Titre: Title:	Estimation de la distribution de la demande pour des familles de produits textiles
Auteur: Author:	Anaëlle Wurm
Date:	2025
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Wurm, A. (2025). Estimation de la distribution de la demande pour des familles de produits textiles [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/71203/

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Estimation de la distribution de la demande pour des familles de produits textiles

ANAËLLE WURM

Département de mathématiques et de génie industriel

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

Génie industriel

Avril 2025

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

Estimation de la distribution de la demande pour des familles de produits textiles

présenté par **Anaëlle WURM**

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

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

Camélia DADOUCHI, présidente
Bruno AGARD, membre et directeur de recherche
Jonathan JALBERT, membre

DÉDICACE

À mes professeurs du Collège International Marie de France, sans lesquels je n'aurais jamais pu aspirer à entrer à l'université.

À Sébastien Rivillon et Pierre Larrouturou, qui ont su me transmettre leur enthousiasme pour les mathématiques.

À mon grand-père défunt, Henri Würm, qui, j'en suis certaine, m'a accompagnée à chaque instant depuis son départ.

REMERCIEMENTS

Merci à Logistik Unicorp pour leur collaboration. Merci en particulier à Philippe St-Aubin, pour son écoute. Merci à Bruno Agard pour son accompagnement, son inépuisable patience, sa bienveillance et son sérieux tout au long du projet. Merci à mes proches pour leur support et leurs encouragements au cours de mes études au cycle supérieur. Merci à Jean-Simon Degagné pour son indéfectible optimisme. Merci à mon père, Daniel Bernard Würm, d'avoir tenu jusqu'à ma soutenance.

RÉSUMÉ

L'évolution rapide de notre population, de nos technologies, de notre climat et de nos besoins nous incite à développer des stratégies d'adaptation à un rythme toujours plus soutenu. La prédiction est un outil qui a le potentiel de faciliter notre progression dans l'environnement incertain dans lequel nous vivons. C'est la raison pour laquelle de nombreuses entreprises comme Logistik Unicorp cherchent à développer des modèles qui leur permettent de prédire la demande de leurs clients afin d'ajuster leur production à cette dernière. Logistik Unicorp est une entreprise québécoise qui se spécialise dans la production d'uniformes, qu'elle distribue aux quatre coins du globe. Désireuse d'améliorer sa gestion d'inventaire, l'organisation souhaite développer un modèle qui lui permette d'obtenir des distributions prédictives de la demande hebdomadaire et mensuelle de ses clients.

Ce mémoire présente une méthode qui exploite des modèles bayésiens descriptifs pour obtenir les prédictions d'une famille de produits dont la demande présente une importante saisonnalité. La simulation par chaînes de Markov est utilisée pour prédire l'incertitude des données résiduelles de la demande, tandis que les composantes saisonnière et tendancielle sont estimées à partir des données historiques. Les prédictions produites sont comparées à celles que peuvent nous offrir des outils comme SARIMA et les modèles d'espaces-états.

Les résultats démontrent que la méthode produit des résultats semblables à ceux que nous offrent nos deux outils de référence. Cependant, la méthode proposée, qui fournit distribution prédictive approximative de la demande, présente un potentiel d'amélioration intéressant vis-à-vis de la gestion de l'approvisionnement du partenaire. Cela car elle lui permettra d'ajuster sa demande avec une estimation probabiliste des risques de rupture de stock.

ABSTRACT

The rapid evolution of our population, our technologies, our climate and our needs is prompting us to develop adaptation strategies at an ever-increasing pace. Prediction is a tool that has the potential to facilitate our progress in the uncertain environment in which we live. That's why many companies like Logistik Unicorp are seeking to develop models that enable them to predict customer demand and adjust their production accordingly. Logistik Unicorp is a Quebec-based company specialising in the production of uniforms, which it distributes to the four corners of the globe. Wishing to improve its inventory management, the organisation wanted to develop a model that would enable it to obtain predictive distributions of its customers' weekly and monthly demand.

This dissertation presents a method that exploits descriptive Bayesian models to obtain predictions for a family of products whose demand exhibits significant seasonality. Markov chain simulation is used to predict the uncertainty in the residual demand data, while the seasonal and trend components are estimated from historical data. The predictions produced are compared with those offered by tools such as SARIMA and state-space models.

The results show that the method produces results similar to those offered by our two reference tools. However, the proposed method, which provides an approximate predictive distribution of demand, has an interesting potential for improvement with regard to the partner's supply management. This is because it will enable them to adjust their demand with a probabilistic estimate of the risk of stock shortages.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES SIGLES ET ABRÉVIATIONS

AICc. Akaike corrected Information Criterion

ARFIMA Autoregressive Fractionally Integrated Moving Average

ARIMA Autoregressive Integrated Moving Average

ARMA Autoregressive Moving Average

CART Classification and Regression Tree

GARCH Generalized Autoregressive Conditional Heteroscedasticity

KPSS Kwiatkowski-Phillips-Schmidt-Shin

LOESS Locally Estimated Scatterplot Smoothing

MAE Mean Absolute Error

MAPE. Mean Absolute Percentage Error

PSIS Pareto-Smoothed Importance Sampling Cross-Validation

SARIMA Seasonal Autoregressive Integrated Moving Average

STL Seasonal-Trend decomposition using Loess

SVR Support Vector Regression

VAR Vector Autoregressive

LISTE DES ANNEXES

CHAPITRE 1 INTRODUCTION

En 2022, les Nations Unies estimaient que la population mondiale atteindrait les 8 milliards d'individus à la fin de l'année, tandis qu'elle frôlerait les 8.5 milliards en 2030 (Department of Economic and Social Affairs, Population Division, 2022). Dans un contexte d'accroissement de la population mondiale et des flux migratoires, nous pouvons présumer qu'une multiplication des infrastructures sur de plus larges étendues est à venir. La croissance démographique exerce en effet une pression sur les villes existantes qui doivent réévaluer leur planification urbaine ainsi que leur gestion des services publics (Hammarberg et al., 2023). D'autre part, en vue de répondre aux objectifs de développement durable universels, le développement de mesures de services publics doit suivre le rythme de l'expansion urbaine (Zhongxu et al., 2022). Il est ainsi raisonnable d'envisager l'émergence prochaine de nouveaux hôpitaux, établissements policiers, ou encore de nouvelles institutions de protection contre les incendies à travers le monde. Nous pouvons de surcroît supposer une hausse de la demande en uniformes institutionnels à moyen et long terme. Dans le cadre d'une telle évolution, il s'avère indispensable pour les industries textiles de suivre la progression des besoins de leurs clients institutionnels. Elles pourront ainsi rester compétitives et contribuer positivement à la gestion infrastructurelle des collectivités grandissantes.

Ce mémoire propose et évalue différentes méthodes de prédiction qui visent à l'obtention de prévisions pour une demande saisonnière à partir d'historiques de données, en partenariat avec l'entreprise Logistik Unicorp.

Logistik Unicorp est une compagnie québécoise qui se démarque depuis plus de 30 ans dans le domaine de l'habillement institutionnel. De la conception de costumes à l'innovation du textile, l'entreprise propose des services qui trouvent grâce aux yeux de clients variés au Canada comme à l'étranger. Cette dernière propose une importante variété de produits personnalisés. Elle approvisionne notamment les Forces Armées Canadiennes, la Ville de Montréal, la Société des établissements de plein air du Québec ainsi que Postes Canada (Champagne, 2022). Sa clientèle s'étend jusqu'en Europe, en Océanie et plus récemment au Togo. En dehors de son bureau central qui se situe au Québec, elle dispose d'une usine au Vietnam, ainsi que d'un centre spécialisé dans les armures, les accessoires de combat ainsi que les costumes de cérémonie en Australie (Logistik Unicorp, 2024). La compagnie dispose d'une importante capacité de production. À titre d'exemple,

celle-ci a été en mesure de livrer 11 millions de blouses de protection médicales en 2020, au cours de la pandémie. La production d'importants volumes de produits personnalisés et innovants livrés dans de courts délais est la véritable marque de fabrication de l'entreprise.

Dans l'optique de parfaire la qualité de ses prestations, notamment en termes de délais de livraison, l'organisation cherche à améliorer la gestion de son inventaire. Pour ce faire, cette dernière mise entre autres sur l'élaboration d'outils d'aide à la décision, dont des méthodes de prédiction qui feront l'objet de cet écrit. Le mandat qui découle de cette volonté de perfectionnement se traduit par l'obtention de distributions prédictives de la demande. L'entreprise souhaite en effet quantifier l'incertitude de la demande de ses clients. Plus particulièrement, nous souhaitons obtenir des intervalles de prédictions pour les quantités commandées de produits constitués d'un textile spécifique. L'entreprise souhaite parvenir à prévoir adéquatement les besoins hebdomadaires, mensuels et annuels de sa clientèle, en les visualisant sous forme de distribution. Les prévisions doivent être réalisées pour des horizons de 12 semaines et 12 mois séparément.

Logistik fait cependant face à un défi, qui est celui d'adopter une stratégie qui soit adaptée à la structure de ses données. L'organisation de ces dernières peut être représentée comme suit :

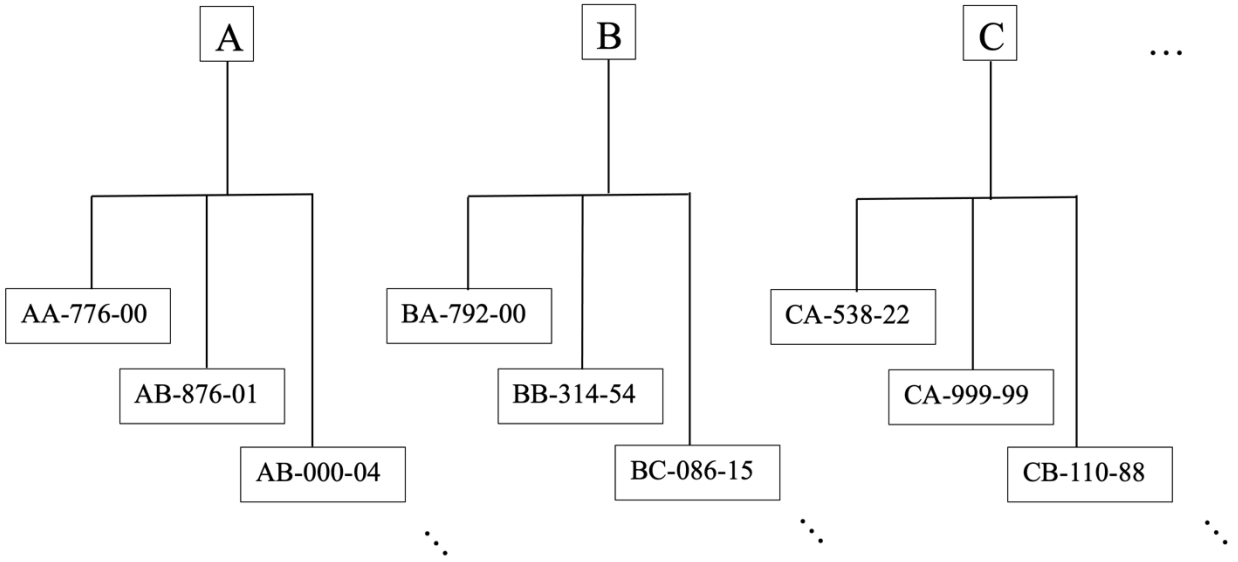


Figure 1.1 : Représentation de la structure des données avec des produits fictifs

Il existe plusieurs familles de produits (A, B, C sur l'image) qui regroupent chacune plusieurs produits (AA-776-00, AB-876-01 et AB-000-04 pour la famille A, sur l'image). Chaque produit

de la même famille est constitué du même tissu, mais peut comporter d'autres caractéristiques différentes (comme la taille, par exemple). Chaque produit a une demande différente, qui est variable et intermittente. La demande par famille est la somme des demandes des produits qu'elle regroupe et se montre conséquemment tout autant variable. Certaines familles présentent également une composante saisonnière. Les quantités commandées sont enregistrées une fois par semaine, ce qui fait que l'entreprise dispose de données hebdomadaires. Outre la structure des données, l'entreprise cherche donc une solution qui soit également compatible avec la quantité d'informations disponibles. Pour certaines familles de produits qui nous intéressent, l'organisation a une historique qui remonte jusqu'à 2013. En revanche, nous n'avons pas de quantités commandées au-delà de Mars 2023, ces dernières ayant été indisponibles au début du projet. De surcroît, les années 2019 à 2023 ayant été marquées par une pandémie, les informations qui y sont associées ne peuvent être considérées comme étant représentatives de la normalité. Nous travaillerons donc sur au plus 312 données par produit (52 semaines par an de 2013 à 2018).

Avec ces informations en tête, nous avons décidé de formuler les objectifs de recherches suivants :

ii. Développer une approche bayésienne de prévision de la demande des familles de produits.
iii. Comparer l'approche aux démarches classiques dans le traitement de séries chronologiques par le biais du MAE, du MAPE et du score quantile.
iv. Émettre une conclusion quant au mode opératoire à privilégier.

Pour mener à bien cette mission, nous débuterons au chapitre 2 par une présentation des connaissances qui nous ont été nécessaires pour aboutir à notre démarche de recherche et notre solution. Nous poursuivrons par une explication détaillée de la méthodologie que nous avons adoptée au chapitre 3, qui sera suivi par une exposition de nos résultats au chapitre 4. Nous réaliserons notamment une analyse des données relatives à la demande, une exposition de nos expérimentations ainsi que de leur évaluation. Nous conclurons par une discussion des limites de notre étude, ainsi que du potentiel d'amélioration du projet.

CHAPITRE 2 REVUE DE LITTÉRATURE

La prédiction est une thématique qui est régulièrement abordée lorsqu'il est question d'efficacité, et plus généralement de performance en matière de gestion. Si la capacité d'appréhender l'avenir sans en connaître la substance ne fait pas l'objet d'un intérêt nouveau, il n'en n'est pas moins croissant. La précédente décennie, marquée par de nombreuses innovations technologiques, s'est entre autres caractérisée par un engouement pour la collecte et l'analyse de données.

Dans le milieu de l'industrie, la gestion de la chaîne d'approvisionnement est envisagée comme un système qui deviendra autonome (Calatayud et al., 2019). Cette ambition repose notamment sur une volonté de rester compétitif en répondant adéquatement à une demande en hausse, de plus en plus diversifiée et incertaine. L'objectif des entreprises se traduit donc par une optimisation de leurs activités, notamment de la gestion des inventaires et de la distribution des livraisons. La personnalisation de masse ainsi que la complexification de notre environnement constituent un défi dont l'issue première est un système de planification hors pair. En effet, selon Gouati et al. (2022), l'évaluation de la disponibilité des ressources et la planification à court terme représentent les solutions les plus explorées par les entreprises. Cependant, l'évaluation des quantités d'approvisionnement et du niveau de stocks optimaux restent un sujet insuffisamment cultivé (Gouati et al., 2022). Cela transparaît dans les difficultés logistiques rencontrées par les entreprises avant et pendant la pandémie. Ces dernières se rapportaient notamment à des problèmes de gestion d'inventaire ainsi que de planification de la demande (Alhasawi et al., 2023). Il est ainsi nécessaire de poursuivre de plus amples recherches en matière de gestion des stocks et de prévision des besoins des clients.

Pour être en mesure de faire face à l'incertitude de la demande et ainsi disposer d'un inventaire approprié, il est nécessaire de se projeter dans le futur. Anticiper les scénarios susceptibles de compromettre la stabilité et la continuité des activités de la chaîne d'approvisionnement contribue au développement de sa résilience (Alhasawi et al., 2023). La prédiction des quantités nécessaires à avoir en inventaire peut toutefois demeurer énigmatique. À titre d'exemple, la survenue impromptue de la pandémie de coronavirus ainsi que de la guerre en Ukraine a induit d'insaisissables changements de la demande. Les quantités sollicitées sont apparues variables et irrégulières dans le temps, ce qui a compliqué la planification de la production des industries

(Benziane et al., 2023). Ces évènements ont contribué à une nouvelle appréhension de la gestion de la chaîne d'approvisionnement, mettant en lumière l'importance de la capacité de s'adapter à de brusques variations. Le recours à un outil de prédiction adapté pourrait de fait constituer une aide à la décision et une amorce de solution vers une meilleure gestion des stocks. De telles considérations sont d'autant plus importantes aujourd'hui que les services publics de première nécessité requièrent l'obtention de leur matériel rapidement en temps de crise. C'est notamment ce que nous avons eu l'occasion de constater au cours de la pandémie. Ainsi, la prédiction de la demande est plus que jamais au cœur des problématiques sociétales du 21ème siècle, tout comme l'incertitude qui la caractérise.

2.1 Méthodes de prédictions

La prédiction est une tâche qui nécessite une définition précise de sa raison d'être et des contraintes qui l'encadrent. Il est en effet bien différent de prévoir la météo, la quantité d'électricité à fournir dans une région spécifique ou encore l'usure d'une machine. Selon ce que nous souhaitons prédire, une méthode de prédiction peut être plus indiquée qu'une autre. De la même manière, un outil peut ne pas être adéquat pour atteindre notre objectif de prévision. À titre d'exemple, si nous souhaitons prédire la météo, nous savons que nous ne pouvons pas utiliser un modèle de régression linéaire simple, parce-que c'est un phénomène qui dépend de plusieurs variables. Il est ainsi indispensable de bien définir notre problématique. Parmi les facteurs qui sont à prendre en considération dans le choix d'une méthode de prédiction, nous avons entre autres le nombre de variables, le type de données et la quantité disponible de ces dernières.

À des fins de simplification, nous décrirons 6 familles générales d'outils qui sont utilisés dans le domaine de la prédiction appliquée aux séries temporelles. Nous ne détaillerons pas les différentes étapes de préparation auxquelles ses séries peuvent être sujettes avant l'application des différents outils exposés. Nous poursuivrons par une revue des méthodes existantes dans la prédiction de données de comptage, puis une description plus détaillée des méthodes qui sont adaptées à notre cas d'étude dans la section suivante.

