisionnement, calculés sur la base d'une politique de gestion des stocks par moyenne mobile sur deux semaines.

— **Capacité et niveau de service :** L'objectif de performance visé est un taux de service client (Order Fulfillment Rate - OFR) compris entre 95% et 99%. Pour garantir que la capacité de transport ne devienne pas un goulot d'étranglement limitant

artificiellement l'OFR, le nombre de camions a été fixé à 1 000.

— **Répétitions :** Pour atténuer la variabilité inhérente aux processus stochastiques du modèle, chaque exécution est répétée cinq fois. Les indicateurs de performance présentés dans ce travail correspondent à la moyenne de ces cinq réplications.

4.2.5 Données en entrées et modélisation des distributions pour la génération des variables aléatoires

Les données d'entrée du modèle de simulation comprennent plusieurs ensembles d'informations essentiels. Tout d'abord, les centres de distribution et les manufacturiers sont définis avec leurs localisations géographiques précises. Ces localisations permettent de modéliser les distances, les temps de transport et les interactions logistiques entre les différents acteurs de la chaîne. Par ailleurs, chaque couple client-produit est identifié, associé à la localisation géographique du client.

Pour modéliser la génération des commandes (volumes commandés) des consommateurs finaux (clients), des distributions de probabilité sont estimées pour chaque couple client-produit, avec leurs paramètres associés. Ainsi, l'ensemble de ces données d'entrée structurelles et probabilistes constitue le socle sur lequel s'appuie la simulation pour reconstituer avec précision les flux et les processus logistiques de la chaîne d'approvisionnement étudiée.

Nous avons proposé de tester deux méthodes pour la génération des variables aléatoires.

Méthode 1 : Distribution Ajustée

Nous proposons dans ce projet de sélectionner la loi de distribution la plus appropriée pour modéliser les quantités commandées en suivant un processus statistique structuré, comme illustré par l'organigramme de la figure 4.1.

L'idée est de déterminer la loi de distribution la plus adaptée pour modéliser les quantités commandées à partir des données réelles, en s'appuyant sur une analyse statistique rigoureuse et sur des critères de qualité de l'ajustement. Les données brutes sont organisées sous forme de couples `client-produit` et quantités commandées. Pour renforcer la qualité de l'analyse, les valeurs aberrantes sont supprimées par la méthode de l'écart interquartile (IQR), ce qui permet de réduire l'influence des données extrêmes et d'assurer la robustesse des ajustements.

Chaque couple `client-produit` est ensuite analysé suivant la taille de son échantillon. Si la taille est inférieure ou égale à 1, ou si la variation des volumes commandés est très faible, la distribution retenue suit une quantité fixe. Pour les échantillons supérieurs à 30, le test de Kolmogorov-Smirnov est appliqué pour permettre une comparaison fiable avec les lois

paramétriques; ce seuil étant choisi pour garantir la puissance du test et la pertinence de l'ajustement. Pour les échantillons compris entre 2 et 30, une distribution discrète est utilisée en raison du manque de robustesse statistique pour les lois paramétriques.

Pour les cas où N>30, plusieurs lois (Pareto, Bêta, Gamma, Exponentielle, Normale, Log-normale, Weibull, Laplace) sont testées, en comparant les données à chacune à l'aide du test de Kolmogorov-Smirnov. Ce test produit une p-value indiquant le degré d'adaptation de la loi à la distribution observée.

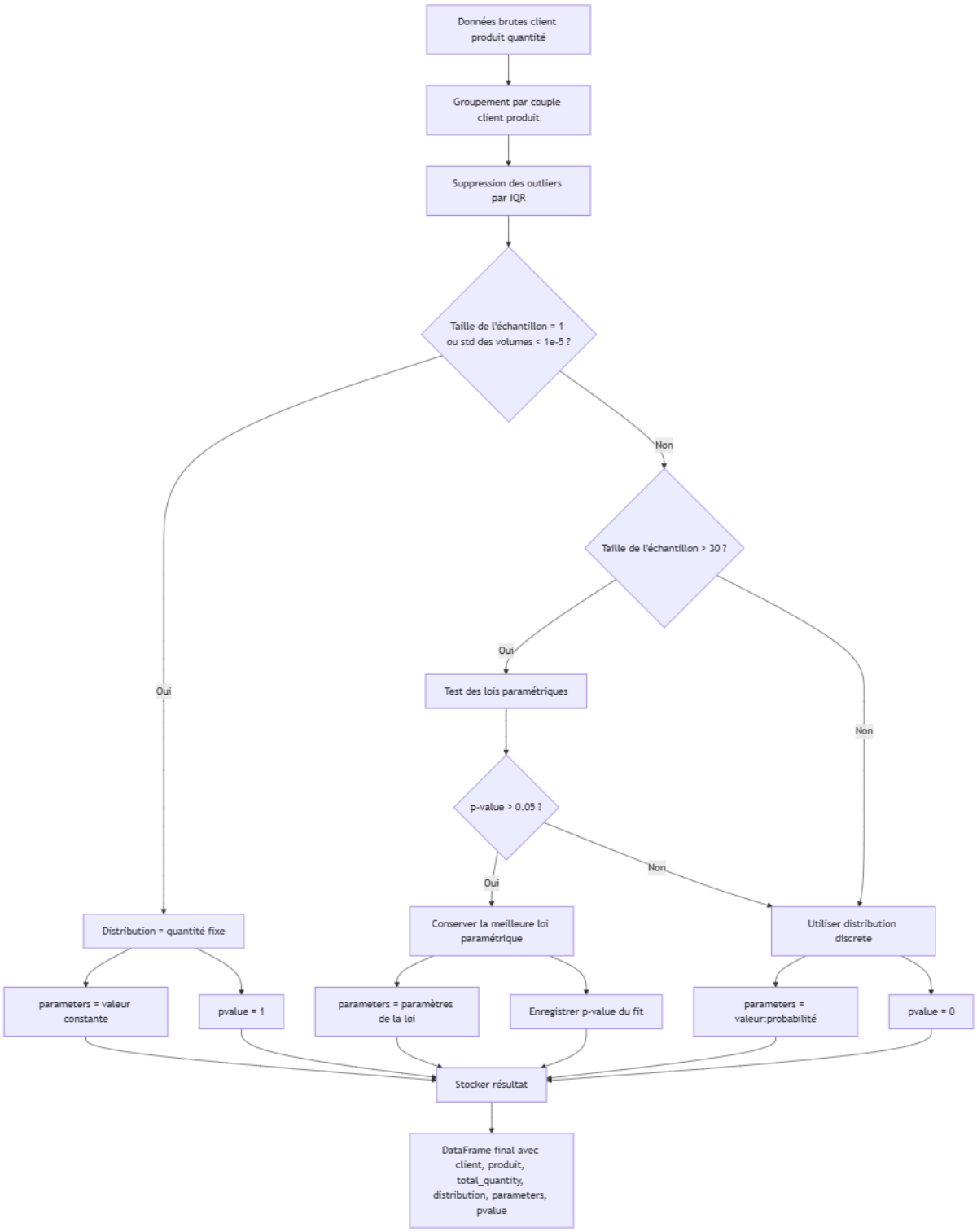


Figure 4.1 Méthodologie pour fixer les distributions dans la Méthode 1

On applique un seuil de p-value de 0.05: si aucune loi ne dépasse ce seuil, on considère que

les données ne peuvent pas être représentées de façon fiable par une loi paramétrique, et on utilise la distribution discrète (empirique). Si une p-value est supérieure ou égale à 0.05, la meilleure loi paramétrique est retenue. Ce seuil standard en statistique permet de contrôler le risque d'erreur et d'assurer la validité de l'ajustement.

Pour chaque couple, la distribution retenue et ses paramètres (valeur fixe, discrète ou paramétrique), la p-value, et le total commandé sont systématiquement enregistrés dans une base de données structurée. Les valeurs générées par ces distributions (les volumes commandés) sont ensuite utilisées pour alimenter le modèle de simulation, afin de générer les comportements de commande les plus réalistes possibles selon les données historiques.

L'extraction et le paramétrage des distributions ont été réalisés par un traitement Python, permettant une correspondance automatisée et fiable entre les environnements de simulation. Le tableau d'analogie des paramètres entre les distributions de SciPy (Python) et AnyLogic (Java), utilisé pour la modélisation, figure en annexe A.1.

Méthode 2 : Loi Normale Tronquée

Dans cette méthode conventionnelle, nous proposons de faire l'hypothèse simplifiée que la taille des commandes suit une loi normale tronquée car elle offre une bonne robustesse, une simplicité de paramétrage et une représentation satisfaisante même lorsque les données ne suivent pas parfaitement une distribution gaussienne. Cette méthode permet de tester l'impact de cette hypothèse simplifiée sur les performances du système.

4.3 Validation du modèle de base (Configuration 1)

La validation des données d'entrée constitue une étape fondamentale pour assurer la fiabilité du modèle de simulation. Cette validation est réalisée en comparant les distributions des indicateurs obtenus avec les données réelles fournies par le partenaire industriel avec les distributions des indicateurs obtenus avec le modèle de base. En particulier, nous nous sommes concentrés sur deux indicateurs critiques : le Total des Volumes Livrés (TVL) agrégé et l'Order Fulfillment Rate (OFR). L'idée consiste à calculer ces indicateurs directement à partir des données historiques brutes et à les comparer avec ceux obtenues après une première exécution du modèle de simulation. Le but est de garantir que le comportement global et les tendances de base du système simulé correspondent fidèlement à la réalité observée en termes de flux et de performance des commandes. Dans le cas contraire, des ajustements doivent être faits selon la méthodologie générale illustrée par la figure 3.3.

4.3.1 Validation des données en entrée avec des distributions suivant la méthode ajustée (Méthode 1)

Dans la Méthode 1 , nous avons testé différentes distributions pour modéliser les tailles des commandes des clients. Les valeurs obtenus pour les distributions des volumes livrés et des couples client/produit sont illustrées dans les graphiques 4.3 et 4.2.

Le figure 4.2 présente le pourcentage des distributions utilisées pour la génération de la demande en termes de couple client/produit. Ici, la distributions à quantité fixe et la distribution discrete sont également les plus dominantes, représentant respectivement 47.7% et 47.6% des commandes, ce qui montre une forte concentration autour de ces deux distributions. Les autres distributions, telles que *Gamma* et *Normale*, représentent des proportions très faibles, en dessous de 1%.

La figure 4.3 montre le pourcentage des différentes distributions en termes de volumes livrés. On observe que la distribution *discrète* domine largement, représentant 49.8% des volumes livrés, suivie de la distribution *à quantité fixe* avec 40.2%. D'autres distributions, telles que *Beta*, *Gamma*, et *Normale*, représentent des parts beaucoup plus petites du total.

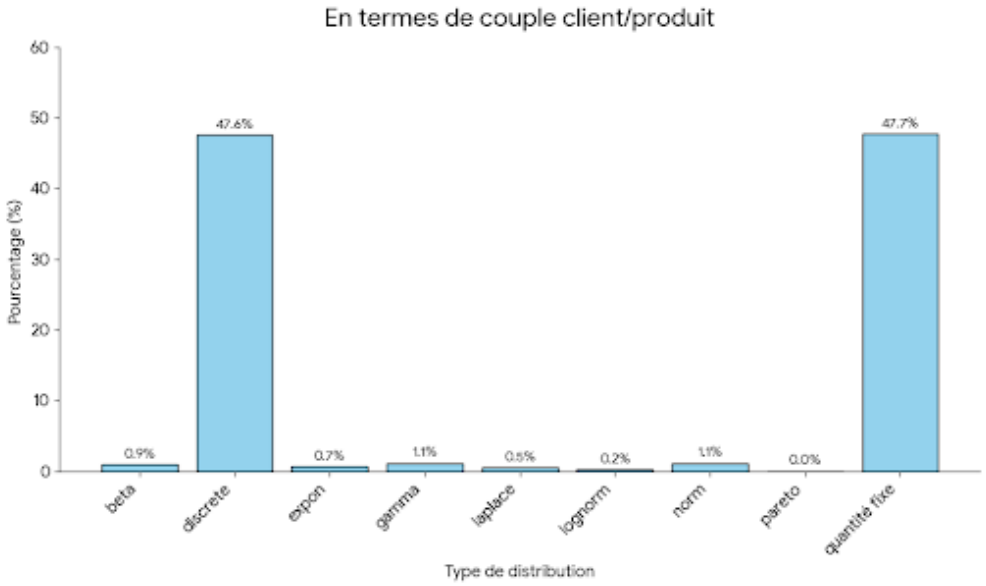


Figure 4.2 Distribution des couples client/produit

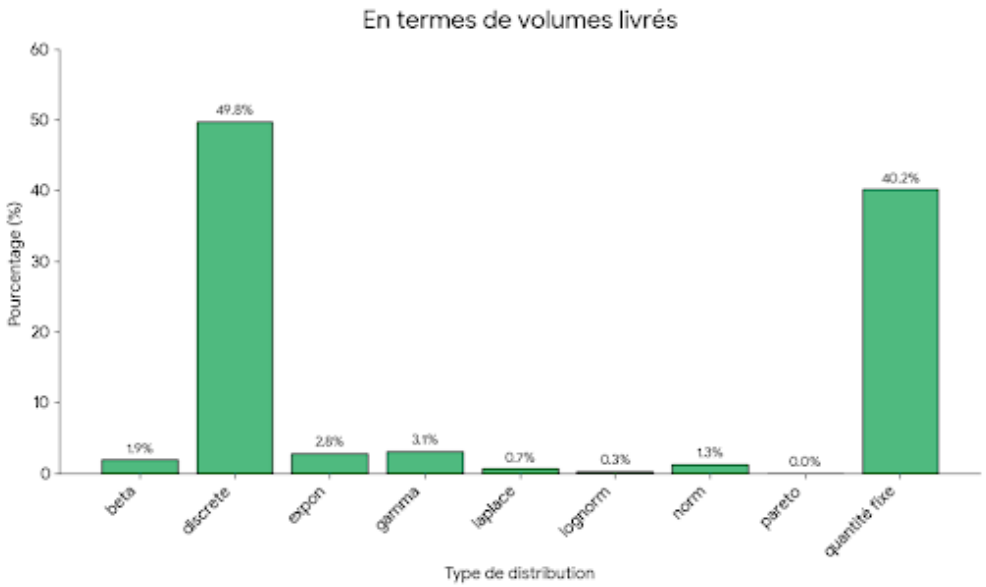


Figure 4.3 Distribution des volumes livrés pour la Méthode 1

Ces constats suggèrent que les distributions *discrète* et *à quantité fixe* sont les plus représentatives de la réalité des commandes dans cette méthode . Elles capturent l'essentiel des volumes livrés et des commandes passées par les clients, tandis que les autres distributions ont un impact marginal.

