Titre: Title:	Valorisation des données de maintenance extraites d'un ERP : développement d'un modèle utilisant l'intelligence artificielle pour la prédiction des consommations
Auteur: Author:	Lucas Franck Frederic Adam
Date:	2024
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Adam, L. F. F. (2024). Valorisation des données de maintenance extraites d'un ERP : développement d'un modèle utilisant l'intelligence artificielle pour la prédiction des consommations [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/58315/

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Valorisation des données de maintenance extraites d'un ERP : développement d'un modèle utilisant l'intelligence artificielle pour la prédiction des consommations

LUCAS FRANCK FREDERIC ADAM

Département de mathématiques et de génie industriel

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

Génie industriel

Avril 2024

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

Valorisation des données de maintenance extraites d'un ERP : développement d'un modèle utilisant l'intelligence artificielle pour la prédiction des consommations

présenté par **Lucas Franck Frederic ADAM**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

a été dûment accepté par le jury d'examen constitué de :

Martin TRÉPANIER, président
Robert PELLERIN , membre et directeur de recherche
Bruno AGARD , membre et codirecteur de recherche
Camélia DADOUCHI , membre

DÉDICACE

À Guillaume, à Sam,

REMERCIEMENTS

Je tiens à remercier tout particulièrement Robert PELLERIN et Bruno AGARD, en leurs qualités de directeur et co-directeur de recherche, pour leurs conseils avisés tout au long de ma maîtrise. Leur accueil et leur sympathie m'ont permis d'effectuer cette étude dans les meilleures conditions possibles dans les locaux de Polytechnique Montréal.

J'aimerais aussi remercier Étienne LE PIRONNEC pour son expertise, sa patience et son suivi régulier de l'évolution de ce projet. Je tiens aussi à remercier M. Timothy AYOUB, à l'origine de ce projet de recherche, Mme Marie-Annie FAUBERT pour sa gestion remarquable des étudiants ainsi que toutes les personnes que j'ai pu rencontrer dans le cadre de ce projet. La confiance et l'entrain qu'ils ont manifesté à l'égard de mon travail ont été particulièrement valorisants.

Enfin, j'aimerais remercier mes collègues de bureau et mes colocataires. Leur soutien indéfectible et leurs encouragements ont été une source inestimable de motivation tout au long de mon projet de maîtrise. Je suis sincèrement reconnaissant d'avoir eu la chance de partager cette aventure avec eux. Pour finir, j'aimerais remercier chaleureusement ma famille, sans qui rien de cela n'aurait été possible.

RÉSUMÉ

La gestion efficace des stocks de pièces de rechange demeure un défi majeur pour de nombreuses entreprises, même avec l'utilisation répandue des systèmes ERP (Enterprise Resource Planning), des SCM (Supply Chain Management) et des technologies de pointe de l'industrie 4.0. Cela s'explique notamment par la nature très irrégulière de la demande en pièces de rechange pour la maintenance. Contrairement aux produits finis dont la demande peut être plus prévisible, les pièces de rechange sont souvent sujettes à des variations importantes et à des intervalles irréguliers entre les consommations successives.

Les classifications traditionnelles basées sur la littérature ont souvent du mal à capturer efficacement ces fluctuations et à fournir une gestion précise des stocks de pièces de rechange. En effet, les périodes prolongées entre les consommations et les variations significatives de quantités rendent difficile l'établissement de profils de demande fiables ainsi que des prévisions efficaces.

Dans cette optique, ce mémoire de maîtrise propose une approche basée sur les données pour la gestion des stocks de pièces de rechange. Cette méthodologie met en lumière l'ensemble du processus de gestion, depuis la collecte et l'analyse des données jusqu'à la gestion des exceptions, le regroupement des articles du catalogue en profils de demande, la prévision de la demande par profil et l'évaluation des performances. Notre attention s’est aussi portée sur l’optimisation des paramètres des méthodes de prévision par profil de demande avec des capacités de calcul limitées grâce à une utilisation intelligente des ressources informatiques à notre disposition. Comme la méthode de prévision miracle ne semble pas être identifiée dans la littérature, notre approche cherche à utiliser les spécificités de chacune d’entre elles en les affectant à des profils de demande différents. Pour cela, nous avons comparé les performances de 16 méthodes de prévision rencontrées dans la littérature scientifique.

Ce qui distingue cette approche, c'est son mariage judicieux entre l'intelligence artificielle et l'intelligence humaine. En combinant algorithmes et l’expérience des experts à la STM, nous avons pu développer une méthodologie robuste et adaptable, capable de fournir des prévisions de demande plus précises et une gestion simplifiée des stocks de pièces de rechange pour la maintenance. Nous avons exploré la piste de la gestion par profil de demande définis par un clustering Kmeans dans le plan ADI/CV^2 pour réduire le nombre de paramètres de gestion dans les opérations. Aussi, un indicateur de risque de rupture est proposé pour permettre aux planificateurs

d’être informé de prévisions potentiellement insuffisantes. De cette façon, ils ne doivent plus gérer des dizaines de paramètres pour des milliers de pièces mais pour quelques dizaines de groupes tout en gagnant de la visibilité sur les problèmes à venir.

Nos résultats montrent que l'application de cette méthodologie avec les données de consommation d’un partenaire industriel peut entraîner une réduction significative de l'erreur absolue moyenne des modèles de prévision, allant jusqu'à 69 % par rapport aux pratiques actuelles basées sur la littérature. La méthode Autoregressive Integrated Moving Average (ARIMA) s’est illustrée comme étant la plus performante sur de nombreux profils de gestion. Cette amélioration notable de la précision de la prévision peut se traduire par des économies substantielles pour les entreprises, en réduisant les coûts liés aux surplus de stocks et aux pénuries de pièces de rechange tout en simplifiant la gestion de leurs inventaires.

ABSTRACT

Effective spare parts inventory management remains a major challenge for many companies, even with the widespread use of ERP (Enterprise Resource Planning) systems, SCM (Supply Chain Management) and the cutting-edge technologies of Industry 4.0. One reason for this is the highly irregular nature of demand for maintenance spare parts. Unlike finished products, whose demand can be more predictable, spare parts are often subject to wide variations and irregular intervals between successive consumption.

Traditional literature-based classifications often struggle to effectively capture these fluctuations and provide accurate spare parts inventory management. Indeed, long periods between consumption and significant variations in quantities make it difficult to establish reliable demand profiles and effective forecasts.

This master's thesis proposes an innovative and comprehensive data-driven approach to spare parts inventory management. This methodology encompasses the entire management process, from data collection and analysis to exception handling, grouping of catalog items into demand profiles, demand forecasting by profile and performance evaluation. We also focused on optimizing the parameters of demand profile forecasting methods with limited computing capacity, by making intelligent use of the computing resources at our disposal. As the miracle forecasting method does not seem to be identified in the literature, our approach seeks to use the specificities of each of them by assigning them to different demand profiles. To this end, we have compared the performance of 16 forecasting methods found in the scientific literature.

What sets this approach apart is its judicious marriage of artificial and human intelligence. By combining algorithms and the experience of experts at STM, we have been able to develop a robust and adaptable methodology capable of providing more accurate demand forecasts and simplified management of spare parts inventories for maintenance. We explored the possibility of managing demand profiles defined by Kmeans clustering in the ADI/CV^2 plan, to reduce the number of management parameters in operations. In addition, a rupture risk indicator is proposed to enable planners to be informed of potentially insufficient forecasts. In this way, they no longer have to manage dozens of parameters for thousands of parts, but for a few dozen groups, while gaining visibility on future problems.

Our results show that applying this methodology with consumption data from an industrial partner can lead to a significant reduction in the average absolute error of forecasting models, by up to 69% compared with current literature-based practices. The Autoregressive Integrated Moving Average (ARIMA) method has proven to be the most efficient on many management profiles. This significant improvement in forecast accuracy can translate into substantial savings for companies, by reducing the costs associated with excess inventory and spare parts shortages, while simplifying inventory management.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES SIGLES ET ABRÉVIATIONS

ANOVA	Méthode d'analyse de variance
ARRSES	Lissage exponentiel simple avec adaptation du coefficient de lissage
AW	Méthode de Winter additive
BM	Méthode de base ou « baseline method »
BootstrapDLP	Variante de bootstrap proposée par Pennings et al. (2017)
BP-NN	Réseau de neurones à rétropropagation
CGSSA-BP	Réseau de neurones à rétropropagation simulant le comportement de groupe des moineaux
DES	Lissage exponentiel double
DLP	Méthode de Pennings et al. (2017)
Dlt	Modèle « Delay time »
DRM	Design Research Methodology
DT	Arbre de décision
EE	Données de l'ingénierie
EMR	Entretien matériel roulant
ES	Lissage exponentiel simple
EWMA	Moyenne mobile pondérée avec tendance
FF-NN	Réseau de neurones à propagation avant
GEV	Loi de l'extrémum généralisé
KNN	Méthode des K plus proches voisins
LR	Régression linéaire
LHD	Échantillonnage par hypercube latin
MA	Moyenne mobile

MSBA	Modification de SBA par Babai et al. (2019)
MTBR	Temps moyen entre chaque remplacement
MW	Méthode de Winter multiplicative
SRM	Modèle de régression saisonnière
NB	Approche de Bayes naïve
RBF-NN	Réseau de neurones à fonction d'activation radiale
RDA	Réparation des autobus
SBA	Modification de la méthode de Croston par Syntetos et Boylan (2005)
SSA-BP	Réseau de neurones à rétropropagation avec algorithme d'optimisation
SVM	Support Vecteur Machine
SVMip	Support Vecteur Machine avec intégration de la chaîne logistique
TAES	Lissage exponentiel à tendance
TSB	Modification de la méthode de Croston par Teunter et al. (2011)
VOTE	Méthode ensembliste utilisée par Sareminia et Amini (2023)
WCDR	Moyenne de demande sur une période en fonction de l'activité
WMA	Moyenne mobile pondérée
WRDF	Régression sur les heures de vol
ZF	Méthode « Zero forecast »
2S	Méthode « Two-step »

LISTE DES ANNEXES

CHAPITRE 1 INTRODUCTION

Pour de nombreuses entreprises, la gestion efficiente des inventaires reste un enjeu de taille malgré la multiplication des outils de traitement de l'information et d'aide à la décision. La plupart du temps, des quantités de données conséquentes sont collectées, mais elles ne sont que très rarement utilisées efficacement. Le problème réside principalement dans le fait que les données ne sont pas structurées correctement et que certaines données soient manquantes ou incompatibles entre elles (Van der Auweraer, Boute, et Syntetos, 2019). Aussi, quand le problème ne vient pas de la qualité des données, il provient du type même des données. Les consommations de pièces détachées et les données de maintenance plus généralement possèdent un comportement très particulier qu'il est compliqué de modéliser avec les méthodes de prévisions conventionnelles. Contrairement à des entreprises de vente, les entreprises assurant la maintenance ne peuvent pas retirer de leurs inventaires des articles qui ne seraient pas rentables, car ils pourraient provoquer des bris de service. La méthode la plus simple pour maximiser la disponibilité des pièces serait de stocker chaque article, mais de nombreux obstacles apparaissent : certaines pièces sont très coûteuses, d'autres font face à l'obsolescence, d'autres sont volumineuses, etc. Beaucoup d'entreprises doivent donc gérer des catalogues contenant des milliers de références et il devient impossible de stocker sans compter. Ces entreprises doivent assurer un service tout en minimisant le coût des inventaires.

La complexité de la gestion des pièces détachées s'est accentuée avec les décisions d'externalisation et de fragmentation des entreprises. Les chaînes logistiques voient intervenir de nombreux acteurs, des locaux de stockage variés et des clients parfois très éloignés de leurs entrepôts, ce qui peut créer ou allonger les délais d'approvisionnement. Ainsi, les quantités transitant d'une entité à une autre peuvent être aussi hétérogènes que les délais de transit. La position dans la chaîne logistique influence significativement les caractéristiques de la demande : dans une structure multi-échelon, un entrepôt central n'aura pas le même volume d'opérations que les ateliers en bout de chaîne d'approvisionnement. Il y a alors une distinction entre les différents secteurs d'affaires et une fragmentation au sein même de ces secteurs avec différents ateliers et entrepôts aux rôles bien différents.

Dans le cas de la maintenance, l'action de réparer ou de changer simplement une pièce détachée ajoute un degré de complexité supplémentaire et est un exemple supplémentaire de la nature

spécifique de la prévision. Par exemple, pour un organisme qui propose un service de transport en commun, la gestion des flottes est spécifique : plusieurs générations d'équipements doivent être entretenues en parallèle puisque la durée de vies des équipements est très importante. Elles peuvent s'étendre sur plusieurs dizaines d'années.

La contribution apportée par ce mémoire est la proposition d'une méthode de collecte des données, d'analyse, de prévision et d'évaluation des résultats applicable pour un organisme de transport en commun. La méthode proposée vise à développer une meilleure compréhension des caractéristiques intrinsèques de la demande en pièces détachées au travers de l'utilisation des méthodes récentes de fouilles de données. Une approche complète, s'étendant de la phase de collecte des données à l'évaluation pratique des résultats en passant par le choix des méthodes de prévision et de leurs paramètres a le potentiel de rendre plus efficace et plus facile la gestion des inventaires en entreprise.

La contribution sera développée dans une optique de réduction du nombre de paramètres de gestion utilisés en prévision de la demande en pièces détachées. L'étude sera réalisée à partir de données réelles, en collaboration étroite avec un partenaire du secteur des transports publics.

Ce mémoire est organisé en 7 chapitres. Nous débuterons par présenter un état de l'art des méthodes de prévision actuelles pour la demande intermittente et grumeleuse au sein du chapitre 2. Cette étape va permettre d'identifier les pistes de recherches que nous pourrons explorer. Le troisième chapitre va détailler les objectifs de recherche et la méthodologie permettant d'y répondre. Le quatrième chapitre est consacré à l'analyse du cas d'étude et à l'exploration des données brutes. Le chapitre 5 détaille l'approche, les modèles et métriques utilisés ainsi que les étapes intermédiaires qui nous ont amenées à ces choix. Le chapitre 6 expose les résultats et étudie plus généralement la performance de la démarche comme ses limitations. Finalement, le chapitre 7 nous permet de rappeler le contexte général de l'étude pour mettre en lumière les bénéfices de notre approche au cours de la conclusion.

CHAPITRE 2 REVUE DE LITTÉRATURE

Dans ce chapitre, nous allons présenter un état de l'art sur le sujet. Ainsi, ce chapitre commencera par identifier et expliquer les termes clés qui seront utilisés tout au long de ce mémoire. Ensuite, le protocole de définition de la stratégie de recherche sera explicité suivi de la présentation des résultats des articles découlant du processus de sélection. Finalement, nous allons conclure ce chapitre sur une revue critique afin de mettre en lumière les faiblesses et les limites des contributions précédentes pour identifier les opportunités de recherche.

2.1 Définition des concepts de base

Cette section introduit les concepts fondamentaux reliés au sujet de ce mémoire. La compréhension des termes clés et de leurs interactions permettra de mieux cerner la démarche de recherche.

2.1.1 Prévision

D'après la définition du dictionnaire Le Robert, la prévision se définit comme l'« Opinion formée par le raisonnement sur les choses futures » (*prévision - Définitions, synonymes, conjugaison, exemples | Dico en ligne Le Robert*). La prévision joue un rôle majeur dans l'industrie puisqu'elle permet une anticipation pour répondre au mieux aux besoins futurs malgré les contraintes d'approvisionnement liées à la chaîne logistique. Dans notre cas, la prévision portera sur la consommation de pièces détachées.

2.1.2 Distinction entre pièce détachée et pièce réparable

Dans le domaine de la maintenance, plusieurs pratiques sont mises en œuvre afin de remettre en service des systèmes. Chacun d'entre eux contient différentes pièces dont certaines peuvent être défectueuses. Il est alors possible de les remplacer par des pièces détachées. Lorsqu'on parle de pièce de rechange, il est important de remarquer à quels groupes, assemblages ou sous-assemblages, elles appartiennent puisque cela peut entraîner des remplacements individuels ou non. Par exemple, si un mécanisme de fermeture de porte est dysfonctionnel, une porte de véhicule peut être changée comme un tout afin de limiter le coût du travail horaire et/ou le temps d'immobilisation du véhicule. La même pièce défectueuse peut tout aussi bien être démontée, inspectée et le remplacement d'une sous composante défectueuse peut être effectué. Le choix de réaliser une opération ou l'autre dépend des compétences des opérateurs, des coûts, de la criticité ou de la complexité de l'opération. En résumé, la réparation peut consister soit à remplacer une ou

plusieurs pièces défectueuses par des pièces en bon état, soit à réparer une pièce endommagée si cela est possible. Par exemple, on peut réusiner ou ressouder une partie de la pièce. Bien que la pièce réparée ne soit pas totalement neuve, elle peut être remise en service. Le choix entre la réparation et le remplacement dépend de divers facteurs et varie selon chaque situation.

2.1.3 Série temporelle

Lorsqu'on cherche à faire de la prévision de demande, on ne se concentre pas seulement sur une anticipation de la quantité demandée, mais aussi sur la date à laquelle on va avoir besoin de cette quantité. Pour cela, il existe plusieurs modèles qui proposent de représenter l'historique de demande sous la forme de série temporelle. Ainsi, on indique à chaque période, la quantité concernée. Les horizons d'agrégation de ces séries temporelles peuvent être multiples, allant de l'heure jusqu'à l'année et même au-delà. Cependant, il est important de mettre le niveau d'agrégation en relation avec le cas étudié pour qu'il ait un sens. Il est clair qu'une agrégation de plusieurs années peut être tout aussi vide de sens qu'une agrégation à la minute suivant les conditions d'observation.

2.1.4 Profil de demande

Il existe différentes classifications qui permettent de distinguer différents profils de demande. Dans leur contribution, Syntetos et Boylan (2005) indiquent qu'il est commun dans l'industrie de segmenter un catalogue en fonction de la structure de la demande pour utiliser des méthodes de prévision de façon plus ciblée et ainsi optimiser le processus de gestion des inventaires. La classification proposée est souvent reprise dans la littérature. Elle définit la demande en fonction de quatre grandes catégories qui sont définies comme :

- Lisse : la quantité demandée et les délais de commandes varient peu ;
- Intermittente : la quantité demandée varie peu, mais les délais fluctuent fortement ;
- Erratique : la quantité demandée fluctue fortement, mais les délais varient peu ; et
- Grumeleuse : la quantité et les délais varient fortement d'une demande à l'autre.

2.1.5 Taux de service

À la différence d'un vendeur qui se contenterait de vendre des produits à des clients, une entreprise d'exploitation d'équipements doit être en mesure de maintenir ces équipements en état de fonctionnement. Elle ne peut pas simplement retirer une pièce de son catalogue parce qu'elle n'est pas rentable. Ainsi, une entreprise de maintenance doit conserver en inventaire les pièces nécessaires à la réparation de ses actifs pendant de nombreuses années dans le but d'assurer un service. C'est alors qu'intervient la mesure du taux de service. Ce terme se définit par la « Proportion de la demande des clients qui pourra, en moyenne, être satisfaite sans rupture de stock ni manquant, ni retard » (Office québécois de la langue française, 1981). Dans notre cas, appliqué à la maintenance d'une flotte de véhicules, « les clients » seront les opérateurs responsables de la maintenance. Cet indicateur est important puisqu'il permet de faire le lien entre l'inventaire et l'influence qu'il a sur les opérations. Cependant, il n'est pas suffisant lorsqu'il est utilisé seul. Un inventaire contenant beaucoup de pièces inutiles peut offrir un taux de service identique à un inventaire ne contenant que les pièces détachées nécessaires. Combiné avec l'indicateur du coût d'inventaire, ces deux mesures sont les plus couramment utilisées dans la gestion des stocks pour l'industrie (Teunter et al., 2010).

2.2 Définition et résultats de la stratégie de recherche

Cette section vise à préciser notre stratégie de recherche bibliographique et analyser les articles retenus.

2.2.1 Définition de la stratégie

Notre recherche bibliographique repose sur la base de données scientifiques Scopus. Elle est la base de littérature interdisciplinaire couvrant le plus de thèmes scientifiques en offrant plus de 94 millions de documents (Elsevier, 2024) . En utilisant une telle base de données, nous cherchons à couvrir le sujet de recherche du mieux possible et à obtenir le maximum d'articles publiés dans la littérature. Notre étude prendra la forme d'une revue narrative. De cette façon, on peut avoir une vue générale du thème puis identifier les problèmes comme les lacunes de la littérature avant d'y apporter une contribution.

La première étape de notre démarche a été d'identifier le besoin d'information auquel va tenter de répondre cette présente revue. On recherche quelles sont les méthodes de prévision de la demande

adaptées à une demande de type grumeleux pour une application industrielle de maintenance d'actifs.

À la suite de cette identification des besoins d'information, nous avons élaboré un plan de concepts. Il cherche à répertorier les mots et synonymes permettant de parler du thème abordé et à séparer les termes dans l'objectif de faciliter la création de la requête utilisée pour consulter la base de données. Les résultats de ce plan de concept sont présentés dans le tableau 2.1.

Tableau 2.1 : Plan de concept

Pièces détachées	**Demande grumeleuse**	**Prévision**
Spare part Spare parts Spares	Lumpy Lumps Intermittent Intermittency Slow-moving Slow-movement	Forecast Forecasts Prediction Predict Predictive Predicting

On a ainsi pu déterminer la requête suivante: *TITLE-ABS-KEY (spare* AND (lump* OR intermitten* OR slow-mov*) AND (predict* OR forecast*))*.

Cette requête, interrogée en date du 29 juin 2023, a permis de dégager 128 articles. La lecture des titres et résumés des 15 premiers articles a mis en évidence que les premiers résultats étaient cohérents avec le sujet de cette recherche, mais qu'ils ne traitaient pas spécifiquement du domaine de la maintenance. Certains utilisaient simplement les pièces détachées comme un exemple de domaine d'application, mais ne considéraient pas les contraintes de la maintenance. Nous avons donc choisi de rajouter le mot clé « Maintenance » pour tenter de nous concentrer encore plus sur ce cas d'étude. La nouvelle requête sera donc: *TITLE-ABS-KEY (spare* AND(lump* OR intermitten* OR slow-mov*) AND (predict* OR forecast*) AND maintenance)*.

Cette nouvelle requête a identifié 28 articles. La lecture des titres et résumés de ces 28 articles nous a conforté dans le choix de cette requête. Un document a été retiré d'office puisqu'il s'agissait d'un doublon. Elle est apparue dans deux revues différentes la même année et seul un tiret variait dans

le titre, ce qui a pu induire en erreur la base de données. Après une lecture plus complète des résumés, un deuxième article a été supprimé puisqu'il n'étudiait que le processus d'élaboration d'une classification sur la criticité des pièces. Aussi, un article présentant de nombreuses figures sans les expliquer, ne respectant aucun formalisme apparent et ne contenant que très peu de texte a été écarté. De plus, un résultat correspondant aux 434 papiers présentés dans une conférence a été écarté pour son manque de pertinence dans notre cas d'étude. Finalement, 24 articles sont retenus pour la suite de notre étude en raison de leur pertinence. Sur ces 24 articles, 4 d'entre eux sont restés inaccessibles malgré une demande d'accès auprès des auteurs. Ce processus de sélection est représenté par la figure 2.1. Finalement, le tableau 2 présente les références, titres et sources des articles retenus. Il nous paraissait important de mentionner les sources des articles pour avoir une première idée de la solidité des informations. Une revue reconnue n'est pas une garantie inconditionnelle de la qualité des articles proposés, mais donne tout de même plus de crédit à l'article.

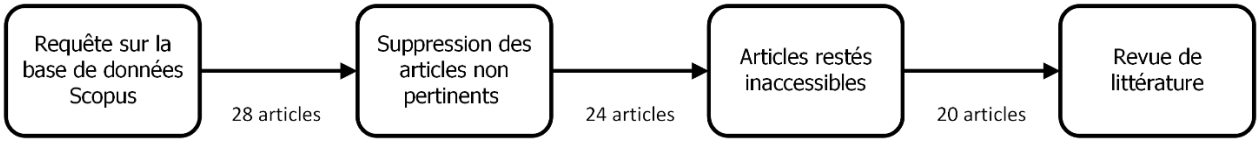

Figure 2.1 : Processus de sélection des articles

Tableau 2.2 : Articles sélectionnés

Référence	**Titre**	**Source**	**Type de contribution**
Guskiva et Dyntar (2023)	Speeding up past stock movement simulation in sporadic demand inventory control	International journal of simulation modeling	Méthode de prévision
Sareminia et Amini (2023	A reliable and ensemble forecasting model for slow-moving and repairable spare parts: Data mining approach	Computers in industry	Méthode de prévision
Wang et Hart (2023)	An Optimal Maintenance Spare Parts Prediction Model and Its Complex Applications	Annual Reliability and Maintainability Symposium (RAMS)	Méthode de prévision
Wei et al. (2022)	Prediction of Maintenance Spare Parts Based on BP Neural Network Optimized by Improved SparrowSearch Algorithm	5th International conference on information systems and computer aided education	Méthode de prévision
Babaveisi et al. (2022)	Integrated demand forecasting and planning model for repairable spare part: an empirical investigation	International journal of production research	Modèle de simulation
Şahin et al. (2021)	Inventory Cost Minimization of Spare Parts in Aviation Industry	Transportation research Procedia	Méthode de prévision
Zhu et al. (2020)	Spare parts inventory control based on maintenance planning	Reliability engineering and system safety	Modèle de simulation
Van der Auweraer et Boute (2019)	Forecasting Spare Part Demand using Service Maintenance Information	International journal of production economics	Modèle de simulation et méthode de prévision
Van der Auweraer et al. (2019)	Forecasting Spare Part Demand with Installed Base Information: a Review	International journal of forecasting	Revue de littérature

Tableau 2.2 : Articles sélectionnés (suite et fin)

Bergman et al. (2017)	A Bayesian approach to demand forecasting for new equipment programs	Robotics and computer - Integrated Manufacturing	Méthode de prévision
Pennings et al. (2017)	Exploiting elapsed time for managing intermittent demand for spare parts	European journal of operational research	Modèle de simulation
Zhang et Guan (2016)	Inventory Management for Spare Parts on engineering Machinery	2015 Chinese Automation Congress	Revue de littérature
Frazzon et al. (2014)	Simulation model concept for evaluating spare parts supply chain planning methods	12th iEEE International conference on Industrial informatics	Modèle de simulation
Verhagen et Curran (2014)	Stochastic Forecasting of Lumpydistributed Aircraft Spare Parts Demand	Advances in Transdisciplinary Engineering (conference)	Modèle de simulation
Romeijnders et al. (2012)	A two-step method for forecasting spare parts demand using information on component repairs	European journal of operational research	Méthode de prévision
do Rego et de Mesquita (2011)	Controle de estoque de peças de reposição: uma revisão da literatura	Producao	Revue de littérature
Wang et Syntetos (2011)	Spare parts demand: Linking forecasting to equipment maintenance	Transportation research part E 47	Modèle de simulation
Chiou et al. (2004)	Grey prediction model for forecasting the planning material of equipment spare parts in navy of Taiwan	Sixth Biannual World Automation Congress	Méthode de prévision
Ghobbar et Friend (2003)	Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model	Computers and operations research	Méthode de prévision
Ghobbar et Friend (2002)	Sources of intermittent demand for aircraft spare parts within airline operations	Journal of Air Transport Management	Méthode de prévision

Pour présenter les résultats de cette revue de littérature, nous débuterons par couvrir le contexte des différentes études. Ensuite, nous allons explorer le contenu de chaque contribution en nous attardant sur la description des modèles utilisés ainsi que les méthodes d'évaluation des résultats. Enfin, nous allons utiliser trois contributions offrant un état de l'art pour couvrir plus largement le sujet de l'étude. Dans chacune des étapes mentionnées, nous prendrons soin de définir les concepts développés et de parcourir les articles pour expliciter leurs objectifs et identifier les pistes laissées inexplorées ou les limites de chaque étude dans une revue critique.