2.1.1 Classification générale

1. La première catégorie de méthodes largement utilisées est celle que nous offrent la statistique. Parmi les outils qui entrent dans cette catégorie, nous retrouvons notamment la régression, les modèles de Box et Jenkins et le lissage exponentiel. Les méthodes statistiques offrent une grande performance lorsque les hypothèses sont satisfaites. Elles présentent l'avantage de ne pas nécessiter une puissance calculatoire aussi importante que les méthodes d'apprentissages automatique et profond (Makridakis et al., 2018). Elles sont en revanche moins adaptées à des données volatiles (Ingle et al., 2021). Lorsque nous souhaitons réaliser des prédictions probabilistes (ce qui est notre cas), nous nous penchons plutôt vers la statistique bayésienne. Tel que nous le verrons en 2.5, le fait de travailler avec des séries temporelles nous empêche de considérer les observations comme indépendantes et nous oblige à adapter la modélisation bayésienne. Tyralis et Papacharalampous (2024), présentent quelques méthodes qui sont mises en pratique dans le cadre d'une démarche bayésienne appliquée aux séries temporelles. Parmi elles, nous retrouvons notamment des modélisations bayésiennes de ARMA, ARFIMA, VAR et GARCH.
2. Nous avons d'autre part un ensemble de méthodes issues de l'apprentissage automatique. Tel qu'expliqué par Guéron (2019b), "l'apprentissage automatique est la science (et l'art) de programmer les ordinateurs de sorte qu'ils puissent apprendre à partir des données" (p. 2). Cette famille de méthodes compte notamment les outils SVR, les perceptrons multicouches, les réseaux de neurones bayésiens, les processus Gaussiens, les arbres de régression CART, la régression kernel, la régression des k plus proches voisins ainsi que les fonctions de base radiales (Makridakis et al., 2018). Ces outils requièrent une quantité de données d'apprentissage importante sans quoi leur performance est limitée. D'autre part, l'implantation de telles méthodes est plus complexe que celles qui reposent sur la statistique et nécessite une puissance de calcul plus élevée. Selon Guéron (2019b), un prédicteur non expérimenté peut aisément construire des modèles ou trop simplistes ou trop complexes, résultant respectivement à un sous- ou sur- ajustement des données (et donc des résultats insatisfaisants). En contrepartie, lorsque les données utilisées sont représentatives, en quantité suffisante et que les variables sont choisies judicieusement, ces outils peuvent être très utiles. Ceux-ci sont notamment adaptés lorsque les données sont fluctuantes et volumineuses (Guéron, 2019b).

3. Il existe également les méthodes de prédictions par apprentissage profond. Ces dernières reposent sur l'usage de réseaux de neurones profonds et opèrent une prédiction à partir de grands volumes de données. Elles peuvent ainsi résoudre des problèmes d'apprentissage automatique complexes à l'aide de milliards de données (Guéron, 2019a). Parmi ces méthodes, nous pouvons retrouver celles qui sont issues de la boîte à outils GluonTS, telles que Deep-AR, Feed-Forward, Transformer et WaveNet (Makridakis et al., 2022). L'inconvénient de ces outils est qu'ils ne peuvent fonctionner qu'avec d'immenses volumes de données (Ingle et al., 2021). Tyralis et Papacharalampous (2024) proposent une revue intéressante des méthodes d'estimation de l'incertitude par apprentissage automatique et par apprentissage profond. Nous n'irons pas davantage dans les détails, car nous ne les utiliserons pas.
4. Il existe par ailleurs des méthodes de prédiction fondées sur l'agrégation. Ces dernières répondent à une volonté de réaliser des prédictions dites "hiérarchiques" pour aider les décideurs de différents niveaux à prendre leurs décisions. L'agrégation peut être temporelle ou "transversale" (auquel cas des produits sont agrégés pour une période temporelle donnée). L'agrégation temporelle peut être avec blocage ou avec rééchantillonnage. L'agrégation dite transversale peut se faire de trois façons différentes. Il est possible, pour une période donnée, de prédire des quantités unitaires qui sont ensuite agrégées pour former les prédictions de la famille de produits (agrégation de bas en haut). Il est possible de faire l'inverse, à savoir de réaliser des prédictions par famille pour ensuite estimer la demande de chaque produits (agrégation de haut en bas). Il existe finalement une approche qui combine les deux. (Babai et al., 2021). Si ces méthodes permettent d'améliorer la précision des prédictions réalisées pour un niveau donné, elles présentent l'inconvénient d'offrir des prévisions incohérentes. C'est-à-dire que si nous réalisons des prédictions à la semaine et au mois en parallèle, la somme des prédictions hebdomadaires ne sera pas nécessairement égale à la prévision mensuelle. Il existe cependant une dernière méthode d'agrégation dite "intertemporelle" qui vise au dépassement de cette problématique (Petropoulos et al., 2022). Finalement, les méthodes fondées sur l'agrégation peuvent faciliter la prédiction lorsque les données sont intermittentes (Babai et al., 2021) et (Petropoulos et al., 2022).
5. Il existe aussi des méthodes de prédiction qualitatives, qui reposent sur le jugement. Ces méthodes, plus subjectives, varient selon les informations dont dispose l'individu. Nous

retrouvons notamment l'heuristique de reconnaissance, l'heuristique de représentativité, les heuristiques d'ancrage et d'ajustement. Couplées à des méthodes statistiques, l'avis d'experts peut être un facteur d'amélioration de la qualité de la prévision. Le désavantage des outils fondés sur le jugement est que la prédiction peut être biaisée par certains mécanismes cognitifs ou par les motivations du prédicteur (Petropoulos et al., 2022).

6. Finalement, il est également possible de réaliser une multitude de combinaisons de ces méthodes de prévisions. La performance de ces combinaisons repose sur une sélection minutieuse de modèles ainsi que de leur pondération (Makridakis et al., 2020).

2.1.2 Prédiction et données de comptage

Nous avons exposé une classification générale de méthodes de prédiction pour les séries temporelles. Il existe cependant différentes sortes de séries. Les données n'ont effectivement pas toujours un espace d'échantillonnage continu. Dans notre cas, elles se présentent sous forme de comptage (nombre de produits commandés). Petropoulos et al, (2022) proposent une revue des méthodes de prédiction applicables aux séries temporelles de comptage. Il en existe qui se fondent sur une démarche bayésienne (ce que nous souhaitons réaliser), et d'autres sur une démarche dite « fréquentiste ». Pour ce genre de données, parmi les méthodes de prédiction probabilistes existantes, nous n'avons pas trouvé d'outil qui se concentre à la fois sur des données d'inventaire régulières, en faible volume (moins de 500), qui présentent une saisonnalité et sans variables explicatives. Dans notre cas de figure, nous ne savons ni ce que nous vendons, ni à qui nous le vendons. Nous n'avons que des quantités commandées régulièrement de façon hebdomadaire et leur date de commande. Nous n'avons aucune information supplémentaire, relative par exemple à des promotions, des vacances, ou le secteur d'activité. Nous n'avons accès à aucun des gestionnaires d'inventaire, et donc aucune expertise de leur part. Notre apport pour la recherche sera donc d'étudier ce cas de figure précis, et de comparer notre solution aux résultats que nous offrent deux méthodes classiques utilisées dans la prédiction de séries temporelles.

Pour de plus amples informations sur la prédiction, nous invitons le lecteur à consulter les références sus-présentées. La prochaine section sera consacrée aux méthodes de prédiction que nous avons sélectionnées pour répondre à notre problématique.

2.2 Méthodes sélectionnées et notions associées

Tel qu'exposé par Petropoulos et al (2022), la prédiction de la demande se concrétise en 3 dimensions. Ces dernières sont le positionnement dans la hiérarchie de la chaîne d'approvisionnement, le niveau dans la hiérarchie du produit et la temporalité. Dans notre cas, nous nous concentrons sur les prédictions d'un centre de distribution, nous travaillons sur des familles de produits (et non en unité de stock) et nous nous intéressons aux commandes hebdomadaires et mensuelles. Nous souhaitons nous pencher sur l'incertitude de la demande et escomptons la caractériser. Plus particulièrement, nous désirons obtenir une distribution prédictive de la demande pour des horizons de 12 semaines et 12 mois, séparément. Pour des motifs énoncés dans la section précédente, nos analyses mensuelles et hebdomadaires seront réalisées en parallèle. Nous n'estimerons donc pas la demande mensuelle à partir des prévisions hebdomadaires.

Pour répondre à notre problématique, nous avons entrepris une recherche sur les méthodes de prédiction axées sur l'obtention d'une distribution prédictive. N'ayant au plus que 312 observations par famille de produits, nous nous concentrerons sur les méthodes qui ne nécessitent pas un important volume de données pour performer. En d'autres termes, nous nous tournerons vers la statistique.

2.3 Statistique classique

Bien que la prédiction ponctuelle soit plus répandue lorsque nous entendons parler de statistique classique, nous verrons qu'il est possible d'intégrer une notion d'incertitude à nos prévisions. C'est-à-dire que nous verrons qu'il nous est possible d'obtenir des intervalles dans lesquels nous supposons que notre prédiction se trouvera pour une certaine probabilité de couverture (80%, 95%, etc). Tel que nous le rapportent McElreath (2020) et Gelman et al. (2021), cela s'apparente à des intervalles de confiance, interprété de façon bayésienne. Comme nous le présenterons au cours des deux prochains chapitres, pour répondre à notre problématique, nous ferons ainsi usage de méthodes statistiques classiques de référence et de méthodes bayésiennes. Plus exactement, nous nous servirons de deux outils de prédiction classiques comme point de comparaison. Ils nous permettront d'évaluer où se situe la méthode que nous proposons. Nous ferons également usage dans notre solution de méthodes de décomposition. Ces différentes méthodes vous seront

présentées à la suite de l'exposition de la démarche statistique sur laquelle nous fonderons notre solution. Cette dernière emprunte à la fois des notions classiques et bayésiennes. Nous nous sommes inspirés de l'approche qui nous est présentée par Hyndman et al. (2021f), aux chapitres 2 à 5. Nous aborderons en détail la statistique bayésienne à la prochaine section.

2.3.1 Le processus de prédiction

Selon Hyndman et al. (2021f), la prédiction est un processus itératif qui s'amorce par une visualisation des données. Il se poursuit par la spécification des modèles et l'estimation des paramètres de ce dernier (lorsqu'applicable), pour se terminer avec l'évaluation de la performance de la modélisation. Lorsque les ajustements mènent à une performance satisfaisante, nous pouvons produire des prévisions.

La visualisation nous aide à comprendre le type de données dont nous disposons. Cette étape peut nous aider à choisir les méthodes de prédictions qui sont les plus adaptées, compte tenu du caractère des informations que nous avons (Hyndman et al., 2021f). Vous observerez aux chapitres 3 et 4 que nous réaliserons cette étape sur R. Le matériel utilisé est celui que nous présentent les auteurs au chapitre 2.

La spécification des modèles consiste en la modélisation de notre problème en tenant compte des informations et contraintes dont nous disposons. L'estimation des paramètres, dans le cadre de la statistique classique, se traduit généralement par la calibration du modèle spécifié sur un échantillon de données d'entraînement. Le modèle paramétré à partir de ces données est ensuite testé sur un échantillon test pour permettre la détection d'un éventuel problème de sur ou sous-ajustement. Cette dernière étape est l'évaluation. Lorsque le modèle présente un défaut significatif, nous pouvons envisager de réitérer l'ensemble des étapes, jusqu'à ce que nous parvenions à un résultat convenable.

Dans notre cas, faisant usage de modèles bayésiens, les deux dernières phases du processus de prédiction diffèrent quelque peu, puisque nous cherchons aussi à caractériser l'incertitude de la demande. C'est-à-dire que nous n'estimons pas dans l'optique d'obtenir une estimation ponctuelle certaine. De fait, l'ajustement de nos modèles fait partie de ceux-ci (puisque le produit de la vraisemblance par l'*a priori* en est une composante). Ainsi, il n'y a pas d'estimation à proprement dit, bien que notre modèle s'ajuste tout de même aux données. Cela impacte directement l'étape

d'évaluation, telle qu'elle est décrite par Hyndman et al. (2021f), puisque nous ne pouvons déterminer l'origine de la mauvaise performance du modèle avec autant d'aisance.

Il existe cependant des méthodes d'évaluation qui nous permettent de nous représenter combien un modèle bayésien est adapté pour répondre aux objectifs qu'il doit servir. Nous les présenterons dans la prochaine section.

Finalement, la phase de prédiction consiste en l'exécution des modèles calibrés et évalués. Nous verrons dans la section suivante que nous pouvons, en appliquant notre modèle bayésien obtenir une distribution prédictive. Selon Hyndman et al. (2021f), il est également possible d'en obtenir une à l'aide de la statistique classique. Elle diffère cependant en plusieurs aspects, tel que nous le verrons dans la prochaine section.

Selon Hyndman et al. (2021c), la plupart des modèles de séries temporelles produisent des prévisions normalement distribuées si les résidus du modèle le sont. Les résidus sont le résultat de la différence entre les valeurs ajustées et les valeurs véritables. Les valeurs ajustées sont des « prédictions » des données historiques réalisées à partir de ces dernières à l'aide du modèle calibré… par celles-ci. Plus exactement, la donnée historique à un instant t est estimée à partir des données historiques précédentes (t-1, t-2, … t-N). Ce ne sont bien évidemment pas de véritables prédictions. Il s'agit simplement d'une façon d'approximer la performance future du modèle ainsi que la variance de la « distribution de probabilité » que nous introduisons au prochain paragraphe. Nous utilisons alors l'ensemble des données dont nous disposons (données d'entraînement et de test confondues).

Ainsi, les auteurs expliquent qu'il est possible de modéliser la quantité future à prédire par une variable aléatoire et d'assumer que les futures valeurs possibles distribuées selon la loi normale. Si les résidus sont non-corrélés, normalement distribués et homoscédastiques, nous pouvons ainsi déterminer une « distribution de probabilité » pour la quantité future et obtenir un intervalle de confiance, que l'auteur appelle « intervalle de prédiction ». Avec suffisamment de données et pour un horizon de 1 période temporelle, nous pouvons estimer que la valeur à prédire suit une loi normale de moyenne $\bar{y}$ et de variance $\hat{\sigma}^2$ égale à celle des résidus, pour tout modèle de série temporelle. En revanche, pour un horizon plus grand, chaque distribution obtenue avec un modèle spécifique dispose d'une variance différente (Hyndman et al., 2021c). Lorsque nous tentons de prédire pour un long horizon, l'intervalle tend à s'agrandir car l'incertitude augmente. En revanche,

tel que nous le mentionne l'auteur, ces intervalles sont généralement étroits, car ils ne considèrent que l'incertitude liée à la part aléatoire qui constitue le modèle. Ils négligent l'incertitude liée au choix du modèle en lui-même ainsi qu'à ses paramètres, ce à quoi nous pouvons remédier en adoptant une approche bayésienne. Il est également nécessaire de garder à l'esprit que l'obtention de ces résultats (intervalles de prédiction et distribution) repose sur des conditions (normalité des résultats futurs, des résidus, etc…) qui ne sont pas un enjeu lorsque nous suivons une démarche bayésienne.

2.3.2 Méthodes de référence

Nous travaillons avec des données qui ne sont pas indépendantes les unes des autres et non identiquement distribuées, car elles dépendent du temps. Nous devons donc faire usage de modèles de prédictions adaptés à ce type données, qui sont communément appelées des séries temporelles (ou chronologiques). Tel que le présente Adjengue (2014), bien qu'il existe des méthodes qui mêlent les deux approches, il existe essentiellement deux démarches d'analyse de séries temporelles. L'une repose sur l'usage de modèles quasi-déterministes, et l'autre sur l'emploi de modèles stochastiques. Nous mettrons en application deux méthodes de décomposition quasi-déterministes. Nous ferons également usage de deux méthodes de prédiction, dont la méthode saisonnière de Holt-Winters (quasi-déterministe) et une méthode générale de Box et Jenkins (approche stochastique), communément appelée ARIMA. Nous verrons que la méthode de Holt-Winters peut être appréhendée de façon stochastique lorsque nous faisons usage de modèles d'espace états.

Méthodes de décomposition

L'analyse des différentes composantes d'une série temporelle peut nous servir à réaliser des prédictions ajustées (selon la tendance et la saison), tout comme cela peut nous permettre d'éliminer certaines composantes du signal d'intérêt. Cette démarche fera partie de notre solution. La décomposition consiste en la séparation des différentes composantes qui constituent une observation, à savoir la tendance, la saison, le cycle et la composante résiduelle (ou bruit). Selon Adjengue (2014), la tendance représente le comportement moyen de la série à long terme. La saison constitue quant à elle le schéma répétitif et périodique de la série, alors que le cycle évoque les

oscillations autour de la tendance sur une longue période. Finalement, la composante résiduelle se rapporte aux irrégularités qui ne peuvent s'expliquer par les autres éléments, et qui sont considérées comme étant aléatoires.

Il existe deux façons de modéliser une série temporelle "décomposée", à savoir les modélisations additive et multiplicative.

Soit :

$Y_t : l'observation\ au\ temps\ t$

$T_t : tendance\ au\ temps\ t$

$C_t : cycle\ au\ temps\ t$

$S_t : saison\ au\ temps\ t$

$r_t : composante\ résiduelle\ au\ temps\ t$

La modélisation additive repose sur la considération que les différentes composantes de la série sont indépendantes, ce qui résulte à l'équation suivante :

$$Y_t = T_t + C_t + S_t + r_t$$

En d'autres termes, l'observation initiale s'obtient par la somme de ses composantes.

La modélisation multiplicative implique quant à elle que les composantes interagissent les unes avec les autres, ce qui se traduit par l'équation ci-dessous.

$$Y_t = T_t \times C_t \times S_t \times r_t$$

L'observation initiale s'obtient par la multiplication des composantes.

Il existe plusieurs méthodes de décomposition qui nous permettent d'obtenir ces deux modélisations. Parmi elles, nous utiliserons celle des moyennes mobiles ainsi que la décomposition dite "STL" ("Seasonal-Trend decomposition using Loess"). Nous avons choisi la première car c'est l'une des plus communes et la seconde car elle présente de nombreux avantages par rapport à la plupart des autres méthodes de décomposition. La technique STL est notamment robuste aux données aberrantes et permet à la composante saisonnière d'évoluer au cours du temps. Nous survolerons ce en quoi consiste ces méthodes sans rentrer dans les détails, car nous ne décomposerons pas manuellement dans le cadre de notre projet. Pour une explication plus détaillée

du fonctionnement de ces deux méthodes, le lecteur est invité à lire les références suivantes : (Adjengue, 2014), (Hyndman et al., 2021f), (Cleveland et al., 1990).