La figure 4.4 présente la distribution des ratios entre les volumes (Total des Volumes Livrés) réels et les volumes (Total des Volumes Livrés) simulés. On observe une certaine dispersion autour de la valeur 1, ce qui indique que les volumes simulés sont globalement proches des volumes réels. Les valeurs extrêmes visibles traduisent des cas où le modèle sous-estime ou surestime les volumes, mais ces cas restent marginaux.

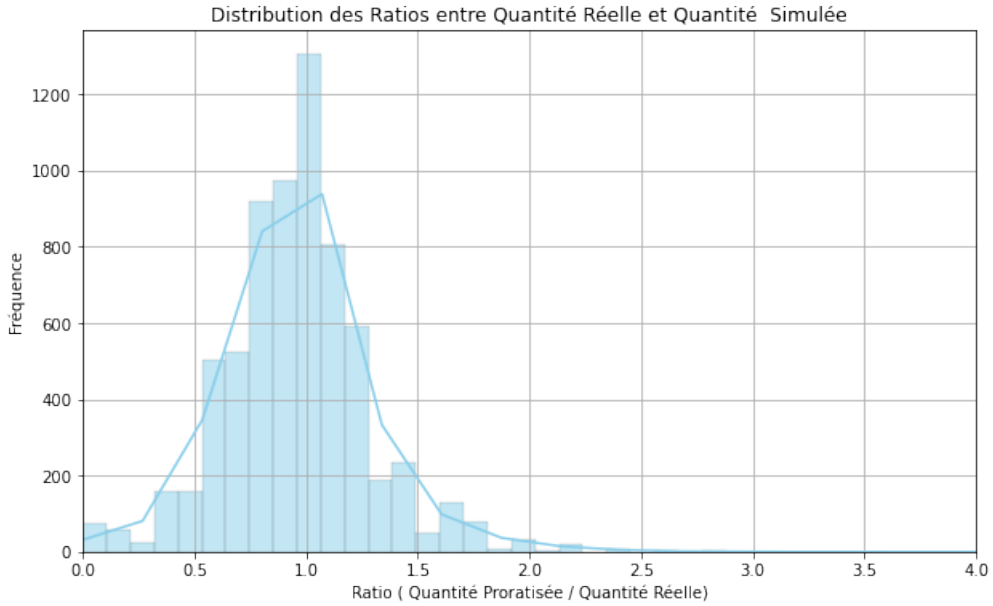


Figure 4.4 Distribution des ratios (volumes réels/volumes simulés) – Méthode 1.

Les figures 4.5 et 4.6 comparent les volumes livrés simulés et réels ainsi que leurs ratios. Les boxplots des volumes réels et simulés montrent des médianes très proches, avec une étendue interquartile réduite. Cela démontre la capacité du modèle à capturer les tendances centrales observées dans les données.

Le boxplot des ratios présente une concentration autour de 1, avec les statistiques suivantes :

- Moyenne : 0.966
- Médiane : 0.962
- Écart-type : 0.740
- Q1 : 0.791, Q3 : 1.114
- Minimum : 0.0, Maximum : 53.29

Ces résultats confirment que les volumes livrés simulés sont, dans l'ensemble, très proches des volumes livrés réels, avec un faible écart pour la majorité des cas. L'écart-type élevé indique une forte variabilité du modèle, bien que quelques ratios très élevés (jusqu'à 53.28 pour le maximum réel) suggèrent des cas isolés de sous-estimation ou de sur-estimation importantes. Toutefois, ces écarts restent marginaux et n'impactent pas significativement la qualité globale de la simulation .

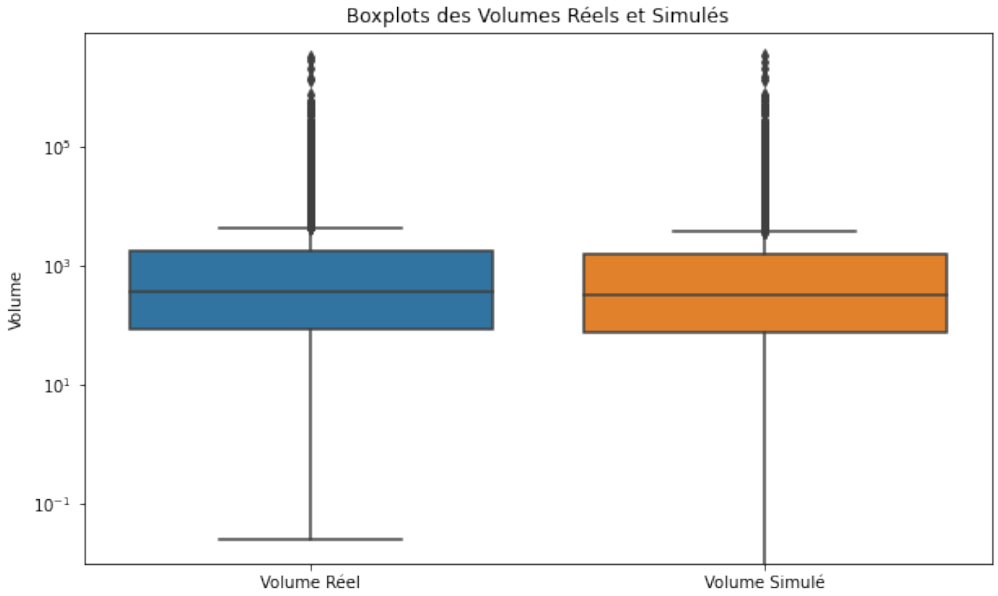


Figure 4.5 Boxplots des volumes livrés réels et volumes livrés simulés – Méthode 1.

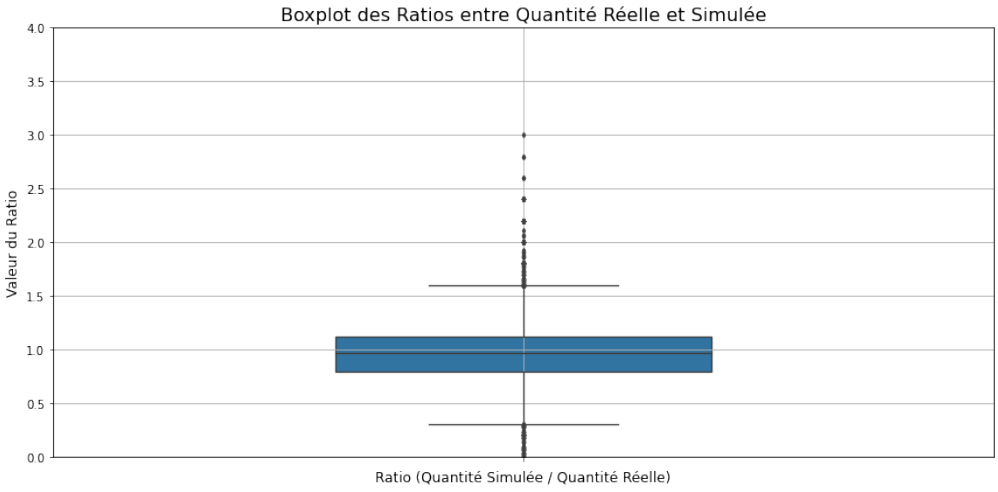


Figure 4.6 Boxplot des ratios (volumes livrés réels/volumes livrés simulés) – Méthode 1.

4.3.2 Validation des données en entrée avec des distributions suivant la loi normale tronquée (méthode 2)

La figure 4.7 montre la distribution des ratios pour la Méthode 2 . On observe une concentration principale autour de la valeur 1, similaire à la méthode 1, mais avec des pics plus

marqués et une dispersion plus importante. Cela suggère que le modèle, bien que toujours globalement précis, présente une plus grande variabilité dans ce cas.

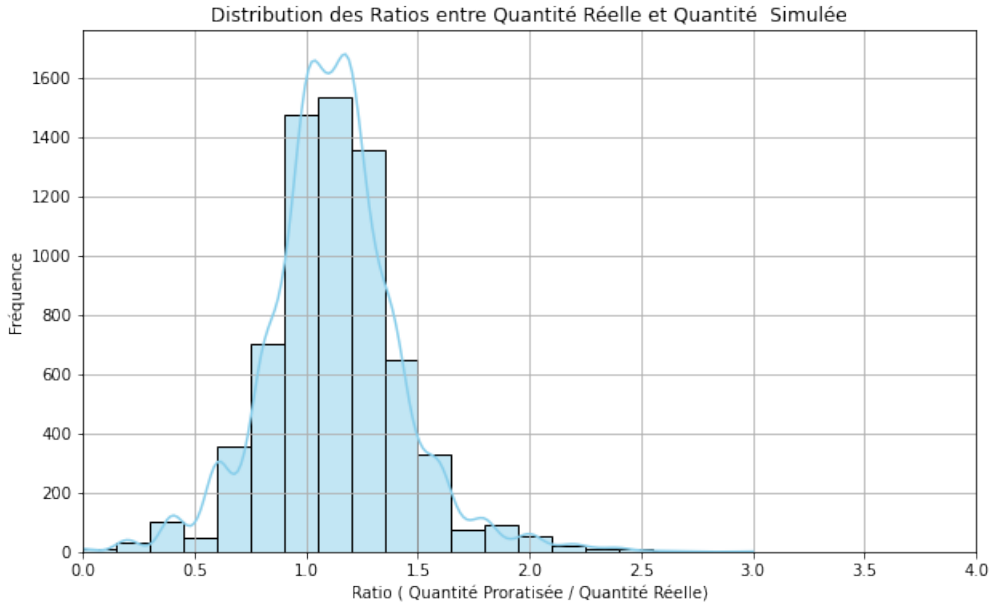


Figure 4.7 Distribution des ratios (volumes livrés réels/volumes livrés simulés) – Méthode 2.

La figure 4.8 compare les volumes livrés réels et simulés. Les médianes sont proches, ce qui montre que le modèle reproduit correctement la tendance centrale, mais l'étendue interquartile des volumes livrés simulés est légèrement plus large, indiquant une dispersion accrue.

La figure 4.9 représente les ratios sous forme de boxplot. Les statistiques associées sont les suivantes :

- Moyenne : 1.122
- Médiane : 1.116
- Écart-type : 0.300
- Q1 : 0.964, Q3 : 1.277
- Minimum : 0.0, Maximum : 3.0

Ces indicateurs révèlent une tendance à la surestimation des volumes réels par le modèle dans cette méthode , combinée à une variabilité modérée. Bien que la moyenne et la médiane restent proches, la présence d'un maximum élevé (ratio = 3) indique des cas extrêmes à surveiller.

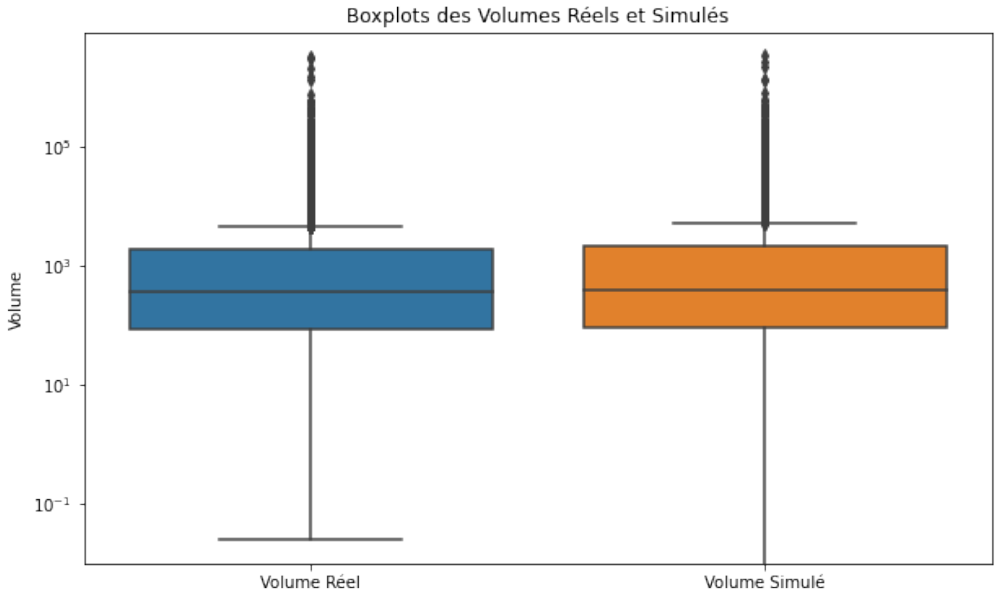


Figure 4.8 Boxplots des volumes réels et volumes simulés – Méthode 2.

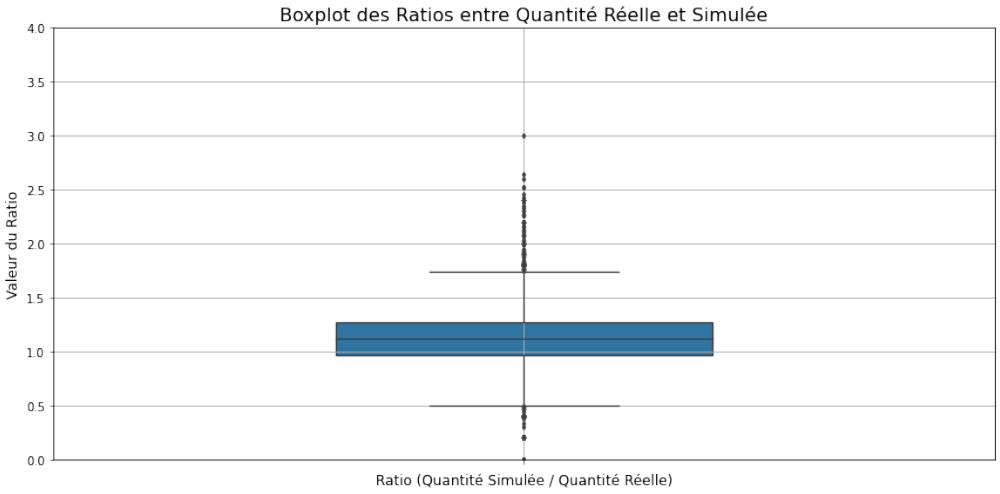


Figure 4.9 Boxplot des ratios (volumes livrés réels/volumes livrés simulés) – Méthode 2.

4.3.3 Comparaison des deux méthodes de génération des variables aléatoires

La comparaison des deux méthodes met en évidence une performance globalement satisfaisante du modèle dans les deux cas. Cependant, plusieurs différences significatives apparaissent :

— La Méthode 1 présente une meilleure précision globale, avec une moyenne et une médiane des ratios plus proches de 1, mais une forte variabilité avec un écart-type élevé (0.740) et des valeurs extrêmes très importantes (jusqu'à 53.29).
— La Méthode 2 montre une tendance à la surestimation des volumes (moyenne des ratios de 1.122), accompagnée d'une variabilité plus modérée (écart-type de 0.300), avec un maximum des ratios de 3.

Dans l'ensemble, la méthode 1 offre une meilleure concordance avec les données centrales mais avec une forte dispersion des cas extrêmes, tandis que la méthode 2 propose une solution plus représentative mais avec une légère tendance à surestimer les volumes.