2.2.2 Contexte des contributions retenues

Un inventaire qui ne permet pas de répondre aux besoins est responsable de la majorité du temps de non-production ou plus généralement, de non-activité (Bergman et al., 2017). Aussi, le manque de pièces détachées peut causer un sentiment de frustration très important pour les opérateurs (Sareminia et Amini, 2023).

Dans leur article, Huskova et Dyntar (2003) soulignent qu'une bonne gestion de l'inventaire de pièces détachées est un enjeu capital pour la plupart des entreprises de production et de services, mais que la complexité de la tâche est d'autant plus importante que la maintenance présente de nombreux cas de consommation sporadiques et grumeleux. Ils indiquent aussi que ce problème de type de consommation est aussi présent dans beaucoup de domaines différents, comme la production automobile, l'aviation ou encore la santé. D'après les mêmes auteurs, la plupart des politiques de gestion d'inventaire reposent sur un seuil de commande. Dès que le niveau d'inventaire passe sous ce niveau, une commande est automatiquement passée auprès du fournisseur. En théorie, ce seuil est défini afin de pouvoir répondre aux demandes sur une période donnée, définie par la consommation estimée ainsi que par le temps de livraison moyen.

Dans un objectif de profitabilité, il est là encore crucial de trouver un équilibre entre la valeur d'inventaire et le coût potentiel d'une rupture de stock. Certaines pièces peuvent par exemple avoir une valeur financière très importante et être aussi très critiques. Ce qui impliquerait qu'une rupture aurait un impact sur la rentabilité encore plus important qu'une pièce dormante (Şahin et al., 2022). Aussi, d'après cette même contribution, même si le problème de gestion d'inventaire est courant dans l'industrie, il est encore plus important dans l'aviation puisque les coûts de stockages tout comme les coûts liés aux ruptures de stock sont très élevés. Au cours de cette revue de littérature,

on peut remarquer que 9 des 20 papiers sélectionnés apportent une contribution grâce à des études liées au domaine de l'aviation. Seulement 2 contributions, celles de Babaveisi et al. (2022) et Frazzon et al. (2014) s'intéressent au cas de l'industrie pétrolière. L'équipement industriel au sens plus large du terme a retenu l'attention de Huskova et Dyntar (2023), Sareminia et Amini (2023), Van der Auweraer et Boute (2019), Pennings et al. (2017), Zhang et Guan (2016). L'application des méthodes de prévision dans un contexte de défense, que ce soit pour de l'aviation militaire ou des systèmes d'armes, a été étudié par Wei et al. (2022), Chiou et al. (2004) et simplement mentionné comme un domaine compatible avec l'étude de Pennings et al. (2017). Finalement, deux papiers retenus ne mentionnent pas de domaine industriel particulier. C'est le cas de Van der Auweraer et al. (2019) puisque les auteurs conduisent une revue de littérature sur les différentes méthodes proposées et se veulent le plus généralistes possible. Quant à eux, Bergman et al. (2017) ne détaillent aucun domaine industriel. On sait simplement que leur étude se déroule sur un cas d'étude « réaliste ».

Lorsqu'il est question du cas d'étude, seulement deux documents proposent des simulations alors que le reste des papiers se base sur l'étude de données réelles. (Wang et Syntetos, 2011; Huskova et Dyntar, 2023).

Comme il a pu être mentionné dans la définition des termes de ce mémoire, il est commun dans l'industrie d'utiliser une catégorisation de la demande afin de pouvoir cibler plus efficacement les méthodes qui vont le mieux performer sur certains groupes de pièces (Syntetos et al., 2005). Dans cette contribution, les auteurs se basent sur une classification de Williams (1984) qui choisissait de séparer les pièces en cinq catégories : lisses, irrégulières, slow-moving, erratique et hautement erratique pour la simplifier en seulement quatre catégories : lisses, erratiques, intermittentes et grumeleuses. Cette classification se base sur le calcul du « ADI » ou « Average demand interval », ce qui se traduit par le temps moyen entre deux demandes non nulles et le « CV^2 » qui se traduit par le coefficient de variation de la quantité commandée. Le ADI se calcule comme le rapport entre le nombre de commandes total sur le nombre de commandes non nulles et le CV^2 comme le rapport entre l'écart type de la demande, noté σ et la moyenne de la demande, notée μ, le tout élevé au carré. Les formules mathématiques de ces quantités sont illustrées par les équations 2.1 et 2.2.

$$\text{ADI} = \frac{\text{Nombre total de périodes}}{\text{Nombre de demandes non nulles}} \tag{2.1}$$

$$CV^2 = \left(\frac{\sigma}{\mu}\right)^2 \quad (2.2)$$

Les valeurs seuils fixées par Syntetos et al. (2005) sont $CV^2 = 0,49$ et $ADI = 1,32$. Alors que l'étude de Williams (1984) faisait apparaître des seuils choisis arbitrairement, Syntetos et al. (2005) se servent des mesures d'erreurs moyennes aux carrés, couramment appelés « MSE » ou « Mean squared error » dans la littérature. Leur étude permet de définir mathématiquement les limites pour lesquelles les méthodes de Croston ou sa révision appelée « SBA » pour « Syntetos and boylan method » (Syntetos et Boylan, 2005) vont offrir les meilleurs résultats.

Au-delà de la représentation mathématique de ces catégories par des limites fixées, Şahin et al. (2021) indiquent que les catégories retenues ont des comportements spécifiques :

- La catégorie Lisse ($ADI < 1,32$ et $CV^2 < 0,49$) est synonyme de faibles occurrences de demandes nulles et de faibles variations des quantités commandées ;
- La catégorie Intermittente ($ADI \geq 1,32$ et $CV^2 < 0,49$) est synonyme de fortes occurrences de demandes nulles et de faibles variations des quantités commandées ;
- La catégorie Erratique ($ADI < 1,32$ et $CV^2 \geq 0,49$) est synonyme de faibles occurrences de demandes nulles et de fortes variations des quantités commandées ; et
- La catégorie Grumeleuse ($ADI \geq 1,32$ et $CV^2 \geq 0,49$) est synonyme de fortes occurrences de demandes nulles et de fortes variations des quantités commandées.

Ces comportements et leur répartition dans le plan ADI et CV^2 sont illustrés par la figure 2.2.

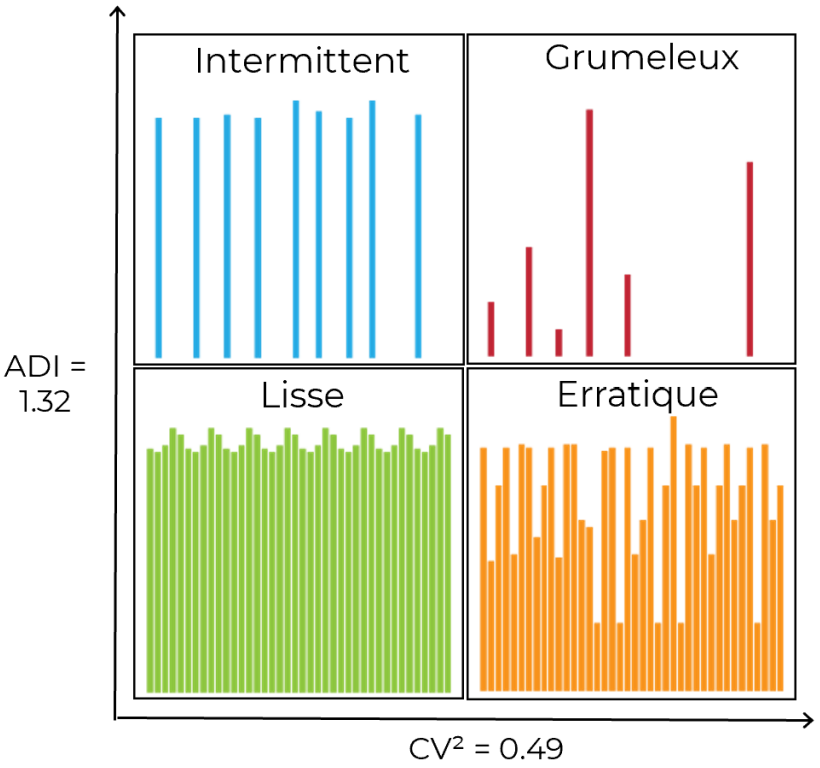

Figure 2.2 : Classification ADI CV^2

L'utilisation de la classification des pièces dans les articles retenus pour ce mémoire est ambivalente. Les contributions de Ghobbar et Friend (2002), Ghobbar et Friend (2003), Verhagen et Curran (2014) ainsi que Babaveisi et al. (2022) utilisent cette classification ADI et CV^2 pour étudier respectivement des pièces au profil grumeleux (3 premières contributions) ou intermittent. Certains auteurs comme Huskova et Dyntar (2023) ou Şahin et al. (2021) l'utilisent pour comparer les résultats entre différents profils de demande. D'une façon similaire, Romeijnders et al. (2012) et Zhu et al. (2020) définissent une classification basée sur l'ADI pour séparer les pièces détachées en 3 catégories. Les données de 10 ans de consommation sont utilisées pour estimer et ajuster les paramètres des modèles de prévision. Les auteurs appellent cette période la « période d'initialisation ». Ces catégories sont :

- « Very slow moving » ou « VSM » : Pièces avec 1 à 5 mois de demande non nulle sur la période d'initialisation ;
- « Slow moving » ou « SM » : Pièces avec 6 à 20 mois de demande non nulle sur la période d'initialisation ; et
- « Fast moving » ou « FM » : Pièces avec 21 à 84 mois de demande non nulle sur la période d'initialisation.

Cette classification découle d'une étude comparative préliminaire qui a permis de mettre en évidence que cette dimension temporelle était la plus cohérente pour l'étude réalisée dans la suite de l'article (Romeijnders et al., 2012). Le détail de cette étude sera donné dans la partie suivante de ce mémoire.

D'autre part, certains auteurs utilisent des classifications plus spécifiques à leur cas d'étude pour définir des profils de demande. Quand Bergman et al. (2017) utilisent une classification beaucoup plus sommaire entre « Low demand » pour des pièces avec moins de trois demandes annuelles et « High demand » pour des pièces avec plus de 3 demandes annuelles. Pennings et al. (2017) explorent une approche encore peu étudiée. Ils proposent une séparation entre les pièces possédant une corrélation croisée positive et une autre négative entre les quantités commandées et les intervalles inter-demandes. Chose que ne font pas les études se basant sur la classification de Syntetos et al. (2005) puisqu'elles supposent une indépendance entre quantité commandée et intervalle de demande. L'approche définie par Van der Auweraer et Boute (2019) utilise la taille moyenne de la demande ou « ADS », le coefficient de variation de la demande ou « CV » ainsi que le pourcentage moyen de semaines sans aucune demande ou « APZ ». Sareminia et Amini (2023) ont quant à eux utilisé la somme des quantités et la variance des demandes pour définir quatre catégories :

- « Low consumption and Turbulent » ou « LCT » pour des pièces dont la somme des consommations est plus faible que le seuil de consommation et la variance des consommations est plus élevée que le seuil de variance ;
- « High consumption and Turbulent » ou « HCT » pour des pièces dont la somme des consommations est plus élevée que le seuil de consommation et la variance des consommations est plus élevée que le seuil de variance ;

- « Low consumption and stable » ou « LCS » pour des pièces dont la somme des consommations est plus faible que le seuil et la variance des consommations est plus élevée que le seuil ; et
- « High consumption and stable » ou « HCS » pour des pièces dont la somme des consommations est plus élevée que le seuil et la variance des consommations est plus faible que le seuil.

Les seuils de quantités et de variance sont fixés en fonction de la distribution des données, mais les auteurs ne donnent pas de justification supplémentaire.

Enfin, cinq contributions, celles de Wang et Hart (2023), Wei et al. (2022), Frazzon et al. (2014), Wang et Syntetos (2011) et Chiou et al. (2004) n'utilisent aucune classification et se contentent simplement de qualifier les demandes des pièces étudiées d'intermittentes ou de grumeleuses.

Dans cette section, nous avons pu détailler le contexte des études retenues afin de mettre en lumière le domaine d'application, si le cas d'étude était réel ou simulé, le profil de demande étudié, la classification mise en œuvre ainsi que le nombre de pièces pour prendre la mesure de l'efficacité des approches sur des inventaires de tailles différentes. Les résultats que nous avons évoqués sont résumés dans le tableau 2.3.

Tableau 2.3 : Contexte des contributions

Référence	**Domaine industriel**	**Cas d'étude**	**Profil de demande considéré**	**Classification utilisée**
Guskiva et Dyntar (2023)	Équipements de production	Simulé	Lisse, Intermittent, Erratique, Grumeleux	ADI, CV^2
Sareminia et Amini (2023)	Industrie de l'acier	Réel	LCT, HCT, LCS, HCS	Somme des quantités, Variance des quantités
Wang et Hart (2023)	Défense et aviation	Réel	Non spécifié	Non spécifié
Wei et al. (2022)	Défense	Réel	Intermittent, Sporadique	Non spécifié

Tableau 2.3 : Contexte des contributions (suite et fin)

Babaveisi et al. (2022)	Industrie pétrolière	Réel	Intermittent	ADI, CV^2
Şahin et al. (2021)	Aviation	Réel	Intermittent, Erratique, Grumeleux	ADI, CV^2
Zhu et al. (2020)	Aviation	Réel	VSM, SM, FM	ADI
Van der Auweraer et Boute (2019)	Équipementier industriel	Simulé	Intermittent	ADS, CV, APZ
Bergman et al. (2017)	Non spécifié	Simulé	Grumeleux	Low demand, High demand
Pennings et al. (2017)	Équipements de production, électronique, aviation, défense	Réel	Intermittent	Nombre, taille, variance, intervalles inter-demande
Frazzon et al. (2014)	Industrie pétrolière	Simulé	Intermittent, Grumeleux	Non spécifié
Verhagen et Curran (2014)	Aviation	Réel	Grumeleux	ADI, CV^2
Romeijnders et al. (2012)	Aviation	Réel	VSM, SM, FM	ADI
Wang et Syntetos (2011)	Aviation	Simulé	Intermittent	Non spécifié
Chiou et al. (2004)	Défense et aviation	Réel	Intermittent	Non spécifié
Ghobbar et Friend (2003)	Aviation	Réel	Lisse, Intermittent, Erratique, Grumeleux	ADI, CV^2
Ghobbar et Friend (2002)	Aviation	Réel	Lisse, Intermittent, Erratique, Grumeleux	ADI, CV^2

2.2.3 Description des modèles

Dans la dernière section, nous nous sommes attelés à une mise en contexte des publications. Dans celle-ci, nous allons détailler les méthodes mises en œuvre pour la prévision de demandes ainsi que les métriques utilisées pour évaluer leurs performances respectives.

Pour répondre à des demandes aux profils très différents, de nombreuses méthodes ont fait leur apparition dans la littérature scientifique. Qu'elles soient plutôt orientées sur les statistiques ou qu'elles répondent des méthodes plus récentes de fouille de données, il n'y a pas de méthode unique permettant de répondre précisément à tous les cas de figure (Sareminia et Amini, 2023). Ces deux auteurs ont proposé un modèle en deux phases. La première repose sur l'utilisation de différentes méthodes comme celle des K Plus Proches Voisins (KNN), les arbres de décision (DT), un réseau de neurones tri-couche à propagation avant (3LFF-NN), un réseau de neurones à fonction d'activation radiale (RBF-NN) et l'approche de Bayes naïve (NB). Une méthode ensembliste nommée VOTE permet de combiner ces résultats pour prédire la prochaine date à laquelle une pièce sera demandée ainsi que sa quantité. En parallèle, ces méthodes sont combinées avec celle du Support Vecteur Machine (SVM) et une régression logistique (LR) puis VOTE dans le but de prédire si lors de la prochaine maintenance, la pièce réparable sera fonctionnelle ou mise au rebut ainsi que la prochaine date à laquelle la pièce devra être jetée. Les sorties de ces 5 méthodes pour prédire la consommation et de ces 7 méthodes pour prédire les résultats de la réparation sont ensuite évaluées par des mesures conventionnelles sur les séries statistiques : l'erreur moyenne absolue (MAE), l'erreur moyenne absolue en pourcentage (MAPE), l'erreur moyenne absolue en pourcentage symétrique (sMAPE) et la racine de l'erreur moyenne au carré (RMSE). Une matrice de confusion contenant les métriques conventionnelles est utilisée pour évaluer la pertinence des prédictions sur le résultat de la prochaine maintenance. Les auteurs apportent une contribution en définissant un nouvel indicateur appelé MARI, combinant un indice de fiabilité et une MAPE. Enfin, ils avancent que leur méthode permet d'améliorer la précision de la prévision jusqu'à 80% par rapport aux méthodes de prévision usuelles.

La même année, Huskova et Dyntar (2023) proposent un modèle visant à réduire le temps de calcul nécessaire à l'optimisation du seuil de commande. Au cours de leur étude, ils suggèrent qu'il est plus efficace de faire une recherche locale d'optimum en bornant le domaine de recherche plutôt qu'une recherche globale. La borne inférieure est une régression linéaire puisqu'elle aurait

tendance à sous-estimer la demande réelle. La borne supérieure est définie par bootstrapping puisque la méthode aurait tendance à surestimer les besoins réels. La recherche d'optimum se calcule sur une fonction définissant le coût d'inventaire. Les auteurs avancent que leur méthode est applicable à un inventaire réel de taille importante et qu'il serait réaliste de l'intégrer dans un progiciel de gestion d'entreprise (ERP).

De leur côté, Wang et Hart (2023) modélisent le processus de maintenance comme une chaîne de Markov continue. On peut définir une probabilité qu'une demande arrive dans une file d'attente plutôt qu'une prise en charge instantanée en fonction de l'utilisation des centres de réparation ainsi qu'une probabilité de rupture de stock de certaines pièces détachées. Dans leur étude, les auteurs supposent que les bris de pièces suivent une distribution exponentielle et que pour un nombre de pièces fixe, le nombre de retours pour des pièces non réparable (c.a.d., le nombre de pièces devant être remplacées plutôt que réparées) suit une loi de poisson. Cette contribution vise à introduire un modèle permettant de déterminer le stock optimal en pièces détachées pour assurer une disponibilité maximum et une satisfaction client optimale tout en minimisant le coût des opérations.

L'utilisation et l'optimisation de réseaux de neurones pour la prédiction de demandes en pièces détachées sont aussi étudiées par Wei et al. (2022). La démarche combine plusieurs étapes permettant de préparer les données d'entrée, d'optimiser les poids pour un réseau de neurones à rétropropagation. Dans un premier temps, la séquence d'occurrence des demandes est lissée par modulation digitale et la séquence de demande est définie par agrégation des données historiques sur les quantités. Ensuite, l'architecture d'un réseau de neurones est définie avant d'optimiser les poids et les seuils du réseau en utilisant une méthode récente qui calque les comportements d'antiprédation et de recherche de nourriture des moineaux (CGSSA-BP). Finalement, les résultats sont démodulés pour obtenir la prédiction. Les auteurs comparent leur méthode avec un réseau de neurones à rétropropagation (BP-NN) ainsi qu'un réseau de neurones à rétropropagation avec un algorithme d'optimisation (SSA-BP). Pour cela, ils utilisent les mesures MAE et RMSE définies précédemment ainsi que le coefficient de corrélation de Pearson (R^2). Les deux premières grandeurs d'erreur servent à déterminer l'écart tandis que le R^2 sert à mesurer le degré d'ajustement du modèle.

Quant à eux, Babaveisi et al. (2022) se sont concentrés sur l'utilisation de modèles paramétriques pour optimiser les prédictions tout en minimisant le coût d'inventaire. Leur étude effectue une

comparaison entre les méthodes de Croston, la révision de cette méthode par Syntetos et Boylan (SBA), la modification de SBA par Babai et al. (2018) (MSBA), SVM et SVM avec l'intégration de la chaîne logistique (SVMip). C'est cette dernière méthode qui est proposée par les auteurs. Ils ont défini mathématiquement la fonction de coût pour la chaîne logistique puis les probabilités de mise au rebut et la répartition des pièces dans les différents centres. Les résultats sont évalués sur le coût d'inventaire, l'erreur moyenne relative (SME), l'erreur moyenne absolue en pourcentage (AMAPE) et l'erreur moyenne absolue relative (MASE). En conclusion de leur papier, ils indiquent que leur méthode est particulièrement efficace pour la prévision à court terme, mais que sur des horizons plus longs, les autres méthodes peuvent être utilisées de façons interchangeables puisque leurs résultats sont similaires.

En utilisant des données provenant d'une compagnie aérienne turque, Şahin et al. (2021) proposent une succession d'étapes menant à l'optimisation du facteur de lissage pour les modèles de prédiction adaptés aux demandes intermittentes. Cette contribution utilise le lissage exponentiel (ES) et les méthodes de Croston et SBA précédemment nommées. Si les prévisions ne sont pas plus efficaces en termes de coût d'inventaire qu'une méthode naïve supposant simplement que la prévision de la prochaine période prend la valeur de la dernière demande, alors différentes valeurs du coefficient de lissage sont utilisées pour trouver un optimum. En conclusion de leur étude, les auteurs indiquent que la méthode naïve permet d'obtenir les meilleurs résultats pour 7 pièces sur 8 ayant un profil de demande erratique, le lissage exponentiel est le plus efficace pour des demandes intermittentes et finalement la méthode de Croston est optimale pour les comportements grumeleux.

Afin de prendre du recul pour intégrer les causes de la demande dans la prévision de pièces détachées, Zhu et al. (2020) utilisent les maintenances planifiées comme sources d'ADI (Advance demand information). La démarche sépare les horizons de prévision : sur l'horizon où des maintenances sont planifiées, la distribution de la demande est estimée par les données historiques et une probabilité de remplacement est associée à chaque inspection planifiée suivant la loi de poisson. Pour la prévision dépassant l'horizon de maintenances planifiées, les auteurs utilisent la méthode SBA. Afin de pouvoir évaluer les bénéfices de la méthode proposée, ses résultats sont comparés avec ceux du lissage exponentiel (ES), de la moyenne mobile (MA), de la méthode de Croston, de SBA et de la méthode de prévision de Teunter and Duncan (2009) (TSB). Cette mesure s'effectue avec l'écart moyen absolu (MAD) et la RMSE. D'autre part, les auteurs cherchent aussi

à optimiser une fonction de coût en y intégrant les méthodes de prévision de la demande. De cette manière, ils peuvent trouver un optimum de commande qui prend en compte les demandes, le coût de stockage, le coût de rupture, le coût des demandes urgentes et le coût d'élimination du stock superflu en fin de cycle de vie. En conclusion de leur étude, les auteurs soulignent la diminution des coûts de pénalités comme des coûts de rebut en fin de vie par rapport à la littérature. Ils font aussi remarquer l'intérêt de leur méthode pour les « very slow moving items » puisqu'ils constituent un défi majeur de planification pour des êtres humains.

Cet intérêt pour la fin de vie des stocks n'est pas quelque chose de nouveau. L'étude de Van der Auweraer et Boute (2019) sépare le cycle de vie d'un produit en 3 phases. La phase initiale où il est progressivement mis en service et donc où la quantité de machines installées croît. La phase dite « mature » où le nombre d'installations décroît progressivement et où les demandes en pièces détachées prennent de l'ampleur. Enfin la phase de fin de vie où il n'y a plus de nouvelles installations et où la taille du parc va progressivement diminuer aux côtés de la demande en pièces détachées. Les ventes de nouvelles machines suivent une loi de poisson, la durée de vie de chaque machine suit une distribution exponentielle. La demande en pièces détachées pour chaque machine peut suivre plusieurs lois, évoluant suivant l'âge des machines installées. Cependant les auteurs ont choisi de modéliser cette demande par une loi de Weibull. Cette proposition est comparée avec deux méthodes classiques de la littérature : lissage exponentiel simple (ES) et SBA. Les résultats des prévisions sont évalués avec l'erreur moyenne (ME), la racine de l'erreur moyenne au carré (RMSE) pour différents taux de service ciblés. Deux études sont réalisées, la première correspondant à une prédiction sur les demandes liées à une maintenance corrective et planifiée et l'autre sur la prévision des demandes correctives seules. La méthode proposée, appelée SMI pour « Service Maintenance Information » surpasse les autres méthodes puisqu'elle permet d'atteindre des taux de service ciblés avec le niveau d'inventaire le plus faible.

Quelques années plus tôt, Bergman et al. (2017) s'intéressent au cas des nouvelles installations. C'est-à-dire celles pour lesquelles il n'y a pas de données historiques et où les méthodes conventionnelles d'étude des séries temporelles ne peuvent pas être utilisées. Dans ce cas, il est coutume dans l'industrie de se baser sur les estimations de l'ingénierie. Cependant, cette approche implique de faire des hypothèses sur la base d'informations très incertaines. Les auteurs vont chercher à proposer une solution pour répondre à cette problématique et à déterminer à partir de quand il devient cohérent d'utiliser les méthodes d'études des séries temporelles. La méthode

détaillée dans la contribution repose sur une approche bayésienne, capable de prendre en compte l'évolution des installations et les premières informations de rupture tout en utilisant les données de l'ingénierie. Elle sera comparée à l'approche basée uniquement sur les données de l'ingénierie (EE) ainsi que sur une méthode dite de « Baseline method » (BM). Cette dernière est initialisée sur les données de l'ingénierie puis le lissage exponentiel simple est utilisé et modulé par des facteurs déterminés par les auteurs. Les résultats des trois méthodes sont comparés en utilisant les mesures MAD, MAPE, RMSE et le rapport MAD/Moyenne. Après avoir simulé les demandes pendant 4 ans, les auteurs tirent les conclusions que leur approche et la BM surpassent les estimations de l'ingénierie. Les deux approches les plus performantes ont des résultats similaires sur les années 3 et 4 en ce qui concerne le taux de remplissage d'inventaire et de précision. Sur l'année 2 cependant, si l'approche Bayésienne est plus performante en termes d'inventaire, les autres méthodes fournissent des prédictions de demande plus précises.

La même année, Pennings et al. (2017) proposent un modèle probabiliste pour faire de la prévision de demandes intermittente. Après avoir défini un modèle général, ils vont le spécifier en supposant que le temps inter-demande suit une distribution géométrique. Leur proposition (DLP) va être comparée avec des méthodes classiques d'étude de séries temporelles : ES, Croston, SBA, TSB et SY. Pour cette dernière, seul le nom est défini, mais aucun calcul n'est explicité. Aussi, ils comparent ces méthodes avec une méthode de boostrap et une variante appelée BootstrapDLP, qui approxime la probabilité d'une demande utilisant une distribution empirique conditionnée par le temps écoulé. Les résultats seront mesurés par MASE, erreur absolue géométrique moyenne (GMAE), taux de service et valeur de stock sur cinq jeux de données provenant de compagnies et de domaines différents. La conclusion de l'étude est que la méthode proposée permet de réduire les coûts d'inventaires de 14%, est robuste dans de nombreux cas d'utilisation et est applicable sur le terrain. Cette contribution a permis de s'intéresser aux effets de corrélations entre les quantités commandées et le temps écoulé entre chaque commande.