La décomposition par moyennes mobiles d'ordre *m* se fait en trois temps. Nous commençons par estimer la tendance en remplaçant chaque donnée par la moyenne arithmétique des *m* observations qui l'avoisinent. Autrement dit, nous réalisons :

$Lorsque\ l'ordre\ m = 2p + 1\ est\ impair$:

$$MM_t = \frac{1}{2p+1} \times \sum_{j=-p}^{p} Y_{t+j} \qquad p+1 \leq t \leq n-p$$

n : $longueur\ de\ la\ série\ chronologique$

$Lorsque\ l'ordre\ m = 2p\ est\ pair$:

$$MM_{t+\frac{1}{2}} = \frac{1}{2p} \times \sum_{j=-p+1}^{p} Y_{t+j} \qquad p \leq t \leq n-p$$

n : $longueur\ de\ la\ série\ chronologique$

Remarquez que tel que l'évoque Adjengue (2014), nous perdons de l'information dans les deux cas, car il y a $m-1$ observations qui ne pourrons être traitées. D'autre part, lorsque notre ordre est pair, les moyennes mobiles calculées correspondent au milieu des temps d'observation des données, ce qui nous pousse à faire usage des moyennes mobiles centrées :

$$MMC_t = \frac{MM_{t-\frac{1}{2}} + MM_{t+\frac{1}{2}}}{2}$$

Par la suite, pour estimer la composante saisonnière, nous ôtons la tendance du signal initial. Dans le cas d'une modélisation additive, nous soustrayons Tt à la donnée t, alors que dans le cas d'une modélisation multiplicative, nous la divisons par Tt. Selon Adjengue (2014), il est admis que la tendance peut être estimée par les moyennes mobiles centrées. C'est la raison pour laquelle nous avons mathématiquement :

$$\widehat{S_t} = \begin{cases} Y_t - MMC_t \; dans \; le \; cas \; d'une \; modélisation \; additive \\ \dfrac{Y_t}{MMC_t} \; dans \; le \; cas \; d'une \; modélisation \; multiplicative \end{cases}$$

$\widehat{S_t} : estimation \; de \; la \; composante \; saisonnière \; au \; temps \; t$

Une fois réalisé, nous obtenons les indices saisonniers en réalisant la moyenne des St pour chaque saison. C'est-à-dire que nous effectuons :

$$\widehat{I_j} = \frac{1}{n_j} \sum_{t \in A_j} \widehat{S_t}\,, \qquad j = 1, \dots m$$

$n_j : \; nombre \; d'observations \; inclues \; dans \; la \; saison \; j$

$A_j : ensemble \; des \; observations \; de \; la \; série \; Y_t \; qui \; font \; partie \; de \; la \; saison \; j$

$\widehat{I_j} : indice \; saisonnier \; estimé \; de \; la \; saison \; j$

Finalement, pour obtenir la composante résiduelle, nous ôtons la saison tout comme nous avons ôté la tendance. Notez que tel que mentionné par Hyndman et al. (2021f), la tendance et le cycle sont généralement combinés sous l'appellation de la tendance, à des fins de simplicité. C'est la raison pour laquelle nous ne l'avons pas formellement exprimée.

La méthode de décomposition STL consiste en la séparation des différentes composantes par filtrage à l'aide d'une séquence d'applications de lissage par Loess ("locally estimated scatterplot smoothing") (Cleveland et al., 1990). Le Loess est une technique de régression non paramétrique qui permet d'estimer la tendance et la saison de la série temporelle à l'aide d'une courbe. Par souci de simplicité et de brièveté, nous ne détaillerons pas les différentes équations qui sont employées pour obtenir une série décomposée.

La décomposition STL est un processus itératif au cours duquel l'estimation de la tendance et de la saison par Loess se précise à mesure que l'estimation de chacune des composantes évolue. C'est-à-dire que l'estimation de la tendance impacte l'estimation de la saison et inversement. Lorsque la tendance et la saison ont été estimées, elles sont soustraites au signal initial pour obtenir la

composante résiduelle. Notez que tel que souligné par Hyndman et al. (2021f), La décomposition STL considère que les données peuvent être modélisées de façon additive. Pour une modélisation multiplicative, il est nécessaire de décomposer le logarithme de nos données (car la somme des logarithmes de a et *b* est égale au logarithme de leur produit).

Pour déterminer le type de modélisation le plus indiqué pour nos données, Hyndman et al. (2021f) nous indiquent qu'une modélisation additive est préférable lorsque les variations saisonnières sont relativement constantes tout au long de la série. Une modélisation multiplicative est quant à elle plus adaptée pour les données dont les variations saisonnières évoluent proportionnellement au niveau de la série. Le processus de décomposition des données doit donc s'amorcer par une analyse de ces dernières, pour établir la modélisation et la méthode conséquemment adaptées.

La méthode de Holt-Winters

La méthode de Holt-Winters reprend le principe du lissage exponentiel simple, à savoir l'estimation d'observations futures par la moyenne pondérée des observations passées. À mesure que nous remontons dans le temps, les données ont un poids de plus en plus faible, laissant une influence plus importante aux informations récentes. Cependant, à la différence du lissage exponentiel simple, Holt-Winters nous permet de tenir compte à la fois de la composante saisonnière et de la tendance. Cela implique donc que nous avons 3 équations de lissage au lieu d'une (une pour les observations, une pour la tendance et une pour la saison). Nous obtenons à partir de ces équations notre équation de prédiction. Notez qu'il existe plusieurs autres méthodes de lissage exponentiel (9 en tout), dont une méthode intermédiaire qui considère la tendance sans tenir compte de la saison. Il s'agit de la méthode de Holt, qui consiste essentiellement en l'ajustement d'une droite. Nous n'approfondirons dans ce mémoire que des variantes qui tiennent comptent de la présence d'une saisonnalité. Pour de plus amples renseignements sur les méthodes de lissage exponentiel, nous recommandons au lecteur de se référer à (Adjengue, 2014) et (Hyndman et al., 2021f). Notez que lorsque nous parlons des méthodes « additive de Holt-Winters » et « multiplicative de Holt-Winters », nous nous référons à la classification proposée par Pegels en 1969, que Hyndman et al. (2021f) nous exposent. Les méthodes qui considèrent une tendance multiplicative ou dite « amortie » seront exclues pour tout ce qui sera présenté dans les prochains paragraphes.

Tel que nous l'expliquent Hyndman et al. (2021f) et Adjengue (2014), il existe deux variantes de la méthode Holt-Winters, dont la différence réside dans l'expression de la composante saisonnière. Nous avons la méthode dite additive, et la méthode dite multiplicative. Ces deux noms ne sont pas sans rappeler les deux types de modélisations exposées précédemment, dont le concept se retrouve ici également. Dans le cas de la méthode additive, la composante saisonnière s'exprime en termes absolus selon l'échelle de la série observée. Dans l'équation de lissage des observations, la série est désaisonnalisée en soustrayant la composante saisonnière. Autrement dit, si chaque mois de Janvier nous savons à l'aide de nos données historiques que les observations sont plus élevées de 1200 unités, il s'agit d'un effet additif. La méthode multiplicative se fonde quant à elle sur la considération que le la saison s'exprime comme une quantité relative (un pourcentage). Dans l'équation de lissage des observations, la série est alors désaisonnalisée en étant divisée par la composante saisonnière. À titre d'exemple, nous aurions recours à une telle méthode si nous observions une augmentation systématique de 27% à une même période donnée. Mathématiquement, tout cela se traduit par les équations suivantes :

$Modélisation\ additive$:

$$\hat{y}_{t+h|t} = \ell_t + hT_t + I_{t+h-m(k+1)}$$
$$\ell_t = \alpha(y_t - I_{t-m}) + (1-\alpha)(\ell_{t-1} + T_{t-1})$$
$$T_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)T_{t-1}$$
$$I_t = \gamma(y_t - \ell_{t-1} - T_{t-1}) + (1-\gamma)I_{t-m}$$
$$0 \leq \alpha \leq 1, \qquad 0 \leq \beta^* \leq 1, \qquad 0 \leq \gamma \leq 1-\alpha$$

$Modélisation\ multiplicative$:

$$\hat{y}_{t+h|t} = (\ell_t + hT_t) \times I_{t+h-m(k+1)}$$
$$\ell_t = \alpha\frac{y_t}{I_{t-m}} + (1-\alpha)(\ell_{t-1} + T_{t-1})$$
$$T_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)T_{t-1}$$
$$I_t = \gamma\frac{y_t}{(\ell_{t-1} - T_{t-1})} + (1-\gamma)I_{t-m}$$

m représente le nombre de saisons, ℓ_t représente une forme de valeur moyenne estimative (valeur lissée), T_t évoque l'estimation de la tendance (donc la pente) de la série au temps t et I_t est la moyenne pondérée entre l'actuelle saison et la même saison de l'année d'avant. Tel que le

mentionnent Hyndman et al. (2021f), l'indice $m(k+1)$ avec k la partie entière de $\frac{h-1}{m}$ nous offre l'assurance que les estimations des indices saisonniers utilisées pour les prévisions proviennent de la dernière année de l'échantillon. Notez que les deux méthodes exigent la connaissance des constantes de lissage α, β^* et γ, des m indices saisonniers ainsi que d'une valeur lissée et une tendance au temps initial (Adjengue, 2014). Tel que nous le mentionnent Hyndman et al. (2021f), nous pouvons les estimer en minimisant la somme des erreurs quadratiques.

La méthode de Holt-Winters nous permet d'obtenir des prédictions ponctuelles. Nous pouvons cependant construire des modèles qui nous permettent d'obtenir des intervalles de « prédiction » autour de cette estimation ponctuelle. Pour des renseignements plus précis sur ce qui suit, nous invitons le lecteur à consulter (Hyndman et al., 2021f)

Tel que mentionné par Hyndman et al. (2021h), nous pouvons générer deux modèles de Holt-Winters grâce auxquels nous pourrons obtenir des intervalles de « prédiction » (et distributions) différents pour une même prédiction ponctuelle. Ces derniers sont qualifiés de « modèles d'espace d'états » (plus connus sous le nom de « state space models »).

Ces modèles sont constitués d'une équation qui décrit les données observées, ainsi que d'équations dîtes d' « état » qui représentent l'évolution du comportement des états qui n'ont pas encore été aperçus. Ces états sont décrits pour la valeur lissée, la tendance et la saisonnalité. Le stochastisme de ces modèles repose sur deux changements. En premier lieu, nous réexprimons les équations précédemment présentées de sorte à laisser transparaître une erreur *e*. Dans un second temps, nous supposons que cette dernière suit une distribution normale centrée en 0. C'est-à-dire que nous spécifions une distribution de probabilité pour l'erreur.

L'incorporation de l'erreur peut être multiplicative, ou additive. Ainsi, pour chaque modélisation présentée plus tôt (saison inclue de manière multiplicative, versus additive) nous avons deux modèles d'états. Si nous reprenons ce que nous avons exposé plus tôt, nous aboutissons au résultat suivant :

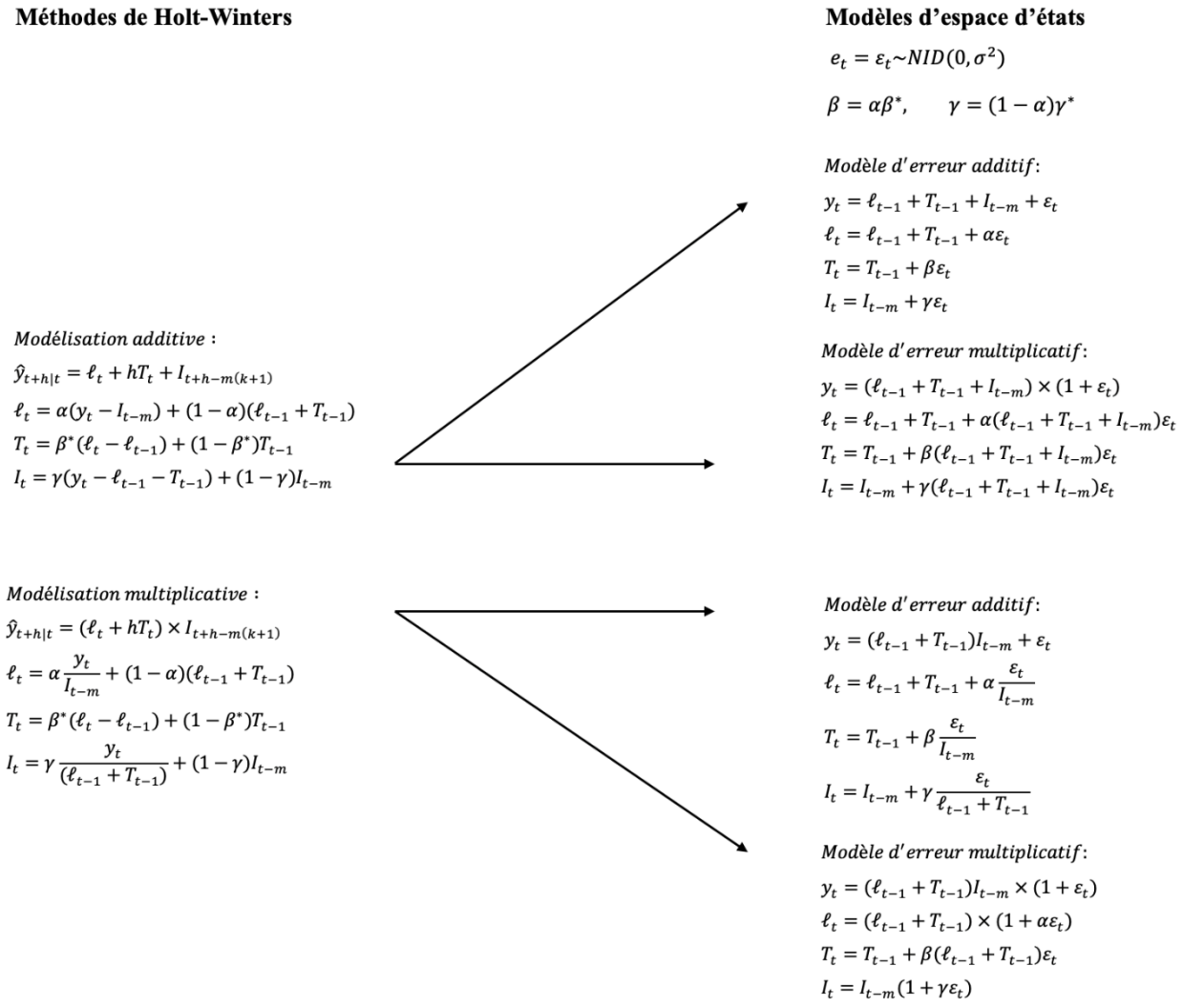


Figure 2.1 : Modèles d'espace états pour les méthodes de Holt-Winters

Tel que suggéré auparavant, ce que nous présentons n'est pas exhaustif. Pour un aperçu plus général, nous invitons le lecteur à consulter (Hyndman et al., 2021f).

Notez que tel que précédemment, nous devons respecter les contraintes suivantes :

$$0 \leq \alpha \leq 1, \qquad 0 \leq \beta^* \leq 1, \qquad 0 \leq \gamma \leq 1 - \alpha$$

D'autre part, « NID » signifie, « normale et indépendamment distribuée ».

Lorsque nous faisons usage de modèles d'espace états, nous avons la possibilité d'estimer les paramètres tout comme les états initiaux $\ell_0, T_0, I_0, I_{-1}, \ldots, I_{-m+1}$, en maximisant la vraisemblance (Hyndman et al., 2021f). C'est-à-dire que nous évaluons combien il est probable que les données que nous observons proviennent d'un certain modèle, avec des paramètres spécifiques. Plus la

vraisemblance est élevée, plus notre modèle est pertinent, ce pourquoi nous cherchons le résultat qui la maximise, tout en respectant cependant les contraintes qui encadrent les paramètres. Ce processus se fera automatiquement à l'aide de la fonction ETS() sur le logiciel R, dans notre cas. Cette fonction nous permettra par ailleurs de sélectionner le modèle le plus approprié compte tenu de nos données. Cette dernière fait usage de l'AICc qu'elle cherche à minimiser pour déterminer le modèle le plus indiqué (Hyndman et al., 2021f).

Lorsque nous avons déterminé notre modèle, nous pouvons en faire usage pour réaliser nos prédictions. Ces dernières peuvent être réalisées à l'aide de la librairie « fable », en utilisant la fonction « forecast ». Si nous utilisons deux modèles dont les paramètres sont identiques, mais dont la modélisation des erreurs diffère, ces derniers nous conduiront tous deux à une même prédiction ponctuelle. En revanche, ils nous fourniront chacun un intervalle de prédiction différent (Hyndman et al., 2021g). Notez qu'il n'est pas toujours possible d'obtenir une formule exacte pour les intervalles. Dans un cas comme celui-ci, ces derniers sont obtenus par simulation (Hyndman et al., 2021g). La librairie que nous utiliserons se charge elle-même de simuler ou d'utiliser une formule, selon ce qui est ou non applicable.

La méthode ARIMA

Le modèle ARIMA est un modèle général de Box et Jenkins qui fait usage des deux processus stochastiques linéaires que sont les processus autorégressifs et moyenne mobile (Adjengue, 2014). L'acronyme ARIMA se rapporte en effet à "Autoregressive Integrated Moving Average". Il est ainsi constitué d'un modèle autorégressif, couplé à un modèle de moyenne mobile, qui prennent respectivement pour intrants des valeurs et des erreurs décalées. "Integrated" fait allusion au processus de différenciation (nous menant à l'obtention d'une série stationnaire) et à celui d'intégration (qui est le processus inverse qui succède à l'ajustement du modèle). Nous aborderons dans cette sous-section chacun des éléments mentionnés pour ensuite expliquer sommairement le principe derrière la méthode ARIMA, et plus particulièrement SARIMA. Pour des explications plus approfondies, le lecteur est invité à consulter les références suivantes : (Adjengue, 2014), (Hyndman et al., 2021f).

La différenciation

La différenciation, telle qu'employée dans les modèles ARIMA, vise à l'obtention d'une série stationnaire. Une série stationnaire est telle que le positionnement dans le temps n'a pas d'incidence sur le comportement observé de cette dernière, qui est imprévisible à long terme. La différenciation consiste à construire une série qui contient les différences des observations consécutives de notre série chronologique originale. Il existe plusieurs sortes de différenciations. Nous les avons résumées dans le tableau suivant :

Tableau 2.1 : Quelques informations sur la différenciation

Différenciation	Processus
Premier ordre	$y'_t = y_t - y_{t-1} = (1-B)y_t$
Second ordre	$y''_t = y'_t - y'_{t-1} = (1-B)y'_t = (1-B)^2 y_t$
Saisonnière	$y'_t = y_t - y_{t-m} = (1-B^m)y_t$

B est l'opérateur de décalage, qui mène simplement à une notation simplifiée des équations présentées. m représente quant à lui le nombre de périodes qui sépare une saison de la même un an auparavant.