Tableau 4.2 Comparaison des indicateurs statistiques des ratios (volumes réels / volumes simulés).

Indicateur	Méthode 1	Méthode 2
Moyenne	0.966	1.122
Médiane	0.962	1.116
Écart-type	0.740	0.300
Q1	0.791	0.964
Q3	1.114	1.277
Min	0.0	0.0
Max	53.29	3

Bien que la méthode 1 offre une représentation plus détaillée des volumes de commandes en considérant plusieurs distributions adaptées à chaque couple client/produit, la comparaison avec la méthode 2 (distribution normale tronquée unique pour l'ensemble des couples) a démontré qu'une granularité supplémentaire complexifie notablement l'implémentation du modèle sans apporter nécessairement une amélioration significative des résultats. Tout en étant plus simple à mettre en oeuvre, la méthode 2 présente des performances comparables avec la méthode 1, notamment en termes de ratios Total des Volumes réels/simulés et de taux de satisfaction des commandes (Order Fulfillment Rate, OFR), ainsi qu'un temps d'exécution identique de 4 minutes par réplication. Cette analyse confirme la validité du modèle avec les deux méthodes 1 et 2, et que dans notre contexte, la complexité additionnelle de la méthode 1 peut être modulée en fonction des objectifs visés et des ressources disponibles pour la modélisation (notamment les données réelles).

4.4 Conclusion

Dans ce chapitre, nous avons couvert l'analyse et la préparation générale des données, puis détaillé la structure du modèle, ses agents, et ses hypothèses de fonctionnement. Finalement, nous avons démontré la validité du modèle peu importe la méthode choisie pour la méthodes de génération de variables aléatoires. Cette validation a permis de définir notre configuration de référence.

Dans le chapitre qui suit, nous utiliserons ce modèle de base comme un modèle de référence pour la phase d'expérimentation et nous détaillerons le plan d'expériences permettant de quantifier l'impact de l'agrégation des données de la demande (càd, les commandes des clients) sur la fidélité du modèle.

CHAPITRE 5 EXPÉRIMENTATION

Ce chapitre présente la mise en œuvre et l'évaluation des deux configurations d'agrégation retenues. Il détaille d'abord la préparation des données industrielles et l'application des méthodes d'analyse ABC croisée et de clustering, puis analyse les résultats des simulations pour comparer l'impact de chaque approche sur les performances du modèle.

5.1 Choix des configurations de test

Afin d'évaluer l'impact de la granularité sur la performance du modèle de simulation, nous avons défini trois configurations distinctes.

Configuration 1 (Baseline): sert de modèle de référence. Elle utilise la granularité maximale en modélisant chaque couple client-produit unique. Cette approche, bien que potentiellement lourde en calcul, est supposée être la plus fidèle à la réalité opérationnelle et sert de point de comparaison pour évaluer la perte de précision des autres modèles.

Configuration 2 (ABC Croisée): adopte une approche d'agrégation managériale. Elle regroupe les produits en huit groupes basés sur une analyse ABC croisée, qui segmente les produits selon le volume et la fréquence de commande. L'objectif est de simplifier le modèle en se basant sur une logique opérationnelle reconnue.

Configuration 3 (Clustering): propose une agrégation statistique. Les produits sont ici regroupés en neuf classes à l'aide d'un algorithme de clustering (K-Means) appliqué sur leurs variables de commande (moyenne, écart-type, fréquence). Cette méthode vise à créer des groupes statistiquement homogènes.

Le tableau 5.1 ci-dessous résume ces trois approches.

Tableau 5.1 Les trois configurations de l'expérimentation

Configuration	**Description**
Configuration 1 : Couple Client-Produit (Baseline)	Utilisation de la granularité maximale (chaque couple unique client-produit est modélisé). Sert de référence pour évaluer l'impact des agrégations.
Configuration 2 : ABC Croisée	Les produits sont agrégés selon 8 clusters basés sur l'Analyse ABC croisée (Volume vs Fréquence). Réduit la complexité tout en conservant les profils comportementaux types.
Configuration 3 : Clustering	Les produits sont agrégés selon 9 clusters basés sur un algorithme de clustering appliqué aux variables de commande. Offre une segmentation statistiquement optimisée.

5.2 Planification des expériences

L'objectif de l'expérimentation est de comparer différentes méthodes d'agrégation en termes de fidélité de la simulation, de stabilité des résultats et d'efficacité computationnelle. Pour ce faire, nous considérons les trois configurations décrites dans la section 5.1 (Baseline, ABC Croisée, et Clustering).

Les indicateurs de performance retenus pour évaluer les performances de chaque configuration sont listés dans le tableau ci-dessous :

Tableau 5.2 Variables de sortie observées

Indicateur	Type	Unité	Interprétation
Order Fulfillment Rate (OFR)	Pourcentage	%	Mesure le taux de commandes livrées à temps. Indicateur clé de la qualité de service.
Le Total des Volumes Livrés(TVL)	Quantité	Cases (CAS)	Quantité totale expédiée par les Centres de distribution. Permet de vérifier la cohérence des flux simulés.

Ces variables sont automatiquement extraites à chaque exécution de configuration dans `AnyLogic`, permettant une analyse comparative rigoureuse entre les différentes configurations testées.

Calcul du Ratio entre le Volume Total Livré réel et le Volume Total Livré simulé

Afin de comparer les volumes réels des produits avec les volumes simulés agrégés par cluster, nous avons défini un *ratio* qui mesure l'écart relatif entre ces deux valeurs. Les étapes mathématiques de ce calcul sont détaillées ci-dessous.

Soit une base de données contenant les informations suivantes :

— P : un produit.

— C : un groupe auquel le produit P appartient.

— $V_{\text{réel}}(P)$: le volume réel observé pour le produit P.

— $V_{\text{total\_cluster}}(C)$: le volume total livré estimé pour tous les produits dans le cluster C.

L'objectif principal est de comparer le volume réel $V_{\text{réel}}(P)$ pour chaque produit avec un *volume proratisé* calculé à partir du volume total du cluster. Ce calcul se fait en deux étapes principales.

■ **Calcul du volume proratisé / désagrégation**

Le volume proratisé $V_{\text{proratisé}}(P)$ pour un produit P dans un cluster C est calculé comme suit :

$$V_{\text{proratisé}}(P) = \left(\frac{V_{\text{réel}}(P)}{\sum_{P' \in C} V_{\text{réel}}(P')} \right) \times V_{\text{total\_cluster}}(C)$$

où $\sum_{P' \in C} V_{\text{réel}}(P')$ est la somme des volumes réels de tous les produits appartenant au cluster C. Ce calcul ajuste le volume réel $V_{\text{réel}}(P)$ en fonction de la proportion du produit dans le total des volumes du cluster, et la multiplie par le volume total estimé du cluster.

■ **Calcul du Ratio de comparaison**

Une fois que le volume proratisé a été calculé, nous définissons le *ratio* $R(P)$ pour chaque produit P comme suit :

$$R(P) = \frac{V_{\text{réel}}(P)}{V_{\text{proratisé}}(P)}$$

Ce ratio permet de mesurer l'écart relatif entre le volume réel et le volume proratisé du produit. Les interprétations possibles de cette mesure sont les suivantes :

— $R(P) = 1$: Le volume réel est égal au volume proratisé.

— $R(P) > 1$: Le volume réel est supérieur au volume proratisé.

— $R(P) < 1$: Le volume réel est inférieur au volume proratisé.

Ce ratio constitue une mesure utile pour évaluer la précision des simulations par rapport aux volumes réels observés, et permet ainsi d'analyser l'efficacité des modèles de simulation dans le cadre de l'agrégation des données par cluster.

5.3 Agrégation et préparation des données des configurations

5.3.1 Préparation des données des configurations

Dans un premier temps, nous avons vérifié la complétude de la base de données. Aucune valeur manquante n'a été détectée, ce qui a considérablement facilité la phase de nettoyage.

Cette étape a ensuite été suivie par une structuration et un enrichissement des données, afin de les rendre plus exploitables dans les étapes ultérieures. Plusieurs variables dérivées ont été construites pour mieux représenter la dynamique des transactions observées :

— *Unique Client* : nombre distinct de clients ayant commandé un produit donné (utilisé notamment pour le clustering des produits) ;

— *Unique Plant* : nombre de centres de distribution ayant expédié un produit (ou inversement, nombre de centres ayant traité un client, dans le cas du clustering des clients) ;

— *Frequency* : nombre total de livraisons pour un produit ou nombre de commandes passées par un client ;

— *Mean Quantity* : moyenne des volumes commandés en *cases*, par commande, pour un produit ou un client.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

où x_i correspond à la quantité commandée lors de la $i^{\text{ème}}$ transaction, et n au nombre total de transactions. Une valeur élevée indique un client à fort volume d'achat ou un produit à forte rotation.;

— *Standard Deviation of Quantity (Std Quantity)* : mesure de la variabilité du volume commandé.

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Cet indicateur permet d'identifier les comportements réguliers ou, au contraire, plus erratiques.

Les deux dernières variables permettent de capturer à la fois le volume moyen et la régularité des commandes. Des valeurs extrêmes (*outliers*) ont été identifiées dans certaines variables,

notamment `mean quantity`, `frequency` et `std quantity`. Celles-ci correspondent souvent à de grands clients du marché ou à des produits à forte rotation. Ces observations n'ont pas été supprimées, car elles peuvent représenter des comportements réels et pertinents pour l'analyse.

Enfin, afin de lisser les effets de fluctuations de court terme, nous avons également pris en compte la fréquence d'occurrence des transactions, permettant d'obtenir une représentation plus stable du comportement d'achat.

Pour garantir une homogénéité d'échelle entre les variables et limiter l'effet des valeurs extrêmes, nous avons appliqué une normalisation via la méthode `StandardScaler`. Cette transformation centre les données autour de zéro avec un écart-type unitaire, ce qui est essentiel pour la performance de l'algorithme de clustering.

Les variables qualitatives ont été transformées par la méthode d'encodage `One-Hot`. Cette technique crée autant de colonnes binaires qu'il existe de modalités, ce qui permet de rendre les données exploitables par les algorithmes numériques tout en conservant l'information catégorielle.

5.3.2 Configuration 2 : ABC Croisée (Matricielle)

L'analyse ABC

L'analyse ABC est une méthode de classification largement utilisée en gestion des stocks et en logistique afin de hiérarchiser les articles selon leur importance relative. Inspirée du principe de Pareto (80/20), elle consiste à diviser les produits en trois catégories :

- Classe A : environ 10–20 % des articles qui concentrent près de 70–80 % de la valeur totale. Ces articles critiques nécessitent une gestion rigoureuse et un suivi étroit (inventaire fréquent, commandes sécurisées).
- Classe B : environ 20–30 % des articles représentant 15–25 % de la valeur. Ces articles ont une importance intermédiaire et peuvent être gérés avec une fréquence moyenne.
- Classe C : environ 50–70 % des articles représentant seulement 5–10 % de la valeur. Ces articles moins critiques peuvent être gérés avec des contrôles moins fréquents [37].

Cette segmentation permet aux entreprises de concentrer leurs efforts et leurs ressources sur les articles à plus forte valeur stratégique.

L'analyse ABC croisée

L'analyse ABC croisée est une extension de l'ABC classique qui intègre une seconde dimension pour affiner la classification des articles. En plus de la valeur ou du volume, une seconde variable est prise en compte, comme :

— la fréquence d'utilisation (nombre d'occurrences),
— la criticité des articles (impact opérationnel),
— ou encore le délai d'approvisionnement.

Cette double segmentation aboutit à une matrice de clusters:

	Fréquence élevée	Fréquence faible
Valeur élevée	AA (priorité maximale)	AB (gestion rigoureuse)
Valeur faible	BA (fréquent mais peu critique)	CC (gestion souple)

L'analyse ABC croisée permet ainsi de développer des politiques différenciées (approche "segmentation des stocks"), en ajustant les niveaux de contrôle et les stratégies d'approvisionnement selon les caractéristiques de chaque cluster.

Ces méthodes offrent un cadre analytique robuste pour hiérarchiser les articles en fonction de leur importance, permettant une allocation optimale des ressources et une réduction des coûts liés à la gestion des stocks.

Analyse ABC: agrégation des produits basée sur les volumes et les occurrences

L'analyse ABC a été utilisée non pas comme un simple outil de gestion des stocks ou de priorisation logistique, mais comme un mécanisme d'agrégation visant à simplifier la représentation des comportements d'achat des produits.

Deux dimensions distinctes ont été considérées pour cette agrégation : le volume livré par produit et la fréquence d'apparition du produit dans les commandes. L'objectif est d'identifier des profils types de produits à intégrer dans la simulation comme entités homogènes, en réduisant la granularité tout en conservant une diversité comportementale suffisante.

La figure 5.1 présente la répartition des produits selon ces deux dimensions : à gauche selon le volume cumulé, à droite selon le nombre d'occurrences dans les commandes.

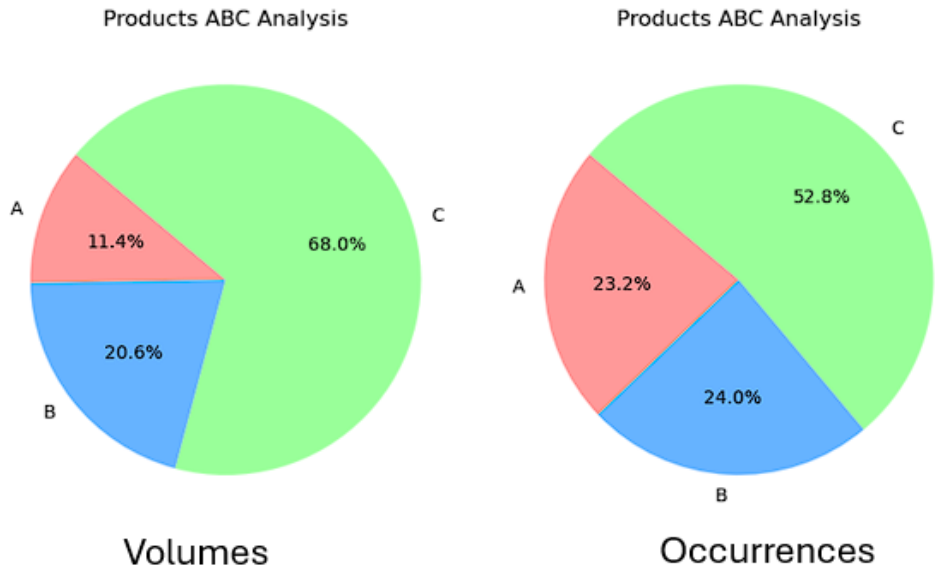


Figure 5.1 Analyse ABC des produits selon les volumes (gauche) et les occurrences (droite).

Cette analyse montre que 68 % des produits appartiennent à la classe C en termes de volume, tandis que seulement 52,8 % y figurent en termes de fréquence. Cette dissymétrie suggère que certains produits, bien que peu volumineux, sont commandés fréquemment — une information utile pour simuler des comportements différenciés dans le modèle.