Frazzon et al. (2014) se concentrent sur la fonction d'approvisionnement et supposent que la demande en pièces détachées suit une loi de poisson pour faire une simulation visant à optimiser les coûts de pénalité et plus généralement le coût du stock. Les auteurs se placent du point de vue fournisseur.

Dans la littérature, les approches paramétriques se basent sur des distributions particulières de la demande. Verhagen et Curran (2014) ont proposé une approche probabiliste basée sur une loi de poisson homogène. Dans le but de s'assurer que leur approche est la plus censée, ils ont utilisé la métrique RMSE pour comparer les résultats de prédiction utilisant des fonctions de densité de probabilités différentes. Ces fonctions de densité sont : la loi de l'extremum généralisé (GEV), la loi de Weibull, loi normale et loi de poisson. Cette dernière permet d'obtenir le RMSD le plus faible en offrant les prévisions les plus proches de la demande réelle. Cependant, sur l'année d'observation, cette méthode n'a pas permis de répondre à un des pics de demandes, correspondant à un mois sur les 12 étudiés.

Deux ans plus tôt, Romeijnders et al. (2012) comparaient les méthodes de Croston, SBA, TSB et la prévision naïve avec leur contribution : la méthode « two-step » (2S). Elle est décrite comme « la seule méthode prenant en compte les informations supplémentaires disponibles ». Afin de faire la prévision sur des assemblages, la méthode fait des prévisions sur les sous-composantes puis elles sont agrégées en tenant compte des nomenclatures des assemblages. Une autre approche, dite de « zero forecast » (ZF) est utilisée à titre de comparaison, mais il est indiqué qu'elle est inutilisable en pratique malgré son efficacité d'après les mesures d'erreurs MSE et MAD. Ces indicateurs sont couplés au RMSE et ME. Prédire une demande nulle à chaque période ne permet pas de gérer un inventaire en maintenant un taux de service élevé, ce qui illustre la nécessité de bien choisir les méthodes de mesures de prévision.

Dans une même optique de prise de recul sur l'origine de la demande, Wang et Syntetos (2014) proposent une approche nommée « Delay Time » (DlT). Les auteurs font remarquer que la maintenance peut être corrective ou préventive. Dans le premier cas, la pièce casse et il faut la changer pour remettre l'actif en service. La maintenance corrective fera donc apparaître des délais inter-demandes variants fortement, mais la quantité demandée sera très souvent unitaire puisque les probabilités de rupture coordonnées sont faibles. Ensuite pour la maintenance préventive, ils indiquent que les délais sont connus en avance, mais que la quantité demandée reste conditionnelle puisqu'il faut inspecter les pièces pour en décider ou non le remplacement. La méthode DT repose sur un processus de rupture en deux temps. Dans le premier temps, un défaut apparaît et dans un second temps, le défaut entraîne une rupture. Si le délai entre l'apparition du défaut et la rupture est suffisant pour atteindre la prochaine date de maintenance, alors elle sera préventive. Sinon elle sera corrective. D'après les hypothèses de l'étude, la demande repose sur une distribution

géométrique et le processus de réparation ou non répond à une épreuve de Bernoulli. En termes d'erreur absolue (AE), cette méthode performe mieux que SBA.

L'idée de différencier les maintenances préventives et correctives apparaît aussi dans les travaux de Chiou et al. (2004). Les auteurs remarquent que dans le cas de la maintenance préventive, il n'est pas toujours nécessaire de conserver des stocks de sécurité puisqu'il est possible de connaître les dates de demande et donc de pouvoir passer des commandes pour que les pièces arrivent à temps. Dans le cas de la maintenance corrective, les stocks de sécurité sont nécessaires, mais les critères sont propres à la structure de l'entreprise, l'obsolescence ou encore les pratiques de gestions mises en place. C'est pourquoi la méthode proposée choisit de ne pas les séparer. Les auteurs la décrivent capable de gérer les cas d'étude avec très peu de données (à partir de 4 occurrences) et d'être plus performantes sur la prévision à court terme que les méthodes de régression, Box-Jenkins (BJ), de lissage simple (ES) et de réseaux de neurones (NN) en termes d'erreur moyenne (ME) ou de racine de l'erreur moyenne au carré (RMSE).

Finalement, dans leurs deux contributions successives, Ghobbar et Friend (2002; 2003) étudient les causes de la génération de la demande grumeleuse en effectuant une analyse de variance (ANOVA) pour déterminer les corrélations entre le type de maintenance, c'est-à-dire corrective, préventive ou prédictive (PMP), le nombre d'heures de vol (AUR), la durée de vie (COL), le CV^2 et le ADI. Cette première étude permet de définir des pratiques de gestion permettant de réduire la caractéristique grumeleuse de la demande. Dans leur deuxième contribution, ils combinent cette étape d'analyse de variance avec diverses méthodes de prévision pour observer leurs influences sur les résultats et définir un modèle de prévision de l'erreur. De cette manière, le modèle va pouvoir proposer la méthode de prévision la plus adaptée sur le critère du MAPE sans avoir besoin de toutes les comparer sur chaque set de pièces. Les méthodes utilisées pour la comparaison sont celles de Winter (la version additive abrégée en AW et la version multiplicative simplifiée en MW), un modèle de régression saisonnier (SRM), le temps moyen entre chaque remplacement (MTBR), une moyenne sur la période en fonction de l'activité (WCDR), une régression sur les heures de vol (WRDF), un lissage exponentiel à tendance (TAES), une moyenne mobile pondérée (WMA), moyenne mobile pondérée avec tendance (EWMA), un lissage exponentiel avec adaptation du coefficient de lissage (ARRSES), Croston, un lissage exponentiel simple et double (ES et DES). La conclusion de l'étude est que les méthodes MTBR et ES ont de mauvaises performances lorsque la variabilité de la demande augmente : il ne faudrait pas les utiliser dans ces cas d'études.

Le résumé des modèles et des méthodes d'évaluation des résultats décrits dans ce chapitre est affiché dans le tableau 2.4.

Tableau 2.4 : Description des modèles

Référence	**Méthodes utilisées**	**Évaluation des résultats**
Guskiva et Dyntar (2023)	Bootstrapping, LR	Coût stock
Sareminia et Amini (2023)	DT, KNN, 3LFF-NN, NB, SVM, LR, VOTE	MAPE, sMAPE, MSE, RMSE, MAD, Matrice de confusion, MARI
Wang et Hart (2023)	Distribution géométrique, Loi de poisson	Coût chaîne logistique
Wei et al. (2022)	Agrégation temporelle, CGSSA-BP, BP-NN, S SA-BP	MAE, RMSE, R^2
Babaveisi et al. (2022)	SBA, MSBA, Croston, SVM, SVMip	AMAPE, MASE, SME, Coût stock
Şahin et al. (2021)	Naïf, ES, Croston, SBA	Coût stock
Zhu et al. (2020)	ADI, MA, ES, Croston, SBA, TSB	RMSE, MAD, Coût stock, Coût rupture, Coût rebut, Coût commande urgente
Van der Auweraer et Boute (2019)	SMI, SBA, ES	RMSE, ME, Taux de service, Niveau stock
Bergman et al. (2017)	Bayes, EE, BM	MAD, MAPE, ratio MAD/Mean, RMSE, Niveau stock, Coût stock
Pennings et al. (2017)	ES, SBA, Croston, SY, TSB, Bootstrap, DLP, Bootstrap DLP	MASE, GMAE, Taux service, Coût stock
Frazzon et al. (2014)	Loi de poisson	Taux service, coût chaîne logistique
Verhagen et Curran (2014)	Loi de poisson, loi normale, loi de Weibull, GEV	RMSE
Romeijnders et al. (2012)	ZF, Naïve, MA, Croston, SBA, TSB, 2S	MSE, MAD, ME, RMSE
Wang et Syntetos (2011)	DlT	AE
Chiou et al. (2004)	Grey, BJ, ES, NN	ME, RMSE
Ghobbar et Friend (2003)	AW, MW, SRM, MTBR, WCDR, WRDF, Croston, SES, EWMA, Holt, TAES, WMA, DES, ARRSES, ANOVA	ANOVA, MAPE
Ghobbar et Friend (2002)	ANOVA	ANOVA

Ainsi, le tableau 2.4 nous permet de mettre en lumière la grande diversité des méthodes de prévision utilisées ainsi que la diversité des critères d'évaluation, rendant toute comparaison objective très difficile. Il n'existe pas de consensus sur la meilleure méthode de prévision ni sur la façon d'évaluer sa performance : tout dépend du cas d'étude.

2.2.4 Analyse critique des documents

Dans cette section, nous allons tenter d'identifier les limites des méthodes proposées afin d'identifier une opportunité de recherche.

Tout d'abord, dans les 20 articles utilisés pour conduire cette revue de littérature, seulement 4 articles, ceux de Sareminia et Amini (2023), Wang et Hart (2023), Babaveisi et al. (2022) ainsi que Frazzon et al. (2014) se positionnent dans un cas d'architecture multi-échelon. Les autres articles ne le mentionnent pas ou se placent délibérément dans un cas simple. Pour des propositions d'améliorations de modèles se basant uniquement sur l'extrapolation des données historiques, le contexte multi-échelon importe peu. Néanmoins, il devient important dans un cas d'application réel puisqu'il fait intervenir d'autres notions comme le temps de livraison, la répartition entre les différents inventaires et les taux de services.

Comme nous avons pu l'évoquer, il est important de différencier les pièces réparables et les pièces consommables puisque leur valeur est souvent très différente (do Rego et de Mesquita, 2011). Aussi, si certaines études supposent une répartition de la demande, ils supposent que les pièces sont neuves et donc que pour deux pièces identiques, les probabilités de rupture seront identiques (Van der Auweraer et Boute, 2019). Il est donc important de savoir si cet aspect de la maintenance est pris en compte dans les études puisque si l'une d'entre elles suppose une répartition particulière des probabilités de ruptures pour des pièces neuves, il est possible que les probabilités de rupture des pièces réparées soient différentes et que ces pièces ne répondent plus aux hypothèses des contributions. Les études intégrant des pièces réparables sont celles de Sareminia et Amini (2023), H.Wang et Hart (2023), Babaveisi et al. (2022), Bergman et al. (2017), Verhagen et Curran (2014), Rojmeijnders et al. (2012) ainsi que Ghobbar et Friend (2002; 2003).

Afin de pouvoir identifier les impacts réels des méthodes de prévision sur les inventaires, il est important de choisir des métriques cohérentes puisque certaines d'entre elles peuvent faire apparaître des résultats biaisés pour des séries avec de nombreuses demandes nulles. Un écart significatif existe entre la littérature et la réalité puisque l'intérêt est plus souvent porté sur

l'évaluation des modèles que sur l'évaluation de leur performance industrielle (Bergman et al., 2017). En effet, sur les 20 articles détaillés, seulement quatre utilisent un indicateur de coût du stock (Bergman et al., 2017; Huskova et Dyntar, 2023; Şahin et al., 2022; Zhu et al., 2020), deux utilisent un indicateur de taux de service (Frazzon et al., 2014; Van der Auweraer et Boute, 2019) et une seule contribution les utilise conjointement (Pennings et al., 2017). Les autres contributions utilisent simplement des indicateurs de mesure d'erreur de séries statistiques.

D'autre part, seulement quatre des contributions que nous avons étudiées sont appliquées sur des inventaires de tailles conséquentes. Sareminia et Amini (2023) s'intéressent à un inventaire de 51 335 pièces détachées, Zhu et al. (2020) comparent leurs méthodes sur deux ensembles de 24 455 et 138 347 pièces, Van der Auweraer et Boute (2019) étudient le comportement de 210 000 pièces. Quant à eux Romeijnders et al. (2012) s'intéressent à 17 012 pièces utilisées dans 3 329 assemblages. Les autres études sont conduites sur moins de 4 000 pièces (Pennings et al., 2017) et moins de 200 pour le reste. Il est donc intéressant de se demander si les méthodes proposées pour des petits inventaires sont cohérentes avec celles contenants plusieurs milliers voire dizaines de milliers de références.

Aussi, nous allons parcourir les trois contributions sous forme de revues de littératures afin d'identifier si les observations adressées dans ce mémoire sont partagées par d'autres auteurs. Cette étape nous permettra aussi d'identifier d'autres pistes de recherche.

La première revue de littérature apparaissant dans notre consultation des bases de données est celle de do Rego et de Mesquita (2011). Elle repose sur 75 contributions de 1962 à 2009. Les auteurs déclarent qu'il est important de différencier pièces consommables et réparables puisque les dernières ont le plus souvent une valeur importante. Cette contribution étudie principalement la prévision de la demande, la classification, la décision de stocker ou non en fonction de plusieurs caractéristiques, les commandes pour les nouveaux programmes d'équipements, la gestion des stocks, l'obsolescence et la fin de vie. Les pistes de recherches identifiées sont les suivantes :

- Il n'y a pas de classification dynamique couplée à de la prévision. Elle est utilisée pour choisir les méthodes de prévision, mais elles n'évoluent pas avec le cycle de vie;
- Il y a encore peu de comparaisons entre les méthodes récentes de fouilles de données (réseaux de neurones par exemple) et les méthodes classiques de prévision;

- Les mesures d'erreur sont principalement axées sur la qualité des prévisions plutôt que sur la performance industrielle;
- La littérature manque de comparaison des méthodes sur des inventaires complets plutôt que sur des jeux de données restreints;
- Il y a parfois un écart entre la théorie et la pratique. Il serait intéressant de continuer d'étudier comment s'appliquent les politiques de gestion au sein des entreprises et quels sont les principaux freins; et
- Les études de cas plus nombreuses pourraient discuter des aspects de l'implantation des systèmes de contrôles avec les gestionnaires et les technologies d'information.

Par la suite, Zhang et Guan (2016) font un résumé des principales pratiques de gestion des pièces détachées dans le cas de machines industrielles. Ils distinguent les entreprises devant gérer leurs inventaires de rechange pour assurer le fonctionnement de leurs machines et celles qui se placent en tant que fournisseur pour des entreprises exploitantes. Les auteurs proposent de coupler des diagnostics de défaillances pour déterminer une distribution des défauts et ensuite utiliser les méthodes de prévision pour déterminer la quantité et la date de la prochaine demande. Les limites de performances des méthodes de lissage, Croston, et Bootstrap sont discutées. Les premières doivent être principalement utilisées pour les demandes régulières sous peine de mauvaise performance. La seconde est applicable si la distribution des demandes suit les hypothèses de Croston, à savoir une loi normale. La troisième affiche des résultats peu stables au cours des différentes prévisions et pourrait sous-estimer la variabilité temporelle de la demande. Les auteurs discutent finalement de la décision d'intégrer des tiers dans la gestion de la maintenance. Ils concluent l'article sur l'opportunité d'une étude comparant les dernières méthodes de prévision avec les plus classiques pour réaliser un résumé général de la littérature.

La revue de littérature la plus récente sur le sujet date de 2019 et est proposée par Van der Auweraer et al. (2019). Elle regroupe les résultats de 39 articles et cinq chapitres de livres datant de 1960 à 2018. Tout d'abord, les auteurs confirment que la demande de pièces détachées en maintenance possède des propriétés spécifiques qui la différencient de beaucoup d'autres produits. Les points faibles de la littérature qu'ils ont relevés sont :

- Les études sont souvent limitées à une certaine phase de vie des produits et cela rend difficile la généralisation au cycle de vie complet;

- Il y a très peu de méthodes combinant beaucoup de facteurs comme les facteurs de conditions extérieures, l'évolution de la taille de flotte et des probabilités de remplacement. Ils supposent une précision accrue de tels modèles;
- Les classifications font difficilement le lien avec les informations sur les quantités d'actifs utilisés. Les auteurs suggèrent fortement une étude plus globale avec des métriques d'évaluations communes pour effectuer une comparaison formelle des méthodes les plus récentes avec les plus anciennes pour identifier les gains réels;
- Le processus de collecte et de gestion des données est rarement mentionné; et
- La littérature étudiée ne prend pas en compte les grandes difficultés des entreprises à utiliser les quantités de données conséquentes provenant du mouvement BigData.

La littérature repose principalement sur l'extrapolation et trop peu sur l'analyse des causes de génération de la demande.

Les résultats de cette revue critique sont présentés dans le tableau 2.5.

Tableau 2.5 : Résumé de la revue critique (ME : Multi-échelon, PR : Pièces réparables, Cs : Coût de stockage, Ts : Taux de service, Cf : Curatif, Pl : Planifié, Ii : Inventaire important)

Référence	**ME**	**PR**	**Cs**	**Ts**	**Cf**	**Pl**	**Ii**
(Huskova et Dyntar, 2023)			X				
(Sareminia et Amini, 2023)	X	X					X
(Wang et Hart, 2023)	X	X					
(Wei et al., 2022)							
(Babaveisi et al., 2022)	X	X					
(Şahin et al., 2022)			X				
(Zhu et al., 2020)			X		X	X	X
(Van der Auweraer et Boute, 2019)				X	X	X	X
(Van der Auweraer et al., 2019)							
(Bergman et al., 2017)		X	X				
(Pennings et al., 2017)			X	X	X	X	
(Zhang et Guan, 2016)							
(Frazzon et al., 2014)	X			X		X	
(Verhagen et Curran, 2014)		X			X		
(Romeijnders et al., 2012)		X			X		X
(Wang et Syntetos, 2011)					X	X	
(Chiou et al., 2004)					X	X	
(Ghobbar et Friend, 2003)		X			X	X	
(Ghobbar et Friend, 2002)		X			X	X	

2.3 Conclusion

Grâce à cette revue de littérature, nous avons pu mettre en lumière de multiples problématiques concernant la prévision de la demande de pièces grumeleuses dans un contexte de maintenance. Nous avons aussi pu identifier les axes de recherche qu'il pourrait être intéressant d'étudier plus en profondeur. De nombreux documents ne mentionnent même pas les étapes de collecte de données et les potentielles incohérences à la source quand d'autres ne le font que partiellement. Aussi, une classification est utilisée en dehors de son cadre d'étude. Des seuils ont été proposés pour définir des domaines de supériorité de deux méthodes de prévision selon un critère d'évaluation. Lorsqu'on utilise d'autres méthodes de prévision ou d'évaluation des résultats, utiliser cette classification sans même mentionner ou vérifier les hypothèses qui la définissent semble discutable. De plus, nous avons remarqué l'absence de comparaison formelle et générale des méthodes, l'une étant comparée à quelques autres, parfois là encore sorties de leurs contextes de prédilection. Finalement, peu d'études semblent mesurer la performance de leur proposition au regard d'indicateurs d'inventaires, mais y préfèrent une évaluation d'écart entre demande et prévision.

Nous allons tenter de résoudre certains des problèmes évoqués ci-dessus en proposant une méthode détaillant collecte, prétraitement, classification, prédiction et évaluation des résultats. Le prochain chapitre détaillera la méthodologie de recherche qui nous permettra de répondre à cette problématique.

CHAPITRE 3 MÉTHODOLOGIE DE RECHERCHE

3.1 Introduction

Dans ce chapitre, nous allons détailler la méthodologie de recherche qui a guidé notre étude. Tout d'abord, nous présenterons la problématique puis les objectifs de recherche qui y sont associés. Nous finirons cette section par l'approche générale qui nous a permis d'y apporter des solutions.

3.2 Objectifs de recherche

L'objectif général de notre étude est de *réduire la complexité de gestion d'un inventaire de pièces détachées dans le domaine de la maintenance* grâce à un modèle utilisant l'intelligence artificielle pour la prédiction des consommations. Pour atteindre cet objectif principal, suite à la revue de littérature, plusieurs objectifs spécifiques ont été définis :

- Définir une approche efficace de classification de l'inventaire;
- Identifier les différentes méthodes de prévision couramment utilisées dans la littérature;
- Mettre en place une mesure d'erreur en cohérence avec les spécificités du partenaire industriel; et
- Déterminer une méthode d'optimisation réaliste pour un cas pratique avec des ressources informatiques limitées.

3.3 Méthodologie de recherche

Ce projet de recherche vise à proposer un cadre de travail permettant une comparaison formelle de diverses méthodes de prévision comme le suggéraient Van Der Auweraer et al. (2019). Pour faire cela, nous définissons une méthodologie s'intéressant à toutes les phases de traitement des données, de la sélection/extraction depuis l'ERP, leur nettoyage, la recherche de schémas intrinsèques pour définir une classification, la recherche de paramètres optimaux pour différentes méthodes de prévision, l'évaluation de leur performance pour finir par la gestion des exceptions et la vérification de cohérence avec les pratiques d'affaires.

Étant donné que ce projet de maîtrise doit être réalisé dans une fenêtre temporelle restreinte et que les conséquences de nos propositions pourraient largement impacter le fonctionnement de notre partenaire, nous ne pouvons pas attendre de mettre en place notre solution pour en mesurer les bénéfices. C'est pourquoi nous choisissons de suivre une approche DRM ou Design Research

Methodology, présentée par Blessing et Chakrabarti (2009). Ceci va nous permettre d'évaluer les bénéfices de l'approche proposée sans avoir besoin d'une implantation complète. La figure 3.1 présente le détail de la méthodologie DRM ainsi que ses implications à notre projet de recherche. Elle se décompose en quatre étapes : revue de littérature pour clarifier les attendus de recherche, analyse du cas d'étude et de ses données pour déterminer leurs contraintes, proposition d'une méthode puis vérification sur le cas d'étude. L'exploration d'une étape peut et va souvent entraîner un retour vers une étape précédente. C'est ce fonctionnement itératif qui va nous permettre de garder la cohérence entre les besoins du cas d'étude et la littérature tout au long de notre étude.

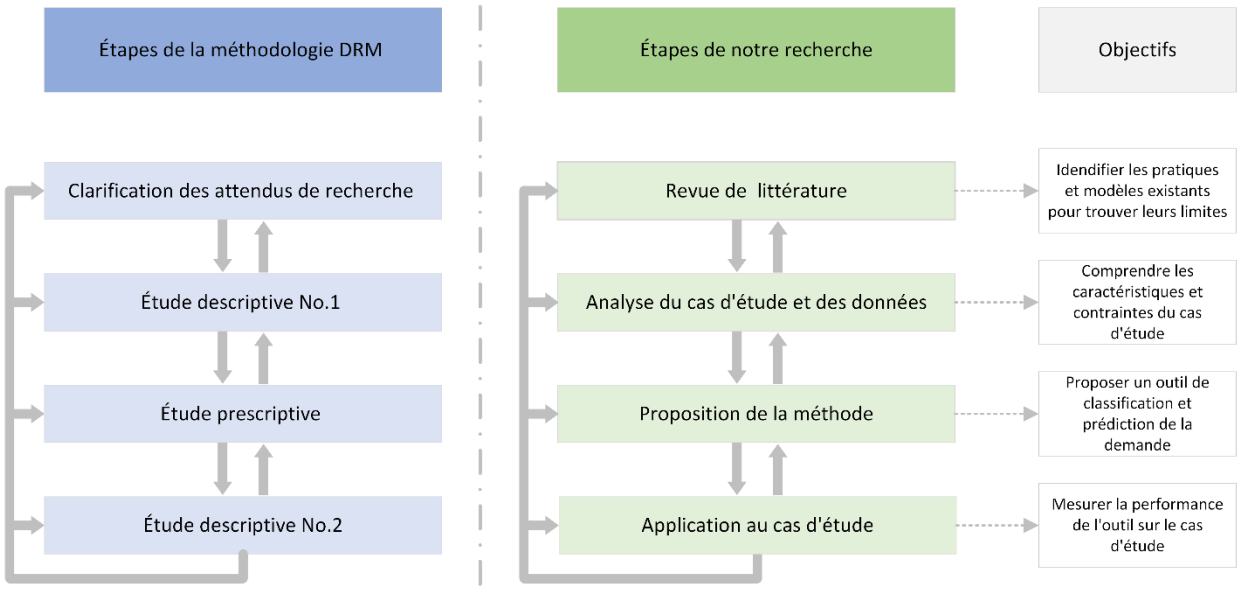

Figure 3.1 : Méthodologie DRM appliquée à notre étude

L'étape de revue de littérature est conduite dans le chapitre 2, l'analyse du cas d'étude et des données dans le chapitre 4, la proposition de la méthode dans le chapitre 5, et finalement, l'application au cas d'étude et les résultats dans le chapitre 6.

3.4 Conclusion

Au cours de notre projet de recherche, nous appliquerons la méthodologie proposée par Blessing et Chakrabarti (2009). Ainsi, nous pouvons nous assurer de rester cohérents tout au long de notre étude, en liant les solutions de la littérature aux besoins et contraintes du partenaire industriel. Nous proposerons une méthode englobant toutes les étapes pour faire une prévision cohérente, depuis la

collecte de données jusqu'à la vérification des pratiques d'affaires. Le prochain chapitre est consacré à la présentation du cas d'étude pour mieux en comprendre les spécificités.

CHAPITRE 4 CAS D'ÉTUDE

Dans ce chapitre, nous allons détailler le problème rencontré en mettant en évidence le fonctionnement de l'entreprise ainsi que les contraintes qui y sont liées.

Le partenaire de cette étude est la STM, la plus grande entreprise de transport en commun sur l'île de Montréal et son agglomération. Elle a notamment permis plus de 165 millions de déplacements utilisateurs en 2021 (STM, 2022). Cette entreprise exploite, répare et produit de nombreuses pièces détachées pour ses bus et son métro. La disponibilité des pièces détachées en temps et en heure est un enjeu crucial puisqu'une immobilisation de véhicule peut impacter lourdement les trajets de centaines d'utilisateurs.

Dans notre étude, nous allons étudier l'architecture de cette entreprise multi-échelon, comprendre les mécanismes de maintenance et de traitement des pièces neuves, réparées et usées, illustrer les défis de gestion liés à la taille conséquente des inventaires ainsi qu'à l'ajout de nouveaux articles et nous terminerons par décrire le processus en place. Nous explorerons le cas de la maintenance du métro pour définir le processus avant d'en illustrer les différences avec la division autobus. C'est pourquoi ces deux types d'actifs seront décrits dans ce chapitre.

4.1 Architecture de l'entreprise

Comme nous avons pu l'écrire dans la partie précédente, l'entreprise partenaire n'est pas constituée d'une seule entité. Elle possède une structure complexe de centres de réparation, entrepôts et usines. Aussi, cette structure organisationnelle est différente entre les divisions métro et bus.

4.1.1 Division Métro

La division métro ou « Entretien Matériel Roulant » (EMR) est une grande entité. Au sein de celle-ci, on retrouve deux types d'ateliers dits de « Petite révision » et de « Grande Révision ». Il y a deux ateliers de « Petite révision », chacun s'occupant d'une génération de rames de métro différente. Dans ces ateliers, les opérateurs vont principalement effectuer des opérations de pose et dépose. C'est-à-dire qu'ils vont utiliser des pièces ou des ensembles fonctionnels pour remplacer les défectueux. Lorsque ces pièces ou assemblages nécessitent un travail de réparation ou de démontage plus poussé, elles sont envoyées aux « Grandes réparations ». Par exemple, une rame de métro arrive dans un atelier de « Petite réparation », un de ses ponts roulants est défectueux. Ce dernier va être retiré et remplacé directement par un pont roulant fonctionnel pour que la rame

puisse reprendre du service dans le délai le plus court. Le pont défectueux va être envoyé vers la « Grande révision » pour qu'il puisse être inspecté plus en détail puis réparé.