Notez que lorsque notre série est un bruit blanc, la différenciation de premier ordre peut être écrite sous la forme ci-dessous, qui représente un modèle de marche aléatoire.

$$y_t - y_{t-1} = \varepsilon_t$$

Le bruit blanc est une série qui peut être représentée par une suite de variables aléatoires non corrélées, de moyenne nulle et de variance finie constante (Adjengue, 2014). Plus simplement, un bruit blanc est une série qui ne présente pas d'autocorrélation.

Pour déterminer si une série est stationnaire ou non (et donc s'il est nécessaire de procéder à une ou plusieurs différenciation), nous pouvons nous servir d'un graphique d'autocorrélation

(Hyndman et al., 2021f). Autrement, nous pouvons également réaliser le test statistique de la racine unitaire (Hyndman et al., 2021f).

Le modèle Autorégressif :

Le modèle autorégressif se traduit par une équation de régression qui est une combinaison linéaire des observations passées de la variable à prédire (d'où la qualification d' "auto-régression"). Ces valeurs passées peuvent être qualifiées de "valeurs décalées", comme le mentionnent Hyndman et al. (2021f), car nous nous décalons dans le passé à partir du temps t. Mathématiquement, le modèle autorégressif d'ordre p se présente comme suit :

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

Qui peut être également écrit sous la forme polynomiale suivante :

$$Soit\ \tilde{y}_t = y_t - \mu, avec\ \mu = E(y_t).$$

$$Nous\ avons\ alors :$$

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \phi_2 \tilde{y}_{t-2} + \cdots + \phi_p \tilde{y}_{t-p} + \varepsilon_t \iff \phi(B)\tilde{y}_t = \varepsilon_t$$

$$O\grave{u}\ \ \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

Notez que le processus stochastique sur lequel repose le modèle est dit inversible, car il respecte la condition suivante :

$$Pour\ y_t = c + \sum_{p=1}^{\infty} \phi_p\, y_{t-p} + \varepsilon_t$$

$$\sum_{p=1}^{\infty} |\phi_p| < \infty$$

L'usage des modèles autorégressifs est généralement restreint aux données stationnaires. Pour nous conformer à la condition générale de stationnarité, les racines du polynôme autorégressif $\phi(B)$ de degré p doivent se trouver en dehors du cercle unitaire sur le plan complexe (Adjengue, 2014) et (Hyndman et al., 2021b).

Lorsque nous faisons usage d'un modèle d'ordre 1, nous pouvons également dire que ce dernier est inversible si $|\phi| < 1$. Lorsque l'ordre augmente, la contrainte est plus difficile à exprimer.

Remarquez que bien que la condition de stationnarité puisse se complexifier lorsque le degré du polynôme augmente, la librairie que nous utiliserons (“fable”) se chargera elle-même de respecter les contraintes associées.

Les moyennes mobiles :

Le processus des moyennes mobiles d’ordre q est fondé sur l’appréhension de la valeur présente de la série chronologique par une somme pondérée des erreurs de prédiction passées. Autrement dit :

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

Tel que le présente Adjengue (2014), les processus des moyennes mobiles est stationnaire mais pas nécessairement inversible. L’inversibilité présente pour ce processus des propriétés mathématiques intéressantes, ce qui nous pousse à nous conformer à la condition d’inversibilité suivante :

$$\text{Soit le polynôme de degré } q \;\; \theta(\mathrm{B}) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

Le processus est dit inversible si les racines du polynôme se trouvent en dehors du cercle unitaire sur le plan complexe. Nous pouvons également dire que le modèle est inversible si $|\theta| < 1$

ARIMA :

Puisque la méthode repose sur l’utilisation de séries que nous rendons stationnaires (ou qui le sont déjà) par différenciation, nous obtenons un signal dont les observations ne dépendent pas du temps. ARIMA ne permet donc pas, en général, d’interpréter la structure de la série temporelle en termes de saisonnalité. En contrepartie, ARIMA permet de modéliser une importante variété de schémas temporels, qui ne peuvent souvent pas être facilement appréhendés par d’autres méthodes (dont le lissage exponentiel). D’autre part, nous verrons à la fin de cette section qu’il existe tout de même des modèles ARIMA saisonniers qui eux, peuvent nous permettre de réaliser des prédictions en tenant compte d’un phénomène saisonnier (malgré la différentiation). C’est d’ailleurs un modèle dont nous ferons usage dans ce mémoire.

Mathématiquement, le modèle général se présente comme suit :

$$y'_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Avec :

y'_t : *les séries différenciées*

Il est ainsi constitué d'une portion autorégressive et d'une portion qui se rapporte aux moyennes mobiles. Plus précisément, notre variable d'intérêt est expliquée par une combinaison linéaire de ses valeurs passées ainsi que des valeurs passées des erreurs de prédiction. Notez que les coefficients des deux portions sont contraints pour assurer les conditions de stationnarité (contraintes sur ϕ) et d'inversibilité (contraintes sur θ).

Les modèles ARIMA, lorsque nous désirons les appliquer, nous exigent de déterminer 3 éléments :

p, l'ordre de la portion autorégressive, d, le degré de la différenciation requise et q, l'ordre de la portion moyenne mobile. Nous pouvons nous servir de la librairie « fable » pour déterminer automatiquement un modèle ARIMA adapté à nos données. Le modèle s'écrit alors « ARIMA(p, d, q) ». L'ingénierie qui se cache derrière cette automatisation est l'application de multiples tests KPSS ainsi qu'une minimisation de l'AICc (Hyndman et al., 2021f). Il reste nécessaire cependant de comprendre comment les différents éléments influencent le modèle.

Les paramètres p et q impactent les prédictions à court terme, alors que l'inclusion d'une constante c et le choix du paramètre d influencent les prédictions à long terme. Les constantes c et d influencent ensemble ce vers quoi vont converger les prédictions. D'autre part, le degré de différenciation impactera directement la taille de l'intervalle de prédiction. L'augmentation de d induit en effet une augmentation de la largeur de l'intervalle. Lorsque d est égal à 0, l'écart-type des prédictions à long terme sera égal à l'écart-type des données historiques. Pour de plus amples renseignements, le lecteur est invité à consulter (Hyndman et al., 2021f).

Lorsque p, d et q ont été déterminés, il est nécessaire de caractériser les paramètres du modèle. Ces derniers sont estimés par la vraisemblance maximale. Notez que comme pour les modèles d'espace états, nous pouvons départager les différents modèles existants en nous servant d'un critère d'information tel que l'AICc.

Lorsque nous avons déterminé notre modèle ARIMA, nous devons avant de l'utiliser pour réaliser des prédictions vérifier si les résidus sont équivalents à du bruit blanc. Pour ce faire, nous pouvons générer un visuel et effectuer un test de Portmanteau (Hyndman et al., 2021f). Notez que tel que

mentionné par Hyndman et al. (2021a), le test offre des résultats plus précis lorsque nous ajustons les degrés de liberté de sorte à tenir compte du nombre de paramètres que comprend le modèle ARIMA. En considérant un tel ajustement, nous réalisons en réalité un test qui est appelé Ljung-box. Ce dernier teste l'hypothèse nulle de l'absence d'autocorrélations.

Lorsque nous avons obtenu un modèle convenable, nous pouvons procéder à la prédiction. Nous utiliserons, tel que précédemment, la fonction « forecast ». La théorie qui se cache derrière la prédiction à partir d'un modèle ARIMA est cependant disponible dans les références présentées. Notez que la fonction nous permet d'obtenir un intervalle de « prédiction », lequel est construit à partir de l'équation suivante :

$$Soit\ \varepsilon_t \sim N(0, \sigma^2), alors :$$

$$\hat{y}_{t+h|t} \pm c \times \sqrt{V_{t+h|t}}$$

$Avec\ c\ qui\ dépend\ de\ la\ probabilité\ de\ couverture$
$V_{t+h|t}\ la\ variance\ estimative\ de\ la\ prédiction$

$Et\ V_{t+1|t} = \hat{\sigma}^2\ \forall\ ARIMA(p,d,q)$

Pour un horizon de prédiction supérieur à 1, la procédure est différente et plus compliquée. Pour quelques approfondissements, le lecteur est invité à consulter (Hyndman et al., 2021f).

ARIMA saisonnière (SARIMA) :

Il existe d'autre part les modèles d'ARIMA saisonniers. Ces derniers sont très semblables aux précédents. À la différence des modèles ARIMA précédents, les saisonniers comportent une seconde partie qui décrit la nature de la saisonnalité de la série chronologique (Hyndman et al., 2021f). Lorsque nous les appliquerons, nous emploierons ainsi la notation suivante :

$$ARIMA(p,d,q)(P,D,Q)_m$$

$P : ordre\ autorégressif\ saisonnier$
$D : degré\ de\ différenciation\ saisonnière$
$Q : ordre\ moyenne\ mobile\ saisonnier$
$m : nombre\ d'observations\ par\ année$

p, q, P, Q et c sont déterminés de sorte que l'AICc soit minimisé. D dépend de la force de la saisonnalité et d est caractérisé à l'aide d'un test KPSS. Lorsque la saisonnalité n'est pas marquée, D est égal à 0 (Hyndman et al., 2021j).

Mathématiquement, tel que présenté par Hyndman et al. (2021j), le modèle se décrit de la façon suivante :

$$\Phi(B)\phi(B)(1-B)^d(1-B)^D y_t = c + \Theta(B)\theta(B)\varepsilon_t$$

Dans le cas où notre série est marquée par une saisonnalité, lorsque nous réalisons la différenciation pour la rendre stationnaire, nous ne procédons pas exactement de la même façon. Au lieu de faire la différence entre une observation et la précédente, nous réalisons la différenciation saisonnière entre une observation et celle d'avant à la même saison. Si nous réalisons une deuxième différenciation, nous pouvons cette fois-ci faire la différence entre une différence et la précédente.

Remarquez que lorsque nous comparons différents modèles ARIMA saisonniers, pour faire preuve de rigueur statistique, nous devons comparer des modèles ayant les mêmes paramètres *d* et *D* (donc le même nombre de différenciations) (Hyndman et al., 2021j).

Tel que pour les modèles précédents, nous pouvons procéder de manière automatique sur R.

2.3.3 Évaluation du modèle

Comme évoqué au début de cette section, avant de procéder à la phase de prédiction, il est nécessaire d'évaluer le modèle. Plusieurs méthodes sont disponibles pour évaluer la performance du modèle, tel que présenté par Hyndman et al. (2021f). Nous en présenterons par ailleurs à la fin de ce chapitre. Pour évaluer le modèle, cependant, nous procéderons plutôt à une analyse des résidus. Notre objectif est ici de déterminer si notre modèle a saisi autant d'informations que disponible. D'autre part, pour obtenir notre « distribution de probabilité » ainsi que nos « intervalles de prédiction », nous devons en premier lieu nous assurer que les conditions qui se rapportent aux résidus sont respectées.

Pour évaluer la normalité des résidus, nous pouvons les visualiser sous forme d'histogramme, tel que nous l'exposent Hyndman et al. (2021f). Pour évaluer la présence d'autocorrélation, nous pouvons également procéder à une analyse graphique. Nous pouvons aussi effectuer un test de Ljung-Box, tel que mentionné précédemment. Lorsque nous réalisons ce test sur R, la fonction

nous renvoie une p-value. Si cette dernière est inférieure à 0.05, alors nous rejetons l'hypothèse que les résidus sont équivalents à du bruit blanc (absence d'autocorrélations). Pour évaluer l'homoscédasticité, nous pouvons aussi procéder à une analyse graphique.

Lorsque les hypothèses d'homoscédasticité, de normalité ou de bruit blanc ne sont pas respectées, Hyndman et al. (2021i) nous mentionnent que cela peut impacter la taille de l'intervalle de prédiction, plus que la qualité de la prédiction ponctuelle.

2.4 Statistique Bayésienne

Contrairement à la statistique classique, la statistique bayésienne nous permet de réaliser des estimations valides pour n'importe quelle taille d'échantillon (McElreath, 2020). Cela peut comporter un avantage, car pour certaines familles de produits, l'historique est plus récente et il y a donc peu de données. D'autre part, la statistique bayésienne nous permet d'obtenir une distribution prédictive de la demande de familles de produits ainsi que des intervalles de « crédibilité ». Nous verrons que ces derniers diffèrent de ceux qui ont précédemment été introduits. Nous exposerons dans la présente sous-section les notions associées à la statistique bayésienne. Les informations qui seront présentées sont issues de (Davidson-Pilon, 2016), (Gelman et al., 2021), (Jalbert, 2023), (Johnson et al., 2021), (Kruschke, 2014), (McElreath, 2020) et (Ross, 2022).

La statistique bayésienne doit son nom au statisticien du 18e siècle Thomas Bayes, précurseur en la matière. L'approche repose sur la considération que l'aléatoire fait partie intégrante des données. Cette pensée s'intègre dans l'idée qu'avec suffisamment d'informations pertinentes, nous serions en mesure de prédire toute chose avec exactitude. Selon ce mode de pensée, l'aléatoire ne représente donc que ce que nous n'avons pas encore la capacité de saisir, nous laissant dans l'incertitude (McElreath., 2020).

Concrètement, la statistique bayésienne sert les mêmes objectifs que la statistique "classique" (aussi appelées "fréquentiste"). C'est-à-dire qu'elle vise à l'élucidation de phénomènes à l'aide de données. Ainsi, les deux disciplines se servent d'informations pour adapter des modèles, tester des hypothèses et réaliser des prédictions (Johnson et al., 2021). En réalité, la différence entre les deux approches peut se résumer à leur définition de ce qu'est une probabilité. En effet, la statistique

bayésienne considère une probabilité comme étant la plausibilité relative (ou crédibilité) de la survenue d'un événement. À l'inverse, la statistique fréquentiste la considère comme étant la fréquence relative d'un événement répétable de façon identique. Dans le second cas, la notion de probabilité s'appuie sur une certitude : nous savons comment l'événement se déroule, et nous tenons pour acquis qu'il se déroulera de la même façon dans le futur. Dans le premier cas, nous considérons plutôt que compte tenu de ce que nous savons, l'événement pourrait se passer d'une certaine façon, sans pour autant en être certain. C'est-à dire que nous nous appuyons sur ce que nous savons du passé tout en gardant à l'esprit que ce que nous obtiendrons à l'avenir risque d'être différent. La notion d'incertitude est donc centrale dans la première approche. En tenant compte de sa définition bayésienne, la probabilité peut être envisagée comme une mesure de l'incertitude d'un événement, ou plus précisément, d'une variable, tel que nous le verrons. La mesure de cette incertitude se présente sous forme de distribution de probabilité, qui peut représenter la valeur d'un paramètre, d'une observation future ou encore d'une hypothèse.

Avant d'obtenir de tels résultats, il est cependant nécessaire de construire un modèle probabiliste adapté, à partir duquel nous pourrons réaliser ces analyses. Tel que mis en lumière par Johnson et al. (2021) et McElreath (2020), la simple modélisation d'une variable pour en comprendre le comportement est dite "descriptive" (c'est l'exercice auquel nous nous prêterons pour répondre à notre problématique). Il est important de noter qu'il existe également des modèles causaux. Ces derniers sont axés sur ce qui induit le comportement d'une variable. C'est-à-dire que l'intérêt est porté sur la relation entre la variable étudiée et celles qui causent son état. La modélisation causale se rapporte ainsi à des exercices de régression ou de classification. Cela ne sera pas couvert dans ce mémoire puisque notre cas d'étude ne s'y prête pas.

Avant de construire le modèle probabiliste, il est nécessaire de déterminer ce que nous souhaitons modéliser. Plus particulièrement, lorsque nous modélisons un phénomène, nous devons nous intéresser aux variables qui le décrivent et faire la distinction entre chacune d'elles. Parmi elles, nous retrouvons des variables observables et non observables. Tel que présenté par McElreath (2020), ce que nous souhaitons inférer est généralement non-observable et s'apparente à un paramètre qui peut être déduit à partir de variables observables. Prenons l'exemple suivant : nous souhaitons étudier la proportion de prisonniers enfermés pour vol à l'étalage dans un échantillon de taille 100 prélevé de façon aléatoire. Le nombre de prisonniers répondant à ce critère est

observable. Le nombre de prisonniers n'y répondant pas ainsi que le nombre total de détenus sont également observables. À l'inverse, la proportion ne l'est pas directement. Ce sont des variables distinctes qui sont pourtant liées, car la proportion peut être déterminée à l'aide des autres variables. Nous utiliserons ces informations plus tard dans nos explications. Cette clarification faite, nous introduirons tout au long de cette section les 3 éléments qui constituent un modèle probabiliste bayésien.

2.4.1 Les trois composantes du modèle probabiliste

Le modèle en question vise à l'obtention d'une distribution de probabilité d'un paramètre (une moyenne, une proportion, etc) qui tienne compte de nos connaissances et des données dont nous disposons. La démarche repose sur la considération que le comportement de la variable étudiée (dans l'exemple précédent, la proportion) est susceptible de se répéter dans le futur tout comme il peut changer (car elle est incertaine). Le modèle tient donc compte des caractéristiques de la variable et de son comportement passé. Lorsque nous parlons de caractéristiques, nous faisons d'abord allusion à la nature de la variable, c'est-à-dire les valeurs qu'elle peut prendre. Pour modéliser cette dernière, nous choisirons donc une fonction adaptée à ses attributs. Ainsi, si nous reprenons l'exemple de la proportion de prisonniers, θ, nous choisirons probablement une loi Bêta, car θ est une variable continue qui prend ses valeurs dans l'intervalle [0;1]. Il nous faut ensuite incorporer ce que nous savons du comportement passé de θ. Pour ce faire, nous ajustons les paramètres de notre fonction de sorte que la distribution soit représentative de nos connaissances initiales. Ce que nous venons de décrire est l' "*a priori*", le premier élément dont nous avons besoin. Cela nous permet en réalité d'exprimer la plausibilité relative initiale de chaque valeur que peut prendre θ. Notez que lorsque nous n'avons aucune connaissance sur le comportement de la variable, nous pouvons utiliser une *a priori* dite "vague" ou "non informative".

Tel que mentionné précédemment, nous considérons que la variable est incertaine. Nous ne pouvons donc pas nous contenter de fonder l'intégralité de nos analyses et autres inférences sur nos seules connaissances *a priori* de la variable. Pour mettre à jour ce que nous savons, nous utilisons les données réelles présentes. L'échantillon prélevé aléatoirement, nous évaluons la compatibilité relative de nos données avec les différentes valeurs que peut prendre notre variable

d'intérêt. Pour quantifier cette compatibilité, nous devons modéliser la relation entre nos données observables et nos connaissances.