La répartition exacte par classe est résumée dans les deux tableaux suivants :

Tableau 5.3 Répartition des produits par classe selon les volumes

Classe	Pourcentage	Nombre de produits
A	11.4%	47
B	20.6%	85
C	68.0%	281

Tableau 5.4 Répartition des produits par classe selon les occurrences

Classe	Pourcentage	Nombre de produits
A	23.2%	96
B	24.0%	99
C	52.8%	218

Pour obtenir une typologie plus fine des produits, une analyse ABC croisée a été réalisée. Chaque produit se voit attribuer un double classement : une lettre pour son importance en volume, et une autre pour sa fréquence. Cette segmentation permet de définir des catégories homogènes de comportements à intégrer comme paramètres dans la simulation.

Le tableau 5.5 présente les effectifs associés à chaque combinaison. Par exemple, les produits `AA` sont à forte intensité comportementale sur les deux dimensions, tandis que les produits `CC` représentent des entités à faible influence globale. Ces profils sont utilisés dans le modèle de simulation pour attribuer des lois de comportement distinctes selon le groupe auquel appartient chaque produit.

Tableau 5.5 Répartition des produits selon l'analyse ABC croisée

Cluster ABC	Nombre de produits
CC	209
CB	65
BA	45
AA	44
BB	31
BC	9
CA	7
AB	3

Ainsi, l'analyse ABC croisée constitue une première approche d'agrégation, orientée vers la définition de profils de comportements types dans le modèle de simulation. Ces profils permettent d'éviter la modélisation individuelle de centaines de produits, tout en conservant une variabilité réaliste dans la génération des commandes.

5.3.3 Configuration 3 : Clustering

En complément de l'analyse ABC, une méthode de clustering a été utilisée afin de regrouper les produits en fonction de leurs comportements de commande.

Choix de l'algorithme

Plusieurs algorithmes de clustering ont été testés et comparés sur la base de critères tels que la complexité algorithmique, la rapidité d'exécution, l'interprétabilité des observations, la robustesse et la forme des clusters produits. Le tableau suivant résume les principales caractéristiques :

Tableau 5.6 Comparaison des algorithmes de clustering testés

Algorithme	Complexité	Rapidité	Interprétabilité	Performances observées	Trade-off général
K-means	Faible (O(nkd))	Très rapide	Excellente (clusters compacts et clairs)	Bons si les clusters sont sphériques, sensible aux outliers	Excellent compromis si les données sont bien préparées
Hiérarchique	Très élevée ($O(n^3)$)	Lente pour grands jeux de données	Moyenne (lecture du dendrogramme)	Performant sur petits jeux complexes, sensible au bruit	Moins adapté aux grands volumes
DBSCAN	Modérée (O(n log n))	Moyenne	Faible (absence de centroïdes)	Très bons pour formes complexes et bruit élevé	Idéal pour formes irrégulières ou données bruitées
GMM	Moyenne (O(nkd))	Moyenne	Moyenne (interprétation probabiliste)	Excellents pour clusters ellipsoïdaux, sensibles aux initialisations	Bon compromis mais plus coûteux que K-means

Sources : MacQueen (1967), Johnson (1967), Ester et al. (1996), Reynolds (2009), Murtagh & Contreras (2012)

Compte tenu de la nature des données (variables numériques continues, sans forme arbitrairement complexe), de la nécessité d'une interprétation rapide, et des contraintes de performance, l'algorithme K-means a été retenu comme le plus adapté à notre contexte.

Principe de l'algorithme K-means

L'algorithme K-means cherche à partitionner n observations en K groupes en minimisant la variance intra-cluster. L'objectif est de minimiser la fonction :

$$J = \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - \mu_k\|^2$$

où :

— K est le nombre de clusters,
— C_k est le $k^{\text{ème}}$ cluster,

— μ_k est le centroïde du cluster k,
— $\|\cdot\|^2$ représente la distance euclidienne au carré.

Détermination du nombre optimal de clusters

L'algorithme K-means nécessite de spécifier le nombre de clusters K à l'avance. Pour identifier une valeur optimale, deux critères complémentaires ont été utilisés :

Méthode du coude Cette méthode repose sur l'analyse de l'inertie intra-cluster, définie comme la somme des distances quadratiques entre chaque observation et le centroïde de son cluster. On trace l'inertie en fonction de K et on observe le point de « coude » à partir duquel l'ajout d'un cluster n'apporte qu'un gain marginal. Ce point représente une transition entre amélioration significative et stabilisation des performances.

Coefficient de silhouette Le coefficient de silhouette mesure simultanément la cohésion intra-cluster et la séparation inter-cluster. Pour chaque observation i :

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

où :

— $a(i)$ est la distance moyenne entre i et les autres points de son propre cluster,
— $b(i)$ est la plus petite distance moyenne entre i et les points d'un autre cluster (le plus proche).

Le score global est la moyenne de $s(i)$ pour toutes les observations. Un score proche de 1 indique une bonne séparation ; un score proche de 0 ou négatif suggère des regroupements peu pertinents.

Dans notre cas, la valeur optimale de K est celle qui maximise le compromis entre une inertie faible et un coefficient de silhouette élevé.

Implémentation technique

L'implémentation a été réalisée à l'aide des bibliothèques Python suivantes :

— `pandas` et `numpy` : pour la manipulation et le traitement des données.

— `scikit-learn` : pour les étapes suivantes :

  — Normalisation des variables numériques avec `StandardScaler`,
  — Encodage des variables catégorielles avec `OneHotEncoder`,
  — Application de l'algorithme `KMeans`,
  — Calcul de l'inertie et du `silhouette_score`.

Détermination du nombre optimal de clusters de produits

Pour la première phase de clustering, trois variables principales ont été retenues : `mean_qty`, `std_qty` et `frequency`. Elles traduisent respectivement le volume moyen commandé, la variabilité des quantités et la récurrence des transactions. Ces dimensions constituent le cœur du processus de segmentation, car elles permettent de caractériser les comportements d'achat selon leur intensité et leur régularité. Les autres variables issues de la base de données sont conservées à des fins d'interprétation des groupes, mais ne sont pas incluses dans la construction initiale des clusters, afin d'éviter tout biais de corrélation.

Les produits dits « occasionnels » — ceux n'ayant été commandés qu'une seule fois sur toute la période analysée — ont été isolés dans un cluster dédié. Cela permet de ne pas biaiser l'analyse portant sur les produits ayant une certaine régularité dans leur consommation.

Après, pour déterminer le nombre optimal de clusters pour les produits restants, deux méthodes d'évaluation ont été mobilisées : la méthode du coude (elbow method) et le score de silhouette.

La figure 5.2 montre que la courbe d'inertie décroît fortement jusqu'à cinq clusters, avant de ralentir sa décroissance, avec une inflexion notable autour de huit clusters. Ce comportement suggère qu'un regroupement en huit clusters permettrait d'obtenir une segmentation fine tout en maintenant une bonne cohérence intra-cluster.

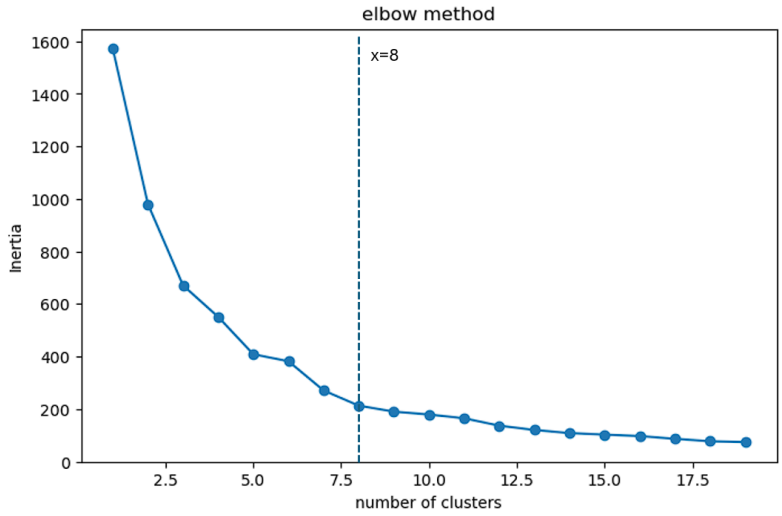


Figure 5.2 Méthode du coude : évolution de l'inertie en fonction du nombre de clusters pour les produits non occasionnels.

Parallèlement, le score de silhouette présenté dans la figure 5.3 reste relativement stable entre 8 et 12 clusters, avec un léger pic à 8 clusters. Toutefois, la différence étant marginale, le choix de 8 clusters reste pertinent car il offre un compromis acceptable entre qualité du regroupement et niveau de détail.

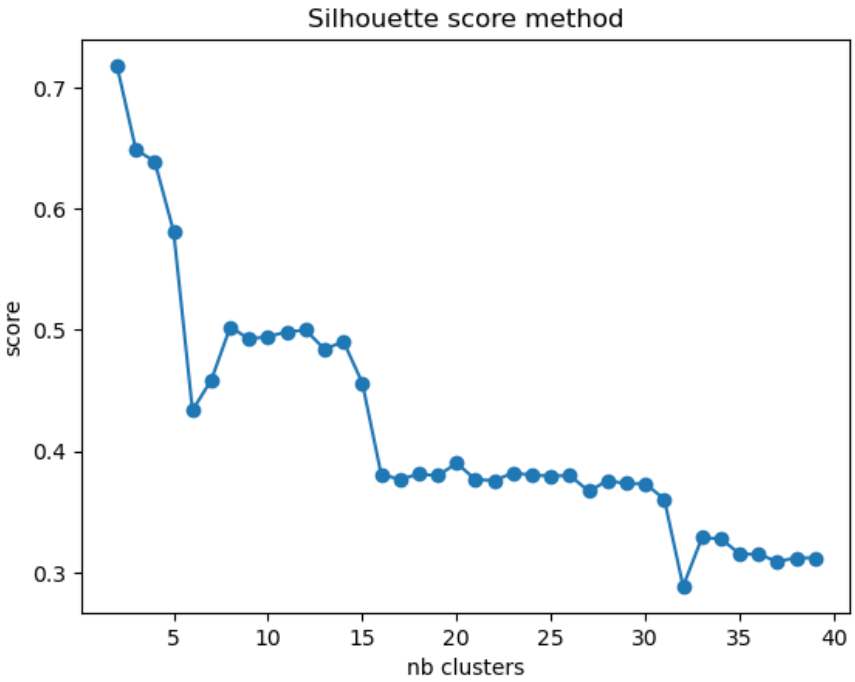


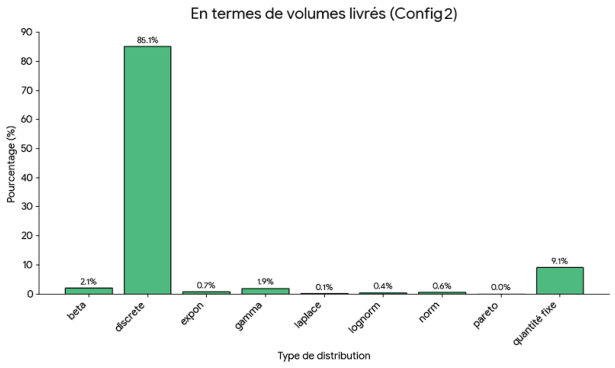
Figure 5.3 Score de silhouette : évaluation de la qualité des clusters en fonction du nombre de clusters.

En conclusion, la configuration retenue pour la suite de l'analyse inclut 8 clusters de produits réguliers, auxquels s'ajoute un cluster spécifique pour les produits occasionnels. Le nombre total de groupes de produits utilisés dans la simulation est donc fixé à 9.

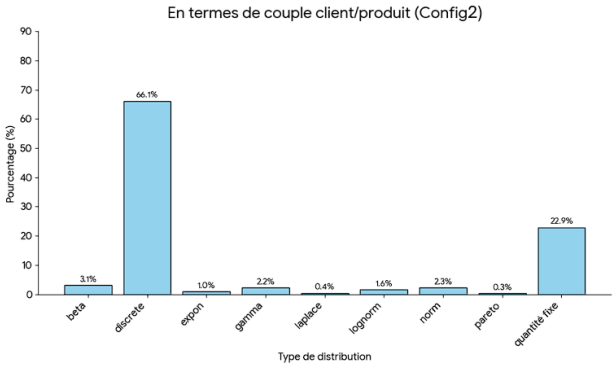
5.4 Analyse de données et choix des lois statistique de description de la demande

Pour la première méthode d'agrégation, basée sur une classification ABC croisée, les résultats sont illustrés dans la figure 5.4. En termes de volumes livrés (figure 5.4a), nous observons une domination de la distribution *discrète*, qui modélise 85.1% de tous les volumes, suivie de la *quantité fixe* avec 9.1%. L'analyse en termes de couple client/produit (figure 5.4b) montre une tendance similaire : la distribution *discrète* reste majoritaire (66.1%), suivie de la *quantité fixe* (22.9%). Les autres distributions comme beta, normale ou gamma représentent des parts marginales.
La seconde méthode d'agrégation, utilisant un algorithme de clustering K-Means, présente des résultats très similaires, visibles à la figure 5.5. L'analyse des volumes livrés (figure 5.5a) montre une forte prédominance de la distribution *discrète* (74.1%), suivie de la *quantité fixe* (18.8%). De même, la répartition des couples client/produit (figure 5.5b) est dominée par la distribution *discrète* (67.0%) et la *quantité fixe* (20.3%). Il est notable que les deux

méthodes d'agrégation, bien que conceptuellement différentes, aboutissent à une structure de distribution très proche, où l'essentiel de la demande est modélisé par des distributions *discrètes* ou à *quantité fixe*.

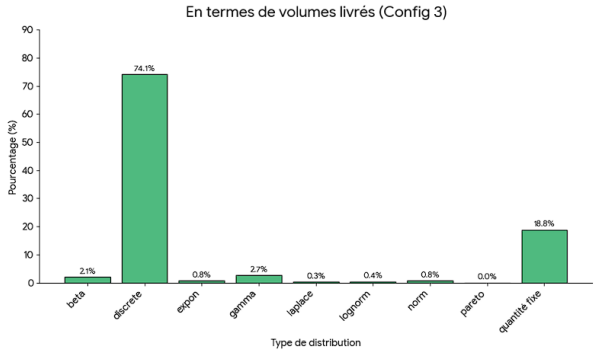


(a) Distribution des volumes livrés (ABC)

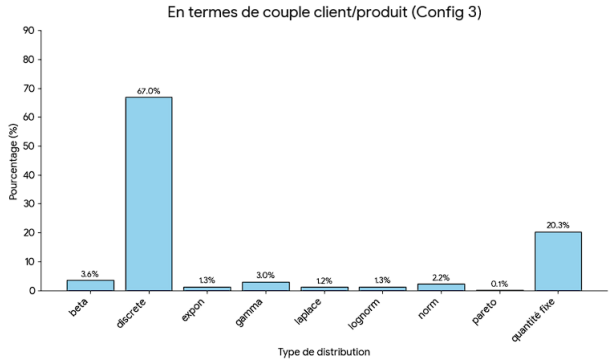


(b) Distribution des couples client/produit (ABC)

Figure 5.4 Résultats de l'ajustement pour la configuration ABC



(a) Distribution des volumes livrés (K-Means)



(b) Distribution des couples client/produit (K-Means)

Figure 5.5 Résultats de l'ajustement pour la configuration K-Means

La différence la plus notable entre la configuration de base et les configurations agrégées réside dans le basculement de l'équilibre des distributions dominantes. Alors que la configuration de base montrait une répartition quasi égale (environ 50/50) entre *quantité fixe* et *discrète*, les deux méthodes d'agrégation provoquent un effondrement de la part de la *quantité fixe* au profit de la distribution *discrète*, qui devient ultra-dominante.