Les assemblages ou pièces volumineuses sont stockés dans un entrepôt situé dans l'atelier « Grandes réparations ». Les assemblages ou pièces les plus utilisés et les moins volumineux peuvent être stockés en partie dans les ateliers de « Petites réparations ».

4.1.2 Division autobus

Le fonctionnement de l'entretien des autobus, appelé « Réparation Des Autobus » (RDA) possède des similarités, mais il n'est pas organisé exactement pareil. Alors que tout ce qui a trait au métro est regroupé dans la « Division Métro », ce qui concerne les bus est réparti dans trois divisions : le « mineur autobus », le « majeur autobus » et l' « entrepôt central ».

Le « Mineur autobus » est constitué de 9 « centres de transport » répartis sur l'île de Montréal. Dans ces ateliers, à la manière des « Petites réparations » pour le métro, on retrouve principalement des opérations de pose et dépose de composants. Des stocks de rechange pour les articles peu volumineux et avec une forte demande sont répartis dans ces centres. Lorsque les composantes ont besoin de réparations plus poussées, elles sont envoyées au « majeur autobus ». Cependant, contrairement au métro qui stockait directement les pièces dans l'atelier « Grande réparation », l'entretien bus lui va stocker les composants les plus volumineux ou les moins utilisés dans l' « entrepôt central ».

L'architecture globale regroupant les points évoqués dans les deux derniers sous chapitres est illustrée dans la figure 4.1.

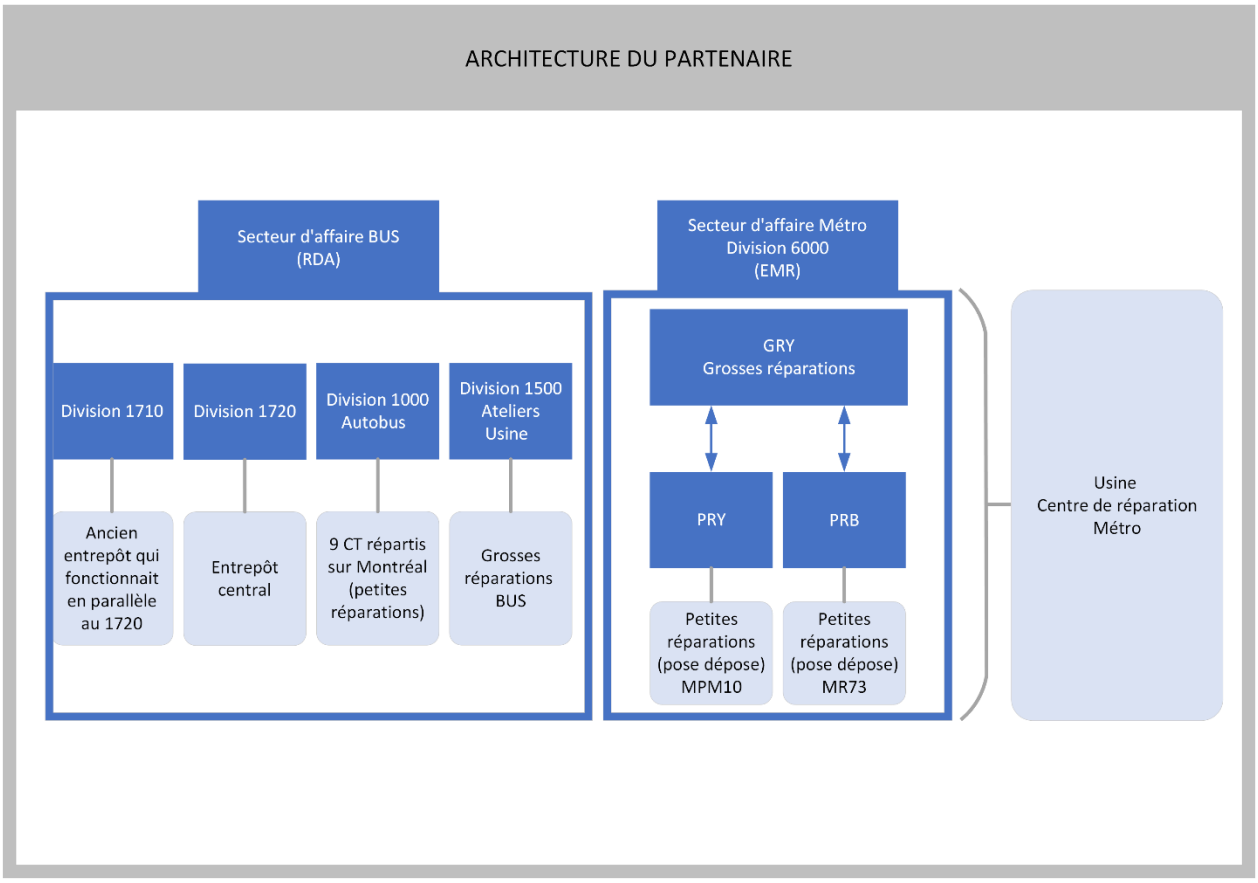

Figure 4.1 : Architecture du partenaire

4.1.3 Mécanisme de réparation des articles

Une des particularités de l'entreprise est sa capacité à réparer en interne les pièces qu'elle consomme pour la maintenance. Afin d'identifier quelles pièces doivent être réparées, lesquelles sont déjà réparées et lesquelles sont neuves, un champ « groupe de valorisation » apparaît sur les fiches articles avec les codes suivant :

- N : Pièce neuve
- R : Pièce réparée
- X : Pièce à réparer
- « » : Laissé vide pour un consommable

Dans les opérations, certaines pièces N et R sont utilisées indifféremment. Certaines exceptions existent, mais elles sont minoritaires. Lors d'une visite d'atelier, nous avons pu avoir une discussion avec un contremaître responsable de la gestion des pneumatiques du métro. Il nous a expliqué que les pneus sur les ponts moteurs étaient montés sur les ponts non-moteurs après un certain nombre de kilomètres d'utilisation. Par la même occasion, la pression de gonflage diminuait et il était impossible de faire le chemin inverse, c'est-à-dire de regonfler à une pression plus importante par mesure de sécurité. Ce degré d'information dépend des opérateurs et il n'est pas retranscrit dans les données.

Les inventaires de l'entreprise présentent des défis considérables à cause de la diversité des articles en stock. Ce sont plus de 30 500 articles pour le métro et 46 500 pour le bus. À la différence d'une entreprise de vente qui pourrait simplement se séparer des articles les moins performants ou les moins souvent vendus, une entreprise de maintenance doit garder un catalogue d'articles lui permettant de répondre à ses besoins de réparation et ce, même si certaines consommations ne s'effectuent pas avant plusieurs années voire dizaines d'années. Pour avoir plus de regards sur ces informations, l'entreprise a mis en place une classification « ABCP » qui repose sur le principe de Pareto. À savoir que « 80% des effets sont le produit de 20% des causes ». Les pièces A sont approximativement les 20% d'articles qui représentent le plus de flux financiers. Elles sont soit des pièces consommées très fréquemment ou relativement fréquemment avec une valeur importante. Les catégories B et C représentent les pièces les moins consommées. La catégorie P correspond aux pièces prioritaires. Ce groupe caractérise des pièces critiques, soit par leur importance, soit par les problèmes qu'elles causent. Par exemple, une pièce critique peut être P tout comme une pièce non critique à l'origine, mais qui n'est pas en stock depuis une grande période et qui risque à présent d'entraîner des conséquences significatives sur les opérations.

D'autre part, un autre défi de gestion apparaît à cause de l'ajout ou le retrait continu de références au catalogue. Certaines pièces ne sont plus utilisées, d'autres ne sont plus disponibles ou d'autres encore entrent en stock pour remplacer d'anciens articles d'un autre fournisseur. Depuis 2019, il y a eu 1250 ajouts d'articles pour la division métro et plus de 7300 pour les autobus. Ces différences de variabilité des inventaires découlent du type de flotte. Nous allons le détailler dans le paragraphe suivant.

Comme nous avons pu le voir, la division métro est celle avec le moins d'ajouts de nouveaux articles. Elle est composée de seulement deux types de véhicules dont nous allons donner les caractéristiques et dates importantes.

Les MR73, qui sont entrés en service à partir de 1976, ont été rénovés entre 2003 et 2007. Ces modifications sont principalement esthétiques et ne changent donc pas radicalement les composantes mécaniques. La flotte est actuellement constituée de 360 voitures assemblées en 120 trains de 3 wagons.

Les MPM10 sont entrés en service en deux phases : 54 livrés entre 2016 et 2018 ainsi que 17 livrés entre 2020 et 2021. Pour le moment, la maintenance de ces trains n'est pas assurée par la STM, mais par le constructeur des trains. Cependant, elle est effectuée dans les locaux de l'entreprise de transport, donc les transactions et les consommations de marchandise sont accessibles dans le logiciel ERP. La flotte est constituée de 71 trains de 9 wagons interconnectés où les passagers peuvent circuler librement.

Cette flotte est amenée à évoluer, mais aucun plan de remplacement des MR73 par des MPM10 n'est prévu avant la fin de ce projet. On pourra donc supposer que le nombre de véhicules dans chaque flotte est constant le temps de notre étude.

La flotte de la division bus est quant à elle beaucoup plus variable. Il y a environ 20 références de bus pour un total de 2 044 bus en 2022. La durée de vie d'une référence d'autobus est de 16 ans. La flotte de bus doit encore évoluer fortement dans les années à venir avec l'arrivée de nouveaux bus hybrides et au minimum 147 bus électriques entre 2025 et 2027 (STM 2024).

Les processus de prévisions des demandes en pièces détachées sont très différents entre le métro (EMR) et les autobus (RDA).

Pour EMR, environ 80% des consommations peuvent être qualifiées de planifiées. En effet, une simulation ARENA permet de définir les dates de réalisation des prochains plans d'entretiens en fonction du nombre de kilomètres parcourus. Dans ces plans d'entretiens, une gamme détaille les opérations à effectuer et est reliée à une nomenclature qui définit quels articles et dans quelles quantités ils seront consommés. Avec l'observation des historiques de consommation, les planificateurs connaissent les fréquences de remplacement de certaines pièces optionnelles ou qui ne seraient pas consommées à chaque fois. Les 20% de consommations curatives correspondent à des réparations qui ne proviennent pas de ces opérations planifiées ou correspondraient à des

articles consommés en nombre plus important que décrits dans la gamme. Les prévisions pour le métro peuvent être plus régulières parce que les actifs roulants sont beaucoup moins soumis aux aléas extérieurs comme les conditions de route ou les forts écarts de température. Cependant les planificateurs éprouvent des difficultés lors de la prévision sur les MPM10 puisque leurs historiques de demandes sont beaucoup plus faibles.

Pour RDA, la planification est beaucoup plus sommaire. En théorie, les prévisions sont effectuées sur le long terme, 3 ans normalement, en fonction des données de consommation moyennes sur la dernière année. Ces chiffres sont ajustés avec les informations sur l'évolution de la flotte. Par exemple, si on sait que la taille de la flotte de bus X va varier de Y%, on va corriger les prévisions avec ce facteur Y%. Les planificateurs se servent des gammes et des nomenclatures pour définir quelles sont les pièces pour lesquelles il faut ajuster la prévision. Les stocks sont gérés avec une stratégie de seuil de commande. Ils sont ajustés presque au jour le jour pour faire face aux ruptures et une période de pic d'ajustements est effectuée en début calendaire, ce qui illustre un certain problème d'organisation et encourage la démarche de prévision de la demande proposée dans ce mémoire. Pour les autobus, les pièces optionnelles n'apparaissent pas dans les gammes et nomenclatures. Au mineur, les quantités renseignées pour les articles optionnels sont à 0 alors qu'au majeur, elles sont de 99. Ici encore, on remarque une différence de fonctionnement entre les différents acteurs de l'organisation.

4.2 Étude des données de maintenance

Dans cette section, nous allons détailler le type de données, le processus de collecte, l'identification des consommations puis l'analyse des consommations avec la classification ADI/CV².

4.2.1 Source de données

Les données sont stockées dans le progiciel de gestion d'entreprise « System Applications and Products in Data Processing », plus connu sous son abréviation « SAP ». Il a été implanté progressivement chez le partenaire depuis 2015. Cette année-là, il prenait en charge le secteur d'affaires EMR. Ce n'est qu'à partir de 2019 que son utilisation a été étendue à RDA. Aucune donnée antérieure n'est disponible pour ces secteurs d'affaires. En faisant l'hypothèse que les données ont beaucoup plus de chance d'être erronées les premières années suivant l'implantation que les années suivantes à cause du temps d'adaptation nécessaire pour adopter les bonnes

pratiques d'utilisation, on comprend mieux pourquoi cette étude est conduite d'abord sur EMR avant de comparer les résultats avec RDA.

4.2.2 Structure des données

Mouvement de stock

Les premières données que nous utilisons proviennent du code de transaction **MB51**. Ces données suivent les mouvements de stock. C'est-à-dire que pour chaque mouvement d'article, que ce soit un déplacement d'une division à une autre, d'un atelier à un autre ou une consommation, chaque demande aura une structure similaire avec une transaction par ligne pour les colonnes :

- Code mouvement
- Article
- Groupe de valorisation
- Division
- Magasin
- Date de saisie
- Quantité
- Unité quantité de base
- Réservation
- Ordre
- Document article

Le code mouvement permet d'identifier si les mouvements correspondent à des déplacements de pièces, des échanges inter-division, des réparations ou des consommations réelles. Pour chaque code mouvement, on code retour y est associé et permet de faire la transaction inverse. Le détail des codes de mouvements est présenté par le tableau 4.1.

Tableau 4.1 : Description des codes mouvements

Code mouvement	Code retour correspondant	Description
201	202	Sortie d'inventaire sur centre de coût
221	222	Sortie d'inventaire pour projet
231	232	Sortie d'inventaire pour commande client
241	242	Sortie d'inventaire pour immobilier
251	252	Sortie d'inventaire ventes
261	262	Sortie d'inventaire sur ordre
281	282	Sortie d'inventaire pour le réseau
291	292	Sortie d'inventaire toutes imputations
961	962	Sortie d'inventaire sur ordre
P61	P62	Sortie d'inventaire pour les outils sur ordre
	531	Entrée matière sous-produit

Le code de l'article nous permet de savoir de quel type de pièce on parle. Ce code ne traduit pas un identifiant unique pour chaque pièce physique, mais pour un type. Par exemple, l'article 12345 correspondrait à un pneu de traction. Ce code ne permet pas de distinguer un pneu de traction d'un autre pneu de traction.

Le groupe de valorisation est le code N-R-X-« » défini dans le paragraphe 4.2.2 de ce mémoire. Il donne une information sur le type de pièce, réparable ou consommable ainsi que son état, à réparer, neuf ou réparé.

Le principe de division est détaillé dans le paragraphe 4.2.1. Il correspond des entités de réparation ou de stockage. Les divisions que nous allons étudier dans ce mémoire sont détaillées dans le tableau 4.2.

Tableau 4.2 : Description des divisions

Code de division	Description
1000	Mineur autobus
1500	Majeur autobus
1710	Ancien entrepôt central
1720	Nouvel entrepôt central
6000	Métro

À l'heure où nous rédigeons ce mémoire, la division 1710 n'est plus active. Elle est remplacée par la division 1720. Cependant les historiques de transactions qui y sont liés doivent être pris en compte pour identifier la demande passée.

Le champ « magasin » permet de différencier les différents ateliers au sein d'une même division. Le détail n'est pas cohérent pour le processus de prévision.

La date de saisie est la date à laquelle la transaction a été rentrée dans le système. Elle a été choisie puisqu'elle représente plus fidèlement la dimension temporelle de la demande qu'une date comptable par exemple.

La quantité va permettre d'identifier les consommations. Elles peuvent être négatives ou positives. Les premières traduisent une sortie de stock alors que les dernières traduisent une entrée. On peut s'assurer de la cohérence des données en croisant le code mouvement (code d'entrée ou de sortie) avec le signe de la quantité.

L'unité quantité de base permet de savoir si on parle d'un article en termes de pièce ou de boîtes d'un certain nombre. Cette information est aussi fondamentale que les quantités puisque commander une pièce ou une boîte de 100 pièces n'aura pas du tout le même impact sur les stocks.

La réservation est l'expression d'un besoin en pièces de rechange. Pour combler ces besoins, il va y avoir d'autres opérations, des confirmations, des transferts de stocks ou des demandes d'achats par exemple. vont être générées au sein du système SAP à la suite du calcul MRP. Une réservation peut être faite sur centre de coût ou sur ordre.

L'ordre de travail va appeler une gamme dans la majorité des cas. Cette gamme renseigne de la suite des activités nécessaires à la réalisation de l'ordre. Si la gamme n'existe pas, une saisie manuelle est possible pour renseigner ces informations. L'ordre est un élément de gestion auquel on peut imputer des codes mouvements pour traduire de la consommation des articles pour cette tâche particulière. Lorsque l'ordre n'est pas disponible, on utilisera le code de réservation pour imputer la consommation des articles à un autre élément de gestion comme un centre de coût. Le centre de coût est une entité logistique, une subdivision de l'entreprise à laquelle on va assigner le coût de certaines opérations lorsque ces dernières ne se font pas sur un ordre de travail.

Le document article permet de retracer l'historique de tous les mouvements d'articles ayant eu lieu dans le logiciel. Il est unique et ne peut pas être supprimé.

Types de maintenances

Le second type de données que nous utilisons provient du code de transaction **IW39.** Il permet d'afficher d'autres informations, non disponibles dans la **MB51** sur le détail du type d'opérations effectuées. Cette extraction sert à distinguer les consommations curatives des consommations planifiées. Pour chaque ligne, on retrouve les informations suivantes :

- Ordre
- Type d'ordre annexe
- Date de référence
- Type de travail
- Groupe de gammes
- Compteur de gammes
- Division

Nous avons déjà évoqué l'ordre, la date de référence et la division dans la première partie de cette section 4.3.1.

Le type d'ordre permet de distinguer les ordres de fabrication, de maintenance ou d'inspection par exemple. Le détail des différents types d'ordres est donné dans le tableau 4.3.

Tableau 4.3 : Description des types d'ordres

Type d'ordre	Description
M1EP	Ordre entretien préventif
M1IN	Ordre d'inspection
M1IQ	Ordre d'inspection avec lot de contrôle
M1PE	Ordre de périodicité

Le type de travail est différent du type d'ordre, il permet un niveau de précision supplémentaire. Par exemple, pour un ordre de remise à neuf, le type de travail va permettre de détailler la raison de cet ordre, que ce soit une révision ou une réparation.

Le groupe de gammes permet de regrouper des gammes. Par exemple, pour une opération de maintenance particulière, on peut faire appel à un groupe de gammes pour les différentes opérations. Pour le moment, d'après le savoir d'experts, à EMR les groupes de gammes n'utilisent qu'une gamme du même nom en théorie.

Le compteur de gamme permet de tenir compte des variantes de gammes. Pour une même gamme, plusieurs variantes sur les pièces à utiliser ou l'ordre des opérations peuvent être renseignées pour des bus différents par exemple. Ce groupe de gammes est principalement utilisé dans la division autobus de l'entreprise.

Pour identifier les pièces associées aux gammes, nous utilisons les données extraites de la transaction **IA17**. Elle définit pour chaque gamme les colonnes suivantes :

- Nom de gamme
- Compteur de gamme
- Désignation de la gamme
- Article
- Désignation du composant
- Quantité
- Unité de mesure

4.2.3 Extraction des données SAP vers Excel

Comme nous avons pu le définir précédemment, les données SAP proviennent des transactions **IW39** et **MB51**. Pour la première, nous avons collecté en les séparant par années calendaires toutes les transactions enregistrées entre le premier 2 novembre 2015 et le 10 juillet 2023. Pour la seconde, nous avons adopté la même stratégie en y ajoutant une discrimination par secteur d'affaire. Les données de EMR s'étendent du 2 novembre 2015 au 10 juillet 2023, les données de RDA s'étendent du 20 octobre 2019 au 10 juillet 2023.

Le secteur d'affaire EMR est celui de la division 6000. Le secteur d'affaire RDA regroupe les divisions 1000, 1500, 1710 et 1720.

Cette politique d'extraction qui sépare les secteurs d'affaires nous permettra de comparer les résultats des prévisions entre eux. Choisir une distinction par années permet de réduire la charge système lors de l'extraction ainsi qu'une actualisation des données historiques sur une base annuelle. Il est à noter que les données ainsi extraites reflètent les mouvements des éléments concernés et ne traduisent pas la consommation de ces derniers, différents traitements devront être mis en place pour cela.

4.2.4 Construction de la matrice des consommations

Dans cette section, nous allons définir comment a été construite la matrice des consommations, illustrée par le tableau 4.4. Elle détaille les consommations pour chaque article jour par jour. Les articles apparaissent dans la première colonne et leurs consommations respectives pour chaque jour de l'horizon d'étude apparaissent dans les colonnes correspondantes.

Tableau 4.4 : Matrice de consommation

Article	2015-11-01	2015-11-02	…	2023-07-31	2023-08-01
1	0	2	…	50	2
2	21	1	.	1	0
3	0	0	.	1	0
4	8000	300	.	50	2

Compensation des sorties par les codes retours

Comme nous avons pu le mentionner précédemment, les données de la transaction **MB51** vont permettre de connaître les quantités réellement consommées jour par jour. Les codes mouvements permettent de se concentrer sur ce qui est consommé et non simplement déplacé. Les codes utilisés sont détaillés dans le tableau 4.1. Ils ont été validés par des spécialistes des opérations de maintenance à la STM.

Les données de transactions sont agrégées par jours puisqu'avec le fonctionnement interne du partenaire, la plupart des transactions de sortie et de retours sont effectuées le jour même. De cette façon, on peut compenser facilement les quantités.

Pour pallier le problème d'un retour effectué un autre jour, une clé unique est définie comme la concaténation du numéro d'article, du numéro de réservation et de l'ordre. Les quantités de sorties sont compensées par les quantités de retours avec une simple addition. La clé unique permet de faire correspondre les codes de sorties avec leurs codes de retours respectifs. Les quantités correspondant à des codes de sortie sont toujours négatives et celles de retours toujours positifs. Si cette quantité compensée est négative, on obtient bien une consommation puisqu'un signe négatif correspond à une sortie d'inventaire. Si la quantité est positive, c'est que plus de pièces ont été retournées que sorties de stock : c'est une erreur, la référence de l'ordre sera notée et la quantité de consommation ramenée à 0. Cette gymnastique permet de quantifier ce qui est réellement consommé le jour de la demande en retranchant les retours futurs.

Pour les consommations qui ne présentent ni ordre ni numéro de réservation, les quantités de sorties et de retour seront simplement ajoutées jour par jour pour chaque article.

Ces identifications des consommations sont faites sur les groupes de valorisations N et R, pour les pièces neuves et réparées. En effet, dans la pratique, on ne va jamais consommer une pièce X (à réparer) pour effectuer une réparation. La logique de l'identification des consommations est illustrée par la figure 4.2.

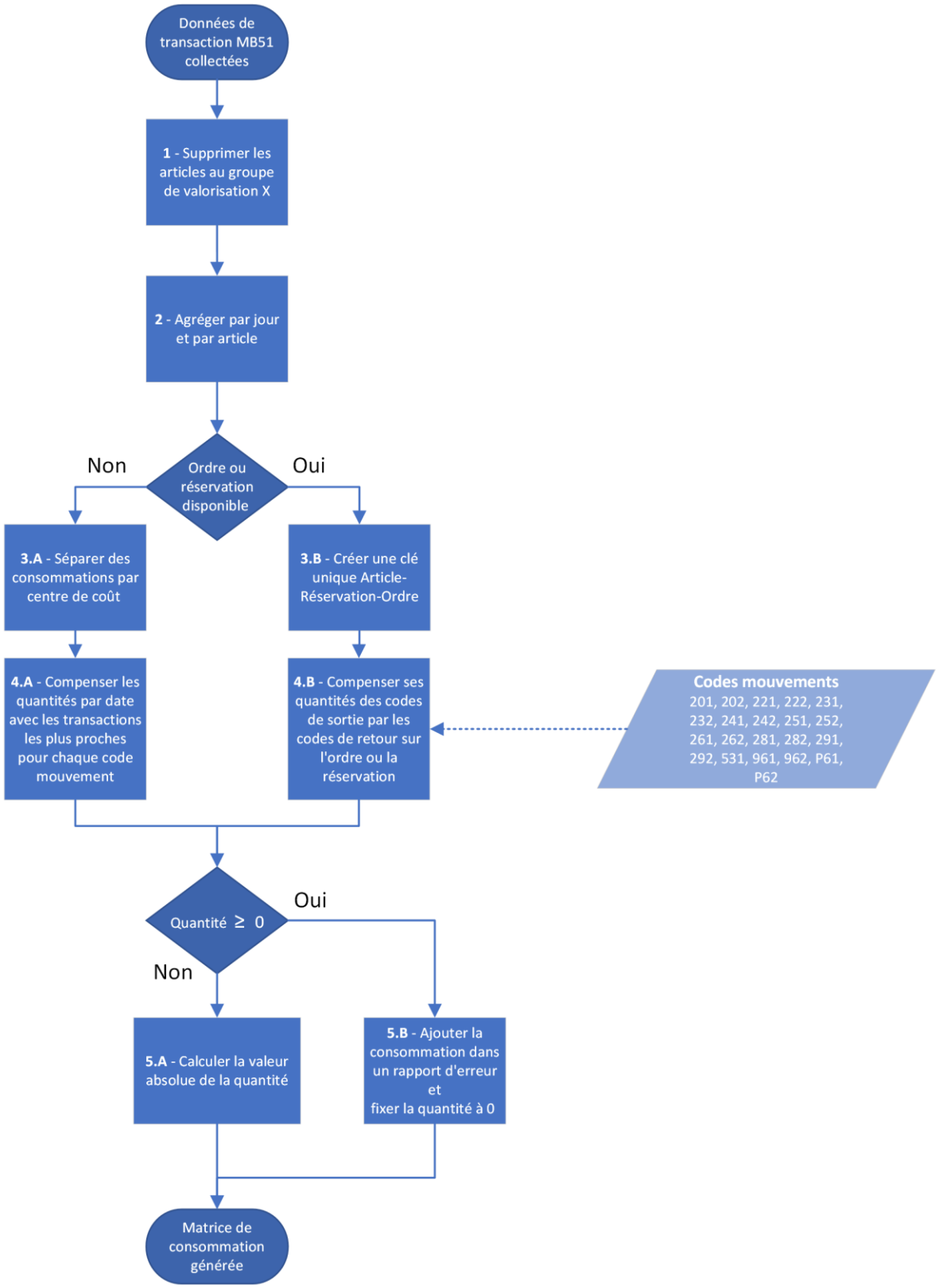

Figure 4.2 : Processus d'élaboration de la matrice de consommation

4.2.5 Exploration des données brutes

La première étape dans cette section concernant l'exploration des données est de segmenter les produits selon leur profil de consommation, pour cela nous allons utiliser la classification ADI / CV^2 mentionnée dans la littérature et notre revue préliminaire.