Comme exposé auparavant, les variables inférées à partir de variables observables peuvent être qualifiées de "paramètres". Les données prélevées n'étant pas prédéterminées, nous les représentons par une variable aléatoire. La dépendance entre nos données et nos connaissances est illustrée par la fonction de probabilité conditionnelle $f(y|\theta)$, avec y notre variable aléatoire et θ notre paramètre. Lorsque nous avons plus d'un paramètre, chacun doit être modélisé.

L'inconvénient est que la véritable valeur de notre paramètre est inconnue. Nous ne connaissons que y. C'est la raison pour laquelle nous utiliserons plutôt la vraisemblance $L(\theta|y)$. C'est cette dernière qui nous permet d'évaluer la compatibilité relative de nos données avec les différentes valeurs de θ. Cette vraisemblance est le deuxième élément dont nous avons besoin. Tel que souligné par Johnson et al. (2021), si $L(\theta|y) = f(y|\theta)$, il ne faut pas oublier que la fonction de vraisemblance n'est pas une fonction de probabilité. Cela signifie que la somme des vraisemblances n'est pas égale à 1. C'est ce qui nous amène à avoir recours à notre dernier élément : la constante de normalisation. Cette dernière fait en sorte que la sommes de probabilités *a posteriori* est égale à 1.

Pour finalement obtenir la distribution dite "*a posteriori*" de notre paramètre θ, nous devons lier nos trois éléments à l'aide du Théorème de Bayes. La logique sur laquelle repose cette approche est la suivante :

La probabilité de n'importe quelle valeur spécifique de θ considérant nos données est égale au produit de la plausibilité relative de nos données sachant θ par la plausibilité relative initiale de θ, le tout divisé par la probabilité moyenne des données (McElreath, 2020).

Autrement dit :

$$A\ posteriori = \frac{Vraisemblance\ \times\ A\ priori}{Constante\ de\ normalisation}$$

$$\Leftrightarrow f(\theta|y) = \frac{L(\theta|y) \times f(\theta)}{f(y)} = \frac{f(y|\theta) \times f(\theta)}{\int f(y|\theta) \times f(\theta).d\theta}$$

$$f(\theta|y) \propto f(y|\theta) \times f(\theta)$$

Notez que lorsque nous avons plusieurs paramètres, nous avons également plusieurs *a priori* à prendre en considération. Dans notre cas, nous en aurons deux : celui de la moyenne et celui de la variance. D'autre part, la mise à jour de nos connaissances peut être ou non séquentielle (une observation aléatoire à la fois), et dans le cas où elle l'est, l'ordre dans lequel nous incluons nos observations n'influence pas le résultat. Cela étant dit, tel que le souligne McElreath (2020), la propriété de l'invariance de l'ordre des données n'est pas toujours applicable. Il est donc parfois nécessaire d'adapter notre modélisation.

Le signe " $\propto$ " signifie "proportionnel à". Grâce à cette relation de proportionnalité, nous ne sommes pas toujours contraints de calculer la constante de normalisation. Nous pouvons en effet, en développant le produit et en éliminant les termes qui ne dépendent pas de θ, reconnaître la forme fonctionnelle de lois de probabilité connues. Cela nous permet de déduire la constante de normalisation, et donc l'*a posteriori*.

Le calcul de l'*a posteriori* peut être simplifié lorsque nous employons des lois conjuguées. C'est-à-dire que lorsque nous choisissons certaines combinaisons de lois pour la vraisemblance et l'*a priori*, nous savons que l'*a posteriori* a la même forme paramétrique que l'*a priori*. Parmi les familles conjuguées, nous retrouvons notamment la combinaison Normale-Normale (la vraisemblance et l'*a priori* sont des lois normales), la combinaison Poisson-Gamma, ainsi qu'Exponentielle-Gamma.

Lorsque nous ne pouvons déduire l'*a posteriori* de l'une de ces deux façons, nous sommes obligés de l'approximer.

2.4.2 La caractérisation de l'*a posteriori* par simulation

Parmi les méthodes dont nous pouvons faire usage pour caractériser l'*a posteriori*, les plus communes sont les techniques d'approximation par Chaînes de Markov, d'approximation quadratique (Gaussienne) ou d'approximation par grille. Notre étude sera restreinte aux Chaînes de Markov. Nous avons jugé inappropriée l'approximation par grille en raison de sa gestion inefficace des modèles impliquant plusieurs paramètres. Nous avons également estimé l'approximation quadratique inadaptée, car, bien que plus efficace que la précédente lorsque nous avons plusieurs paramètres, elle n'est pas optimale pour les modèles hiérarchiques (multiniveaux). De surcroît, elle suppose que l'*a posteriori* peut être approximée par une Gaussienne, ce qui peut

ne pas toujours être l'approche la plus indiquée. Bien que nous ne fassions pas usage de modèles hiérarchiques, nous avons estimé qu'il était plus judicieux d'employer des chaînes de Markov, car la technique est plus facilement généralisable. Pour de plus amples renseignements sur l'approximation quadratique, le lecteur est invité à consulter la référence suivante : (McElreath, 2020)

L'usage de Chaînes de Markov peut nous permettre d'estimer une distribution de probabilité *a posteriori*. Concrètement, nous pouvons, à l'aide de celles-ci obtenir une image d'une haute résolution de l'*a posteriori* sans avoir à calculer d'intégrale (Kruschke, 2014). Tel que le présente McElreath (2020), la chaîne de Markov peut être décrite comme un processus stochastique. Tel qu'exposé par Johnson et al. (2021), la simulation MCMC (par chaînes de Markov) consiste en l'échantillonnage de N valeurs dépendantes de notre paramètre qui constitueront une chaîne de taille N. Chaque échantillon dépend du précédent et est issu d'un modèle qui n'est pas celui de l'*a posteriori*. Nous recommandons au lecteur de lire les références sus-nommées, car elles offrent une explication digeste d'une matière compliquée que nous ne ferons qu'approcher. Notez simplement que, "mathémagiquement" (Johnson et al., chap6), cela fonctionne.

Il existe plusieurs algorithmes de MCMC. Nous utiliserons dans notre cas l'algorithme Métropolis, ce pourquoi nous n'aborderons brièvement que ce dernier. Nous aurions aussi pu considérer l'échantillonnage de Gibbs, qui aurait par ailleurs été nécessaire si nous avions eu plus de deux paramètres à modéliser. Tel que synthétisé par Jalbert (2023), l'algorithme de Métropolis peut se résumer de la façon suivante :

Soit $g(\theta|y)$ la forme fonctionnelle de notre densité *a posteriori* et C la constante de normalisation inconnue. Nous avons alors :

$$f(\theta|y) = \frac{g(\theta|y)}{C}$$

Étape 1) L'algorithme commence par une valeur initiale $\theta^{(1)}$ à l'état 1 que nous fixons pour θ.

Étape 2) Nous définissons par la suite une marche aléatoire δ (pas) qui nous permettra d'explorer les valeurs de θ. Dans notre cas, nous utilisons la fonction pm.Metropolis pour la déterminer. L'algorithme utilise nécessairement une « distribution de sauts/propositions » symétrique (Gelman

et al., 2021) pour générer des candidats δ, telle qu'une loi normale (Kruschke, 2014,). Notez que cela n'est pas un impératif pour la forme généralisée de l'algorithme (Métropolis Hastings).

Étape 3) Un candidat $\tilde{\theta}$ est alors proposé pour devenir la valeur suivante de θ à l'itération $t+1$, avec $\tilde{\theta} = \theta^{(t)} + \delta$, t étant le numéro de l'itération. Nous évaluons si le candidat est recevable considérant la forme fonctionnelle. C'est à dire que nous calculons :

$$p = min\left\{\frac{g(\tilde{\theta}\,|\mathrm{y})}{g(\theta^{(t)}|y)}, 1\right\}$$

Étape 4) Nous attribuons la réalisation suivante :

$$\theta^{t+1} = \begin{cases} \tilde{\theta}\ avec\ une\ probabilité\ p \\ \theta^{(t)}\ avec\ une\ probabilité\ 1-p \end{cases}$$

Nous recommençons ces étapes pour N itérations, qui vont représenter les N maillons de la chaîne. Notez qu'il est nécessaire d'avoir plus d'une chaîne pour observer une convergence (et donc une simulation réussie). Dans notre cas, nous en aurons 3. La convergence signifie que chaque chaîne explore la bonne et la même distribution. Ce que nous recherchons, c'est que chaque chaîne, en commençant par des valeurs différentes, converge vers la même distribution *a posteriori*.

Lorsque nous parlons de mise en pratique d'une simulation par chaîne de Markov, il est indispensable d'aborder les mesures qui nous permettent d'en évaluer la qualité. Cette étape est importante, car nous utilisons le modèle probabiliste obtenu pour réaliser nos inférences. Ainsi, si la simulation est défectueuse, nos inférences le seront également. Nous présentons donc à présent ce que nous utiliserons pour mesurer la qualité des simulations, à savoir les tracés de la chaîne, la statistique de Gelman-Rubin, l'ESS ainsi que le MCSE.

Les tracés permettent de vérifier visuellement s'il y a eu convergence vers une distribution stationnaire (c'est-à-dire qui se conserve dans le temps). Une bonne chaîne présente des fluctuations aléatoires autour d'une moyenne stable après la période de chauffe. Les différents auteurs que nous avons consultés évoquent qu'elle ressemble alors à une grosse chenille poilue. Lorsqu'il y a convergence, toutes les chaînes devraient à peu près être superposées (Kruschke, 2014). Le graphique de densité, qui est également présenté lorsque nous générons les tracés, peut aussi nous renseigner sur la représentativité des échantillons. Là encore, nous nous attendons à ce

que les différentes courbes soient sensiblement alignées. L'évaluation des tracés demeure cependant insuffisante pour conclure que les échantillons offrent un aperçu représentatif de la distribution *a posteriori*.

Le $\hat{R}$ (statistique de Gelman-Rubin) est un outil qui nous permet de diagnostiquer la convergence (et ainsi la représentativité des échantillons). Il s'agit du ratio entre la variance moyenne à l'intérieur des chaînes et la variance totale. Lorsque le ratio est égal à 1, il y a de grandes chances que les chaînes convergent. Tel que mentionné par McElreath (2020), bien que nous ayons un $\hat{R}$ de 1, il est toujours possible que la chaîne ne soit pas valide. Cette mesure est utile pour signaler un danger lorsqu'elle est inférieure à 1. Cependant, elle ne garantit pas une simulation de qualité à elle seule. Comme nous l'expose Kruschke (2014), un $\hat{R}$ supérieur à 1 peut indiquer que la chaîne est orpheline (qu'elle explore un environnement inhabituel), voir qu'elle est coincée. Dans un cas comme celui-là, nous ne pouvons conclure qu'elle converge.

L' "ESS" (ou estimé du nombre d'échantillons efficaces) évoque une approximation de la longueur qu'aurait la chaîne si elle était parfaitement non corrélée. C'est-à-dire qu'elle représente une mesure de la quantité d'informations indépendantes dans une chaîne autocorrélée (Ross, 2022). Tel que mentionné par McElreath (2020), le nombre approximatif d'échantillons efficaces est habituellement plus petit que le nombre d'échantillons que nous avons généré. Lorsque l'ESS est largement plus petit que le nombre d'itérations auquel nous soustrayons les itérations de rodage, la métrique indique que la chaîne n'est pas efficiente. Pour obtenir un aperçu précis et plutôt stable des régions les moins denses de la distribution (telle que les limites d'un intervalle de densité *a posteriori* le plus élevé à 95%), Kruschke (2014) recommande un nombre d'échantillon efficaces d'au moins 10 000.

L'erreur standard de Monte Carlo (MCSE) est une mesure de la précision des estimations obtenues à l'aide des chaînes de Markov. Elle mesure la variabilité d'une exécution à l'autre des valeurs de la variable que nous souhaitons estimer (dans le cas de notre exemple précédent, la proportion) sur de nombreuses exécutions de la chaîne. Concrètement, elle aide à quantifier l'incertitude de notre estimation. Si la mesure est grande, alors cela signifie que le résultat est imprécis. Comme l'expose SAS Institute Inc. (2017), lorsque le ratio $\frac{MCSE}{Déviation\ standard}$ est petit, alors cela signifie que la chaîne

de Markov (dans notre cas, ce sera la combinaison des 3 mélangées) s'est stabilisée et que l'estimation du paramètre varie peu.

Notez que nous aurions également pu nous servir du graphique de l'autocorrélation. Ce dernier présente l'autocorrélation à l'intérieur d'une chaîne. Tel que mentionné par (Ross, 2022, chap18), lorsque la fonction d'autocorrélation (ACF) requiert beaucoup de temps pour décroître jusqu'à 0, alors la chaîne présente un niveau élevé de dépendance. Concrètement, cela signifie que la chaîne aura tendance à tourner en rond. Cette mesure peut nous aider à déterminer si notre pas est trop petit. Dans notre cas, nous avons décidé de ne pas l'utiliser à des fins de simplifications, car nos résultats ne laissent entrevoir aucun problème majeur au niveau de l'échantillonnage. Nous convenons cependant que de faire usage de tels graphiques est toujours intéressant, particulièrement lorsque les tracés présentent des anomalies.

2.4.3 Évaluation du modèle

Lorsque nous avons déterminé notre *a posteriori* (et donc notre modèle probabiliste), nous devons procéder à son évaluation. Tel que nous l'expose McElreath (2020), l'évaluation du modèle bayésien visent à la vérification de deux éléments. Nous souhaitons en premier lieu vérifier que le modèle s'est convenablement ajusté aux données (sur le simple plan pratique/informatique). Dans un second temps, nous voulons nous assurer que le modèle est adéquat pour répondre aux objectifs qu'il est censé servir (prédiction, test d'hypothèse, etc). Bien que l'ajustement ait pu se dérouler sans encombre d'un point de vue informatique, il est possible que ce dernier ne rende pas suffisamment compte du comportement réel des données (auquel cas le modèle est inadéquat). Dans de telles circonstances, il en découle de mauvaises inférences.

Nous ne conserverons pour notre étude que deux des méthodes exposées par Gelman et al. (2021) et McElreath (2020). Il s'agit d'une analyse graphique et de la méthode « PSIS » (« Pareto-Smoothed Importance Sampling Cross-Validation »). Pour de plus amples renseignements sur le sujet, le lecteur est invité à consulter les références susnommées.

Pour évaluer si l'ajustement a fonctionné, nous simulons des données à partir de l'*a posteriori* pour vérifier que nous obtenons des données semblables à celle qui étaient utilisées pour construire le modèle. Pour ce faire, nous prélevons des échantillons de la fonction de vraisemblance, paramétrée à l'aide de l'*a posteriori*. Les données issues de l'*a posteriori* sont obtenues dans des proportions

qui illustrent leurs plausibilités relatives. Les données simulées qui en découlent sont alors issues de ce qui s'appelle la distribution prédictive *a posteriori*. Si le modèle fonctionne convenablement, nous devrions obtenir des données semblables à celles utilisées pour la construction du modèle (McElreath, 2020). Nous jugeons de cette similitude en procédant à une analyse graphique.

Pour évaluer l'adéquation du modèle, nous pourrions tout simplement réaliser nos prédictions (en procédant de la façon présentée dans la prochaine section) et mesurer l'écart entre les valeurs réelles et prédites. Nous pourrions alors juger de l'efficacité du modèle par validation externe. Cela dit, tel que mis en lumière par Gelman et al. (2021), il est dans certains cas souhaitable d'évaluer le modèle avant de l'exécuter sur de nouvelles données. C'est la raison pour laquelle nous pouvons avoir recours à une méthode comme celle de PSIS. Cette dernière nous permet d'obtenir un aperçu de la performance moyenne que notre modèle est susceptible d'avoir. La méthode nous offre la possibilité d'approximer un score de validation croisée sans avoir à effectuer les calculs répétés impliqués sur de multiples échantillons.

Pour obtenir le score de validation croisée, il est commun de faire usage de la méthode « Leave-one-out-cross-validation ». Nous n'en expliquerons que très sommairement le concept. Le lecteur est invité à consulter les références précédemment présentées pour davantage de renseignements. En résumé, cette dernière consiste en l'ajustement du modèle à évaluer sur $\mathrm{N} - 1$ des données dont nous disposons, et le test de ce dernier sur la dernière donnée disponible. Le tout est répété N fois, de sorte que chaque point soit une fois la donnée test, tandis que les autres sont les données de calibration. Le modèle est donc calibré avec $\mathrm{N} - 1$ données, puis nous tentons de prédire la dernière donnée par la suite pour en tester la performance. Après N répétitions, nous utilisons les résultats pour scorer la supposée performance du modèle. Chaque répétition implique le calcul chronophage de l'*a posteriori*.

La méthode PSIS, nous permet d'éviter de calculer plusieurs fois l'*a posteriori*. Pour ce faire, des poids sont utilisés pour donner plus ou moins d'importance à certaines données. En limitant le poids d'une donnée (sans l'éliminer pour autant), nous pouvons approximer le résultat que nous aurions obtenu si le modèle avait été ajusté sans ce point. Nous pouvons alors répéter la procédure pour chaque point, sans avoir à réajuster le modèle. Un lissage de Pareto est utilisé pour éviter que certains poids ne nuisent à l'estimation de la performance du modèle (lorsqu'ils sont trop importants) (McElreath, 2020).

Mathématiquement, PSIS se décrit de la façon suivante :

$$lppd_{IS} = \sum_{i=1}^{N} log\left(\frac{\sum_{s=1}^{S} r(\theta_s)p(y_i|\theta_s)}{\sum_{s=1}^{S} r(\theta_s)}\right)$$

Avec :

$$r(\theta_s) = \frac{1}{p(y_i|\theta_s)}$$

Et :

$$R \sim GDP(u, \sigma, k)$$

$$p(r|u, \sigma, k) = \begin{cases} \frac{1}{\sigma}\left(1 + k\left(\frac{r-u}{\sigma}\right)\right)^{-\frac{1}{k}-1} & , \ k \neq 0 \\ \frac{1}{\sigma}\, e^{\frac{r-u}{\sigma}} & , \ k = 0 \end{cases}$$

$u : paramètre\ de\ localisation$

σ : $échelle$

$k : paramètre\ de\ forme$

Dans la mesure où nous nous servirons d'une fonction qui approximera le score de validation croisée automatiquement sur python, nous n'irons pas davantage dans les détails théoriques.