Parallèlement à ce changement, on observe un effet sur les lois paramétriques continues (telles que *beta*, *gamma* ou *normale*). Bien que leur part totale en termes de volumes livrés reste stable ou diminue, leur part en termes de couples client-produit a plus que doublé, passant d'un total d'environ 4.5% dans la configuration de base à plus de 10% (et même 12.6% pour le K-Means).

Cependant, et c'est le point essentiel, même si leur proportion a augmenté, ces distributions paramétriques restent très minoritaires dans l'ensemble. Elles ne parviennent pas à capter une part significative de la demande, que ce soit au niveau fin ou agrégé.

Cette différence globale s'explique par la nature même de l'agrégation. Au niveau de base (client-produit), un grand nombre de couples ont un comportement simple de type *quantité fixe*. Cependant, lorsque ces couples sont regroupés (dans un cluster ou une classe ABC), leurs différentes quantités fixes s'additionnent. Le nouveau groupe agrégé n'a plus une seule quantité fixe, mais un ensemble de valeurs distinctes.

Cet ensemble de valeurs est alors modélisé comme une distribution *discrète*, ce qui augmente mécaniquement sa part au détriment de la distribution à *quantité fixe*. Le processus d'agrégation tend donc à "fusionner" de nombreux comportements individuels simples (quantité fixe) en une distribution *discrète* plus large et plus générale. Autrement dit, ce qui apparaissait comme des commandes constantes et distinctes au niveau le plus fin est absorbé dans un comportement discret plus complexe au niveau agrégé, tandis que l'influence des lois paramétriques continues demeure négligeable.

5.5 Conclusion

Ce chapitre a permis de comparer le potentiel des deux configurations distinctes d'agrégation des produits dans le contexte de la simulation:

- Clustering : une méthode qui offre une segmentation fine, fondée sur les caractéristiques statistiques (volume, fréquence, variabilité), augmentant potentiellement la finesse descriptive au prix d'une complexité de calibration plus forte.
- ABC croisée : une approche d'agrégation plus simple et intuitive, fournissant des profils de comportement synthétiques et favorisant l'équilibre entre réalisme structurel et simplicité de mise en œuvre.

Les observations issues des deux configurations sont comparées selon les critères suivants :

- correspondance entre le total des volumes livrés simulés et le total des volumes livrés réels ;

— stabilité et variabilité des comportements de commande selon la segmentation ;
— impact sur le taux de satisfaction des commandes (OFR).

La configuration basée sur le clustering offre une meilleure représentation des comportements hétérogènes de commande mais entraîne une complexité plus élevée lors de la calibration du modèle. À l'inverse, la configuration ABC croisée, plus synthétique, conserve des performances comparables tout en réduisant le nombre d'entrées à paramétrer.

CHAPITRE 6 ANALYSE DES RÉSULTATS ET DISCUSSION

Ce chapitre présente et analyse les résultats obtenus à partir de la simulation pour deux configurations de segmentation des produits. L'objectif est d'évaluer l'impact de la méthode d'agrégation choisie sur la précision et la stabilité du modèle de simulation.

Les configurations testées sont les suivantes :

- Configuration 2 – Agrégation selon l'analyse ABC croisée : basée sur la segmentation des produits selon leur volumes et leur fréquence de consommation, cette approche allie logique opérationnelle et régularité statistique.
- Configuration 3 – Agrégation par clustering K-means : fondée sur des critères statistiques (moyenne, dispersion et fréquence des commandes), cette approche non supervisée met en évidence la diversité des comportements produits.

Ces deux approches permettent d'examiner le compromis entre réalisme opérationnel, robustesse statistique et finesse d'analyse. Les résultats sont évalués principalement selon deux axes :

1. Le ratio entre les quantités réelles et les quantités simulées.
2. L'analyse des erreurs et la stabilité statistique du modèle.

6.1 Analyse des résultats

L'objectif central de cette phase expérimentale est d'évaluer l'impact des méthodes d'agrégation sur la performance du modèle. Pour garantir une comparaison rigoureuse et équitable entre la configuration de référence (Baseline) et les configurations agrégées (ABC Croisée et Clustering), toutes les expériences ont été menées dans des conditions identiques.

Conformément aux paramètres définis dans la section 4, chaque scénario a été testé sur un horizon de 730 jours avec une phase de préchauffage de 28 jours. La capacité logistique (1 000 camions) et le nombre de réplications (5 par scénario) sont demeurés constants. Cette uniformité du protocole expérimental assure que les variations observées dans les indicateurs de performance résultent exclusivement du changement de niveau de granularité, et non de facteurs externes liés à la configuration du moteur de simulation.

Il est à noter que le temps d'exécution par réplication n'a pas dépassé 2 minutes, ce qui représente moins de la moitié du temps d'exécution observé dans la configuration de base. Cette réduction significative du temps de calcul montre que le modèle a été efficacement allégé d'un point de vue computationnel, validant ainsi l'intérêt des méthodes d'agrégation

pour optimiser la réactivité du modèle.

L'analyse des écarts entre les volumes livrés réels et les volumes livrés simulés (ou désagrégés) constitue un élément central de la validation du modèle. L'objectif est d'évaluer la capacité de chaque configuration à reproduire fidèlement la réalité tout en maîtrisant la variabilité induite par la granularité et les choix de modélisation.

Trois configurations sont comparées : la configuration de référence (Baseline), l'agrégation selon l'analyse ABC croisée, et le clustering non supervisé K-means. Pour chacune, deux méthodes distinctes pour la génération des variables aléatoires sont testées :

— Méthode 1 (Distribution Ajustée): lois probabilistes adaptées à chaque segment ou cluster ;
— Méthode 2 : génération par loi normale tronquée.

Tableau 6.1 Plan de synthèse des expérimentations

Configuration	**Méthode**	**Objectif de l'Expérience**
Config 1: Baseline (Granularité maximale)	Méthode 1 Méthode 2	Établir une performance de référence (précision et stabilité) sans agrégation.
Config 2: ABC Croisée	Méthode 1	Tester l'adéquation d'une méthode ajustée (M1) avec une agrégation managériale.
Config 2: ABC Croisée	Méthode 2	Tester l'adéquation de la loi normale tronquée (M2) avec une agrégation managériale.
Config 3: K-means	Méthode 1	Tester l'adéquation d'une méthode ajustée (M1) avec une agrégation statistique.
Config 3: K-means	Méthode 2	Tester l'adéquation de la loi normale tronquée (M2) avec une agrégation statistique.

L'analyse suivante présente successivement les résultats pour les configurations 2 et 3, avant de proposer une discussion comparative et des recommandations.

6.1.1 Configuration 2 : Agrégation selon l'analyse ABC croisée

La configuration 2 repose sur une segmentation fondée sur une analyse ABC croisée à huit classes. Cette approche regroupe les produits selon leur importance macro (volume total et fréquence totale), créant des clusters homogènes sur le plan managérial.

Méthode 1 : Distribution Ajustée

L'application de lois probabilistes spécifiques à chaque produit *au sein* des clusters ABC aboutit à une stabilité remarquable. Les ratios entre quantités réelles et simulées-désagrégées présentent une dispersion très faible :

— Moyenne : 0.863 ;
— Médiane : 0.907 ;
— Écart-type : 0.07 ;
— Q1 : 0.816 ; Q3 : 0.949 ;
— Minimum : 0.674 ; Maximum : 0.949.

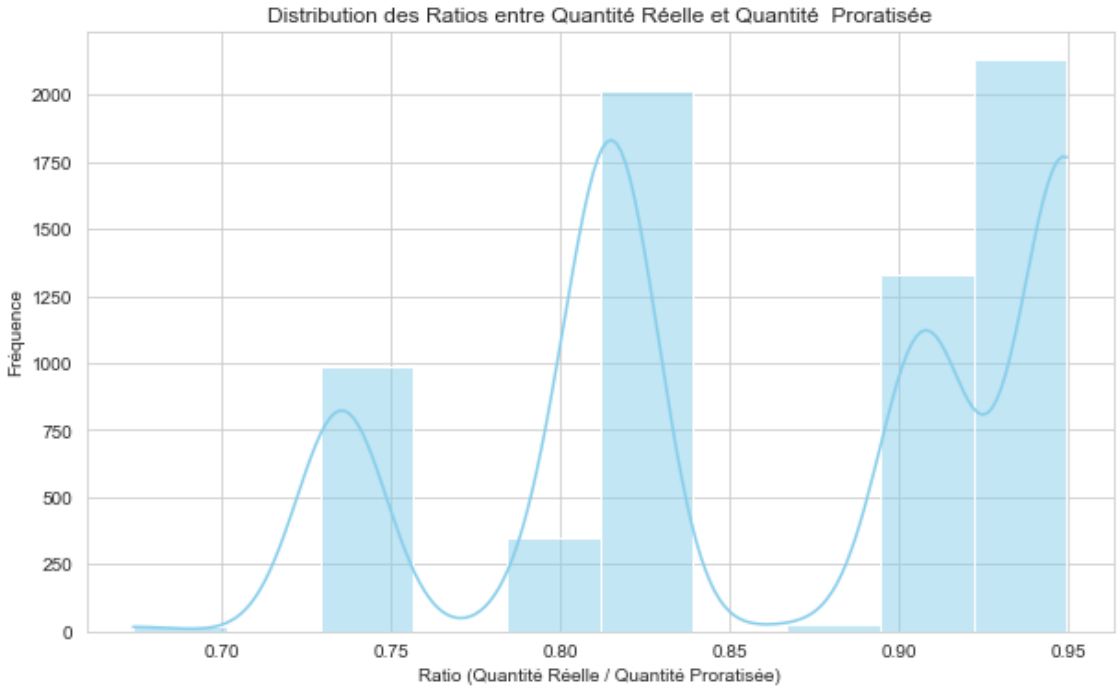


Figure 6.1 Distribution des ratios (Volumes livrés réels / volumes livrés désagrégés) – Configuration 2, Méthode 1

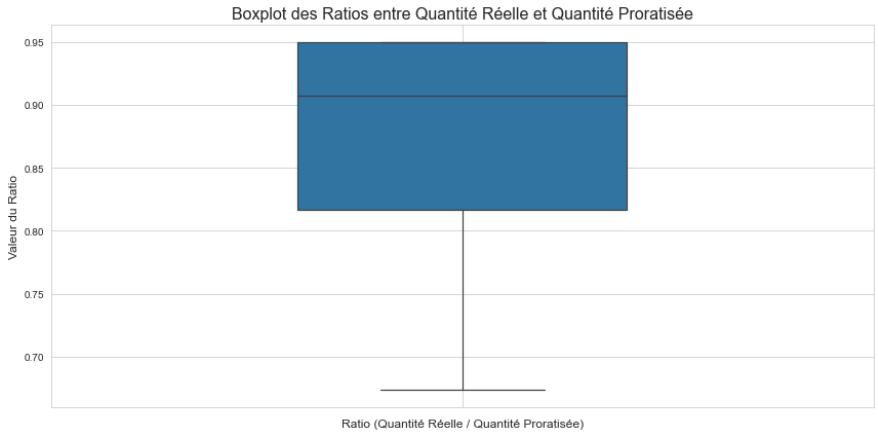


Figure 6.2 Boxplot des ratios (Volumes livrés réels / volumes livrés désagrégés) – Configuration 2, Méthode 1

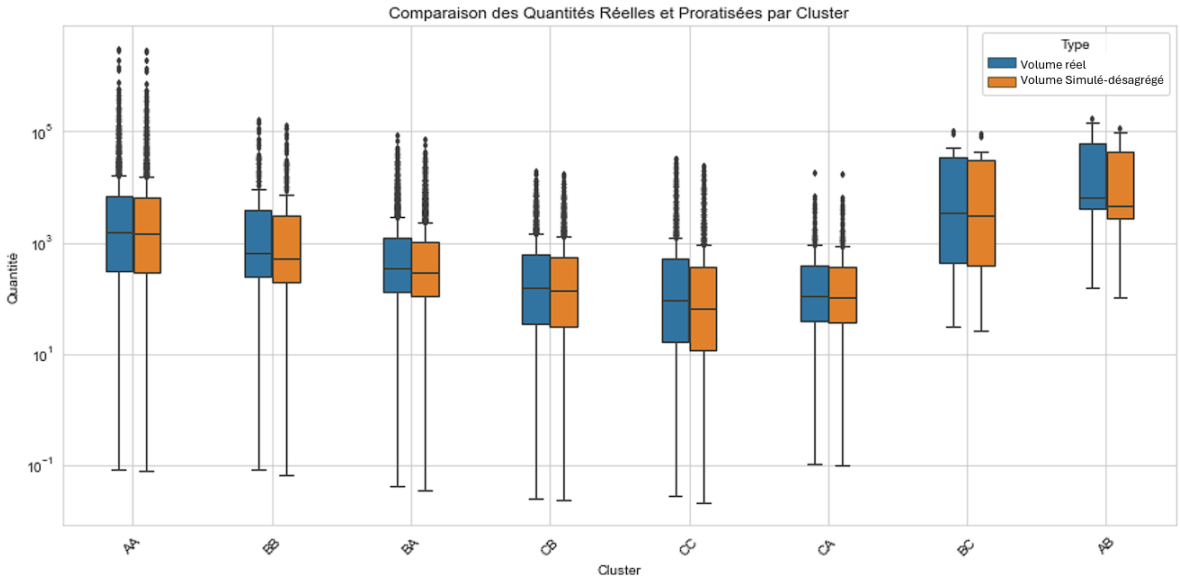


Figure 6.3 Boxplots des volumes livrés réels et désagrégés par cluster (log-scale) – Configuration 2, Méthode 1

Les figures confirment la cohérence des résultats obtenus par le modèle et les données réelles. Cette performance s'explique par la flexibilité de la Méthode 1. Bien que les clusters ABC soient homogènes au niveau macro, ils restent hétérogènes au niveau micro (comportement de commande). La Méthode 1 respecte cette diversité en modélisant chaque groupe de produits individuellement (loi constante, loi gamma, etc.), ce qui mène à une simulation stable et fiable.