Pour rappel, les profils selon Syntetos et al. (2005) sont définis comme suit :

- Lisse (ADI < 1.32 et CV^2 < 0.49) : Caractérise des articles avec une demande qui présente peu de variation de quantité et de temps moyen inter demande;
- Intermittent (ADI ≥1.32 et CV^2 < 0.49) : Caractérise des articles avec une demande qui présente peu de variation de quantité, mais des variations de temps moyen inter demande;
- Erratique (ADI < 1.32 et CV^2 ≥ 0.49) : Caractérise des articles avec une demande qui présente de fortes variations de quantité, mais un temps moyen inter demande variant peu; et
- Grumeleux (ADI ≥ 1.32 et CV^2 ≥ 0.49) : Caractérise des articles avec de fortes variations de quantités et de temps moyen inter demande.

Dans notre cas d'étude, nous avons pu comparer les répartitions de profils pour des agrégations journalière, hebdomadaire, mensuelle et annuelle. Ces informations sont présentées dans le tableau 4.5.

Tableau 4.5 : répartition des profils pour des agrégations journalière, hebdomadaire, mensuelle et annuelle

		Profil				
		Lisse	Intermittent	Erratique	Grumeleux	Total
Agrégation	Journalière	0	0	0	100 %	100%
	Hebdomadaire	0.04 %	0	0.32 %	99.64 %	100%
	Mensuelle	0.84 %	0	1.58 %	97.58 %	100%
	Annuelle	8.00 %	0	7.50 %	84.5 %	100%

Le tableau 4.5 montre que l'horizon d'agrégation joue un rôle important dans la classification d'inventaire par profil de demande lorsqu'on utilise les seuils de la littérature. Lorsqu'on regarde le détail des consommations de façon journalière, l'inventaire entier est considéré comme ayant un comportement grumeleux.

On remarque que le profil intermittent n'est jamais identifié dans un inventaire de plus de 13000 références. Aussi, on observe une proportion croissante de profils lisses et erratiques ainsi qu'une diminution des profils grumeleux lorsque l'horizon d'agrégation change du jour vers l'année. Cette observation est cohérente puisqu'en agrégeant plus de données, le nombre d'observations par période augmente.

4.3 Conclusion

Cette étude préliminaire nous conforte sur la mauvaise performance des seuils définis par Syntetos et Boylan (2005) pour notre cas d'étude. Il faudra redéfinir une approche permettant de regrouper les articles pour ensuite être capable de faire leur prévision par profil de demande dans le domaine de la maintenance. C'est ce que nous allons faire dans le chapitre suivant en détaillant les obstacles rencontrés et les solutions choisies pour d'aboutir à l'approche proposée. Remplacer ce texte par votre texte.

CHAPITRE 5 DESCRIPTION DE L'APPROCHE PROPOSÉE

Dans ce chapitre, nous allons détailler le processus de traitement des données. Comme nous avons pu le remarquer précédemment, une première étape de traitement a été nécessaire afin d'établir une matrice de consommation à partir des codes de transaction de l'ERP. Cependant, cette matrice fait apparaître des articles qui pourraient ne plus être utilisés ou dont les consommations ne sont pas exploitables pour notre étude. C'est pourquoi nous allons effectuer plusieurs étapes de traitement. L'ordre de ces étapes a été défini dans une optique d'entonnoir afin de pouvoir identifier les exceptions à chaque étape et pouvoir y appliquer les politiques de gestion appropriées. Cette logique est illustrée par la figure 5.1.

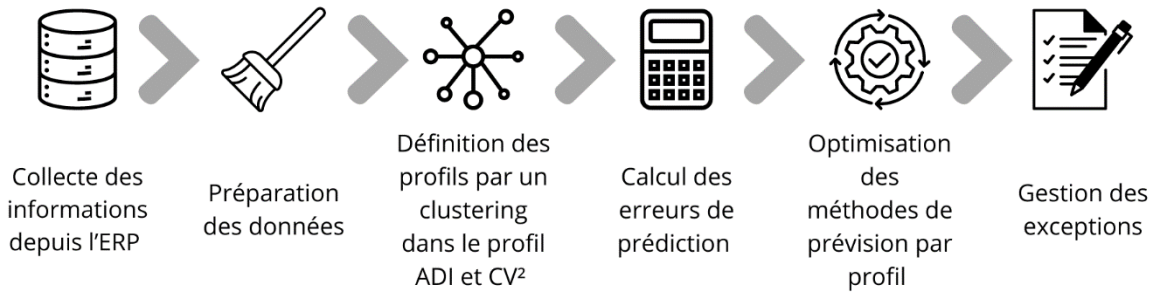

Figure 5.1 : Logique de traitement

5.1 Préparation des données

La préparation des données est composée de quatre étapes. Chacune est expliquée ci-dessous.

5.1.1 Filtrage par les items actifs

La première étape de prétraitement est le filtrage des articles actifs. Au fil du temps, la flotte d'actifs de notre partenaire industriel évolue et cela implique nécessairement l'ajout ou le retrait de références au catalogue. Cependant, dans l'ERP, les articles qualifiés d'inactifs ne sont pas simplement retirés. Un statut « inactif » leur est affecté. Si on sait qu'on ne va plus utiliser certaines références, il semble logique de retirer ces références de nos prévisions pour ne pas alourdir le calcul inutilement.

Afin d'identifier les articles inactifs, nous avons utilisé le code de transaction **SE16N**. Ensuite, nous utilisons la table **MARC** pour accéder aux fiches de données qui nous sont utiles. On peut alors définir un filtre sur les statuts d'activité. Dans notre cas, les statuts considérés sont les **01**,**02**, **A0** et **A1.** Cette étape permet d'identifier 28 637 articles encore actifs dans l'ERP.

Une fois ces données extraites sous la forme d'un tableau Excel, nous allons croiser les codes d'articles actifs avec ceux apparaissant dans la matrice de consommation. Cette étape nous permet d'identifier 13 184 articles inactifs et de les isoler. Nous continuerons notre étude sur les articles actifs.

5.1.2 Nettoyage de la matrice de consommation

La deuxième étape du traitement est le nettoyage de la matrice de consommation des articles pour lesquels aucune consommation n'a été effectuée sur l'ensemble de l'historique. Cette opération nous permet de réduire l'étude de 13 184 articles vers 12 572 articles. Les 612 articles n'ayant aucune consommation peuvent être considérés comme du stock mort, c'est-à-dire des références en stock pour lesquels il n'y a eu que des transactions de mouvement, mais aucune de consommation ou alors que les consommations et les retours balancent parfaitement. Nous continuerons notre étude sur les articles avec au moins une consommation sur l'historique.

5.1.3 Identification des comportements de « démarrage à froid »

La troisième étape de traitement est l'identification des articles en « démarrage à froid ». Ce terme permet de qualifier des articles pour lesquels les premières consommations apparaissent seulement depuis quelques mois au moment de faire la prévision. Comme il est difficile d'identifier à partir de combien de consommation un article est dans cette catégorie ou dans celle des articles « normaux », nous avons choisi d'aligner la période de démarrage à froid avec notre horizon de validation, soit trois mois. Nous allons donc considérer que tout article qui ne présente des consommations que pendant les trois derniers mois est en « démarrage à froid ». Il sera isolé et nécessitera un traitement spécifique. Cette étape nous permet de réduire la matrice de consommation de 12 572 à 12 497 articles, nous permettant d'identifier 75 articles en démarrage à froid. Nous continuerons notre étude sur les articles dans un cycle de demande normal.

5.1.4 Filtrage des articles avec moins de deux consommations

Dans cette quatrième étape, nous allons identifier les articles présentant moins de deux consommations sur l'historique. Il est difficile de définir le nombre de valeurs à partir duquel on peut réaliser une étude prédictive grâce à l'étude de l'historique de consommation. Dans notre travail, nous avons défini ce seuil à deux, car certaines des librairies utilisées retournaient une erreur de calcul pour les articles avec une seule consommation. Cette approche est discutable; il serait intéressant d'étudier le nombre minimum d'occurrences dans les séries temporelles en fonction des différentes méthodes de prévision pour avoir des résultats cohérents. Cette étape nous a permis de séparer 3 228 articles pour arriver finalement à 9 269 articles.

À la suite de toutes ces étapes de filtrage, la matrice de consommation contient 9 269 articles et nous allons pouvoir y appliquer les autres étapes de la méthodologie proposée.

5.2 Regroupement des articles par profil de demande

Comme nous avons pu le détailler dans la revue de littérature, un certain nombre de contributions utilisent les seuils définis par Syntetos et al. (2005) pour caractériser un inventaire dont un nombre encore plus faible l'applique à la maintenance. Ces profils définissent des zones de prédilection pour deux méthodes de prévision selon leur performance. Les études s'appuyant sur ces profils contestent rarement les seuils définis dans un contexte différent ou les utilisent avec d'autres méthodes de prévision. Aussi, il nous semble peu efficace de regrouper un inventaire de plusieurs dizaines de milliers de références en seulement quatre catégories. Dans notre cas particulier, plus de 97% du catalogue d'articles est considéré comme grumeleux, ce qui laisse deux solutions de gestion : considérer une méthode de prévision sur ce groupe entier ou entrer dans le détail de chaque article. Dans le premier cas, la littérature a, à de nombreuses reprises, souligné que la méthode miracle efficace dans tous les cas est une illusion. Aussi, les articles grumeleux du catalogue étudié s'étendent sur une grande surface du plan ADI / CV^2, ce qui laisse penser qu'il est possible d'identifier plusieurs groupes au sein du profil grumeleux, comme l'illustre la figure 5.2.

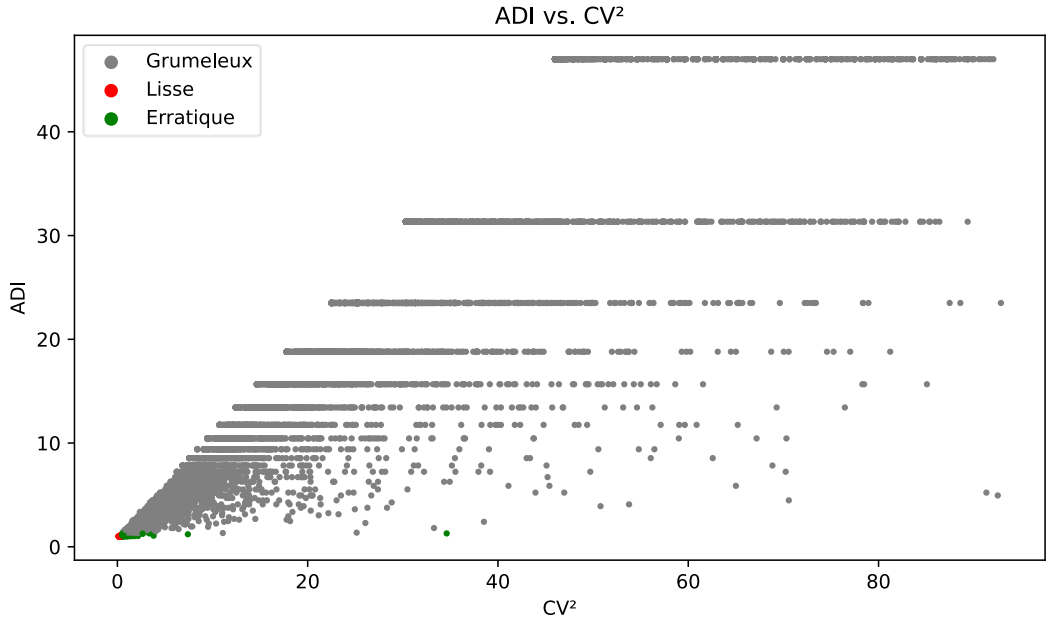

Figure 5.2 : Classification mensuelle du catalogue d'articles dans le plan ADI vs CV².

La deuxième solution de gestion, qui était de considérer chaque article individuellement, implique des contraintes de gestion monumentales pour un inventaire de grande taille. Les deux approches ne sont pas satisfaisantes. C'est pourquoi nous proposons de rechercher un nombre intermédiaire de groupes grâce à des méthodes de clustering dans le plan ADI /CV².

Pour segmenter les points dans l'espace ADI/CV², nous avons choisi la méthode K-means puisqu'elle s'appliquait particulièrement bien à notre cas d'étude, à savoir : utilisable sur un grand nombre de références et s'applique pour une géométrie plate (Ikotun et al., 2022). Cette méthode vise à regrouper les points en un nombre de groupes k, définis par l'utilisateur qui minimisent la distance entre chacun de ses points et le centre du groupe. Les différentes étapes d'exécution d'un clustering K-means sont les suivantes :

- Initialisation des centres :

  L'algorithme va choisir aléatoirement k points. Ils seront considérés comme les centroïdes de chacun des groupes.

- Calcul de la distance euclidienne et attribution au groupe :

Pour chacun des points de l'espace, la distance euclidienne est calculée avec tous les centres définis précédemment par l'équation 5.1.

$$d_{i,k} = \sqrt{(x_i - \mu_k)^2 + (y_i - \gamma_k)^2} \quad (5.1)$$

Avec (x_i, y_i) les coordonnées du point i, (μ_k, γ_k) les coordonnées du centroïde k et $d_{i,k}$ la distance du point i au centroïde k.

Chaque point est ensuite attribué au centroïde dont il est le plus proche.

- Actualisation des centroïdes :

Une fois que les points sont attribués à leurs centroïdes, l'algorithme calcule la moyenne de leurs coordonnées pour définir le nouveau centroïde.

- Répétition jusqu'à convergence :

Les 2 étapes précédentes sont répétées jusqu'à ce que la phase d'actualisation des coordonnées des centroïdes ne varie plus de manière significative.

Pour comparaison avec la littérature sur notre cas d'étude, le clustering pour trois clusters est représenté sur la figure 5.3. Étant donné qu'on cherche à regrouper les articles par profil de demande similaire, on remarque que la répartition semble plus équilibrée en termes de nombre de pièces par profils et que l'étendue des profils sur le plan ADI/CV² est réduite par rapport aux pratiques de la littérature. Les détails et une comparaison plus précise de cette observation seront réalisés dans le chapitre 6.

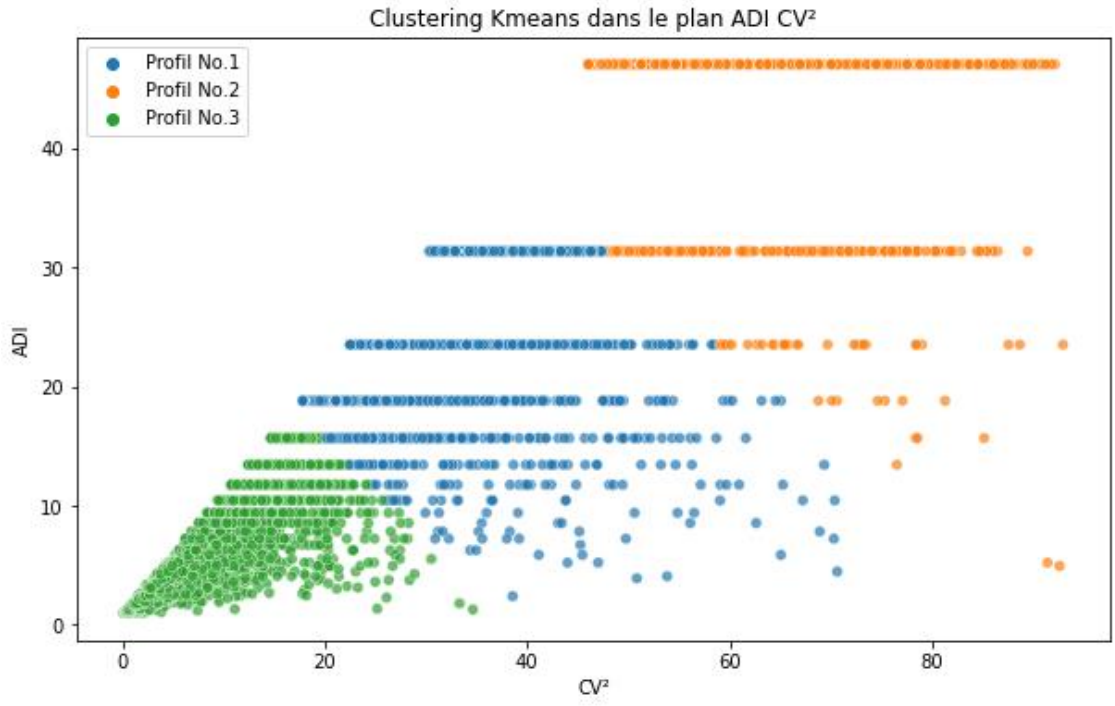

Figure 5.3 : Répartition des profils de clustering pour 3 clusters

5.3 Calcul des prévisions par méthode

Comme nous avons pu le détailler précédemment dans ce document, nous cherchons à effectuer les prévisions par profil de demande en attribuant un couple méthode/paramètres dans le but de réduire les contraintes de gestion opérationnelles. Nous avons recensé 14 méthodes de prévision différentes dans notre étape de revue de littérature : naïve, naïve non nulle, moyenne, moyenne mobile, moyenne mobile pondérée, Croston original, Croston SBA, Croston SBJ, Croston TSB, Lissage exponentiel simple, lissage exponentiel double (Méthode de Holt), lissage exponentiel triple (Méthode de Holt-Winters), ARIMA et SVM. Les premières méthodes ont été choisies pour leur simplicité d'implantation dans un contexte de prévision réel pour que des planificateurs n'aient pas besoin de formation poussée. ARIMA et SVM ont été ajoutés puisqu'elles apparaissent de plus en plus souvent dans la littérature et leur comparaison à grande échelle avec des méthodes plus conventionnelles n'a que rarement été proposée. Le détail de chacune de ces méthodes est disponible en annexe A en fin de document.

Cette comparaison de méthodes de prévision selon un même indicateur a été encouragée par la littérature. Notre étude cherche à donner plus de résultats à ce sujet, notamment dans un cas

identifié comme « extrêmement grumeleux » avec plus de 95% de l'inventaire identifié comme tel. La section suivante va détailler grâce à quel indicateur nous allons mesurer les performances de ces méthodes de prévision.

5.4 Définition de mesures d'erreurs

Dans cette section, nous allons définir les mesures d'erreur que nous utiliserons pour comparer les différentes méthodes de prévision. Tout d'abord, nous devons identifier les contraintes imposées par le contexte d'utilisation afin de pouvoir utiliser une mesure d'erreur quantifiant bien les résultats en fonction de notre besoin.

Dans un contexte de prédiction de la demande intermittente, nous ne pouvons pas choisir une mesure d'erreur divisant simplement une quantité par la demande réelle. Le cas de demande intermittente fait intervenir des consommations nulles et le calcul d'erreur devient mathématiquement impossible. Aussi, comme nous allons mesurer les erreurs par article sur chacune des séries temporelles avec des quantités commandées variant très fortement, il faut choisir une méthode à l'échelle. Cela veut dire qu'il faut ramener chacune des erreurs à une quantité représentant la série comme une moyenne. On comprend aisément qu'une erreur de prévision de deux pièces sur une demande de cinq est beaucoup moins bonne qu'une erreur de dix articles sur une demande de 5 000. Aussi, il faut tenir compte des besoins d'une entreprise en charge de la maintenance d'actifs. Une prévision supérieure à la consommation entraîne un coût de stockage supplémentaire tandis qu'une prévision inférieure à la demande peut causer des ruptures et entraîner des surcoûts à cause d'une impossibilité de fournir le service proposé. Il faudrait également que la mesure d'erreur travaille différemment avec les erreurs positives et négatives. Comme la mesure d'erreur a pour vocation d'être utilisée pour l'optimisation des hyperparamètres et pour les planificateurs, il faut qu'elle soit compréhensible facilement. On choisit de ne pas utiliser une mesure différente pour l'optimisation et pour les planificateurs puisqu'optimiser l'une d'elles ne garantit pas que l'autre le soit. À l'image d'un candidat au permis de conduire, s'il améliore sa conduite pour tourner à gauche, rien ne nous dit qu'il sera aussi efficace pour tourner à droite. Peut-être qu'il va développer des biais qui vont dégrader ses résultats.

Afin de déterminer quelle mesure utiliser, nous avons exploré la littérature pour essayer d'identifier une métrique qui corresponde à nos exigences. Les résultats de cette étude sont renseignés dans le Tableau 5.1.

Tableau 5.1 : Respect des critères de l'étude pour les différentes métriques de performance

Mesure d'erreur	Mesure à l'échelle	Fonctionne avec des demandes nulles	Traite différemment les erreurs positives et négatives	Interprétable facilement	Calculable facilement
MAPE	X			X	X
sMAPE	X				X
MAE		X		X	X
MASE	X				X
ME			X	X	X
sME	X		X	X	X
MSE		X			X
RMSE		X			X
MAD		X			X
R^2	X	X			X
GMAE					X
AE		X		X	X
Taux de service	X	X	X	X	
MAAPE	X	X			X
sPIS	X	X			X
sAPIS	X	X	X		X
sMSE	X	X			X
sMAE	X	X		X	X
MUES	X	X	X	X	X
MOES	X	X	X	X	X

À première vue, seules les mesures de moyenne de sous-estimation (MUES) et moyenne de surestimation (MOES) semblent posséder toutes les caractéristiques recherchées. Elles mesurent respectivement la proportion de prévision inférieure et supérieure à la demande réelle sur un horizon. Leurs formules sont données dans les équations 5.2 et 5.3.

$$MUES = \frac{1}{H}\sum_{h=1}^{H} \max(signe(Y_h - P_h), 0) \tag{5.2}$$

$$MOES = \frac{1}{H}\sum_{h=1}^{H} \max(signe(P_h - Y_h), 0) \tag{5.3}$$

Avec H : l'horizon de prévision, Y_h : valeur de consommation à la période h, P_h : valeur de la prévision à la période h.

Dans notre cas, on s'intéresserait plutôt à la MUES pour suivre le nombre de ruptures potentielles à cause d'une prévision insuffisante. Cependant, elle est insuffisante seule. Si on veut minimiser les prévisions qui ne répondent pas à la demande, on pourrait se contenter de commander en large excès.

Comme aucune des mesures d'erreur ne répond à notre besoin, nous avons essayé d'en définir une personnalisée au cas d'étude. Pour faire cela, on s'est inspiré des équations de modélisation de chaîne logistique couramment utilisées dans les chaînes de Markov dans l'objectif de minimiser le coût des prévisions. Nous avons commencé par poser les hypothèses suivantes :

- On suppose qu'une prévision correcte implique de parfaitement pouvoir répondre à la demande. On aura donc les bonnes pièces en stock quoi qu'il arrive;
- On suppose que les pénalités utilisées reflètent les coûts réels de rupture. À savoir 10 000$ par période pour une pièce critique et 200$ pour une pièce non critique; et
- On suppose qu'une rupture de pièce A ou P est critique puisqu'elle entraînerait une immobilisation

On peut ensuite définir les équations de coût de stockage (5.4) et de coût de rupture (5.5) .

$$Hc = \frac{0{,}2 * Pc}{12} * (2 - CT) \tag{5.4}$$

$$IMc = CT * 10000 + (1 - CT) * 200 \tag{5.5}$$

Avec Hc : coût de stockage au mois, Pc : prix de la pièce, CT : criticité valant 1 pour une pièce A ou P et 0 pour une pièce B ou C, IMc : coût de rupture.

Pour mieux comprendre l'utilisation de cette mesure, nous allons utiliser un exemple avec pour conditions initiales le seuil de sécurité fixé à trois, le stock de sécurité fixé à cinq et le stock initial fixé à cinq. La figure 5.4 illustre le comportement de la mesure d'erreur lorsque la consommation est égale à la prévision. Le coût global est incrémenté d'un montant proportionnel au coût de stockage et au nombre de pièces en stock.

	Apprentissage						Validation		
Mois	1	2	3	4	5	6	7	8	9
Prévision	3	5	0	4	2	1	2	3	5
Consommation	3	2	2	5	1	7	3	4	3
Ecart	0	3	-2	-1	1	-6	-1	-1	2
Stock	5	8	6	5	6	0	0	0	0
Stock de sécurité	5	5	5	5	5	5	4	3	5

Tant qu'il y a du stock et du stock de sécurité :

$$coût = coût + Hc * (x_t + Ssecu_t)$$

$$Hc = \frac{0{,}2 * Pc}{12} * (2 - CT)$$

Figure 5.4 : Utilisation de la mesure d'erreur pour une consommation égale à la prévision

Lors de la période suivante, on remarque une prévision supérieure à la consommation. Ce surplus est stocké et on continue d'incrémenter le coût total comme dans l'étape précédente. Cette logique est présentée à la Figure 5.5.

	Apprentissage						Validation		
Mois	1	2	3	4	5	6	7	8	9
Prévision	3	5	0	4	2	1	2	3	5
Consommation	3	2	2	5	1	7	3	4	3
Ecart	0	3	-2	-1	1	-6	-1	-1	2
Stock	5	8	6	5	6	0	0	0	0
Stock de sécurité	5	5	5	5	5	5	4	3	5

Prévisions > Consommation :
Surplus à stocker

$x_t = x_{t-1} + \epsilon_t$ avec ϵ_t l'écart, x_t le stock au temps t

Figure 5.5 : Utilisation de la mesure d'erreur pour une consommation inférieure à la prévision

Après plusieurs périodes de prévisions insuffisantes, le niveau de stock a atteint 0. Si on continue de cette manière, on va devoir piocher dans le stock de sécurité. Ceci ne provoque pas une rupture de stock, car on s'autorise à piocher dans le stock de sécurité lorsque nécessaire. On va seulement imposer une pénalité pour forcer la méthode de prévision à prédire plus et ne plus avoir besoin de piocher dans cette réserve. Cette étape est illustrée par la Figure 5.6.

	Apprentissage						Validation		
Mois	1	2	3	4	5	6	7	8	9
Prévision	3	5	0	4	2	1	2	3	5
Consommation	3	2	2	5	1	7	3	4	3
Ecart	0	3	-2	-1	1	-6	-1	-1	2
Stock	5	8	6	5	6	0	0	0	0
Stock de sécurité	5	5	5	5	5	5	4	3	5

On doit piocher dans le stock de sécurité car le stock est nul et la consommation est supérieure à la prévision.

On pénalise de 50$ (Montant arbitraire pouvant peut-être changé d'une pièce à l'autre)

$Ssecu_t = Ssecu_t + \epsilon_t$

avec ϵ_t l'écart au temps t

$coût = coût + Hc + 50$

Figure 5.6 : Utilisation de la mesure d'erreur à la suite d'une pioche dans le stock de sécurité

La figure 5.7 montre que, lorsqu'on franchit le seuil de sécurité fixé à 3, nous allons pénaliser plus fortement la fonction de coût pour traduire le risque de rupture qui est associé à cette situation.

	Apprentissage						Validation		
Mois	1	2	3	4	5	6	7	8	9
Prévision	3	5	0	4	2	1	2	3	5
Consommation	3	2	2	5	1	7	3	4	3
Ecart	0	3	-2	-1	1	-6	-1	-1	2
Stock	5	8	6	5	6	0	0	0	0
Stock de sécurité	5	5	5	5	5	5	4	3	5

Le niveau du stock de sécurité passe en dessous du seuil critique et on risque une rupture suivant la prochaine demande : on pénalise fortement le coût

$Ssecu_t < Seuil_{secu}$
$coût = coût + Hc + IMc$

$IMc = CT * 10000 + (1 - CT) * 200$

$Hc = \frac{0{,}2 * Pc}{12} * (x_t + Ssecu_t) * (2 - CT)$

Figure 5.7 : Utilisation de la mesure d'erreur à la suite du franchissement du seuil de sécurité

Finalement, comme la prévision dépasse la consommation lors du mois neuf, on va alors pouvoir remplir le stock de sécurité pour qu'il retrouve son niveau initial.