La fonction qui nous permet d'exécuter la méthode PSIS nous renvoie une estimation de la densité prédictive ponctuelle logarithmique attendue (« elpd_loo »), le nombre effectif de paramètres (p_loo) et le diagnostic de Pareto « k » (pareto_k). Le premier élément nous donne une estimation de la précision de la prédiction. Une valeur élevée indique une meilleure précision. Un score de validation croisée estimatif peut être obtenu en multipliant elpd_loo par -2.

Le second élément est, tel que le nomme McElreath (2020) une forme de pénalité de surajustement. Lorsque le nombre effectif de paramètres est supérieur au nombre de paramètres (du modèle) et au nombre total d'observations, cela peut nous indiquer que le modèle dispose d'une faible capacité de prédiction. Le dernier extrant nous renseigne sur la fiabilité de l'approximation. Une valeur inférieure à 0.5 indique que l'approximation est probablement fiable. Une valeur comprise entre

0.5 et 0.7 évoque une instabilité moyenne, alors qu'une valeur k supérieur à 0.7 exprime une instabilité significative (McElreath, 2020) et (Vehtari et al., 2024).

2.4.4 La distribution prédictive

Lorsque nous avons notre distribution *a posteriori* sur les paramètres du modèle, nous pouvons finalement réaliser nos analyses *a posteriori* (ou inférences). Ce mémoire n'aborde que la prédiction, car c'est le seul objet de notre cas d'étude. Plus précisément, nous souhaitons obtenir la distribution prédictive d'une observation future $\tilde{y}$. Pour l'obtenir, nous devons à la fois considérer l'incertitude liée à notre paramètre et celle de l'échantillonnage. Il existe, pour chaque valeur que peut prendre notre paramètre, une distribution de probabilité de l'observation n+1. Nous savons cependant grâce à l'*a posteriori* que certaines valeurs sont plus plausibles que d'autres. C'est pourquoi la distribution prédictive s'obtient en réalisant la moyenne des distributions d'échantillonnages (distributions obtenues pour chaque valeur du paramètre), pondérée par les plausibilités de l'*a posteriori* pour chaque valeur du paramètre. Cela se traduit par l'équation suivante :

$$f(\tilde{y}|y) = \int f(\tilde{y}|\theta) \times f(\theta|y).d\theta$$

Naturellement, le choix de l'*a priori* et de la fonction de vraisemblance influence ce que sera l'*a posteriori,* et donc les inférences qui seront réalisées à partir de cette dernière. Il y a une part de subjectivité dans le choix des composantes du modèle probabiliste. Le modèle que nous construisons n'est pas nécessairement parfaitement représentatif de la réalité. Cela étant dit, tel que mis en lumière par McElreath (2020), les modèles que nous construisons n'ont pas besoin d'être une reproduction parfaite de la réalité pour produire des inférences hautement précises et/ou utiles. En outre, comme nous l'expose (Johnson et al., 2021, chap4), à mesure que la quantité de données augmente, l'influence de l'*a priori* diminue. Cela signifie qu'avec suffisamment de données, deux modèles n'ayant pas les mêmes *a priori* peuvent offrir des résultats semblables.

Finalement, nous pouvons obtenir un intervalle de « crédibilité ». Ce dernier peut nous permettre d'estimer une gamme de valeurs plausibles pour une certaine probabilité de couverture. À l'inverse des intervalles de « prédiction » que nous avons présenté dans la section précédente, les intervalles de crédibilité nous présentent une incertitude qui tient compte de l'influence des paramètres.

Tel que nous le mentionne McElreath (2020), deux types d'intervalles similaires sont couramment utilisés. L' « intervalle de densité *a posteriori* le plus élevé » est, selon l'auteur, l'intervalle le plus étroit contenant la probabilité de masse spécifiée. Ce dernier représente les valeurs des observations futures les plus cohérentes avec les données. Le deuxième est l'intervalle percentile. Il s'agit d'un intervalle qui assigne des probabilités de masses égales aux deux extrémités (ou queues) de la distribution. À l'inverse du précédent, ce dernier est susceptible d'exclure certaines des valeurs les plus plausibles, tel que nous l'expose McElreath (2020). En contrepartie, l'intervalle percentile est moins coûteux sur le plan informatique, plus intuitif et se montre moins sensible au nombre d'échantillons prédictifs *a posteriori* prélevés. Dans notre cas d'étude, nous utiliserons l'intervalle percentile.

Selon que le modèle soit approché par simulation ou d'une forme analytique connue, nous pouvons procéder de deux façons différentes pour obtenir nos intervalles. Dans le premier cas, nous utilisons une fonction qui classe les échantillons simulés en ordre croissant et renvoie l'intervalle de notre choix (lequel est obtenu en calculant des fréquences). Dans le second cas, si la forme analytique est celle d'une distribution symétrique, nous pouvons nous servir des tables usuelles pour calculer nos deux intervalles (percentile et de densité *a posteriori* le plus élevé), qui sont alors équivalents. Lorsque la distribution est asymétrique, en revanche, il est plus difficile de calculer l'intervalle de densité *a posteriori* le plus élevé. Nous ne verrons pas ce dernier cas de figure dans notre cas d'étude.

2.5 Évaluation de la prédiction

Lorsque le modèle a été spécifié, évalué et que nous avons réalisé nos prédictions, il est temps d'analyser la qualité de ces dernières. Nous opterons pour des méthodes d'évaluation facilement interprétables et applicable tant pour des modèles bayésiens que classiques. Plus exactement, nous ferons usage de deux métriques ponctuelles ainsi que d'une métrique quantile. Il existe cependant bien d'autres métriques, dont des métriques qui nous permettent de comparer la performance de deux modèles (telles que les critères d'informations, à titre d'exemple). Nous adressant à des industriels, nous avons opté pour des métriques qui leurs seront certainement plus intuitives. Cela étant dit, pour de plus amples renseignements sur l'évaluation de la prédiction, le lecteur est invité

à consulter les références suivantes : (Hyndman et al., 2021f), (McElreath, 2020) et (Koutsandreas et al., 2021).

Le « Mean Absolute Error » - MAE :

Le MAE est une métrique d'erreur qui représente l'écart moyen entre la valeur prédite et la valeur réelle, en valeur absolue. Simple à saisir et à calculer, elle s'exprime selon l'unité de l'objet de prédiction et se présente comme suit :

$$MAE = \frac{1}{T} \times \sum_{t=1}^{T} |y_t - \hat{y}_t|$$

L'inconvénient du MAE, c'est qu'il dépend d'une échelle. Ainsi, lorsque nous souhaitons comparer la performance du modèle sur deux séries d'unités différentes, il est nécessaire de faire usage d'une autre métrique, comme le MAPE.

Le « Mean Absolute Percentage Error » - MAPE :

Tel que nous le mentionne Hyndman et al. (2021e), le MAPE est à la fois facilement interprétable et l'une des métriques les plus utilisées, car il ne dépend pas d'une échelle. Lorsque nous utilisons une métrique comme le MAE ou le RMSE et que nous souhaitons les utiliser sur plusieurs séries temporelles, il est difficile de pouvoir comparer et interpréter les résultats, car les échelles peuvent varier. Le MAPE permet de remédier à cette problématique. Il est en effet assez facile de se représenter une erreur en pourcentage. Cette métrique se calcule de la façon suivante :

$$MAPE = \frac{100 \times \sum_{t=1}^{T} \frac{|y_t - \hat{y}_t|}{y_t}}{T}$$

Bien que le MAPE se présente comme une métrique simple à appliquer comme à comprendre, elle s'accompagne de certains inconvénients. Ainsi, lorsque $y_t = 0$, le MAPE est indéfini. D'autre part, lorsque y_t tend vers 0, nos résultats peuvent devenir extrêmes. Faisant usage de plusieurs métriques

différentes, nous ne devrions pas nous retrouver dans l'incapacité d'évaluer nos résultats en raison de ces désavantages.

Le score quantile :

Moins évidente que les deux premières, la métrique du score quantile demeure accessible et permet à l'industriel d'évaluer la qualité de la « distribution » que nous obtenons en réalisant nos prédictions. Plus exactement, cette métrique nous permet d'évaluer la qualité de nos quantiles. Elle se présente comme suit :

$$Q_{p,t} = \begin{cases} 2(1-p)\left|y_t - f_{p,t}\right|, & si\ y_t < f_{p,t} \\ 2p\left|y_t - f_{p,t}\right|, & si\ y_t \geq f_{p,t} \end{cases}$$

$Avec :$

$f_{p,t} : valeur\ du\ quantile\ prédite\ avec\ une\ probabilité\ p\ au\ temps\ t$

$y_t : observation\ au\ temps\ t$

$p : P(y_t < f_{p,t})$

Un score quantile plus petit correspond à une meilleure performance. Pour les quantiles supérieurs à 50, une observation supérieure à la valeur quantile est davantage pénalisée. Pour les quantiles inférieurs, c'est l'inverse.

Pour une probabilité p et un temps t donnés, Hyndman et al. (2021f) expliquent que nous devrions nous attendre à ce que nos observations y_t soient plus petites que notre valeur quantile $f_{p,t}$ avec une probabilité p. Ainsi par exemple, pour $p = 0.95$, nous nous attendons à ce que nos observations soient plus petites que la valeur du 95ème centile 95% du temps. La probabilité que notre observation soit supérieure étant plus faible, si nous observons ce phénomène, alors notre score quantile sera plus élevé. Cela car il s'agit d'un résultat qui est peu probable, en supposant que notre valeur quantile soit représentative de la réalité. Inversement si nous souhaitons évaluer la valeur quantile du 40ème centile, alors nous serions davantage pénalisés si notre observation était inférieure à ladite valeur. Cela car il serait plus plausible que nous obtenions une observation supérieure, dans l'optique où 60% des observations seraient plus grande que notre valeur quantile.

Notez que la présence du 2 en facteur vise à faciliter l'interprétation du score quantile lorsque $p = 0.5$. Nous sommes en mesure, grâce à ce facteur, d'interpréter le résultat comme une erreur absolue (Hyndman et al., 2021d).

CHAPITRE 3 MÉTHODOLOGIE

Les étapes présentées dans ce chapitre sont réalisées pour une famille de produits test. Il est à noter qu'il est bien évidemment possible de les réaliser pour d'autres familles de produits, ce qui ne sera cependant pas couvert dans ce mémoire.

Dans le but d'atteindre notre objectif, nous allons suivre la démarche représentée à la figure suivante :

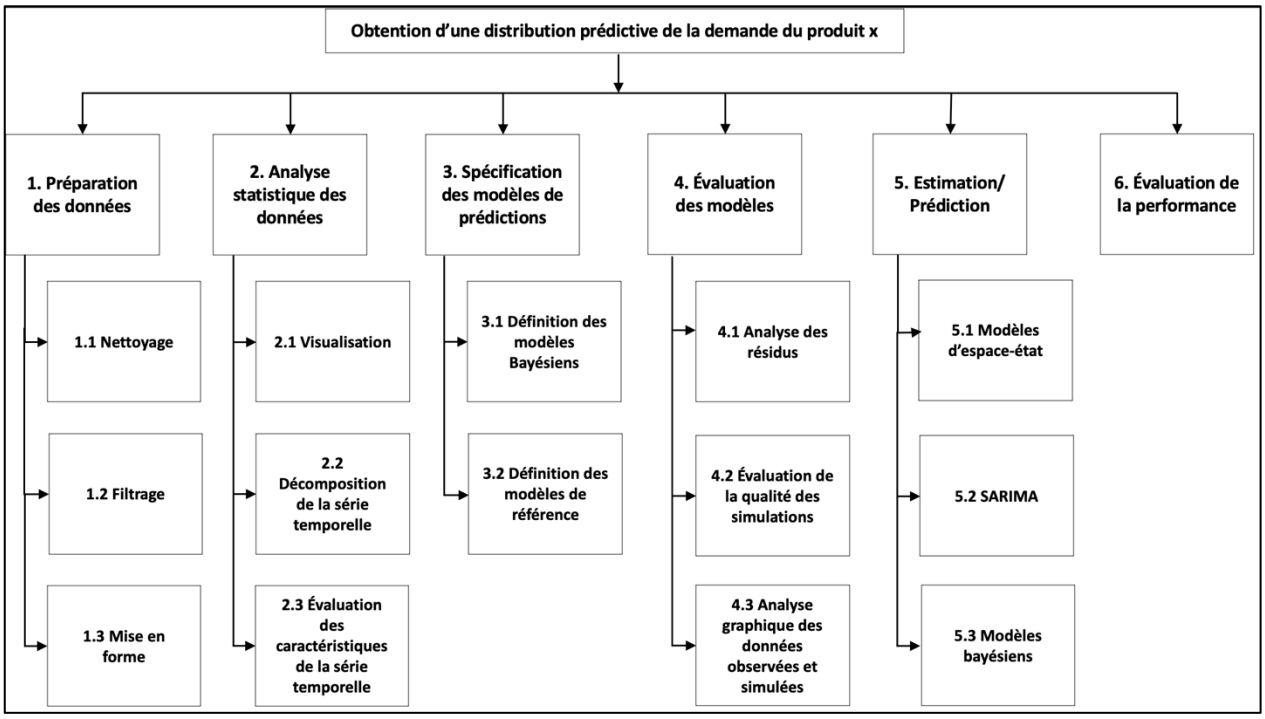


Figure 3.1 : Méthodologie

La stratégie présentée ci-dessous résulte d'un processus itératif qui ne sera pas formellement présenté. Seul l'aboutissement de ce cheminement fera l'objet d'une exposition détaillée. Les étapes principales sont présentées ci-dessous.

3.1 Préparation des données

Cette première étape vise à l'obtention de données adaptées au traitement qu'elles devront subir pour atteindre notre objectif. Il s'agit en effet de constituer l'intrant de nos modèles de prédiction, à savoir des séries temporelles, dans l'optique de prédire une demande future de manière efficace. Cet exercice comprend lui-même plusieurs processus. Avant de formater les données sous forme de séries temporelles, nous devrons les nettoyer et les organiser afin de nous assurer de n'avoir que des informations authentiques et complètes.

3.1.1 Nettoyage

Le nettoyage des données consiste en l'évaluation de la présence de doublons, d'informations incomplètes et la vérification de l'homogénéité du format pour l'ensemble des éléments de chaque colonne. Au cours de cette sous-étape, nous prêtons également attention à la présence d'éventuelles informations aberrantes (erreurs de saisie, par exemple).

3.1.2 Filtrage

La phase de filtrage consiste en la sélection des informations associées aux familles de produits critiques pour la période temporelle souhaitée en vue de la tenue de notre analyse. Nous créons ainsi un fichier qui ne regroupe que les données qui concernent une ou plusieurs années spécifiques pour les familles de produits les plus importantes pour l'entreprise. Notez que la criticité de ces derniers nous a été confidentiellement communiquée au préalable.

3.1.3 Mise en forme

La mise en forme des données vise à organiser les informations pour ne conserver que ce dont nous avons besoin pour notre étude, dans le format désiré. Ce processus est constitué de 3 sous-étapes qui s'entremêlent, qui sont le regroupement des données, la transformation des colonnes et le filtrage secondaire. À l'issue de ces trois sous-étapes, la mise en forme s'achève par la création de deux fichiers individuels pour chaque famille de produits, dont un comporte la demande hebdomadaire et l'autre la demande mensuelle. Le prochain paragraphe détaillera le processus qui sera illustré à l'aide de notre cas d'étude dans le chapitre suivant.

Nous commençons par regrouper chaque produit de la même famille par date. C'est-à-dire que pour chaque date et chaque famille, nous réalisons la sommation des demandes des produits spécifiques. Nous obtenons ainsi pour chaque famille autant de lignes qu'il y a de dates. La colonne des quantités commandées contient alors la demande agrégée de tous les produits spécifiques d'une même famille par semaine. Étant donné qu'il y a une date par semaine, nous devons, pour obtenir la demande mensuelle, effectuer un second regroupement. Nous remplaçons la colonne des dates par une colonne qui indique le numéro de la semaine et un autre qui indique le numéro du mois pour la période sélectionnée à l'étape précédente. Nous créons ensuite une colonne qui contient pour chaque famille la demande mensuelle (donc la somme des demandes de 4 semaines consécutives). Les données regroupées comme souhaité, nous éliminons les informations qui ne présentent pour nous aucun intérêt, pour ne conserver que les colonnes suivantes : {Nom de la famille ; Numéro de la semaine ; Numéro du mois ; Quantité commandée par semaine ; Quantité commandée par mois}. Pour terminer notre processus de mise en forme des données, nous regroupons les informations issues des colonnes précédentes dans deux fichiers distincts, et ce pour chaque famille de produits. Nous avons alors un fichier texte contenant les quantités commandées par semaine, et un autre contenant celle qui sont commandées par mois.

3.2 Analyse statistique des données

L'analyse statistique de nos données vise à nous aider à déterminer les modèles les plus à même de nous fournir de bonnes prédictions. De fait, en évaluant les caractéristiques de notre série temporelle, nous serons en mesure de constituer notre information *a priori* et de choisir une distribution de probabilité adaptée à cette dernière. Cela nous aidera dans la construction de nos modèles bayésiens.

Ce processus se divise en plusieurs étapes qui ne seront réalisées que sur les données historiques des années dont nous nous servirons pour prédire la demande future. Autrement dit, nous n'analyserons pas l'année que nous souhaitons prédire afin de ne pas risquer de minimiser l'incertitude réelle de la demande. Cette étape sert à construire les connaissances *a priori* nécessaires pour construire le modèle probabiliste.

3.2.1 Visualisation

Nous débutons par une visualisation des données historiques considérées afin d'apercevoir les caractéristiques de la série temporelle, notamment sa saisonnalité. Nous produisons pour ce faire un graphique de la demande dans le temps (dates en abscisse, quantités commandées en ordonnée). Nous évaluons également l'évolution du schéma saisonnier présent dans notre série (éventuel changement drastique et/ou progressif dans le comportement saisonnier). Nous générons ainsi un graphique sur lequel nous pouvons observer le schéma saisonnier de chaque année de façon individuelle, et un autre sur lequel les signaux sont superposés. Nous évaluons également l'autocorrélation présente dans les données à l'aide d'un graphique de décalage et de corrélogrammes. Finalement, nous observons également le comportement saisonnier de la demande agrégée par mois et par quadrimestre. Pour obtenir les différents graphiques, nous nous servons des codes qui sont présentés par Hyndman et al. (2021f), que nous appliquons sur R. Pour plus de renseignements, le lecteur est invité à consulter la référence.