Malgré cette stabilité, la moyenne du ratio (0.863) indique une tendance à la sous-estimation. Ce biais demeure constant, ce qui permettrait de le corriger aisément.

Méthode 2 : Génération par loi normale tronquée

Lorsque la même structure ABC est utilisée avec une loi normale tronquée (Méthode 2), le comportement du modèle se dégrade fortement. Les indicateurs statistiques révèlent une dispersion extrême :

- Moyenne : 4.106 ;
- Médiane : 1.670 ;
- Écart-type : 7.786 ;
- Q1 : 0.100 ; Q3 : 4.364 ;
- Minimum : 0.100 ; Maximum : 39.484.

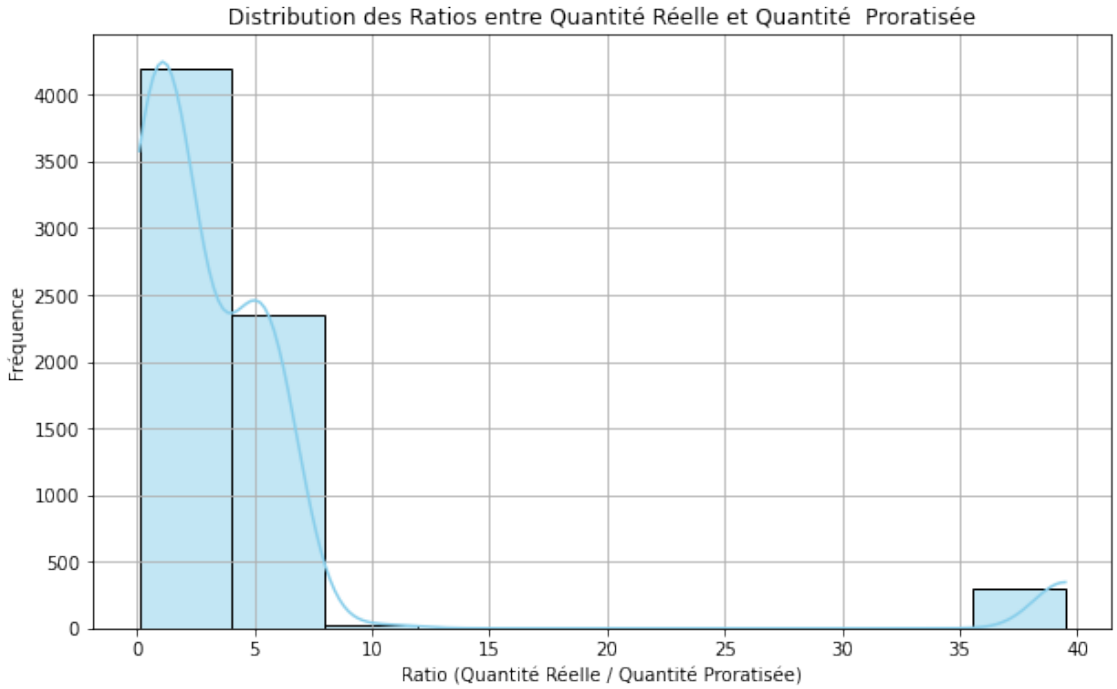


Figure 6.4 Distribution des ratios (Volumes livrés réels / volumes livrés désagrégés) – Configuration 2, Méthode 2

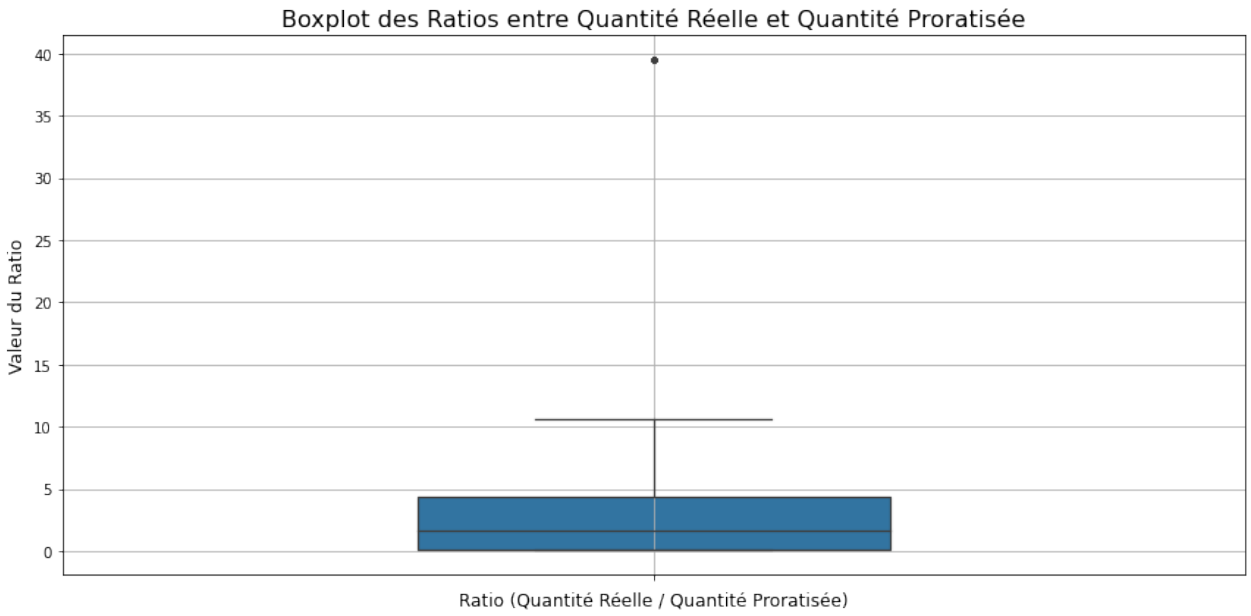


Figure 6.5 Boxplot des ratios (Volumes livrés réels / volumes livrés désagrégés) – Configuration 2, Méthode 2

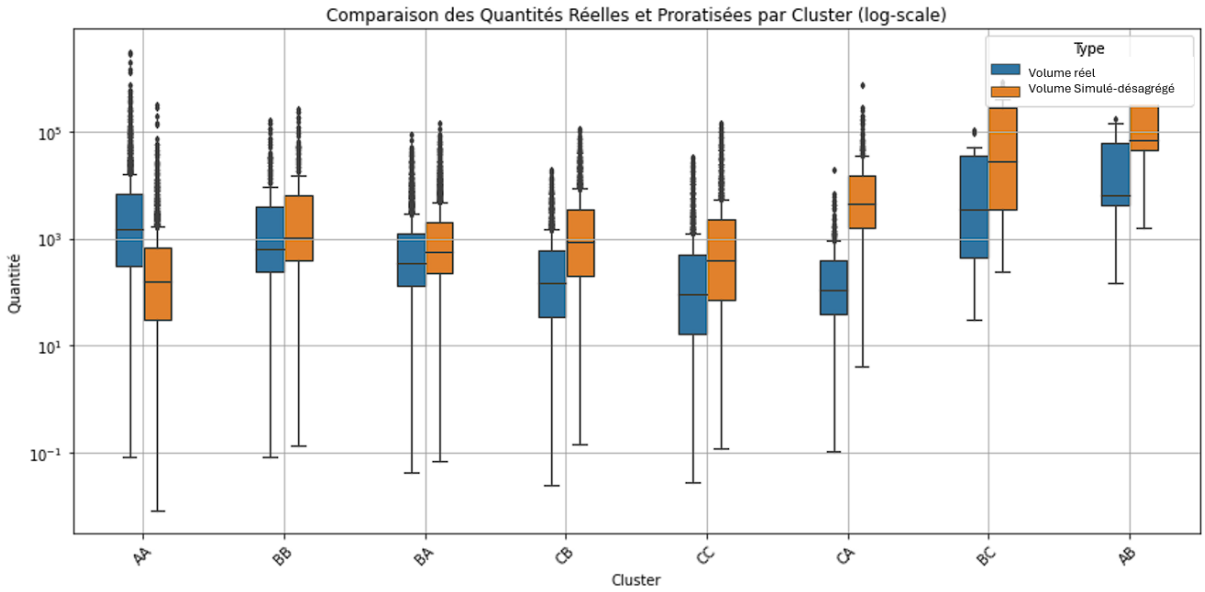


Figure 6.6 Boxplots des volumes livrés réels et désagrégés par cluster (log-scale) – Configuration 2, Méthode 2

L'échec de cette méthode s'explique par une incompatibilité fondamentale. La Méthode 2 est rigide : elle tente d'appliquer une seule loi normale (une moyenne, un écart-type) à un cluster

qui, bien qu'homogène au niveau macro, est statistiquement hétérogène au niveau micro (mélange de produits "constants", "variables", etc.). Tenter de moyenner des comportements de commande si différents est un non-sens statistique qui génère une volatilité forte.

6.1.2 Configuration 3 : Clustering K-means (9 clusters)

La configuration 3 repose sur une segmentation K-means. Contrairement à l'ABC, elle regroupe les produits sur la base de leurs caractéristiques micro (moyenne par commande, écart-type par commande, etc.). Elle crée donc des clusters statistiquement homogènes au niveau du comportement de commande.

Méthode 1 : Lois probabilistes par segment

Paradoxalement, l'application de la Méthode 1 (ajustée) à ces clusters homogènes aboutit à des résultats peu performants. Les indicateurs révèlent une forte variabilité :

- Moyenne : 1.577 ;
- Médiane : 1.213 ;
- Écart-type : 1.885 ;
- Q1 : 0.523 ; Q3 : 1.822 ;
- Minimum : 0.000075 ; Maximum : 6.566.

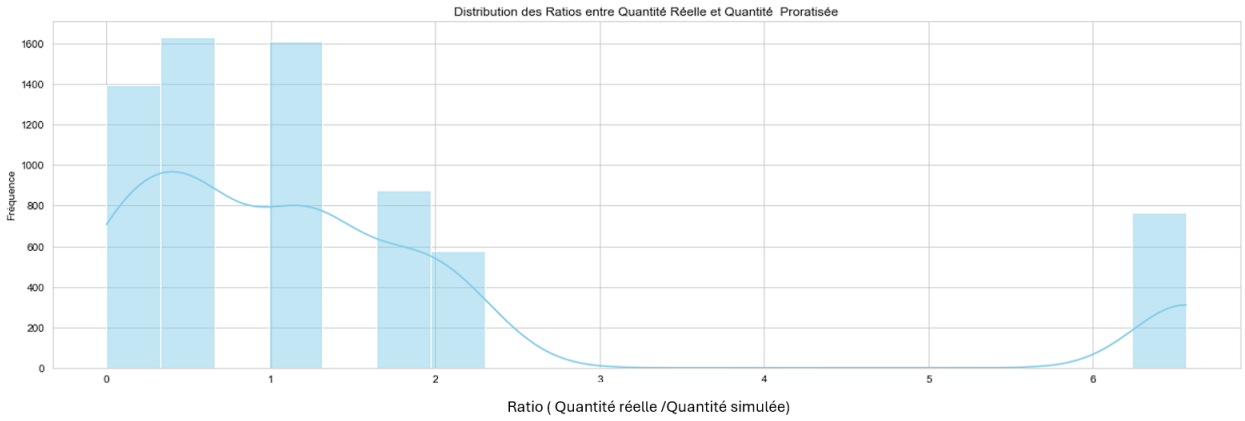


Figure 6.7 Distribution des ratios (Volumes livrés réels / volumes livrés désagrégés) – Configuration 3, Méthode 1

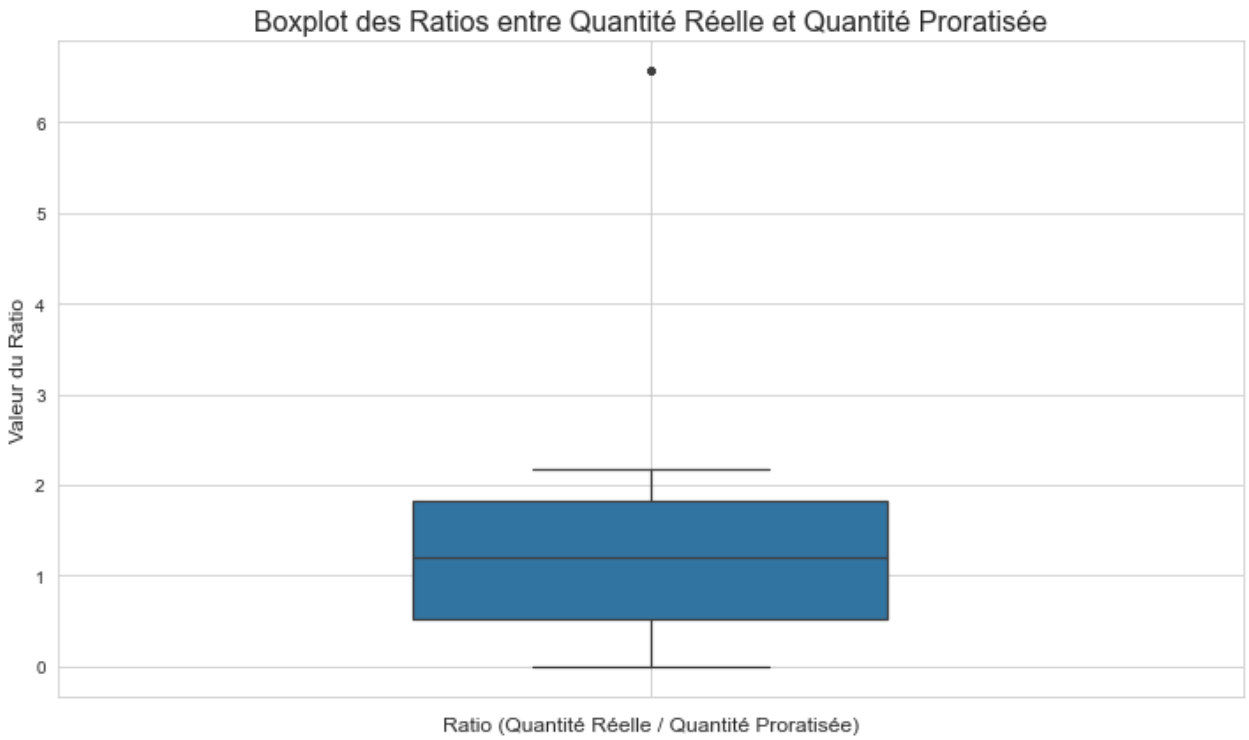


Figure 6.8 Boxplot des ratios (Volumes livrés réels / volumes livrés désagrégés) – Configuration 3, Méthode 1