Le principal problème de cette mesure d'erreur est qu'il est très difficile d'estimer le coût de rupture d'une pièce, son coût de stockage, tout comme le montant de la pénalité lorsqu'on utilise le stock de sécurité. Cette mesure d'erreur semble beaucoup trop subjective pour être utilisée en pratique.

Notre dernière solution est de coupler la MUES avec une autre mesure d'erreur comme la sMAE. En utilisant les deux mesures, nous sommes capables de répondre à toutes les contraintes de notre cas d'étude. La sMAE est définie par l'équation 5.6.

$$sMAE = \frac{1}{H}\sum_{h=1}^{H}\frac{|Y_{N+h} - F_h|}{\frac{1}{N}\sum_{t=1}^{N} Y_t} \tag{5.6}$$

Avec H : horizon de prévision, N : nombre de demandes observées, Y_{N+h} : demande observée pour la h-ième période de prévision et F_h : quantité prévue à la période h.

Une modification de la MUES va nous permettre de définir une proportion de ruptures tolérée par les opérations sur l'horizon de prévision. Lorsqu'elle n'est pas respectée, on va pouvoir informer le planificateur du risque. Nous allons la calculer de la façon suivante :

- Utiliser la criticité des articles pour définir un seuil;
- Récupérer le stock initial;
- Ajouter au stock la valeur des prévisions et retrancher la consommation réelle; et

- Si le seuil n’est pas respecté, informer le planificateur d’une rupture potentielle.

5.5 Optimisation des prévisions par profil de demande

Dans cette section sur l’optimisation des hyperparamètres des méthodes de prévision par profil de demande, nous allons détailler les enjeux, les problèmes et les solutions que nous avons pu apporter.

L’objectif de notre travail d’optimisation par cluster est de réduire les contraintes de gestion pour les équipes opérationnelles. Grâce aux différentes techniques de prévision disponibles et aux puissances de calcul considérables disponibles à la location pour les professionnels, il serait théoriquement possible de réaliser des prévisions avec des méthodes et paramètres différents pour chacun des articles. Cependant, en pratique, les planificateurs ne peuvent pas se reposer uniquement sur les prévisions puisque l’étude des séries temporelles présente toujours des biais de calcul et il n’existe pas encore, à notre connaissance, de méthode de calcul permettant de prendre en compte toutes les contraintes de notre partenaire industriel. C’est pourquoi nous avons choisi de les regrouper par profil de demande en explorant l’option d’un clustering Kmeans dans le profil ADI/CV² : on suppose que les pièces au sein d’un même groupe possèdent des comportements de demande similaire et qu’on peut réaliser des prévisions avec un seul couple méthode-paramètres.

Une méthode de prévision peut être plus ou moins adaptée à un comportement, mais sa performance dépend aussi des hyperparamètres qui lui sont associés. Pour comparer ces méthodes de prévision de manière cohérente, il semble important de le faire au meilleur de leurs performances. C’est pourquoi nous devons d’abord optimiser les hyperparamètres de prévision par méthode. Il existe de nombreuses façons d’optimiser ces paramètres, mais peu de détail est donné dans la plupart des articles spécifiques à la demande grumeleuse que nous avons rencontrés. Nous avons donc dû étendre la recherche sur l’optimisation à d’autres domaines et définir quelles approches pourraient nous être utiles.

Dès lors, plusieurs techniques d’optimisation sont disponibles : la recherche aléatoire, la recherche par grille, l’essaim de particules, la programmation génétique. Ces quatre méthodes sont celles que nous avons le plus rencontrées dans nos différentes lectures.

5.5.1 Recherche aléatoire

Cette forme d'optimisation est la plus simple à mettre en place. Elle repose sur l'essai aléatoire de valeurs dans l'espace de recherche. Avec un nombre de tests assez important, on peut considérer être dans le cas d'une simulation de Monte Carlo et donc de balayer suffisamment correctement l'espace des solutions. Son principal point faible est qu'elle ne garantit pas de trouver un extrema intéressant. En effet, ses performances sont totalement dépendantes des points choisis et donc, du hasard.

5.5.2 Recherche par grille

La recherche par grille repose sur l'essai d'un nombre de valeurs renseigné par l'utilisateur. On va ensuite pouvoir essayer toutes les combinaisons entre ces différents paramètres pour effectuer une recherche exhaustive des solutions. Cependant, il faut apporter un soin particulier à la définition de ces valeurs puisque seules les valeurs renseignées par l'utilisateur seront essayées.

5.5.3 Optimisation par essaim de particules

Comme son nom l'indique, cette technique d'optimisation repose sur le comportement des essaims pour trouver un optimum. Initialement, un nombre de particules N est réparti sur l'espace de recherche. Chacune des particules se voit attribuer une direction et une vitesse initiale aléatoire. Les particules vont interagir entre elles de proche en proche pour converger vers les meilleures solutions. Il existe aussi dans certaines versions de l'algorithme une influence de la particule ayant les meilleurs résultats vers celles étant moins performantes. La meilleure solution en fin de cet algorithme est déterminée par les coordonnées de l'espace regroupant le plus grand nombre de particules. L'avantage de cette approche est d'être simple à paramétrer, qu'elle converge rapidement et qu'elle permet de trouver un optimum global dans la plupart des cas. Cependant, dans certains rares cas, elle peut ne pas converger et rester bloquée dans un optimum local. Aussi, il existe relativement peu de librairies python pour faire ce type d'optimisation, qu'elle est initialement définie pour des problèmes continus et qu'elle ne gère pas l'optimisation par contrainte.

5.5.4 Algorithme génétique

L'optimisation par algorithme génétique consiste à simuler l'évolution et la sélection naturelle au sein d'une population. L'utilisateur va définir chaque paramètre d'une méthode de prédiction

comme étant un gène d'un individu. Initialement, on définit un nombre P d'individus avec une combinaison de gènes aléatoires. Les individus les plus performants (dans notre cas, les combinaisons de paramètres donnant l'erreur de prévision la plus faible) vont pouvoir mélanger leurs gènes dans une étape de reproduction. Aussi, on va introduire des mutations de façon aléatoire sur certains gènes (en modifiant légèrement certains paramètres) pour s'assurer de balayer convenablement l'espace de recherche. Si on gardait simplement les gènes initiaux, on pourrait rester dans un cas d'extrema local. Les individus nouvellement créés constituent la génération suivante. Ces opérations sont répétées sur un nombre G de générations pour converger vers une solution optimale. Cette approche a l'avantage d'offrir une bonne diversité des solutions et d'avoir une complexité faible, la rendant particulièrement efficace pour une optimisation dans un espace de dimension importante ainsi que de permettre la programmation par contrainte. Toutefois, cette méthode peut avoir des problèmes avec une convergence prématurée à cause d'un nombre de gènes trop faible dans la population initiale ou au contraire trop longue à cause de mutations trop nombreuses. Aussi, les mutations peuvent potentiellement causer une perte d'optimums. De nombreuses versions de cet algorithme existent, certaines conservent les meilleurs candidats de chaque génération, d'autres réalisent cette évolution naturelle sur plusieurs îles en parallèle ou ajoutent encore des comportements de prédation. Notre étude s'est basée dans un premier temps sur un algorithme simple, utilisant simplement reproduction et mutation.

Dans notre cas d'étude, nous devons travailler avec des techniques capables de traiter d'optimisations avec des variables continues comme discrètes. Il faut aussi être capable de reproduire les optimisations pour chacune des méthodes de prévisions sans que leurs performances soient totalement dépendantes du hasard. Nous avons écarté la recherche aléatoire puisque rien ne nous permettait de définir le nombre de valeurs à essayer avant de conclure que la meilleure solution trouvée soit effectivement la meilleure. Nous n'avons pas non plus choisi la recherche par grille basique puisqu'elle nécessitait de choisir arbitrairement des paramètres à essayer et rien ne nous permettait de conclure sur le choix de bonnes valeurs ou non.

5.5.5 Utilisation de l'optimisation par algorithme génétique

La première démarche a été de considérer toutes les méthodes de prévision au sein d'une seule optimisation. La sortie de l'algorithme donnerait potentiellement le couple méthode et paramètres qui permettent d'avoir une prévision minimisant l'erreur sur un cluster. Avec les 15 méthodes

retenues, nous avons eu besoin d'optimiser les paramètres, soit 33 dimensions pour notre espace de recherche de solution au total. L'optimisation par algorithme génétique est efficace lorsque le nombre de variables devient important, ce qui la rend particulièrement intéressante aux premiers abords. De plus, elle est capable d'utiliser des fonctions objectives et des fonctions contraintes. Nous avons choisi la bibliothèque python « deap », car elle permet de mettre en place facilement ce type d'optimisation. Ensuite, nous avons commencé par tenter d'optimiser les quatre variantes de Croston. Chaque individu possède deux gènes : « variante » et « lissage ». Le premier contient une chaîne de caractère « sba », « sbj », « tsb » ou « original » et le deuxième est un flottant dans l'intervalle]0,1].

Nous avons essayé de faire une optimisation sans mutations sur 5 générations de 20 individus et pour une matrice de consommation de 250 articles. Le calcul a duré 74 secondes avec un processeur 16 cœurs de fréquence 2.2GHZ ainsi que 16GB de RAM. Pour obtenir des résultats convaincants, il est de coutume d'utiliser beaucoup plus d'individus ainsi qu'augmenter le nombre de générations. Pour 10 générations et 100 individus, le calcul dure près de 5h. Étant donné que le temps d'exécution de l'algorithme est lié au nombre d'articles et donc de séries temporelles pour lesquelles on effectue une prévision, il faudrait au minimum 260 heures pour effectuer l'optimisation sur tout l'inventaire d'une division. Ce n'est pas une solution envisageable.

En analyse de données, les problèmes en dimensions élevés souffrent de la « malédiction de la dimension ». Elle se caractérise par une mauvaise répartition des points dans l'espace de recherche. En général, les régions au centre de l'espace sont moins densément peuplées que celles aux limites, ce qui peut créer des difficultés de convergence ou, au contraire, un blocage sur un extrema local. La solution la plus simple et la plus commune est de réduire le nombre de dimensions. Dans notre cas, il suffirait de découpler les méthodes de prévision pour lancer une optimisation pour chacune d'entre elles avant de comparer leurs performances. Seulement, lorsque certaines méthodes n'ont besoin que d'un paramètre à optimiser, la programmation génétique perd tout son sens et devient trop gourmande en ressources alors que d'autres techniques d'optimisation peuvent être utilisées.

5.5.6 Utilisation de l'optimisation par essaim de particules

Nous avons ensuite tenté d'utiliser l'optimisation par essaim de particules. Un des principaux obstacles à son utilisation est qu'il fonctionne sur des paramètres continus. Dans notre cas, certains couples de paramètres doivent faire apparaître des valeurs entières. Pour tenter de résoudre ce

problème, nous avons utilisé l'algorithme Particle Swarm Optimization (PSO) proposé par la bibliothèque « pyswarm » sur les indices d'une série. Pour arriver à faire cela, on a d'abord réparti au hasard des points dans l'espace de recherche dont on garde les coordonnées dans une liste. Ensuite, on définit les bornes du PSO en fonction du nombre de valeurs choisies. Si on répartit 2000 points, la recherche d'optimum se fera entre 0 et 2000. À cause de sa nature continue, le PSO va parcourir toutes les valeurs continues entre 0 et 2000, nous avons choisi de garder la partie entière de ces valeurs pour réaliser notre optimisation sur des entiers. Ensuite, la valeur de la liste correspondant à cet indice est utilisée comme coefficient de lissage de Croston, dont on va mesurer l'erreur de prévision. On rappelle qu'on cherche à n'utiliser qu'une seule technique d'optimisation pour toutes les méthodes de prévision dans le but de réduire les contraintes de gestion réelles. En réalisant nos tests sur une méthode comme celle de Croston avec des indices de listes, on s'assure de pouvoir faire de l'optimisation discontinue, car nous avons discrétisé l'espace de recherche. La Figure 5.8 illustre le comportement de la méthode de Croston sur les consommations d'un article choisi au hasard dans le catalogue. On remarque que les prévisions à la suite des deux techniques d'optimisation permettent d'obtenir des résultats très similaires. Sur les 3 mois prédits, l'erreur sMAE des deux approches vaut 0.16. Les écarts de performances sont de l'ordre de $10^{-9.}$ On peut considérer qu'avec un pas de discrétisation suffisant, le PSO est aussi efficace en discret qu'en continu. Leurs temps de calculs respectifs sont de 2.24 secondes et 2.16 secondes, ce qui semble indiquer que nous pouvons potentiellement utiliser cette méthode d'optimisation.

Cependant, des problèmes de convergence sont apparus lors de l'utilisation de cet algorithme. En effet, lorsqu'on arrondit les valeurs du PSO pour les faire coïncider avec des indices de liste, on crée des 'marches' dans la surface de réponse. Certaines fois, les particules vont pouvoir rester coincées dans une région et y osciller sans jamais pouvoir en sortir.

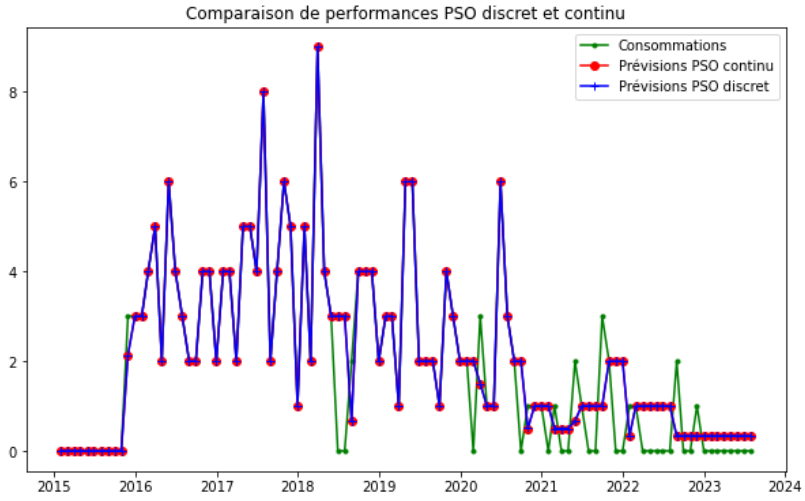

Figure 5.8 Prévisions de Croston optimisées par PSO discret et continu

Aussi, malgré de nombreuses lectures sur le sujet, nous n'avons pas trouvé de proposition convaincante permettant de réaliser une optimisation multi-objectifs, ni même avec une fonction contrainte.

Si on se contente de pondérer les différentes fonction objectifs pour se placer dans un cas mono objectif, il faut être capable de déterminer les différents poids. Chose que nous ne sommes pas capables de faire de façon objective, surtout lorsque les différentes mesures d'erreur possèdent des échelles différentes.

Dans un cas où il est impossible de définir l'ordre de préférence sur les objectifs, on ne peut plus comparer les différentes solutions entre elles. C'est alors qu'intervient le principe du front de Pareto, représentant une ligne à partir de laquelle il n'est plus possible d'améliorer un objectif sans dégrader l'autre. Dès lors, nous ne trouvons plus un « meilleur » couple de paramètres par méthode, mais plusieurs. On crée ainsi un nouveau problème : celui de choisir la meilleure solution.

Une autre solution est d'utiliser une fonction objectif et une fonction contrainte. La première est optimisée tant que la deuxième est respectée. Si elle ne l'est plus, on impose une pénalité sur la valeur de l'objectif. La nouvelle question est de savoir de combien on pénalise le non-respect de la contrainte. Aussi, si on pénalise trop grandement la fonction objectif, on peut créer des régions de l'espace inaccessibles pour un PSO, à l'image d'un « cirque naturel » représenté par la figure 5.9. Certaines particules peuvent être coincées dans la Zone 1 et ne jamais arriver à surpasser la Zone 2 pour arriver dans la Zone 3, le minimum qui nous intéresse.

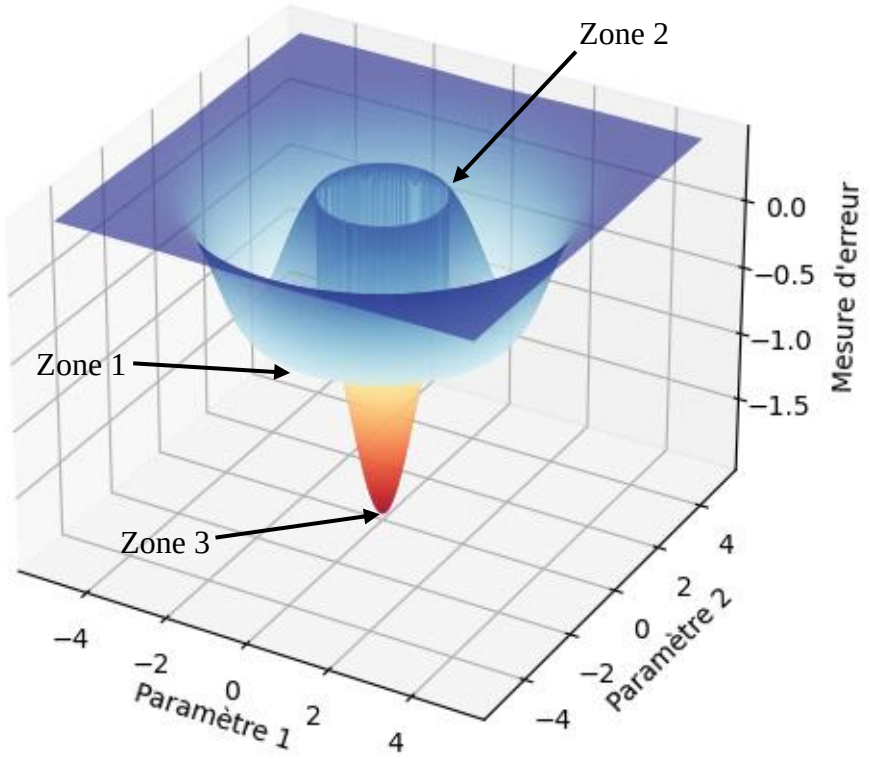

Figure 5.9 : Zones inaccessibles pour un PSO

Ces expérimentations avec l'algorithme génétique et le PSO nous ont permis de tirer plusieurs conclusions :

- L'optimisation multi-objectifs est encore un problème nécessitant plus de recherche;
- Définir une méthode d'optimisation est une contribution à part entière, on ne peut pas simplement passer du continu au discret pour notre cas d'étude;

- Il faut à tout prix réduire le nombre de dimensions lors de l'optimisation; et
- Le temps de calcul peut être un frein lors du choix d'une méthode d'optimisation.

5.5.7 Utilisation du Latin Hypercube Sampling

Par suite des conclusions précédentes, nous avons exploré une solution plus sommaire telle que la recherche par grille. Comme mentionné précédemment, il est important de pouvoir choisir les domaines de recherche et le pas de discrétisation de façon intelligente pour ne pas avoir à calculer des milliers de valeurs inutiles.

Nous avons commencé par essayer manuellement différentes valeurs de pas pour jauger de la sensibilité de chacune des méthodes de prévision. Pour les méthodes de Croston et ses variantes, il n'y avait pas de variation significative de l'erreur pour des modifications de l'ordre de 0.01 du coefficient de lissage. Nous allons donc chercher à placer 100 points sur l'intervalle de recherche du coefficient de lissage tout en uniformisant au maximum sa couverture. Une première façon de faire est de diviser l'intervalle par 100 puis placer chaque point à une distance identique des autres. Mais imaginons qu'il existe un phénomène d'une périodicité 0.01, on pourrait par hasard se placer dans un cas problématique et penser que cette méthode de prévision n'est pas efficace alors qu'il s'agirait simplement d'un mauvais choix de paramètres. Utiliser une répartition uniforme pour chacun des axes n'est pas non plus efficace puisque certaines régions de l'espace pourraient être plus denses que d'autres et fausser les calculs.

La solution la plus intéressante pour pallier ce problème est d'utiliser un échantillonnage par hypercube latin (LHD). Cette technique permet de choisir judicieusement les paramètres de chacune des méthodes de prévision pour garantir une couverture équilibrée de l'espace de recherche. Le principe est simple, pour une dimension 1, on découpe une ligne en N nombres de segments (correspondants au nombre de points à placer) puis on place aléatoirement un point dans ce segment. En dimensions 2, on doit placer ce point dans une case, en dimension 3 dans un cube et ainsi de suite. De cette façon, on s'assure de répartir le plus uniformément possible les points dans l'espace pour avoir une évaluation efficace des performances des méthodes de prévision. La figure 5.10 représente une comparaison entre une distribution uniforme et un LHD en dimension deux. Les deux premiers graphiques représentent, pour les deux dimensions de l'espace de recherche, le nombre de points présents en fonction de la plage de valeur. Par exemple, pour le LHD sur le Paramètre 1, on peut lire qu'il y a six points entre 0.4 et 0.45, 4 points entre 0.45 et 0.5

puis 5 points pour chacune des tranches entre 0.5 et 1.0. Que ce soit pour le Paramètre 1 ou Paramètre 2, on remarque que le LHD permet effectivement une meilleure couverture de la plage de valeur puisque, comparé à une distribution uniforme, il ne présente pas de zones beaucoup plus denses que d'autres. Le troisième graphique représente la répartition des points de coordonnées (Paramètre 1, Paramètre 2). Ici, on observe bien que le LHD permet de placer des points proches des limites du domaine tout comme au centre.

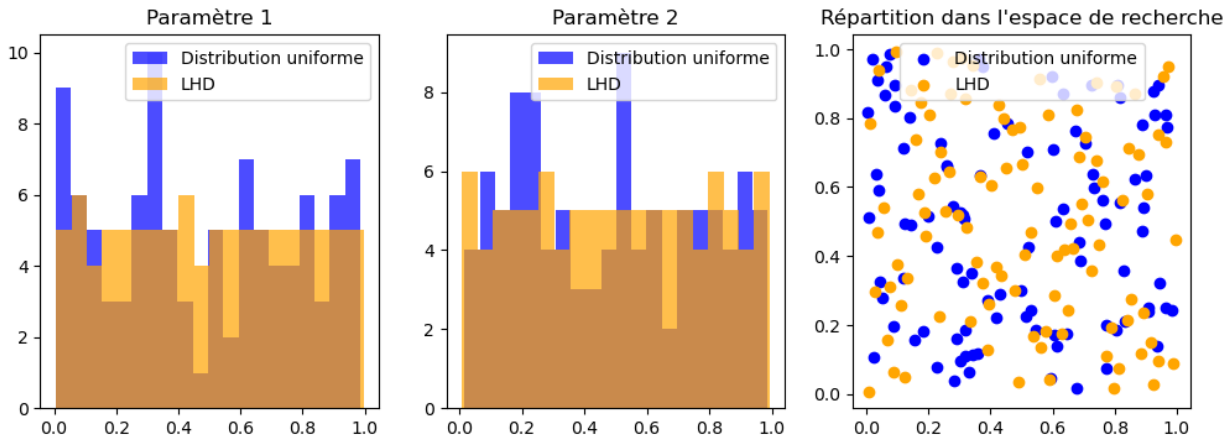

Figure 5.10 : Comparaison de performance : LHD contre distribution uniforme

L'optimisation de chacune des méthodes va donc se faire en évaluant chaque couple de paramètres définis de manière successive. Le meilleur couple sera celui qui permettra de minimiser la fonction objectif, à savoir la sMAE. S'il faut définir des flottants, nous allons utiliser le LHD, s'il faut utiliser des entiers, nous allons simplement réaliser une recherche par grille. La deuxième mesure d'erreur servira simplement de point de vigilance pour les planificateurs en les informant que nous n'étions pas capables de répondre aux demandes sur la période de validation et qu'il existe donc un risque de rupture sur les périodes de prévision suivantes.

5.6 Choix de la meilleure méthode de prévision par profil

Jusqu'à présent, nous avons détaillé comment regrouper les articles de l'inventaire par profil de demande grâce à l'utilisation de kmeans. Nous avons ensuite détaillé les méthodes de prévisions disponibles, les mesures d'erreurs permettant d'évaluer leurs performances ainsi que les différentes techniques permettant de trouver les paramètres optimaux pour chacune des méthodes. Ces différentes étapes ont été présentées de manière individuelle par souci de compréhension. Dans cette section, nous allons détailler comment elles vont être amenées à fonctionner ensemble.

Après avoir défini différents groupes d'articles, nous voulons déterminer quelle méthode de prévision permet d'obtenir l'erreur sMAE la plus basse pour chacun d'entre eux. Nous avons d'abord pensé à utiliser un point représentatif du groupe sur lequel déterminer les paramètres optimaux. Ces paramètres seraient ensuite simplement utilisés pour faire la prédiction sur les autres articles du groupe. Cette approche a l'avantage d'être peu coûteuse en temps de calcul puisqu'on n'utilise l'algorithme d'optimisation qu'une seule fois. Cependant, rien ne garantit que les paramètres optimaux sur cet article soient ceux qui permettent d'obtenir l'erreur minimale pour tous les autres articles du cluster. De plus, il faut arriver à déterminer ce qu'on entend par article représentatif.

Par la suite, nous avons pensé utiliser plusieurs articles sélectionnés dans le profil puis choisir le couple de paramètres minimisant l'erreur moyenne sur ces derniers. À nouveau, le problème de la sélection de ces points apparaît : faut-il considérer les points en bordure de cluster, ceux au centre ou des points au hasard ? En répétant une sélection aléatoire un nombre important de fois, on pourrait dire que ces résultats sont représentatifs du cluster.

Si la section précédente nous a appris une chose, c'est bien de commencer au plus simple. C'est pourquoi nous allons simplement essayer tous les couples de paramètres sur chacun des articles du cluster.

Afin de réduire au maximum le temps d'exécution de l'algorithme tout en nous assurant de ne pas avoir à faire de calculs inutiles, nous utilisons un LHD pour générer les couples de paramètres (ou une grille d'entiers lorsqu'il le faut) puis nous calculons les erreurs pour tous les articles de l'inventaire. Nous nous servons ensuite du clustering kmeans pour identifier les articles de chaque groupe. Il ne suffit plus que de calculer les moyennes des erreurs par cluster. De cette manière, on peut faire varier le nombre de groupes sans avoir à refaire le calcul à chaque fois. Une fois que nous avons pu déterminer les couples de paramètres optimaux par profil et par méthode de prévision, une comparaison des résultats de ces dernières permet de choisir la plus performante pour chaque groupe.

Comment déterminer l'optimal par cluster ?

Dans les paragraphes précédents, on parle souvent d'optimum par cluster sans avoir pour autant défini de quoi il s'agit. Encore une fois, il n'existe pas une seule façon de choisir comment évaluer

la performance d'une méthode sur un groupe. Parmi les indicateurs les plus utilisés en statistiques, on peut citer la médiane, le mode ou encore la moyenne.

La médiane est la valeur qui partage une série statistique ordonnée en deux groupes de même effectif. Cette mesure peut être efficace dans le cas où les erreurs par cluster sont très variées. Cependant, dans notre cas d'étude, des méthodes de prédiction permettent une erreur de prédiction nulle sur certaines séries. Dans quelques clusters, la médiane est à 0. Il est alors impossible de comparer deux méthodes ayant toutes les deux une médiane à 0 en utilisant ce seul indicateur.