3.2.2 Décomposition

La décomposition des données se réalise de façon automatique sur Python comme sur R. Tel qu'expliqué au chapitre précédent, nous faisons usage des méthodes de décomposition STL et moyennes mobiles, qui s'obtiennent sur Python respectivement à l'aide des fonctions suivantes :

```
STL(data, period = 52)
```

```
seasonal_decompose(data, model = 'multiplicative', period = 52,
extrapolate_trend = 'freq')
```

L'argument « period » sert à définir la longueur du schéma saisonnier. Le 52 se rapporte ici à la taille de la période pour la demande hebdomadaire. L'argument « model » prend la valeur 'additive' pour une décomposition additive.

Nous réalisons également une décomposition STL sur R, pour pouvoir faire usage de fonctions qui nous permettent d'analyser certaines caractéristiques de notre série temporelle. Bien qu'il existe probablement une équivalence de ces fonctions sur Python, nous avons décidé d'appliquer ce que

nous présentent Hyndman et al. (2021f). Pour plus de précisions, le lecteur est invité à consulter la référence. La théorie sur laquelle repose ces fonctions est présentée au chapitre 2 ainsi que dans les références qui s'y rapportent.

3.2.3 Évaluation des caractéristiques de la série temporelle

Nous évaluons ici quelques caractéristiques de base d'une série : son minimum, son premier quartile, sa médiane, son troisième quartile ainsi que son maximum. Nous pouvons ainsi avoir une meilleure idée du comportement des données. Notez que pour des raisons de confidentialité, bien que nous procédions à l'analyse de ces informations, nous ne les présentons pas. La fonction utilisée est disponible via la référence (Hyndman et al., 2021f). Nous estimons également la facilité de prédiction de la série à l'aide de l'entropie spectrale de Shannon (Hyndman et al., 2021f). Une entropie proche de 0 indique que la série est facile à prédire (et donc qu'elle a probablement une forte tendance et/ou saisonnalité). Nous évaluons finalement la force de la tendance et de la saison sur la série décomposée. Nous utilisons ainsi des fonctions que Hyndman et al. (2021f) regroupent sous l'appellation « STL features ». L'ensemble de ces analyses sont réalisées sur R. Ces caractéristiques connues, nous pourrons sélectionner les modèles qui nous semblent les plus adaptés pour obtenir une précision prédictive satisfaisante.

3.3 Spécification des modèles de prédiction

Au cours de cette section, nous définissons les modèles et méthodes que nous jugeons adéquats pour obtenir des prévisions considérant les caractéristiques de la série temporelle à l'étude. Nous détaillons leur structure ainsi que la raison pour laquelle nous les avons sélectionnés. Nous expliquons en 3.3.1 comment nous obtenons des prédictions à l'aide de nos modèles bayésiens, ce qui va au-delà de la définition de ces derniers. Nous le faisons pour faciliter la compréhension du lecteur. La définition des modèles se restreint en réalité à ce que nous présentons dans le tableau présenté à l'annexe B et au calcul des paramètres à partir des données historiques. C'est-à-dire que la définition de nos modèles se restreint au choix de leurs composantes et au paramétrage des *a priori*.

3.3.1 Définition des modèles Bayésiens

Pour choisir nos méthodes, nous prenons en compte à la fois les caractéristiques de nos données et les exigences de l'entreprise. Pour répondre à la demande de l'organisation, nous mettons en place des modèles bayésiens. Leurs composantes sont exposées à l'annexe B. Tel que nous le présenterons ultérieurement, certaines expériences ont été infructueuses et nous ont conduit à adapter l'approche bayésienne que nous souhaitions mettre en place. Nos données étant dépendantes du temps, la propriété d'invariance n'est pas applicable. Ces dernières suivent en effet une séquence précise, qui est dictée par le temps.

Comme nous aurons l'occasion de le détailler au chapitre suivant, la série temporelle à l'étude est marquée par une importante saisonnalité. La demande est ainsi en grande partie expliquée par la saison. Nous avons décidé de nous concentrer sur l'incertitude qui encadre la composante résiduelle de notre série. En ôtant la tendance et la saison, nous avons tenu pour acquis que nous respections dorénavant la propriété d'invariance des données. Nous sommes néanmoins conscients que notre approche néglige l'incertitude qui est propre à la saison et à la tendance. La qualité de nos prédictions est de cette façon compromise. Aussi, tel que cela sera mis en évidence dans les graphiques dans le chapitre des résultats, les résidus ne semblent pas parfaitement indépendants, et ce peu importe la méthode de décomposition. Il sera nécessaire de le considérer dans les études subséquentes.

Les modèles bayésiens adaptés se construisent ainsi selon le processus général suivant :

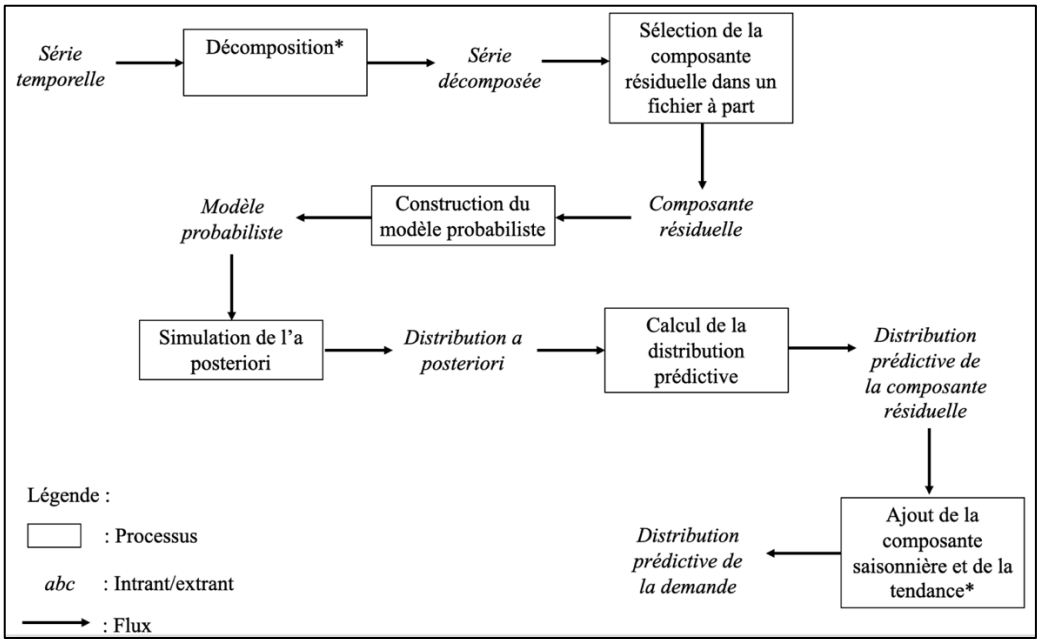


Figure 3.2 : Processus général nous conduisant à l'obtention d'une distribution prédictive

Ce processus est le même d'une expérience à l'autre. Cependant, selon l'essai, la méthode de décomposition ainsi que la manière dont sont ajoutées la tendance et la saison changent. C'est-à-dire que les cases dans lesquelles se trouvent un astérisque diffèrent d'une expérience à l'autre. Tel que présenté dans le tableau B.1 de l'annexe B, la décomposition (**étape 2.2**) peut être classique-additive, classique-multiplicative, STL-additive ou STL-multiplicative. Cela change directement la façon dont nous incorporons la saison et la tendance (par addition ou multiplication). De plus, nous testons deux manières d'ajouter la tendance (ce qui n'est pas présenté dans le tableau 4.8 par manque d'espace). D'une part, nous expérimentons l'inclusion de la tendance à l'aide d'une moyenne historique, et d'autre, nous tentons plutôt de l'extrapoler. Illustrons la chose :

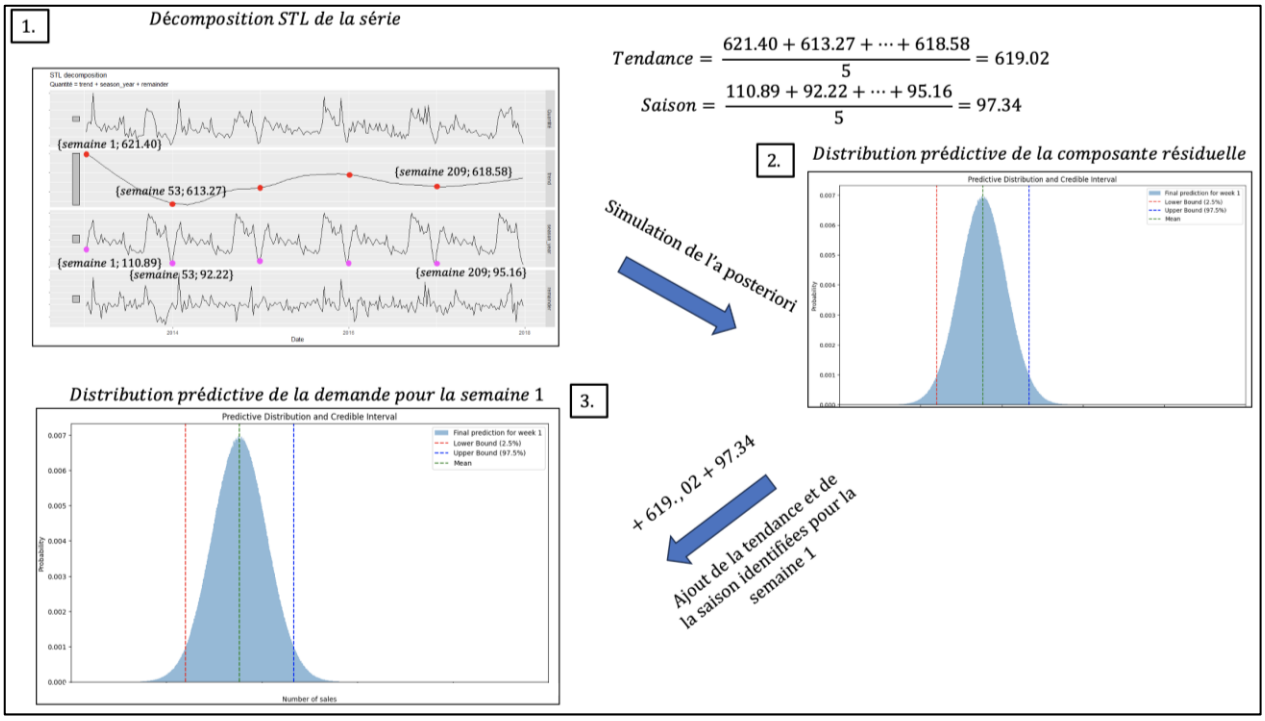


Figure 3.3 : Inclusion de la tendance à l'aide d'une moyenne pour une décomposition STL

La figure ci-dessus représente une expérimentation avec décomposition STL et inclusion de la tendance à l'aide d'une moyenne des tendances passées pour une période donnée sur Python. En supposant que nous souhaitons réaliser une prédiction pour la demande de la première semaine de l'année à venir, nous procédons ainsi : la série décomposée, nous réalisons une moyenne des composantes saisonnières des semaines 1 des cinq dernières années (points roses sur le graphique en haut à gauche de la figure). Sur le graphique, le résultat de cette moyenne est de 97.34. En parallèle, nous réalisons la moyenne des tendances de la semaine 1 des cinq années précédentes (points rouges). Sur le graphique, le résultat de cette moyenne est de 619.02. Nous obtenons ainsi deux coefficients différents. Pour chaque point obtenu par simulation (pour le calcul de la distribution prédictive de la composante résiduelle identifiée en haut à droite de l'image, au point 2) nous ajoutons les deux composantes (97.34 et 619.02). Nous obtenons ainsi ce que nous identifions sur notre précédent schéma comme la distribution prédictive de la demande, dont un exemple est présenté en bas à gauche de la figure ci-dessus (point 3).

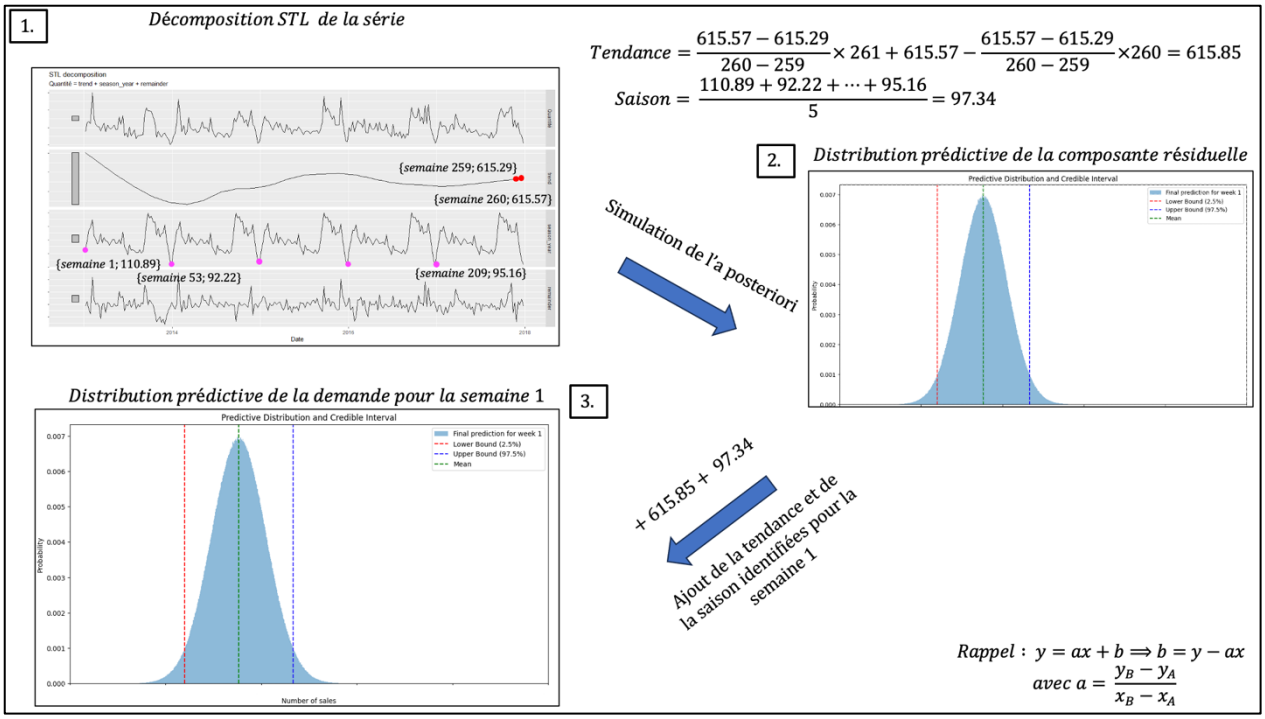


Figure 3.4 : Inclusion de la tendance par extrapolation

La figure ci-dessus représente une expérimentation avec décomposition STL et inclusion de la tendance par extrapolation. En supposant que nous souhaitons réaliser une prédiction pour la demande de la première semaine de l'année à venir, nous procédons cette fois de la façon suivante : la série décomposée, nous réalisons une moyenne des composantes saisonnières des semaines 1 des cinq dernières années (points roses sur le graphique en haut à gauche de la figure). En parallèle, nous extrapolons la tendance à l'aide d'une équation de droite que nous obtenons à partir des deux derniers points rouges. Nous appliquons ensuite les deux coefficients obtenus de la même façon que précédemment.

Lorsque nous procédons à une décomposition multiplicative, notre approche est la même, mais nous multiplions les points simulés par les coefficients.

Les décompositions sont programmées sur Python. Notez que pour la simulation, nous avons utilisé 3 chaînes de Markov, avec une phase de chauffe de taille 10 000 et un échantillonnage de taille 20 000.

3.3.2 Définition des modèles de référence

En parallèle, nous avons décidé de faire usage de modèles statistiques plus communément employés, afin de comparer notre solution à ce qui se fait de façon plus répandue. En tenant compte de l'importante saisonnalité de la série, nous avons décidé de faire usage des méthodes Holt-Winters (plus particulièrement des modèles d'espaces-états qui s'y rapportent) et SARIMA. Plus spécifiquement, nous considérerons les modèles de lissage exponentiel suivants : {ETS(A,A,A), ETS(A,A,M), ETS(M,A,A), ETS(M,A,M), ETS(A,N,A), ETS(A,N,M), ETS(M,N,A), ETS(M,N,M)}. Pour toute question relative à la notation employée, le lecteur est invité à se référer à (Hyndman et al., 2021f). La théorie sur laquelle repose les différents modèles est exposée au chapitre précédent. Nous nous servirons des codes qui sont présentés par Hyndman et al. (2021f), que nous mettons également en annexe. Ces modèles seront donc programmés sur R.

3.4 Évaluation des modèles

Selon le modèle à évaluer (bayésien ou autre), nous n'adoptons pas la même démarche, tel que cela a été mentionné dans la revue de littérature.

Pour les modèles de Holt-Winters et SARIMA, nous effectuons une analyse des résidus (**Étape 4.1**). En premier lieu, nous programmons nos différents modèles sur R, comme nous le présentent Hyndman et al. (2021f). Par la suite, pour chacun de nos modèles, nous calculons les résidus à l'aide de la fonction « augment() », tel que mentionné par Hyndman et al. (2021f). Les résidus obtenus, nous générons les graphiques qui nous permettent de poser des diagnostiques en termes de normalité, d'homoscédasticité et d'autocorrélation. Pour ce faire, nous faisons usage de la fonction « gg_tsresiduals() », tel que suggéré par Hyndman et. (2021f). Cette fonction nous permet de générer simultanément un graphique d'autocorrélation, une figure des résidus en fonction du temps ainsi qu'un histogramme. Nous appliquons également un test de Ljung-Box à l'aide de la fonction « features() ».

Pour les modèles bayésiens, nous effectuons en premier lieu une analyse de la qualité de la simulation (**étape 4.2**). Pour ce faire, nous visualisons les chaînes de Markov ainsi que les statistiques qui nous permettent d'en évaluer la performance (tel que présenté au chapitre 2). Nous utilisons également la méthode PSIS. Par la suite, nous menons une analyse graphique (**Étape 4.3**). Ainsi, lorsque nous obtenons notre première distribution prédictive, nous la comparons à un

histogramme des quantités commandées des années 2016 et 2017. Pour appliquer la méthode PSIS (**Étape 4.2**), nous faisons usage du code suivant :

```
az.loo(trace)
```

« trace » étant nos échantillons simulés.

Nous évaluons alors ce que la fonction nous renvoie après exécution.