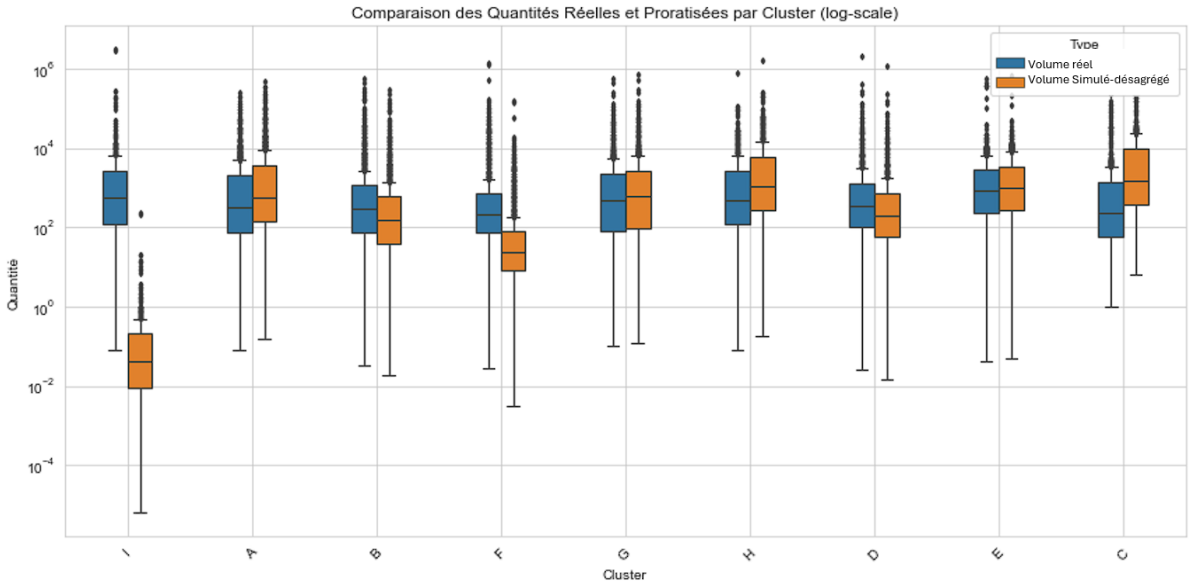


Figure 6.9 Boxplots des volumes livrés réels et désagrégés par cluster (log-scale) – Configuration 3, Méthode 1

Cette contre-performance, c'est-à-dire l'échec du modèle à produire des résultats précis et stables malgré l'homogénéité apparente des groupes, s'explique probablement par un sur-apprentissage (overfitting). Face à un groupe de produits déjà très similaires, la Méthode 1 (ajustée) tente de modéliser les variations aléatoires mineures (le bruit) de chaque produit comme s'il s'agissait d'un comportement distinct. En "sur-interprétant" le bruit, elle réintroduit une volatilité et des biais inutiles.

Méthode 2 : Loi normale tronquée

L'application de la loi normale tronquée (rigide) dans cette structure K-means (homogène) conduit à un résultat opposé et particulièrement satisfaisant. Les indicateurs sont les suivants :

- Moyenne : 1.008 ;
- Médiane : 0.825 ;
- Écart-type : 0.768 ;
- Q1 : 0.515 ; Q3 : 1.181 ;
- Minimum : 0.0001 ; Maximum : 2.309.

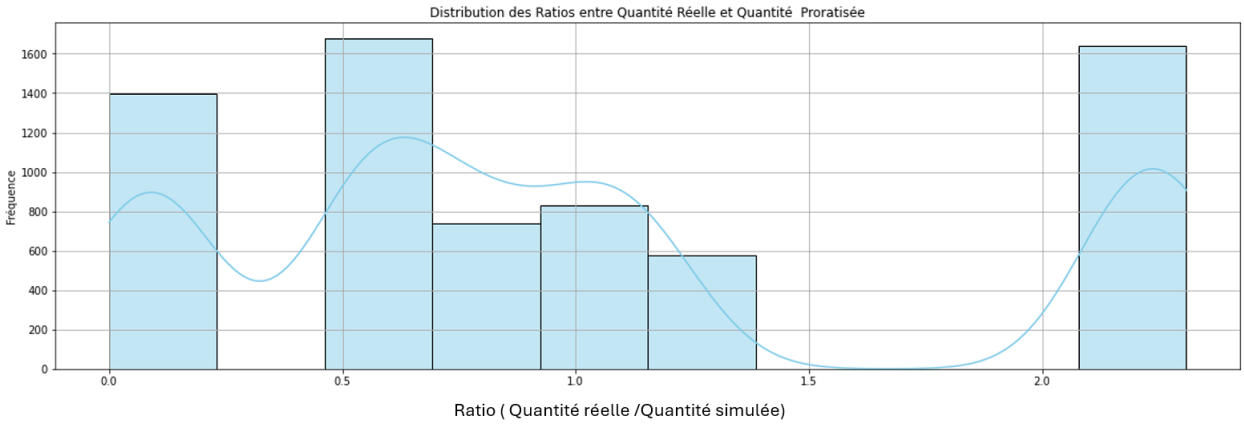


Figure 6.10 Distribution des ratios (Volumes livrés réels / volumes livrés désagrégés) – Configuration 3, Méthode 2

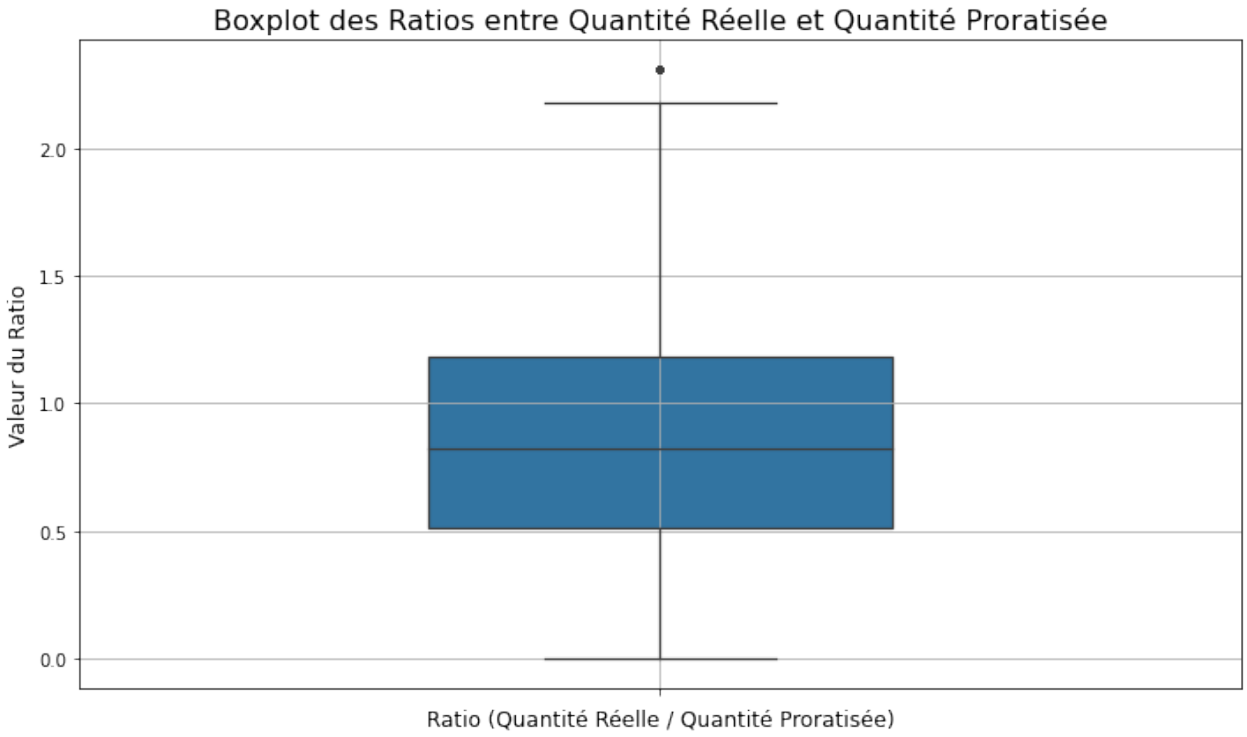


Figure 6.11 Boxplot des ratios (Volumes livrés réels / volumes livrés désagrégés) – Configuration 3, Méthode 2

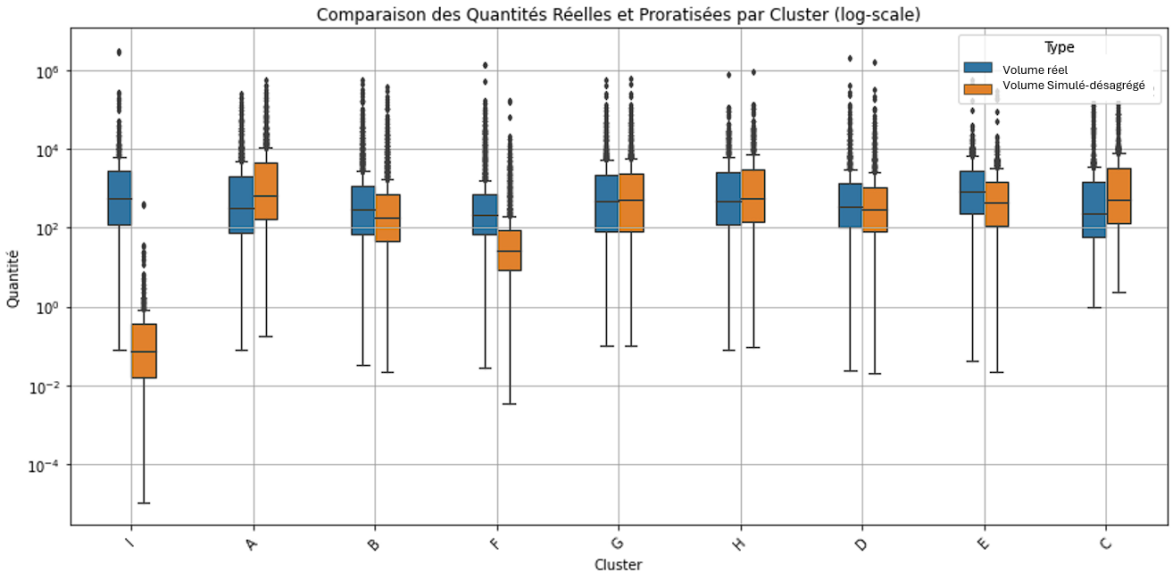


Figure 6.12 Boxplots des volumes livrés réels et désagrégés par cluster (log-scale) – Configuration 3, Méthode 2

La moyenne du ratio (1.008) est quasiment parfaite. Cette performance s'explique par une adéquation parfaite : la Méthode 2 est l'outil statistiquement valide pour modéliser un groupe déjà homogène au niveau micro. Elle capture la tendance centrale du cluster K-means sans sur-interpréter le bruit, ce qui mène à une précision et une stabilité optimales.

6.2 Discussion et recommandations

L'analyse met en évidence une dépendance forte entre la méthode d'agrégation (macro vs micro) et la méthode de génération de variables aléatoires (ajustée vs rigide). Une méthode efficace dans une structure donnée peut devenir inopérante dans une autre.

Cette inversion de performance s'explique par la nature fondamentale de chaque agrégation :

— Pour la Configuration 2 (ABC croisée) : L'analyse ABC crée des clusters homogènes au niveau macro (importance managériale) mais qui restent hétérogènes au niveau micro (comportement de commande).
La Méthode 1 (Probabiliste) excelle ici car sa flexibilité lui permet de respecter et de modéliser cette diversité interne.
La Méthode 2 (Normale Tronquée) échoue car sa rigidité tente d'appliquer une loi unique à un mélange de comportements, ce qui est statistiquement inconsistante.

— Pour la Configuration 3 (K-means) : Le clustering K-means crée des groupes homogènes au niveau micro (comportement de commande).
La Méthode 2 (Normale Tronquée) est ici la plus performante car sa rigidité est parfaitement adaptée. Elle capture la tendance centrale de ce groupe déjà homogène, ce qui est une opération statistiquement valide.
La Méthode 1 (Probabiliste) est moins efficace car sa flexibilité tend à "sur-apprendre" (overfit) le bruit aléatoire au sein du groupe.

L'expérimentation démontre donc que la méthode de génération de variables aléatoires doit être en adéquation avec la nature de l'agrégation : une méthode ajustée (M1) pour des clusters hétérogènes (ABC), et une méthode centralisée (M2) pour des clusters homogènes (K-means).

Tableau 6.2 Synthèse comparative des ratios (Réel / Simulé) – Meilleures méthodes par configuration

Indicateur	Config. 1 (Baseline)	Config. 2 (ABC + Méthode 1)	Config. 3 (K-means + Méthode 2)
Moyenne	0.966	0.863	1.008
Médiane	0.962	0.907	0.825
Écart-type	0.740	0.077	0.768
Q1	0.791	0.816	0.515
Q3	1.114	0.949	1.181
Max	53.287	0.949	2.309

La configuration 1 (Baseline) sert de référence. Elle présente une bonne précision en moyenne (0.966) mais souffre d'une volatilité forte (écart-type 0.74, maximum à 53.29), liée à l'absence d'agrégation et donc à la sensibilité aux cas extrêmes.

La configuration 2 (ABC croisée avec lois probabilistes) est la plus stable de toutes : l'écart-type (0.077) et la plage des ratios (de 0.674 à 0.949) traduisent une régularité exemplaire. En revanche, elle tend à sous-estimer les volumes d'environ 14 % (moyenne à 0.863), ce qui limite sa précision absolue mais pas sa cohérence interne.

La configuration 3 (K-means avec loi normale tronquée) obtient le meilleur compromis global : la moyenne (1.008) traduit une précision quasi parfaite, et la variabilité (écart-type 0.768, maximum 2.309) reste contenue, bien inférieure à celle de la Baseline. Cette configuration combine donc précision, flexibilité et robustesse.

Pour un objectif de robustesse et de facilité de pilotage, la configuration 2 constitue un choix particulièrement sûr. Elle offre une excellente prévisibilité, un comportement stable et une

interprétation intuitive des classes ABC.

En revanche, pour un objectif de performance prédictive et d'équilibre entre précision et variabilité, la configuration 3 s'impose comme la meilleure option. Elle exploite efficacement la richesse statistique des données sans sacrifier la stabilité.

La configuration 1, bien qu'utile comme point de comparaison, reste peu opérationnelle en raison de sa sensibilité excessive aux valeurs extrêmes.

La configuration 3 (Clustering K-means associé à une loi normale tronquée) se distingue comme la solution la plus performante dans une perspective d'agrégation optimale. Elle permet de simplifier la structure tout en améliorant simultanément la fidélité statistique et la stabilité du modèle. La configuration 2 demeure toutefois une alternative solide dans des contextes où la régularité et la lisibilité priment sur la précision absolue.

6.3 Conclusion

Ce chapitre a exploré les conséquences de la réduction de la granularité structurelle du modèle, en agrégeant des centaines de produits en un petit nombre de clusters. L'objectif était de simplifier le modèle tout en préservant sa validité. L'expérimentation a démontré que cette simplification n'est pas un simple compromis, mais qu'elle dépend de la granularité informationnelle capturée par la méthode d'agrégation.