Le mode correspond à la valeur possédant l'effectif le plus élevé dans une série statistique. Le problème avec notre cas d'étude est que chacune des séries temporelles, malgré qu'elle puisse être regroupée avec d'autres au sein d'un même profil de demande, possède une consommation qui lui est propre sur la période d'évaluation. Ainsi, sur la plupart des clusters, comme presque toutes les erreurs sont différentes, le mode identifie des valeurs n'apparaissant qu'un nombre très faible de fois. Le mode n'est donc pas adapté pour notre étude.

La moyenne géométrique correspond à la racine n-ième des produits des erreurs du cluster. Elle ne peut pas être utilisée, car elle est incalculable lorsque l'erreur est nulle pour ne serait-ce qu'un seul article du cluster. La moyenne arithmétique est la somme des erreurs du cluster divisé par le nombre de ces valeurs. Un de ses principaux inconvénients est qu'elle n'est pas vraiment représentative d'une série statistique lorsque cette dernière ne suit pas une distribution normale. En effet, une moyenne peut être identique pour deux séries de nature très différentes. Elle reste néanmoins l'approche la plus cohérente avec nos besoins. La figure 5.11 illustre ce problème grâce à une distribution normale (à gauche) et une distribution exponentielle (à droite) possédant toutes les deux une moyenne à 5.04.

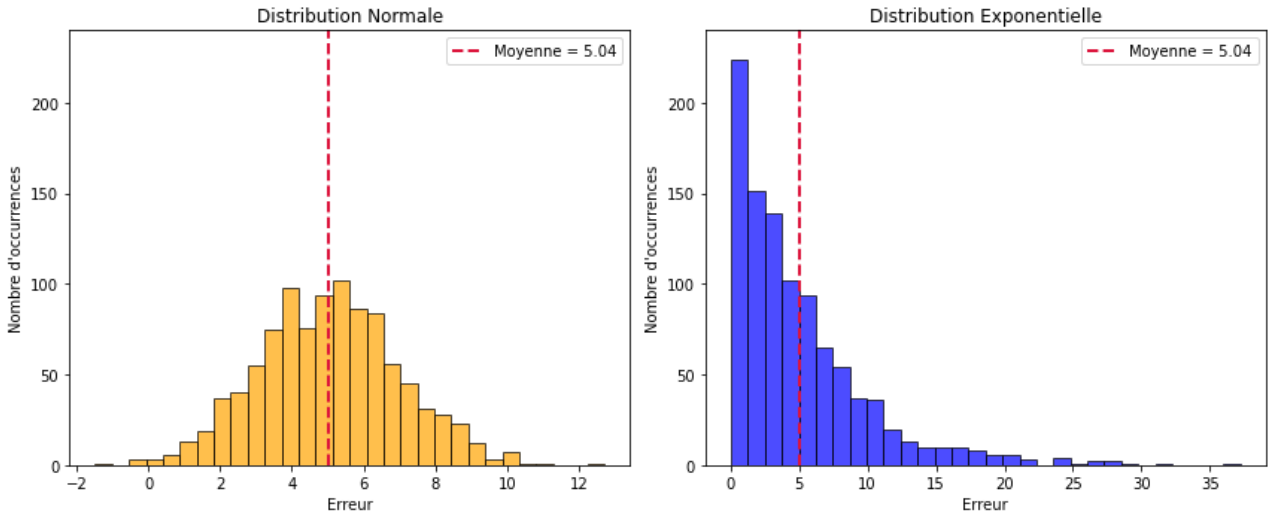

Figure 5.11 : Exemples de moyenne identiques avec différentes distributions

Comme nous avons pu en discuter dans la section sur les techniques d'optimisation, nous utilisons la mesure d'erreur sMAE comme fonction objectif, c'est donc une moyenne de cette mesure d'erreur par article qui permet d'évaluer la performance d'un couple de paramètres et d'une méthode de prévision sur l'ensemble du cluster.

5.7 Conclusion

Ce chapitre reflète l'ampleur et la complexité des étapes entreprises pour élaborer une méthodologie de prévision robuste et adaptée à notre cas d'étude. Il a permis de mettre en évidence les étapes de préparation des données, regroupement par profil de demande, calcul des prévisions, choix des mesures d'erreur et des méthodes d'optimisation pour obtenir des résultats significatifs.

La préparation des données a été une étape fondamentale, impliquant plusieurs phases comme le filtrage des articles actifs, le nettoyage de la matrice de consommation et l'identification des cas de démarrage à froid. Le regroupement des articles par profil de demande a été réalisé avec la méthode de clustering Kmeans dans le plan ADI et CV^2, offrant ainsi une approche encore peu explorée pour définir des catégories de gestion pertinentes dans le domaine de la maintenance. Cette approche a permis de remettre en question les seuils établis par la littérature, offrant ainsi une perspective enrichissante sur la segmentation de l'inventaire. Le calcul des prévisions a été effectué à l'aide d'un comparatif comprenant 14 méthodes de prévision, permettant ainsi une comparaison

approfondie de leurs performances. Après plusieurs essais, la sélection de la mesure d'erreur sMAE s'est avérée la plus pertinente, bien que la définition d'une mesure basée sur le coût des prévisions et du stockage ait été tentée sans succès. En complément de la mesure d'erreur principale, une mesure secondaire a été définie pour informer les planificateurs d'un potentiel de rupture, avec des seuils adaptés à la criticité de la pièce. Cette approche holistique de l'évaluation des performances vise à fournir une vision plus complète de la fiabilité des prévisions. Le choix de la méthode d'optimisation par échantillonnage latin hypercube est basé sur sa capacité à explorer efficacement l'espace des paramètres. Cette approche s'est révélée être un compromis idéal entre la complexité et la performance des prévisions sur des données incertaines. Enfin, le travail sur la meilleure façon d'évaluer les performances par cluster, en privilégiant une moyenne plutôt que d'autres mesures statistiques, a permis une évaluation plus représentative et robuste des résultats obtenus.

En somme, ce chapitre met en lumière les défis et les choix stratégiques effectués pour élaborer une méthodologie de prévision rigoureuse et adaptée, ouvrant la voie à des perspectives prometteuses pour son déploiement et son utilisation dans un contexte opérationnel. Après avoir défini notre approche dans ce chapitre, nous allons en illustrer les performances dans le chapitre suivant.

CHAPITRE 6 RÉSULTATS ET DISCUSSIONS

Dans cette section, nous allons explorer les résultats de la méthodologie proposée par rapport à la littérature et comparer les performances des différentes méthodes de prévisions sur notre cas d'étude.

6.1 Résultats

Nous avons choisi de faire les calculs pour un nombre de clusters entre trois et dix. La valeur basse est utilisée pour comparer les performances de notre méthodologie par rapport à la littérature, la valeur haute est définie par le nombre approximatif de planificateurs disponibles à la STM. On suppose donc qu'au maximum, on peut attribuer un profil à chaque planificateur. Les valeurs intermédiaires permettent d'identifier la présence d'une tendance en fonction du nombre de profils identifiés.

Le Tableau 6.1 présente les résultats de la littérature sur les différents profils identifiés. Pour rappel, Syntetos et al. (2005) avaient défini des seuils permettant de segmenter un inventaire en quatre groupes en fonction des domaines de performance optimale de deux méthodes de prévision : Croston et sa variante SBA. On observe que les prévisions sur les profils Lisse et Erratique présentent une erreur presque quatre fois plus faible sur le profil Grumeleux. Ces résultats semblent cohérents puisque le dernier profil regroupe les articles aux demandes les plus problématiques, à savoir une très grande variation de quantités commandées ainsi qu'une grande variation des délais entre deux demandes. Aussi, on peut constater que cette classification éprouve des difficultés à découper l'inventaire en groupes de tailles similaires. Les profils Lisse et Erratique regroupent respectivement 111 et 206 articles tandis que le profil Grumeleux contient 8 952 articles. D'un point de vue gestion d'inventaire, il est presque inutile de se donner la peine de séparer en différents profils de gestion de cette manière puisqu'un groupe contient plus de 96% de l'inventaire.

Tableau 6.1 : Performances de la littérature sur l'inventaire

Profil	Méthode de prévision	Erreur moyenne	Nombre d'articles	Articles ne respectant pas la fonction contrainte
Lisse	Croston	0,52	111	0
Erratique	SBA	0,55	206	7
Grumeleux	SBA	2,27	8952	14
Total	Littérature	2,21	9269	21

Nous allons ensuite pouvoir évaluer la performance de la méthode proposée face aux résultats de la littérature. Le Tableau 6.2 présente nos résultats pour un nombre de profils identiques à ce que la littérature a identifié dans notre inventaire. Nous avons utilisé les 91 premiers mois d'historique pour initialiser les méthodes de prévision et nous avons calculé les erreurs de prévision sur les 3 mois suivants. Cet horizon de prévision correspond aux besoins des opérations de notre partenaire. La première observation est que la méthode ARIMA semble être la plus performante pour chacun des trois profils avec une erreur moyenne par cluster valant au maximum 0,88. L'erreur moyenne par article pour tout l'inventaire vaut 0,67, soit une réduction de presque 70% comparée à la littérature. Le nombre d'articles est quant à lui plus équilibré au sein de chacun des profils, ce qui facilitera les pratiques de gestion comme la répartition des articles parmi les planificateurs. Le nombre d'articles ne respectant pas la fonction contrainte augmente de 21 à 33. En effet, comme on cherche à réduire au maximum la surconsommation d'articles, on peut créer quelques cas de rupture potentielle supplémentaires. L'impact de cette augmentation sur les opérations est difficile à déterminer avec une simple étude statistique et on suppose que la réduction de 70% des erreurs de prédiction permettra une économie de coût suffisante pour choisir notre méthodologie plutôt que la littérature.

Tableau 6.2 : Performances de notre méthodologie avec 3 clusters

Profil	Méthode de prévision	Erreur moyenne	Nombre d'articles	Articles ne respectant pas la fonction contrainte
Profil 1	ARIMA	0,88	2848	0
Profil 2	ARIMA	0,68	2257	0
Profil 3	ARIMA	0,52	4164	33
Total	Kmeans 3	0,67	9269	33

Nous avons continué l'exploration en faisant varier le nombre de clusters de trois à dix afin de chercher s'il existe un nombre de clusters optimal permettant de minimiser l'erreur. Ces résultats sont présentés dans le tableau 6.3. L'erreur moyenne par article diminue lorsque le nombre de profils augmente de trois à six, mais il semble stagner, voire augmenter au-dessus de six profils. Pour notre cas d'étude, il semble que six groupes soient le nombre optimal. Rien n'indique que l'erreur ne continuerait pas de baisser pour un nombre de profils plus important, mais on suppose qu'au-dessus de 10 profils, les contraintes de gestion deviendraient trop importantes. La figure 6.1

permet d'illustrer cette évolution plus visuellement. La méthodologie proposée permet de diminuer entre 69% et 72% l'erreur moyenne par rapport à la littérature.

Tableau 6.3 : Évolution des erreurs en fonction du nombre de clusters

Nombre de profils	Erreur moyenne par article
3 (littérature)	2,21
3	0,67
4	0,64
5	0,63
6	0,62
7	0,64
8	0,64
9	0,64
10	0,64

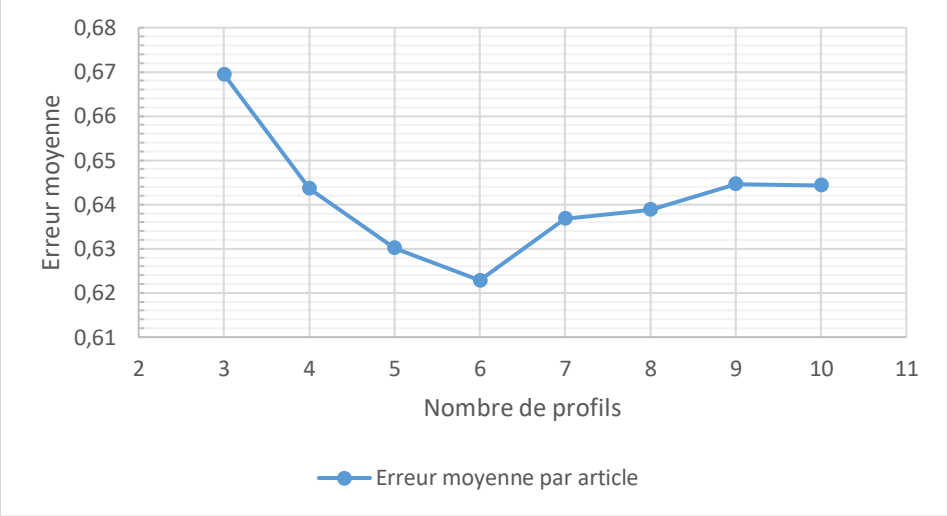

Figure 6.1 : Performance de la littérature par rapport à la méthodologie proposée

Au cours de nos lectures, nous avons remarqué que le temps de calcul était problématique pour certaines approches. En effet, certains algorithmes étaient déjà très longs à exécuter pour une dizaine voire une centaine d'articles. Leur utilisation pour la gestion d'un inventaire complet de dizaines de milliers de références peut poser problème. Nous avons donc mesuré les temps d'exécution pour chacun des algorithmes de prédiction pour évaluer leur pertinence dans un contexte industriel. Le tableau 6.4 présente ces résultats de temps de calcul en minutes. Ces temps de calcul peuvent être variables étant donné qu'ils dépendent de la capacité de calcul de chaque

ordinateur ainsi que de la charge processeur au moment de lancer les algorithmes. On peut néanmoins remarquer que certaines méthodes comme les moyennes ou la prédiction naïve sont particulièrement rapides. Le temps de calcul pour la moyenne mobile pondérée est plus important, car nous avons essayé 1 950 couples de paramètres (nombre de périodes à considérer ainsi que les poids de ces différentes périodes entre deux et six mois). Les méthodes les plus longues à exécuter sont Holt-Winters dans leurs variantes additives et multiplicatives suivies par ARIMA. Dans notre cas, comme les prévisions vont être effectuées au maximum une fois par mois, les 54h de calcul au total ne sont pas un obstacle à l'implantation de notre méthodologie dans les opérations.

Tableau 6.4 : Temps de calcul des différentes méthodes de prédiction

Méthode de prédiction	Temps de calcul (minutes)
Moyenne	0.1
Naive	0.23
Moyenne mobile	4
SVM	12
Croston SBJ	18
Croston original	19
Croston SBA	19
Holt	27
Croston TSB	61
Lissage exponentiel simple	64
Moyenne mobile pondérée	80
Holt amorti	198
ARIMA	507
Holt Winters multiplicatif	1089
Holt Winters additif	1116
Total	3214 (54 heures)

6.2 Contraintes matérielles

Dans tout projet travaillant avec un ensemble de données volumineux, les contraintes matérielles peuvent jouer un rôle significatif dans le choix des techniques employées. Bien que l'optimisation du code, la création de pipelines de calcul ou le déploiement à grande échelle d'algorithmes

nécessitent une expertise particulière en génie logiciel, nous avons pu mettre en œuvre quelques principes simples pour permettre l'exécution de notre code sur un ordinateur portable milieu de gamme. Notre matériel possède un processeur 16 cœurs de fréquence 2.2GHZ ainsi que 16GB de RAM.

Afin d'accélérer les calculs pour rendre viable notre approche, nous avons exploré plusieurs possibilités : améliorer notre algorithme pour ne pas faire de calculs inutiles, utiliser du calcul parallèle ou encore utiliser des services de calcul cloud.

Pour réduire le nombre de calculs, nous avons cherché à prendre du recul sur la façon de procéder. Lorsqu'on fait un clustering puis qu'on optimise toutes les méthodes par cluster, nous allons dans tous les cas devoir effectuer les calculs sur chacun des articles de l'inventaire. Cette approche nous oblige à refaire les calculs d'optimisation pour chaque nouvelle valeur de kmeans. Pour avoir à faire les calculs une seule fois, nous allons simplement les faire avant de faire le clustering puis sauvegarder les résultats pour chaque couple de paramètres. De cette façon, on peut ensuite identifier les articles de chaque cluster avant de calculer leurs erreurs.

Utiliser du calcul parallèle nécessite une certaine mise en forme du code, mais nous a permis de diviser le temps de calcul jusqu'à 16 fois. Il est important de réfléchir aux opérations à paralléliser pour que le programme puisse s'exécuter correctement. Par exemple, traiter les calculs par cluster en parallèle est beaucoup moins efficace que traiter les articles en parallèle. Dans le premier cas, il faut attendre que les opérations parallèles soient terminées pour passer au calcul suivant. Si un cluster contient beaucoup moins d'articles qu'un autre, il faudra attendre la fin des opérations sur le plus gros avant de continuer. En travaillant par articles, cette contrainte disparaît puisque tous les calculs ont un temps très similaire pour différents paramètres d'une même méthode de prévision. Nous n'avons pas utilisé de calcul sur carte graphique puisqu'il nécessite une architecture plus complexe et que nous ne disposons pas de tel matériel.

Nous n'avons pas utilisé de ressources de calcul cloud puisque nous cherchions à produire une méthode exécutable sur un ordinateur qui pourrait être disponible au niveau d'un planificateur. Cependant, comme nos algorithmes utilisent du calcul parallèle, ils pourraient potentiellement être exécutés par des serveurs de calcul cloud pour une utilisation non plus à l'échelle d'une division, mais de l'entreprise entière et de son inventaire de plus de 30 000 références.

D’autre part, une attention toute particulière a dû être apportée à la gestion de la mémoire vive ou RAM. Dès qu’on va vouloir exécuter un programme python, il faut pouvoir garder à portée de calcul toute information utile en la chargeant dans la RAM. Une fois que les calculs pour une méthode de prévision ont été effectués, nous avons pris soin d’enregistrer les résultats puis de libérer l’espace mémoire pour la méthode suivante. Cette gestion intelligente nous a permis de pouvoir lancer tous nos algorithmes successivement à partir d’un programme « maître ».

6.3 Perspectives de déploiement chez le partenaire

Au cours de ce projet, nous avons eu la chance de travailler en collaboration très étroite avec un partenaire industriel. En plus de développer une méthodologie en accord avec leurs contraintes et principes de gestion, nous avons pu réfléchir à son opérationnalisation.

Les planificateurs ont l’habitude de se répartir les articles par domaine d’expertise. Par exemple, l’un va s’occuper de la visserie, un autre des pneumatiques, etc… Cette façon de fonctionner à l’avantage d’utiliser leurs compétences en gestion des contrats, délais fournisseurs ou encore délais d’approvisionnement sur des articles avec lesquels ils ont l’habitude de travailler. On pourrait alors créer des catégories de profil de demande au cœur de chaque domaine d’expertise. En fonction de ces catégories et des résultats des différentes mesures d’erreur, les planificateurs vont pouvoir juger du niveau de confiance à accorder à des prévisions qui ne sont pas dénuées de biais.

Changer cette façon de fonctionner pour travailler par profil de demande peut leur permettre de se spécialiser sur des comportements particuliers qui nécessiteraient une gestion de contrat spécifique ou des discussions avec les fournisseurs sur les quantités ou délais de livraison. La dernière approche demande néanmoins un effort important puisqu’il requiert un changement majeur des habitudes de chacun.

Pour leur simplifier la tâche, il pourrait être intéressant de déployer des pipelines de calcul au cœur d’un lac de données. Avec ce type de procédés, les planificateurs n’auraient qu’à utiliser les indicateurs et les prévisions depuis une seule interface PowerBI par exemple. Les modifications et améliorations du code python seraient externalisées vers des équipes spécialisées dans la gestion logicielle. Aussi le fonctionnement dans un lac de données permet un accès à de nombreuses données supplémentaires comme des évolutions de flottes, délais fournisseurs, contrats et autres. Il peut même permettre une mise en place de maintenance prédictive pour les cas où l’étude des séries temporelles se montre particulièrement inefficace.

6.4 Conclusion

Cette section a permis d'explorer les résultats obtenus à travers la méthodologie proposée et de les confronter à ceux de la littérature, tout en examinant les performances des différentes méthodes de prévision dans le cadre de l'étude. Les résultats ont révélé que la méthodologie présentée offre des performances nettement améliorées par rapport à la littérature, notamment en réduisant l'erreur moyenne de prédiction de près de 70%. De plus, une analyse a été menée pour déterminer le nombre optimal de clusters, aboutissant à la conclusion que six groupes semblent être le nombre idéal pour minimiser l'erreur de prédiction dans notre cas d’étude.

Par ailleurs, des considérations matérielles ont été prises en compte, notamment en évaluant les temps de calcul des différentes méthodes de prévision, ce qui a permis de s'assurer de la viabilité de l'approche proposée dans un contexte industriel. Des stratégies pour accélérer les calculs ont été envisagées, telles que l'utilisation de calculs parallèles ou l'optimisation du code.

Enfin, des perspectives de déploiement chez le partenaire industriel ont été discutées, mettant en avant la possibilité d'intégrer la méthodologie dans un environnement opérationnel existant. Malgré les défis et les limites rencontrés tout au long de l'étude, cette section a mis en lumière des avancées réalisées dans la prédiction de la demande et a ouvert la voie à de futures améliorations et applications dans le domaine de la gestion d'inventaire.

CHAPITRE 7 CONCLUSION ET RECOMMANDATIONS

Ce dernier chapitre rappelle le cadre général de notre étude, l'étendue de notre contribution ainsi que les perspectives à plus long terme de ce type de méthodologies.

Au cours de notre projet de recherche dédié à l'amélioration de la prévision de la demande grumeleuse en pièces détachées, nous avons pu mettre en lumière les nombreux défis rencontrés, tant matériels que stratégiques.

La première phase du projet a été consacrée à la collecte et au traitement des données, visant à garantir leur cohérence et leur fiabilité. Ensuite, nous avons utilisé le clustering Kmeans pour une répartition plus équitable de l'inventaire au sein des groupes, facilitant ainsi la gestion et la planification dans les opérations. Cette méthode de clustering a aussi permis de définir un nombre de groupe optimal pour minimiser l'erreur de prévision (sous objectif 1). L'optimisation des méthodes de prévision préalablement identifiées dans la littérature (sous objectif 2), souvent négligée dans la littérature, a été menée afin d'identifier la technique la plus adaptée à notre cas d'étude (sous objectif 4). La comparaison de diverses méthodes couramment utilisées dans la littérature a ainsi constitué un élément essentiel de notre projet. Cette étape a permis d'affiner les prévisions en fonction des caractéristiques spécifiques des groupes identifiés, améliorant ainsi la précision globale du processus de prévision. La détermination d'une mesure d'erreur cohérente avec les contraintes du partenaire a été explorée pour évaluer l'efficacité des prévisions de notre modèle (sous objectif 3). Cela a non seulement fourni une base objective pour évaluer les performances, mais a également permis une comparaison à grande échelle de méthodes de prévisions. Enfin, la gestion des exceptions pour coller au mieux aux besoins d'affaires souligne la pertinence et l'applicabilité pratique de la méthode proposée. Au final, cette méthodologie est parfaitement alignée avec les pratiques d'affaires et les systèmes d'information de notre partenaire industriel. Les contributions se mesurent tant en termes d'efficacité opérationnelle, grâce à la réduction du nombre de paramètres de gestion, que de précision dans la prévision de la demande en pièces détachées. Les résultats obtenus montrent d'ailleurs une amélioration significative de la performance de prévision par profil de demande.

Au cours de notre étude, nous avons remarqué de nombreuses limitations ou améliorations possibles pour certaines étapes.

Premièrement, lors de l'étape de filtrage des données, nous avons déterminé un seuil à partir duquel on considère que le nombre d'observations est insuffisant. Nous avons arbitrairement défini la valeur de ce seuil à deux, mais on peut supposer que trois ou quatre consommations sur 4 ans ne suffisent pas non plus à avoir une prévision cohérente avec l'étude de séries temporelles. Il pourrait être intéressant d'étudier le nombre d'observations en fonction du type de prédiction utilisée (Réseaux de neurones, méthodes paramétriques, maintenance prédictive…).

Ensuite, nous avons pu identifier des limitations sur la caractérisation de la demande dans le plan ADI / CV^2. Comme l'illustre le Tableau 7.1, deux séries temporelles vont avoir le même ADI et le même CV^2 lorsqu'elles possèdent le même nombre de demandes et les mêmes quantités par demande, mais ces caractéristiques ne changent pas en fonction de l'ordre de la demande. Cette caractérisation ne prend pas en compte le cycle de vie d'un article. La première série temporelle présente une demande seulement à partir de la quatrième période tandis que la deuxième série présente une demande nulle entre chaque demande. Une classification dans ce plan peut regrouper deux articles qui vont évoluer de manière différente.

Tableau 7.1 : Limite de la caractérisation ADI CV^2

Série temporelle	ADI	CV^2
[0 0 0 1 2 1]	2	1.25
[2 0 1 0 1 0]	2	1.25

Aussi, on suppose que le clustering avec Kmeans permet de séparer efficacement les articles, mais rien ne nous a permis de conclure sur une influence directe entre la métrique de ressemblance utilisée et la performance des méthodes de prévision. On a simplement remarqué que cette approche fonctionne mieux que la littérature dans notre cas d'étude par sa capacité à regrouper de façon plus équilibrée les articles dans les différents groupes (sous objectif 1). Il pourrait être intéressant de changer la métrique de ressemblance ou même d'utiliser une autre technique de clustering.

D'autre part, notre méthodologie a mis en œuvre un nombre fini de méthodes de prédiction. La littérature regorge de méthodes qui pourraient s'avérer plus efficaces. Nous n'avons pas pu toutes

les essayer notamment à cause de leur complexité de mise en œuvre. Certaines nécessitent des informations complémentaires à l'étude des consommations et d'autres sont tellement précises qu'elles pourraient être des projets de recherche à part entière. On pense notamment aux différentes technologies de réseaux de neurones. Dans une optique d'amélioration de notre méthodologie, une étape de stacking pourrait être ajoutée avant de réaliser les prévisions.

Lors de l'étape d'optimisation, nous avons utilisé un LHD pour répartir au mieux possible les paramètres dans l'espace de recherche. Malgré cela, nous avons dû faire un choix sur le nombre de points à placer. Ce nombre repose sur une étude « à l'œil » sur un pas qui ne change pas de façon considérable les résultats de prévision entre deux paramètres successifs. On pourrait par exemple améliorer la recherche d'optimum en redéfinissant un LHD avec un maillage plus fin autour du meilleur couple de paramètres trouvé lors du balayage préliminaire.

Finalement, nous avons remarqué un changement de niveau de détail entre certaines méthodes. Notre objectif est de réduire le nombre de paramètres et de méthodes à un couple « hyperparamètres optimaux – Méthode » par cluster. Pour une méthode comme ARIMA, les couples d'entiers en entrée définissent successivement le nombre d'observations à considérer, le nombre de soustractions successives pour rendre la série stationnaire et le nombre de termes d'erreur. Néanmoins, les nombres d'observations et d'erreurs sont pondérés et une optimisation s'effectue à l'échelle de chaque série temporelle. Donc même si, à première vue, les paramètres de prévision sont les mêmes, une optimisation au niveau de chaque série temporelle est effectuée. Cela pourrait expliquer la meilleure performance de cette méthode de prévision par rapport aux autres. Nous avons essayé de définir ces poids sur un article représentatif du cluster pour ensuite les utiliser pour la prédiction des autres articles du groupe. Nous avons choisi le centroïde de chaque cluster. Cette fonction disponible dans l'algorithme Kmeans de la librairie utilisée donne les coordonnées d'un point « fictif » dont les abscisse et ordonnée viennent d'une moyenne des coordonnées de chaque article du groupe. Pour faire l'optimisation, nous avons choisi l'article ayant les coordonnées les plus proches du centroïde selon une distance euclidienne. Après avoir optimisé ARIMA sur cet article, on utilise ses paramètres sur le reste du cluster. Les performances de la méthode ne semblent pas donner plus de logique en fonction du nombre de groupes. L'erreur la plus faible est celle d'un kmeans avec trois groupes. Cette observation semble illustrer que le point choisi pour l'optimisation n'est pas vraiment représentatif et que les performances de chaque prévision sont très variables. Aussi, on ne peut plus faire les calculs avant de faire le clustering comme expliqué

dans la section 6.2 puisqu'il dépend du nombre de groupes. Avec cette approche, on doit relancer dix fois le calcul pour comparer les dix nombres de clusters.

Une utilisation conjointe de notre méthodologie et de la maintenance prédictive nous apparaît comme une solution d'avenir dans une industrie de plus en plus portée par la valorisation de données et l'intelligence artificielle, notamment pour les nouveaux articles et ceux présentant des comportements trop complexes pour l'étude des séries temporelles.

RÉFÉRENCES

Babai, M., Dallery, Y., Selmen, B., et Kalai, R. (2018). A new method to forecast intermittent demand in the presence of inventory obsolescence. *International Journal of Production Economics, 209, 30-41*. https://doi.org/10.1016/j.ijpe.2018.01.026

Babaveisi, V., Teimoury, E., Gholamian, M. R., et Rostami-Tabar, B. (2022). Integrated demand forecasting and planning model for repairable spare part: an empirical investigation. *International Journal of Production Research*. 61(20), 6791-6807. https://doi.org/10.1080/00207543.2022.2137596

Bergman, J. J., Noble, J. S., McGarvey, R. G., et Bradley, R. L. (2017). A Bayesian approach to demand forecasting for new equipment programs. *Robotics and Computer-Integrated Manufacturing, 47,* 17-21. https://doi.org/10.1016/j.rcim.2016.12.010

Blessing, L., et Chakrabarti, A. (2009). *DRM*, a Design Research Methodology. Dans *Springer eBooks*. https://doi.org/10.1007/978-1-84882-587-1

STM (2024), *Bus Électriques*. Tiré de https://www.stm.info/fr/electrification/bus-electriques#id_premiere, [consulté le 08 janvier 2024].

Chiou, H., Tzeng, G., Cheng, C. C., & Liu, G. (2004). Grey prediction model for forecasting the planning material of equipment spare parts in Navy of Taiwan, *Soft Computing with Industrial Applications - Proceedings of the Sixth Biannual World Automation Congress*, 17, 315-320. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1439385

Elsevier (2024), *Scopus Content*, Tiré de https://www.elsevier.com/solutions/scopus/how-scopus-works/content, [consulté le 05 mars 2024].

do Rego, J. R., et de Mesquita, M. A. (2011). Spare parts inventory control: a literature review. *Producao, 21*(4), 645-655. https://doi.org/10.1590/S0103-65132011005000002

Frazzon, E. M., Albrecht, A., Israel, E., Hellingrath, B., Cordes, A. K., & Saalmann, P. (2014). Simulation model concept for evaluating spare parts supply chain planning methods, *2014 12th IEEE International Conference on Industrial Informatics (INDIN)*, 560-565. https://doi.org/10.1109/INDIN.2014.6945574

Ghobbar, A. A., et Friend, C. H. (2002). Sources of intermittent demand for aircraft spare parts within airline operations. *Journal of Air Transport Management, 8*(4), 221-231. https://doi.org/10.1016/S0969-6997(01)00054-0

Ghobbar, A. A., et Friend, C. H. (2003). Evaluation of forecasting methods for intermittent parts demand in the field of aviation: A predictive model. *Computers and Operations Research, 30*(14), 2097-2114. https://doi.org/10.1016/S0305-0548(02)00125-9

Huskova, K., et Dyntar, J. (2023). Speeding up past stock movement simulation in sporadic demand inventory control. *International Journal of Simulation Modelling, 22*(1), 41-51. https://doi.org/10.2507/IJSIMM22-1-627

Ikotun, A., Ezugwu, A., Abualigah, L., Abuhaija, B., et Heming, J. (2022). K-means Clustering Algorithms: A Comprehensive Review, Variants Analysis, and Advances in the Era of Big Data. *Information Sciences, 622, 178-210*. https://doi.org/10.1016/j.ins.2022.11.139

Pennings, C. L. P., van Dalen, J., et van der Laan, E. A. (2017). Exploiting elapsed time for managing intermittent demand for spare parts. *European Journal of Operational Research, 258*(3), 958-969. https://doi.org/10.1016/j.ejor.2016.09.017

Le Robert (2023), *prévision - Définitions, synonymes, conjugaison, exemples | Dico en ligne Le Robert*. Tiré de https://dictionnaire.lerobert.com/definition/prevision, [consulté le 26 novembre 2023]

STM (2022), *Rapport annuel STM*. Tiré de https://www.stm.info/fr/a-propos/informations-entreprise-et-financieres/rapport-annuel-2021#id_troisieme, [consulté le 05 mars 2024].

Romeijnders, W., Teunter, R., et Van Jaarsveld, W. (2012). A two-step method for forecasting spare parts demand using information on component repairs. *European Journal of Operational Research, 220*(2), 386-393. https://doi.org/10.1016/j.ejor.2012.01.019

Şahin, M., Eldemir, F., et Turkyilmaz, A. (2022). Inventory Cost Minimization of Spare Parts in Aviation Industry. Transportation Research Procedia, 59, 29-37. https://doi.org/10.1016/j.trpro.2021.11.094

Sareminia, S., et Amini, M. (2023). A reliable and ensemble forecasting model for slow-moving and repairable spare parts: Data mining approach. *Computers in Industry, 145*, Article 103827. https://doi.org/10.1016/j.compind.2022.103827

Seabold, S., et Perktold, J. (2010). Statsmodels : Econometric and statistical modeling with python. *Proceedings Of The 9th Python In Science Conferences*. https://doi.org/10.25080/majora-92bf1922-011

Syntetos, A. A., et Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting, 21(2),* 303-314. https://doi.org/10.1016/j.ijforecast.2004.10.001

Syntetos, A. A., Boylan, J. E., et Croston, J. D. (2005). On the categorization of demand patterns. *Journal of the Operational Research Society, 56*(5), 495-503. https://doi.org/10.1057/palgrave.jors.2601841

Office québécois de la langue française (1981), *taux de service*. Tiré de https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/2096398/taux-de-service, [consulté le 05 mars 2024].

Teunter, R. H., Babai, M.Z. et Syntetos, A.A. (2010), ABC Classification: Service Levels and Inventory Costs. *Production and Operations Management, 19,* 343-352. https://doi.org/10.1111/j.1937-5956.2009.01098.x

Teunter, R. H., et Duncan, L. (2009). Forecasting intermittent demand: A comparative study. *Journal of the Operational Research Society, 60(3),* 321-329. https://doi.org/10.1057/palgrave.jors.2602569

Van der Auweraer, S., et Boute, R. N. (2019). Forecasting spare part demand using service maintenance information. *International Journal of Production Economics, 213,* 138-149. https://doi.org/10.1016/j.ijpe.2019.03.015

Van der Auweraer, S., Boute, R. N., et Syntetos, A. A. (2019). Forecasting spare part demand with installed base information: A review. *International Journal of Forecasting, 35*(1), 181-196. https://doi.org/10.1016/j.ijforecast.2018.09.002

Verhagen, W. J. C., et Curran, R. (2014). Stochastic forecasting of lumpy-distributed aircraft spare parts demand, *Advances in Transdisciplinary Engineering*, 1, 706-715. http://dx.doi.org/10.3233/978-1-61499-440-4-706

Wang, H., et Hart, B. (2023). An optimal maintenance spare parts prediction model and its complex applications, *2023 Annual Reliability and Maintainability Symposium (RAMS)*, 1-7. https://doi.org/10.1109/RAMS51473.2023.10088217

Wang, W., et Syntetos, A. A. (2011). Spare parts demand: Linking forecasting to equipment maintenance. *Transportation Research Part E: Logistics and Transportation Review, 47*(6), 1194-1209. https://doi.org/10.1016/j.tre.2011.04.008

Wei, Z., Chen, G., Xu, L., et Gao, C. (2022). Prediction of Maintenance Spare Parts Based on BP Neural Network Optimized by Improved Sparrow Search Algorithm, *2022 IEEE 5th International Conference on Information Systems and Computer Aided Education, (ICISCAE),* (p. 872-878). https://doi.org/10.1109/iciscae55891.2022.9927513

Williams, T. M. (1984). Stock Control with Sporadic and Slow-Moving Demand. *The Journal of the Operational Research Society, 35*(10), 939-948. https://doi.org/10.2307/2582137

Zhang, F., et Guan, Z. (2016). Inventory management for spare parts on engineering machinery. *Proceedings - 2015 Chinese Automation Congress (CAC)*, 1951-1955. https://doi.org/10.1109/CAC.2015.7382824

Zhu, S., Jaarsveld, W. V., et Dekker, R. (2020). Spare parts inventory control based on maintenance planning. *Reliability Engineering and System Safety, 193*, Article 106600. https://doi.org/10.1016/j.ress.2019.106600

ANNEXE A DÉTAIL DES MÉTHODES DE PRÉVISIONS UTILISÉES

Méthode ARIMA

Le terme ARIMA est l'acronyme de « AutoRegressive Integrated Moving Average ». Cette méthode peut être décomposée en 3 composantes :

Une composante autorégressive qui cherche à prévoir la demande future comme une combinaison linéaire des observations passées. Elle est notée AR(p) avec p le nombre d'observations considérées. Lorsque la composante AR est la seule considérée, sa formule est :

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t$$

Avec Y_t la valeur de la série temporelle à l'instant t, c une constante, $\phi_1, \phi_2, \ldots, \phi_p$ les coefficients autorégressifs et ϵ_t l'erreur à l'instant t.

Une composante d'intégration qui cherche à rendre la série temporelle stationnaire pour que ses propriétés statistiques restent constantes au cours du temps. Elle est notée I(d) avec d le nombre de soustractions nécessaires avec la valeur de l'observation précédente afin de rendre la série stationnaire. La formule de la composante I est :

$$Y_t^{(d)} = Y_t^{(d-1)} - Y_{t-1}^{(d-1)}$$

Avec $Y_t^{(d)}$ la valeur de la série temporelle différenciée à l'instant t, $Y_t^{(d-1)}$ la valeur précédente de la série temporelle différenciée à l'instant t, $Y_{t-1}^{(d-1)}$ la valeur précédente de la série temporelle différenciée à l'instant t-1.

Une composante de moyenne mobile qui cherche à prédire la demande future comme une combinaison linéaire des erreurs passées. Elle est notée MA(q) avec q le nombre d'erreurs considéré. La formule générale de la composante MA est :

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

Avec Y_t la valeur de la série temporelle à l'instant t, μ la moyenne de la série temporelle, $\epsilon_{t-i}, i \in [0, q]$ les erreurs aux instants t-q et $\theta_1, \theta_2, \ldots, \theta_q$ les coefficients de moyenne mobile.

Finalement, le modèle ARIMA (p, d, q) est défini par la formule générale :

$$Y_t^{'} = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

Avec $Y_t^{'}$ la valeur de la série temporelle rendue stationnaire à l'ordre d, c une constante, $\phi_1, \phi_2, \dots, \phi_p$ les coefficients autorégressifs, $\theta_1, \theta_2, \dots, \theta_q$ les coefficients de moyenne mobile, ϵ_t l'erreur à l'instant t.

L'implémentation de cette méthode en python est disponible dans la bibliothèque statsmodels (Seabold et Perktold, 2010).

Méthode de Croston (originale)

Cette méthode est conçue spécifiquement pour les séries temporelles présentant des comportements intermittents et sporadiques. Elle se caractérise par un lissage séparé d'une composante de demande et d'une composante de fréquence dont les termes ne sont actualisés qu'à l'occurrence d'une demande non nulle.

Composante de demande :

$$Si\ D_t > 0, \qquad z_t = \alpha * D_t + (1 - \alpha) * z_{t-1}$$

$$Sinon, \qquad z_t = z_{t-1}$$

Avec z_t la composante de demande au temps t, D_t la demande réelle au temps t, $0 < \alpha < 1$ le coefficient de lissage.

Composante de fréquence :

$$Si\ D_t > 0, \qquad p_t = \alpha * q_t + (1 - \alpha) * p_{t-1}$$

$$Sinon, \qquad p_t = p_{t-1}$$

Avec p_t la composante de fréquence au temps t, q_t le compteur d'intervalles depuis la dernière demande non nulle au temps t, $0 < \alpha < 1$ le coefficient de lissage.

Prévision finale :

$$F_{t+1} = \frac{z_t}{p_t}$$

Avec F_{t+1}la valeur prévue pour la période t+1, z_t la composante de demande au temps t, p_t la composante de fréquence au temps t.

Méthode de Croston (SBA)

La variante SBA de la méthode de Croston originale est l'acronyme de « Syntetos and Boylan Approximation ». Les auteurs de cette proposition ont montré que la contribution de Croston présentait un biais de prévision. Ils ont alors ajouté un terme de compensation lors du calcul final de la prévision. Les étapes qui précèdent le calcul de la prévision restent identiques :

Prévision finale :

$$F_{t+1} = \left(1 - \frac{\alpha}{2}\right) * \frac{z_t}{p_t}$$

Avec F_{t+1} la valeur prévue pour la période t+1, z_t la composante de demande au temps t, p_t la composante de fréquence au temps t et $0 < \alpha < 1$ le coefficient de lissage.

Méthode de Croston (SBJ)

La variante SBJ de la méthode de Croston originale est l'acronyme de « Shale Boylan Johnston ». À l'image de Syntetos et Boylan, les auteurs de la méthode SBJ ont cherché à corriger le biais de prévision en affinant le terme de compensation. Encore une fois, seul le calcul de la prévision finale est affecté :

Prévision finale :

$$F_{t+1} = \left(1 - \frac{\alpha}{2 - \alpha + \epsilon}\right) * \frac{z_t}{p_t}$$

Avec F_{t+1} la valeur prévue pour la période t+1, z_t la composante de demande au temps t, p_t la composante de fréquence au temps t et $0 < \alpha < 1$ le coefficient de lissage et ϵ un terme d'erreur.

Méthode de Croston (TSB)

La variante TSB de la méthode de Croston originale est l'acronyme de « Teunter, Syntetos et Babai ». À la différence des contributions précédentes autour de la méthode de croston, les auteurs

vont transformer le terme de fréquence en probabilité, l'actualiser à chaque période puis calculer la prévision finale d'une nouvelle manière :

Composante de demande :

$$Si\ D_t > 0, \qquad z_{t+1} = \alpha * D_t + (1 - \alpha) * z_t$$

$$Sinon, \qquad z_{t+1} = z_t$$

Avec z_t la composante de demande au temps t, D_t la demande réelle au temps t, $0 < \alpha < 1$ le coefficient de lissage.

Composante de probabilité :

$$Si\ D_t > 0, \qquad p_{t+1} = \alpha + (1 - \alpha) * p_t$$

$$Sinon, \qquad p_{t+1} = (1 - \alpha) * p_t$$

Avec p_t la composante de probabilité au temps t, $0 < \alpha < 1$ le coefficient de lissage.

Prévision finale :

$$F_{t+1} = z_{t+1} * p_{t+1}$$

Avec F_{t+1} la valeur prévue pour la période t+1, z_{t+1} la composante de demande au temps t+1, p_{t+1} la composante de probabilité au temps t+1.

Pour toutes les approches dérivées de Croston, si la prévision veut être effectuée sur un horizon plus important, la valeur de prévision sera donc simplement copiée sur le nombre de périodes désiré.

<u>Lissage exponentiel simple</u>

Le lissage exponentiel simple est une des méthodes les plus répandues dans l'analyse des séries temporelles. Elle se base sur l'étude des observations passées et des prévisions en les ajustant avec un coefficient de lissage selon la formule :

$$z_t = \alpha * D_{t-1} + (1 - \alpha) * z_{t-1}$$

Avec z_t la composante de demande au temps t, D_t la demande réelle au temps t, $0 < \alpha < 1$ le coefficient de lissage.

Pour une prévision sur un nombre h de périodes, le calcul devient :

$$z_{t+h} = \alpha * D_t + (1 - \alpha) * z_t$$

Lissage exponentiel double (Holt)

La méthode de Holt convient à l'étude des séries temporelles à tendance, mais ne présentant pas de saisonnalité. Elle consiste en un double lissage exponentiel, le premier effectué sur le niveau de demande et le deuxième sur la tendance. Le modèle général est donné par :

$$z_0 = D_0$$

$$a_0 = D_1 - D_0$$

$$z_t = \alpha * D_t + (1 - \alpha) * (z_{t-1} + a_{t-1})$$

$$a_t = \beta * (z_t - z_{t-1}) + (1 - \beta) * a_{t-1}$$

Avec z_t la composante de demande au temps t, D_t la demande réelle au temps t, $0 < \alpha < 1$ le coefficient de lissage de la demande, a_t la composante de tendance au temps t, $0 < \beta < 1$ le coefficient de lissage de la tendance

Pour une prévision sur un nombre h de périodes, le calcul devient :

$$z_{t+h} = z_t + h * a_t$$

Lissage exponentiel double amorti (Holt amorti)

La méthode de Holt amortie convient à l'étude des séries temporelles à tendance, ne présentant pas de saisonnalité, mais une décroissance du facteur de tendance avec le temps. Elle consiste en un double lissage exponentiel, le premier effectué sur le niveau de demande et le deuxième sur la tendance. Un coefficient d'amortissement de la tendance est utilisé pour actualiser le facteur de tendance au cours du temps. Le modèle général est donné par :

$$z_0 = D_0$$

$$a_0 = D_1 - D_0$$

$$z_t = \alpha * D_t + (1 - \alpha) * (z_{t-1} + \phi a_{t-1})$$
$$a_t = \beta * (z_t - z_{t-1}) + (1 - \beta) * \phi a_{t-1}$$

Avec z_t la composante de demande au temps t, D_t la demande réelle au temps t, $0 < \alpha < 1$ le coefficient de lissage de la demande, a_t la composante de tendance au temps t, $0 < \beta < 1$ le coefficient de lissage de la tendance et $0 < \phi < 1$ le coefficient d'amortissement.

Pour une prévision sur un nombre h de périodes, le calcul devient :

$$z_{t+h} = z_t + (\phi + \phi^2 + \cdots + \phi^h) * a_t$$

Lissage exponentiel triple (Holt-winters)

La méthode de Holt-Winters étend la méthode de Holt pour prendre en compte à la fois la tendance et la saisonnalité dans les séries temporelles. Elle convient particulièrement aux données qui présentent une tendance et une saisonnalité. Le modèle général est défini par trois composantes : le niveau (z_t), la tendance (a_t), et la saisonnalité (s_t).

$$z_t = \alpha * \left(D_t - s_{t-p}\right) + (1 - \alpha) * (z_{t-1} + a_{t-1})$$
$$a_t = \beta * (z_t - z_{t-1}) + (1 - \beta) * a_{t-1}$$
$$s_t = \gamma * (D_t - z_{t-1} - a_{t-1}) + (1 - \gamma) * s_{t-p}$$

Avec z_t la composante de demande au temps t, D_t la demande réelle au temps t, $0 < \alpha < 1$ le coefficient de lissage de la demande, a_t la composante de tendance au temps t, $0 < \beta < 1$ le coefficient de lissage de la tendance, s_t la composante saisonnière au temps t, $0 < \gamma < 1$ le coefficient de lissage de saisonnalité et p la période de saisonnalité.

Le calcul de la prévision sur un nombre h de périodes s'effectue avec la formule :

$$F_{t+h} = z_t + h * a_t + s_{t+h-p(k+1)}$$

Ici, $k = \frac{h-1}{p}$. De cette manière, on s'assure d'utiliser la valeur saisonnière correspondant à la dernière année d'observation.

Variantes multiplicatives

Lorsque les valeurs de saisons ou de tendance s'effectuent de manière proportionnelle entre des périodes successives, les calculs de ces deux composantes peuvent être ajustées grâce aux équations de la variante multiplicative :

- Tendance additive et saisonnalité multiplicative

$$z_t = \alpha * \left(\frac{D_t}{s_{t-p}}\right) + (1-\alpha) * (z_{t-1} + a_{t-1})$$

$$a_t = \beta * (z_t - z_{t-1}) + (1-\beta) * a_{t-1}$$

$$s_t = \gamma * \left(\frac{D_t}{z_t + a_t}\right) + (1-\gamma) * s_{t-p}$$

$$F_{t+h} = (z_t + h * a_t) * s_{t+h-p(k+1)}$$

- Tendance multiplicative et saisonnalité additive

$$z_t = \alpha * \left(\frac{D_t}{s_{t-p}}\right) + (1-\alpha) * (z_{t-1} + a_{t-1})$$

$$a_t = \beta * \left(\frac{z_t}{z_{t-1}}\right) + (1-\beta) * a_{t-1}$$

$$s_t = \gamma * (D_t - z_{t-1} - a_{t-1}) + (1-\gamma) * s_{t-p}$$

$$F_{t+h} = (z_t + h * a_t) * s_{t+h-p(k+1)}$$

- Tendance multiplicative et saisonnalité multiplicative

$$z_t = \alpha * \left(\frac{D_t}{s_{t-p}}\right) + (1-\alpha) * (z_{t-1} + a_{t-1})$$

$$a_t = \beta * \left(\frac{z_t}{z_{t-1}}\right) + (1-\beta) * a_{t-1}$$

$$s_t = \gamma * \left(\frac{D_t}{z_t + a_t}\right) + (1-\gamma) * s_{t-p}$$

$$F_{t+h} = (z_t + h * a_t) * s_{t+h-p(k+1)}$$

Méthode moyenne

L'une des méthodes les plus basiques est celle de la moyenne. Elle consiste simplement à calculer la moyenne des consommations sur l'historique et supposer que la consommation de chaque période sur l'horizon aura cette valeur. Voici la formule générale :

$$F_{\mathrm{t+h}} = \frac{1}{t}\sum_{i=1}^{t} D_i$$

Méthode moyenne mobile

Une version plus élaborée de la moyenne consiste à ne considérer que les m dernières occurrences de la demande afin de prendre en compte des tendances récentes :

$$F_{t+h} = \frac{1}{m}\sum_{i=0}^{m-1} Y_{t-i}$$

$$Si\ t < t + h : Y_t = D_t$$

$$Si\ t > t + h : Y_t = F_t$$

Méthode moyenne mobile pondérée

Lorsqu'on suppose une influence plus importante de certaines valeurs sur les prévisions, il est possible de pondérer chaque observation par un poids w_i :

$$F_{t+h} = \frac{\sum_{i=0}^{m-1} w_{t-i} * Y_{t-i}}{\sum_{i=0}^{m-1} w_{t-i}}$$

$$Si\ t < t + h : Y_t = D_t$$

$$Si\ t > t + h : Y_t = F_t$$

Méthode naïve

La méthode de prévision la plus empirique est appelée « méthode naïve ». Elle suppose simplement que la prochaine observation sera égale à la précédente :

$$F_{t+h} = F_{t+h-1}$$

Méthode naïve non nulle

La méthode naïve non nulle peut-être efficace dans un contexte où il est nécessaire de garder un certain nombre de pièces en stock pour être capable de répondre à coup sûr à une demande future en supposant que la dernière demande non nulle Y_t soit représentative de la quantité :

$$F_{t+h} = Y_t$$

ANNEXE B MESURES D'ERREUR RÉPERTORIÉES

Dans le tableau suivant, on note H : horizon de prédiction, N : nombre total d'observations, Y_t : valeur de consommation à la période t, P_t : valeur de la prévision à la période t et $\overline{P_t}$: valeur moyenne de la prédiction.

Tableau B.1 : Formulations mathématiques des mesures d'erreurs

Nom de la mesure	Définition	Formulation mathématique
MAPE	Mean absolute percentage error	$MAPE = \frac{100}{H}\sum_{t=1}^{H} \frac{\lvert Y_t - P_t \rvert}{\lvert Y_t \rvert}$
sMAPE	Symmetric mean absolute percentage error	$sMAPE = \frac{100}{H}\sum_{t=1}^{H} \frac{\lvert Y_t - P_t \rvert}{(\lvert Y_t \rvert + \lvert P_t \rvert)/2}$
MAE	Mean absolute error	$MAE = \sum_{t=1}^{H} \frac{\lvert Y_t - P_t \rvert}{H}$
MASE	Mean absolute scaled error	$MASE = \frac{1}{H}\sum_{h=1}^{H}\left(\frac{\lvert Y_{N+h} - P_h \rvert}{\frac{1}{N-1}\sum_{t=2}^{N}\lvert Y_t - Y_{t-1} \rvert}\right)$
ME	Mean error	$ME = \frac{1}{H}\sum_{t=1}^{H} P_t - Y_t$
sME	Scaled mean error	$sME = \frac{1}{H}\sum_{h=1}^{H} \frac{P_{N+h} - Y_t}{\frac{1}{N}\sum_{t=1}^{N} Y_t}$
MSE	Mean squarred error	$MSE = \frac{1}{H}\sum_{t=1}^{h} (Y_t - P_t)^2$
RMSE	Root mean squared error	$RMSE = \sqrt{\frac{\sum_{t=1}^{H}(Y_t - P_t)^2}{H}}$
MAD	Mean absolute deviation	$MAD = median(\lvert Y_t - P_t \rvert$

Tableau B.1 : Formulations mathématiques des mesures d'erreurs (suite et fin)

R^2	Coefficient de détermination de Pearson	$R^2 = 1 - \frac{\sum_{t=1}^{N}(Y_t - P_t)^2}{\sum_{t=1}^{N}(Y_t - \bar{P}_t)^2}$
GMAE	Geometric mean absolute error	$GMAE = \sqrt[H]{\prod_{t=1}^{H}(Y_t - P_t)}$
AE	Absolute error	$AE = \lvert Y_t - P_t \rvert$
MAAPE	Mean absolute arctangent percent error	$MAAPE = \frac{1}{H}\sum_{h=1}^{H} \tan^{-1}\left\lvert \frac{Y_{N+h} - P_h}{Y_{N+h}} \right\rvert$
sPIS	Scaled period in stock	$sPIS = \frac{\sum_{h=1}^{H}\sum_{j=1}^{h}(P_j - Y_{N+j})}{\frac{1}{N}\sum_{t=1}^{N} Y_t}$
sAPIS	Scaled absolute period in stock	$sAPIS = \frac{\sum_{h=1}^{H}\sum_{j=1}^{h}\lvert P_j - Y_{N+j} \rvert}{\frac{1}{N}\sum_{t=1}^{N} Y_t}$
sMSE	Scaled mean squared error	$sMSE = \frac{1}{H}\sum_{h=1}^{H}\left(\frac{Y_{N+h} - P_h}{\frac{1}{N}\sum_{t=1}^{N} Y_t}\right)^2$
sMAE	Scaled mean absolute error	$sMAE = \frac{1}{H}\sum_{h=1}^{H}\frac{\lvert Y_{N+h} - P_h \rvert}{\frac{1}{N}\sum_{t=1}^{N} Y_t}$
MUES	Mean underestimation share	$MUES = \frac{1}{H}\sum_{h=1}^{H} \max\left(signe(Y_h - P_h), 0\right)$
MOES	Mean overestimation share	$MOES = \frac{1}{H}\sum_{h=1}^{H} \max\left(signe(P_h - Y_h), 0\right)$