L'analyse révèle une interaction fondamentale : le succès de la réduction structurelle (moins de produits) est conditionné par l'adéquation entre la logique de l'agrégation et la méthode de génération de variables aléatoires.

En conclusion, la réduction de la granularité structurelle est une stratégie viable, mais son succès n'est pas automatique. La configuration K-means (Config. 3) s'est imposée comme l'approche supérieure, car sa préservation de l'homogénéité informationnelle a permis d'obtenir le meilleur équilibre entre simplification et fidélité au réel. La configuration ABC (Config. 2) demeure une alternative robuste lorsque la granularité informationnelle recherchée est d'ordre managérial et que la stabilité prime sur la précision absolue.

CHAPITRE 7 CONCLUSION ET RECOMMENDATIONS

L'objectif principal de ce travail était d'étudier le niveau de granularité des produits à adopter dans un modèle de simulation d'une chaîne d'approvisionnement, afin d'obtenir un compromis entre la qualité de la représentation, l'efficacité computationnelle et la simplicité structurelle du modèle.
Il s'agissait de tester différents degrés de détail pour restituer fidèlement les dynamiques opérationnelles sans alourdir inutilement la simulation, garantissant ainsi des performances rapides et des résultats fiables pour l'aide à la décision opérationnelle en gestion de chaîne d'approvisionnement.

Dans la configuration de base (Configuration 1), la simulation a permis de modéliser le processus de manière détaillée. Les résultats ont montré une bonne précision moyenne (ratio Total des Volumes Livrés réel/Total des Volumes Livrés simulé de 0.966), démontrant la capacité du modèle à reproduire la moyenne de la demande. Cependant, cette configuration a révélé une volatilité forte (écart-type de 0.740 et des maximums très élevés), la rendant peu fiable pour des analyses opérationnelles. Cette première étape a confirmé la pertinence de l'agrégation pour réduire la complexité et maîtriser la variabilité.

Les configurations d'agrégation ont ensuite permis d'explorer différentes stratégies de réduction de granularité, en testant deux méthodes de génération de volume (Méthode 1 : Distributions ajustées ; Méthode 2 : loi normale tronquée) :

- La configuration 2 (agrégation ABC croisée), en utilisant la méthode des Distributions ajustées (Méthode 1), a montré la meilleure stabilité de toutes les approches. Elle affiche une dispersion des ratios extrêmement faible (écart-type de 0.077) et un comportement très prévisible. Cependant, elle introduit un biais de sous-estimation notable (ratio moyen de 0.863). L'utilisation de la loi normale tronquée (Méthode 2) sur cette structure s'est révélée statistiquement invalide, générant une volatilité extrême.
- La configuration 3 (clustering K-means) a révélé une interaction cruciale. L'approche par Distributions ajustées (Méthode 1) a échoué, générant une forte dispersion (écart-type de 1.885) due au sur-apprentissage. En revanche, l'approche par loi normale tronquée (Méthode 2) s'est révélée être la combinaison optimale : elle atteint une précision quasi parfaite (ratio moyen de 1.008) tout en maîtrisant la variabilité (écart-type de 0.768), une performance bien supérieure à la configuration de base.

Ainsi, l'analyse comparée des configurations montre que le niveau de granularité optimal

n'est pas unique, mais dépend de l'adéquation entre la logique d'agrégation et la méthode de génération de volume (flexible ou rigide). La configuration 3 (K-means + loi normale tronquée) offre le meilleur équilibre global entre précision, stabilité et simplification. La configuration 2 (ABC + Distributions ajustées) représente le choix de la robustesse maximale, au détriment d'un biais sur la précision.

Les points forts du modèle résident dans sa robustesse, sa flexibilité d'adaptation et sa capacité à concilier précision statistique et pertinence. De plus, les comparaisons de configurations ont permis d'identifier des métriques claires (ratio moyen, écart-type) pour mesurer l'impact de la granularité et guider les choix futurs en modélisation.

Les limites principales concernent les hypothèses simplificatrices, notamment l'absence de saisonnalité, de gestion dynamique et de contraintes de capacité détaillées (notamment le transport et la production). Enfin, la sensibilité du modèle K-means au choix des variables et du nombre de clusters constitue également une source potentielle d'instabilité.

Les perspectives de ce travail s'orientent vers des améliorations tant méthodologiques qu'opérationnelles :

- Introduire des facteurs dynamiques (saisonnalité, promotions, contraintes de production et de transport) pour accroître le réalisme du modèle.
- Explorer des approches hybrides associant les avantages de la segmentation ABC croisée (lisibilité managériale) et des méthodes non supervisées (précision statistique) afin d'optimiser le compromis entre stabilité et finesse d'analyse.
- Implémenter une boucle d'optimisation continue pilotée par les indicateurs de performance (ratio moyen, écart-type) pour ajuster les stratégies de réapprovisionnement.
- Élargir le cadre du modèle à des systèmes multi-niveaux de la chaîne d'approvisionnement, intégrant la planification collaborative et le transport partagé.

En conclusion, ce projet a permis de démontrer qu'un modèle de simulation peut atteindre un excellent équilibre entre précision et simplicité à condition de maîtriser le niveau de granularité intégré et de lui associer la bonne méthode de génération statistique. La configuration 3 (K-means avec loi normale tronquée) représente le compromis idéal pour des usages prédictifs, offrant la meilleure précision tout en maîtrisant la volatilité. La configuration 2 (ABC croisée avec Distributions ajustées) demeure une alternative de grande valeur lorsque la stabilité, la prévisibilité et la lisibilité managériale sont les objectifs prioritaires. Le modèle dans son ensemble offre ainsi une base robuste pour la simulation, la planification et la prise de décision stratégique dans un contexte logistique complexe.

RÉFÉRENCES

[1] PwC, “Digital trends in supply chain survey 2023,” 2023, consulté en mai 2025. [En ligne]. Disponible: https://www.pwc.com/us/en/services/consulting/business-transformation/digital-supply-chain-survey.html

[2] S. Inc., “Supply chain complexity: How to manage and simplify it,” 2025, consulté en mai 2025. [En ligne]. Disponible: https://syntacticsinc.com/news-articles-cat/supply-chain-complexity/

[3] D. Ivanov, A. Dolgui et B. Sokolov, “The impact of digital technology and industry 4.0 on the ripple effect and supply chain risk analytics,” *International Journal of Production Research*, vol. 57, n°. 3, p. 829–846, 2019.

[4] A. Johnson, B. Smith et C. Lee, “The impact of predictive analytics on supply chain and inventory management,” *Journal of Business Logistics*, vol. 43, n°. 3, p. 234–250, 2022.

[5] T.-M. Choi et J. H. Lambert, “Supply chain risk management in the era of big data,” *Production and Operations Management*, vol. 26, n°. 10, p. 1647–1654, 2017.

[6] C. M. Macal et M. J. North, “Introductory tutorial: Agent-based modeling and simulation,” dans *Proceedings of the 2014 Winter Simulation Conference*, 2014, p. 6–20.

[7] F. Tao, H. Zhang, A. Liu et A. Nee, “Digital twin in industry: State of the art,” *IEEE Transactions on Industrial Informatics*, vol. 15, n°. 4, p. 2405–2415, 2019.

[8] F. Lee, H. Zhang et J. Wang, “Digital twins in industry: Operational enhancement through real-time simulation,” *IEEE Transactions on Industrial Informatics*, vol. 16, n°. 4, p. 2568–2577, 2020.

[9] M. Schönemann, M. Kück et S. Wenzel, “Digital twin applications for production logistics: A literature review,” *Procedia CIRP*, vol. 93, p. 253–258, 2020.

[10] K. Park, Y. Son et S. Noh, “The architectural framework of a cyber physical logistics system for digital twin based supply chain control,” *International Journal of Production Research*, vol. 59, n°. 19, p. 5721–5742, 2021.

[11] L. Monostori, “Cyber physical production systems: roots from manufacturing science and technology,” *Automatisierungstechnik*, vol. 63, n°. 10, p. 870–887, 2015.

[12] J. Tavčar et I. Horváth, “A review of the principles of designing smart cyber-physical systems for runtime adaptation,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, n°. 5, p. 1620–1630, 2018.

[13] J. Han, M. Kamber et J. Pei, *Data Mining: Concepts and Techniques*. Elsevier, 2011.

[14] L. Cao, “In-depth behavior understanding and use: The behavior informatics approach,” *Information Sciences*, vol. 180, n°. 17, p. 3067–3085, 2010.

[15] L. Cao, P. S. Yu, C. Zhang et Y. Zhao, “Agent mining: The synergy of agents and data mining,” *Artificial Intelligence*, vol. 173, n°. 5-6, p. 650–664, 2009.

[16] D. Zhang, J. Shen, X. Wang et X. Zeng, “Data-driven agent-based modeling in supply chain management: A review and case study,” *Procedia Computer Science*, vol. 134, p. 310–317, 2018.

[17] A. Maier, C. Eckert et P. Clarkson, “Model granularity in engineering design - concepts and framework,” *Research in Engineering Design*, vol. 28, n°. 3, p. 197–220, 2017.

[18] H. A. Simon, *The Architecture of Complexity*. MIT Press, 1962.

[19] H. Kim et J. Park, “Finite element analysis and granularity trade-offs,” *International Journal of Computational Mechanics*, vol. 41, n°. 2, p. 315–332, 2024.

[20] K. Karlsen, B. Dreyer, P. Olsen et E. Elvevoll, “Traceability standards and challenges in the food industry,” *Journal of Food Engineering*, vol. 103, n°. 2, p. 127–134, 2011.

[21] L. Qian, Z. Yang et M. Li, “Supply chain visibility and data aggregation techniques,” *Logistics Research*, vol. 10, n°. 1, p. 87–102, 2017.

[22] B. Alkan et D. Kocaoglu, “Modular supply chain management strategies,” *International Journal of Production Economics*, vol. 231, p. 107920, 2021.

[23] R. Franceschini, S. Van Mierlo et H. Vangheluwe, “Towards adaptive abstraction in agent based simulation,” dans *2019 Winter Simulation Conference (WSC)*. IEEE, 2019, p. 2725–2736.

[24] G. Morvan et G. Kubera, “Temporal coherence in multi-level agent-based models,” *Journal of Artificial Societies and Social Simulation*, vol. 20, n°. 2, p. 5–22, 2017.

[25] O. Baqueiro, Y. J. Wang, P. McBurney et F. Coenen, “Integrating data mining and agent based modeling and simulation,” dans *Industrial Conference on Data Mining*. Springer, 2009, p. 220–231.

[26] R. Bemthuis, S. Leemans et W. van der Aalst, “A crisp-dm methodology for assessing agent-based simulation models using process mining,” *International Journal of Data Science and Analytics*, vol. 15, n°. 3, p. 210–227, 2024.

[27] F. Tao, Q. Qi, A. Liu et A. Y. C. Nee, “Digital twins and cyber–physical systems: Industry 4.0 and beyond,” *Engineering*, vol. 5, n°. 5, p. 653–661, 2019.

[28] S. Sowe, S. Nakajima et Y. Oyama, “A conceptual framework for digital twins of multi-agent systems,” *arXiv preprint arXiv:2402.02630*, 2024. [En ligne]. Disponible: https://www.researchgate.net/publication/391178685_A_Conceptual_Framework_for_Digital_Twins_of_Multi-Agent_Systems

[29] S. Mariani, M. Picone et A. Ricci, “About digital twins, agents, and multiagent systems: a cross-fertilisation journey,” *arXiv preprint arXiv:2206.03253*, 2022. [En ligne]. Disponible: https://arxiv.org/abs/2206.03253

[30] M. Ritou, F. Belkadi, Z. Yahouni, C. Da Cunha, F. Laroche et B. Furet, “Knowledge-based multi-level aggregation for decision aid in the machining industry,” *CIRP Annals*, vol. 68, n°. 1, p. 475–478, 2019.

[31] S. Wogrin, “Time series aggregation for optimization: One-size-fits-all?” *IEEE Transactions on Smart Grid*, vol. 14, n°. 3, p. 2489–2492, 2023.

[32] I. David et D. Bork, “Infonomics of autonomous digital twins,” p. 563–578, 2024.

[33] Y. Ngatilah, N. Rahmawati, C. Pujiastuti, I. Porwati et A. Hutagalung, “Inventory control system using distribution requirement planning (drp)(case study: Food company),” dans *Journal of Physics: Conference Series*, vol. 1569, n°. 3. IOP Publishing, 2020, p. 032005.

[34] I. Rizkya, K. Syahputri, R. Sari, I. Siregar, M. Tambunan *et al.*, “Drp: Joint requirement planning in distribution centre and manufacturing,” dans *IOP Conference Series: Materials Science and Engineering*, vol. 434, n°. 1. IOP Publishing, 2018, p. 012243.

[35] E. Domingos, J.-M. Frayret et M. B. Ali, “Critical analysis of digital supply chain twin concepts and technological requirements,” dans *IISE Annual Conference. Proceedings*. Institute of Industrial and Systems Engineers (IISE), 2024, p. 1–6.

[36] R. J. Hyndman et G. Athanasopoulos, *Forecasting: Principles and Practice*, 3e éd. OTexts, 2018, accessed July 2025. [En ligne]. Disponible: https://otexts.com/fpp3/

[37] E. A. Silver, D. F. Pyke et R. Peterson, *Inventory Management and Production Planning and Scheduling*, 3e éd. John Wiley & Sons, 1998.

ANNEXE A ANALOGIE DES PARAMÈTRES DE DISTRIBUTION : SCIPY (PYTHON) ET ANYLOGIC (JAVA)

Tableau A.1 Analogie des Paramètres de Distribution : SciPy (Python) et AnyLogic (Java)

Distribution	**Fonction AnyLogic**	**Paramètres AnyLogic**	**Paramètres SciPy**	**Formule de Conversion**
Normale	`normal(`μ`, `σ`)`	μ, σ	loc, scale	μ = loc σ = scale
Normale Tronquée	`normal(min, max, shift, stretch)`	min, max, shift, stretch	a, b, loc, scale	shift = loc stretch = scale min = loc + (a × scale) max = loc + (b × scale)
Log-normale	`lognormal(`μ`, `σ`, min)`	μ, σ, min	s, loc, scale	μ = ln(scale) σ = s min = loc
Gamma	`gamma(`α`, `β`, min)`	α, β, min	a, loc, scale	α = a β = scale min = loc
Weibull	`weibull(`α`, `β`, min)`	α, β, min	c, loc, scale	α = c β = scale min = loc
Beta	`beta(`α`, `β`, min, max)`	α, β, min, max	a, b, loc, scale	α = a, β = b min = loc max = loc + scale
Laplace	`laplace(`μ`, `β`)`	μ, β	loc, scale	μ = loc β = scale
Pareto	`pareto(`α`, min)`	α, min	b, loc, scale	α = b min = scale *(Ajouter loc au résultat de la fonction)*