3.5 Estimation/prédiction

La phase de prédiction consiste en l'exécution des différents modèles sélectionnés en vue d'obtenir des prédictions. Pour ce faire, nous utiliserons les données historiques analysées précédemment comme intrant pour obtenir une estimation de la demande hebdomadaire et mensuelle. Les données résiduelles des années 2013 à 2015 son utilisées pour paramétrées les *a priori*. Les données de 2016 et 2017 sont celles que nous utilisons comme observations. Cette phase se divise en deux répétitions, pour chaque méthode (modèles d'espace-état, SARIMA, modèles bayésiens). La première consiste en l'exécution des modèles en utilisant des données historiques hebdomadaires pour obtenir une estimation des quantités commandées pour les 12 premières semaines de l'année à prédire. La seconde ne diffère que par l'usage de données historiques mensuelles, qui permettra une estimation des 12 mois de l'année. Nous présenterons les deux exécutions de la méthode Holt-Winters, suivi de celles de SARIMA, puis enfin celles des modèles bayésiens.

L'application des modèles Normale-Normale-Inverse Gamma et Normale-Exponentielle-Inverse Gamma (modèles bayésiens sélectionnés) s'accompagne d'un processus pré et post-simulation, tel qu'illustré par les précédentes figures. Avant la simulation, nous décomposons les données pour séparer la composante résiduelle de la tendance et de la saison. La simulation réalisée, nous réincorporons la tendance et la saison. Pour réaliser notre simulation, nous avons recours aux Chaînes de Markov. Nous faisons plus particulièrement usage de l'algorithme de Métropolis. Pour effectuer notre simulation par chaînes de Markov, nous avons utilisé la librairie pymc3. Plus précisément, nous avons eu recours au code que nous offre l'ouvrage de (Davidson-Pilon, 2016). Notez que l'ouvrage ne faisait pas usage de la dernière version de pymc, ce pourquoi nous nous sommes inspirés du code mis à jour par certains contributeurs sur Github. Pour en savoir plus, le lecteur est invité à consulter les références.

L'application des autres modèles consiste en la simple exécution des codes (présentés en annexe) que nous rédigeons dans la phase de spécification. Les étapes 5.1 et 5.2 sont identiques mais s'appliquent l'une sur les données hebdomadaires (**Étape 5.1**) et l'autre sur les données mensuelles (**Étape 5.2**).

3.6 Évaluation de la performance

Nous évaluons ici les résultats des prédictions réalisées à l'aide des différents modèles. Pour chaque période prédite, nous calculons l'erreur absolue, le pourcentage d'erreur, ainsi qu'un score quantile. Pour obtenir le MAE, le MAPE, et le score quantile moyen, nous réalisons respectivement la moyenne des erreurs absolues, des pourcentages d'erreurs et des scores enregistrées pour les 12 périodes prédites. Nous classons par la suite dans un tableau les méthodes en ordre décroissant de performance. Nous adoptons cette démarche tant pour la demande mensuelle qu'hebdomadaire, séparément.

CHAPITRE 4 RÉSULTATS

Nous détaillons au cours de ce chapitre la mise en application des différentes étapes de la méthodologie que nous avons exposée au chapitre précédent.

4.1 Préparation des données

Les données dont nous disposons sont structurées comme illustré dans le tableau 4.1.

Tableau 4.1 : Structure originale des données de l'entreprise

Numéro de contrat	Famille de produits	Produit spécifique	Date	Quantité
12	A	AA-776-00	2013-01-07	46
12	A	AB-876-01	2013-01-07	9
12	A	AC-888-54	2013-01-07	75
12	A	AD-971-22	2013-01-14	71
12	A	AE-190-33	2013-01-14	54
…	…	…	…	…
12	B	BA-444-21	2013-01-28	66
12	B	BB-234-88	2013-01-28	18
12	B	BC-103-22	2013-02-04	25
12	B	BD-404-53	2013-02-04	32
…	…	…	…	…

Nous avons ainsi une identification du contrat, de la famille de produits, du produit spécifique, la date de commande ainsi que la quantité demandée. Nous avons ces informations pour 1973 familles de produits, qui contiennent chacune plusieurs produits spécifiques. De plus, nous disposons de l’historique des commandes de Janvier 2013 à Mars 2023.

4.1.1 Nettoyage

Avant de mettre nos données dans le format requis pour réaliser les prédictions souhaitées, nous devons évaluer la qualité des informations dont nous disposons. Nous avons évalué la présence de données manquantes, de doublons et nous avons vérifié l'uniformité du format des données. Ayant le privilège de travailler auprès d'une entreprise dont la culture des données est bien établie, nous n'avons eu aucune correction à apporter aux informations. Nous avons cependant observé la présence de quantités commandées négatives. Ces dernières correspondent aux retours effectués par les clients. Nous avons pris la décision de les exclure de notre analyse, car notre mandat consiste en la prédiction de futures commandes. Bien que les articles retournés puissent être considérés comme des commandes nulles et non avenues, nous avons estimé que ces derniers devraient faire l'objet de prévisions distinctes. L'estimation des retours ne sera donc pas abordée dans le présent ouvrage et les informations associées seront soustraites des données à analyser.

4.1.2 Filtrage

L'entreprise nous a fourni dans un fichier contenant des scores de criticité pour chaque famille de produits. Nous avons sélectionné les 27 familles les plus importantes pour la suite de l'étude. Concrètement, cela signifie que nous avons éliminé toutes les lignes du tableau précédent qui correspondent aux familles les moins critiques. Ainsi, si par exemple B n'était pas critique, notre tableau ressemblerait alors à ceci :

Tableau 4.2 : Tableau des données filtrées

Numéro de contrat	Famille de produits	Produit spécifique	Date	Quantité
12	A	AA-776-00	2013-01-07	46
12	A	AB-876-01	2013-01-07	9
12	A	AC-888-54	2013-01-07	75
12	A	AD-971-22	2013-01-14	71
12	A	AE-190-33	2013-01-14	54
…	…	…	…	…

D’autre part, nous avons sélectionné pour les familles restantes les données de 2013 à 2017 et 2018 séparément (des explications sont fournies dans une section ultérieure). Autrement dit, nous avons supprimé toutes les entrées associées aux années suivantes (2019, 2020… 2023), et nous avons séparé les données des années 2013 à 2017 de celles de 2018.

4.1.3 Mise en forme

Ce processus est réalisé pour le tableau des informations de 2013 à 2017 et pour celui de 2018, toujours séparément. L’entreprise désire réaliser des prédictions par famille, cela dans le but de prévoir la quantité totale de tissus nécessaire pour répondre aux besoins de ses clients. Nous devrons donc agréger la demande des produits spécifiques d’une même famille. Nous négligerons la présentation détaillée des étapes intermédiaires expliquées au chapitre précédent. Nous exposerons cependant les extrants de certains sous-processus.

Pour chaque famille (A, B…) nous créons une colonne des quantités totales commandées regroupant la demande de tous les produits spécifiques faisant partie du même groupe (A, B…). Cela nous conduit à obtenir un résultat structuré tel que présenté dans le tableau ci-dessous, pour une demande hebdomadaire. Pour la demande mensuelle, le résultat est semblable. Il y a simplement moins de lignes car nous agrégeons les données au mois.

Tableau 4.3 : Quantités commandées par famille de produits

Famille de produits	Date	Quantité
A	2013-01-07	130
A	2013-01-14	125
…	…	…
B	2013-01-28	84
B	2013-02-04	57
…	…	…

Dans la mesure où nous devons prédire par famille de produits, il est nécessaire que nous séparions les données des différentes catégories afin de les traiter séparément. Ainsi, nous ne conservons

pour notre étude que les séries temporelles (quantités commandées en fonction du temps) de la forme suivante :

Tableau 4.4 : Série temporelle de la famille A après formatage

A	
Date	Quantité
2013-01-07	130
2013-01-14	125
…	…

Nous enregistrons dans un fichier texte les informations sus-présentées pour chaque famille de produit, pour la période 2013-2017 et 2018.

Pour les étapes suivantes, nous ne conserverons que l'une des 27 familles les plus critiques, qui sera notre série test. Nous ne considérerons pas les autres, qui devront faire l'objet d'études prochaines. La mise en forme achevée, nous procédons à une analyse statistique de notre série temporelle test. Cela nous permettra de la caractériser pour en retirer autant d'informations utiles que possible.

4.2 Analyse statistique des données

Ce processus se divise en plusieurs étapes qui seront uniquement réalisées sur les années 2013, 2014, 2015, 2016 et 2017. Dans la mesure où nous avons l'intention d'utiliser la statistique bayésienne, nous devons nous abstenir d'analyser les données de l'année 2018 qui sera notre période de prédiction. Prendre connaissance du comportement de cette année risquerait de nous conduire à minimiser l'incertitude de la demande en ayant une connaissance *a priori* que nous n'aurions pas si nous prédisions le futur dans le présent. D'autre part, nous avons estimé que les années ultérieures ne seraient pas représentatives de ce à quoi nous pouvons nous attendre à l'avenir en raison de l'événement exceptionnel que représente la pandémie de covid-19. C'est la raison pour laquelle les données de la période 2019-2023 ont été éliminées.

Pour débuter notre analyse, tel que présenté dans la méthodologie, nous produirons des graphiques qui nous offrirons un premier aperçu des caractéristiques de notre série temporelle. La demande analysée est ici hebdomadaire.

4.2.1 Visualisation

Commençons par visualiser la série temporelle entre les années 2013 et 2017.

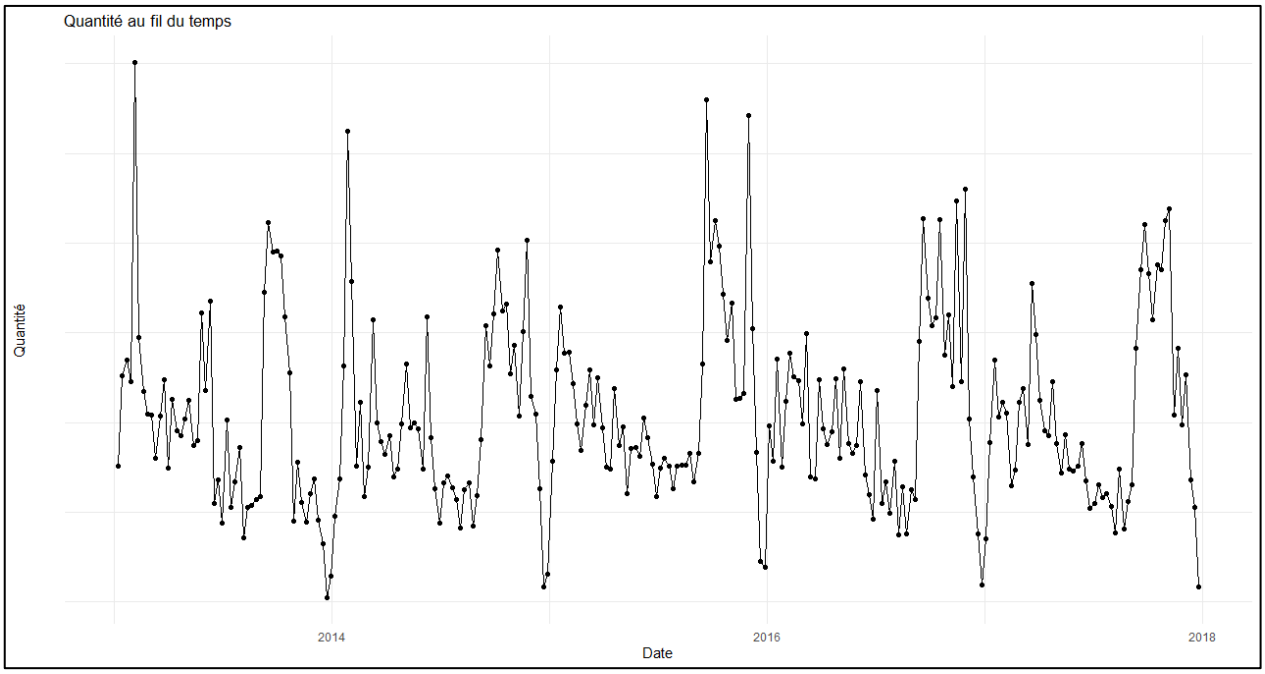


Figure 4.1 : Série temporelle de la famille test

En observant ce premier graphique, il nous semble raisonnable d'affirmer que la demande dispose d'une composante saisonnière. Il semblerait en effet qu'il y ait un pic en début d'année, suivi d'une brusque diminution qui tend par la suite à osciller autour de 500 unités jusqu'à la fin de l'automne où nous percevons un nouveau pic. Le schéma saisonnier tend à se faire plus petit au cours des deux dernières années à l'étude. Cependant, ce simple aperçu ne suffit pas à caractériser notre série. Pour approfondir notre étude, seront présentés des graphiques saisonniers ainsi qu'une analyse visuelle de l'autocorrélation présente dans les données.

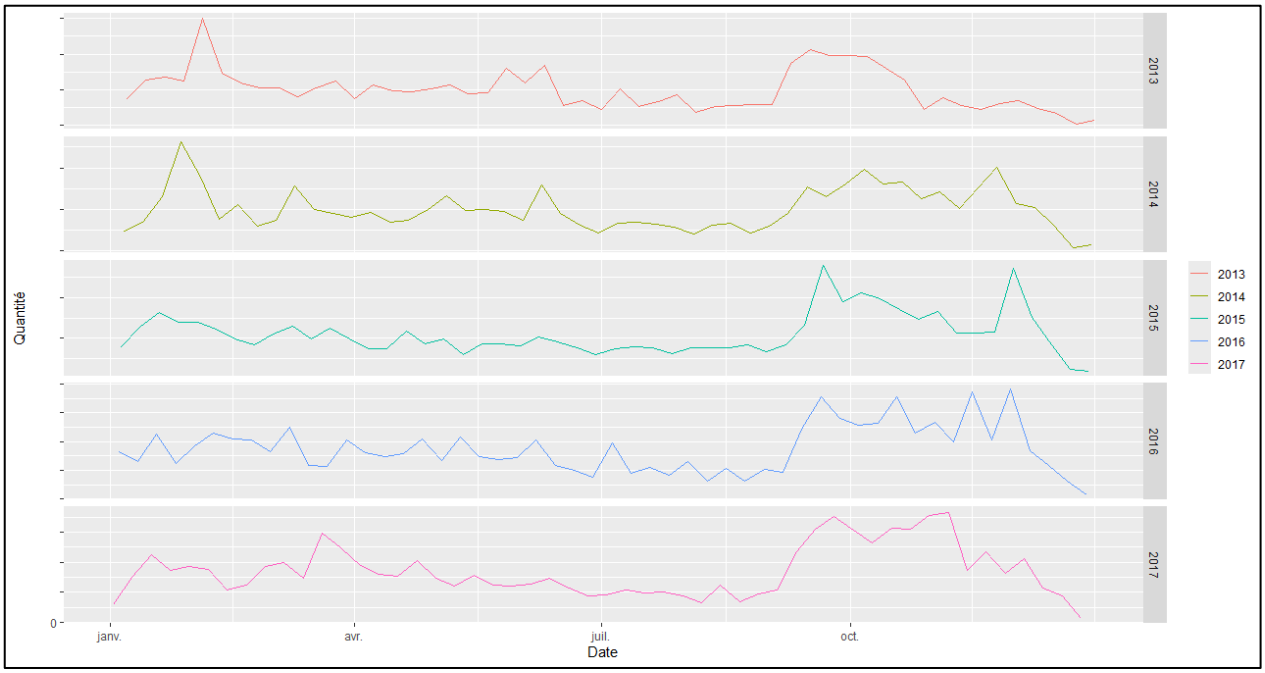


Figure 4.2 : Graphique saisonnier

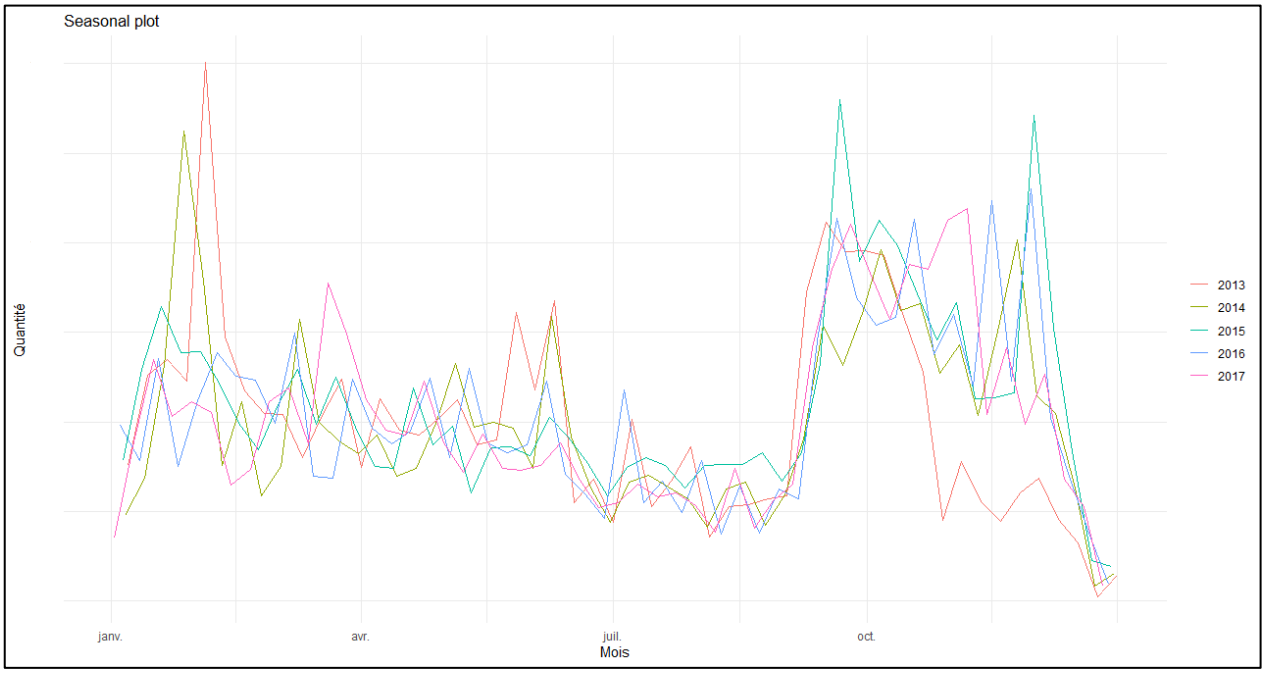


Figure 4.3 : Graphique saisonnier superposé

Les deux graphiques exposés ci-dessus mettent davantage en évidence la présence d'une saisonnalité. Bien que l'amplitude des signaux se montre variable d'une année à l'autre, nous

remarquons le même comportement : un pic, une diminution graduelle puis une brusque augmentation en automne.

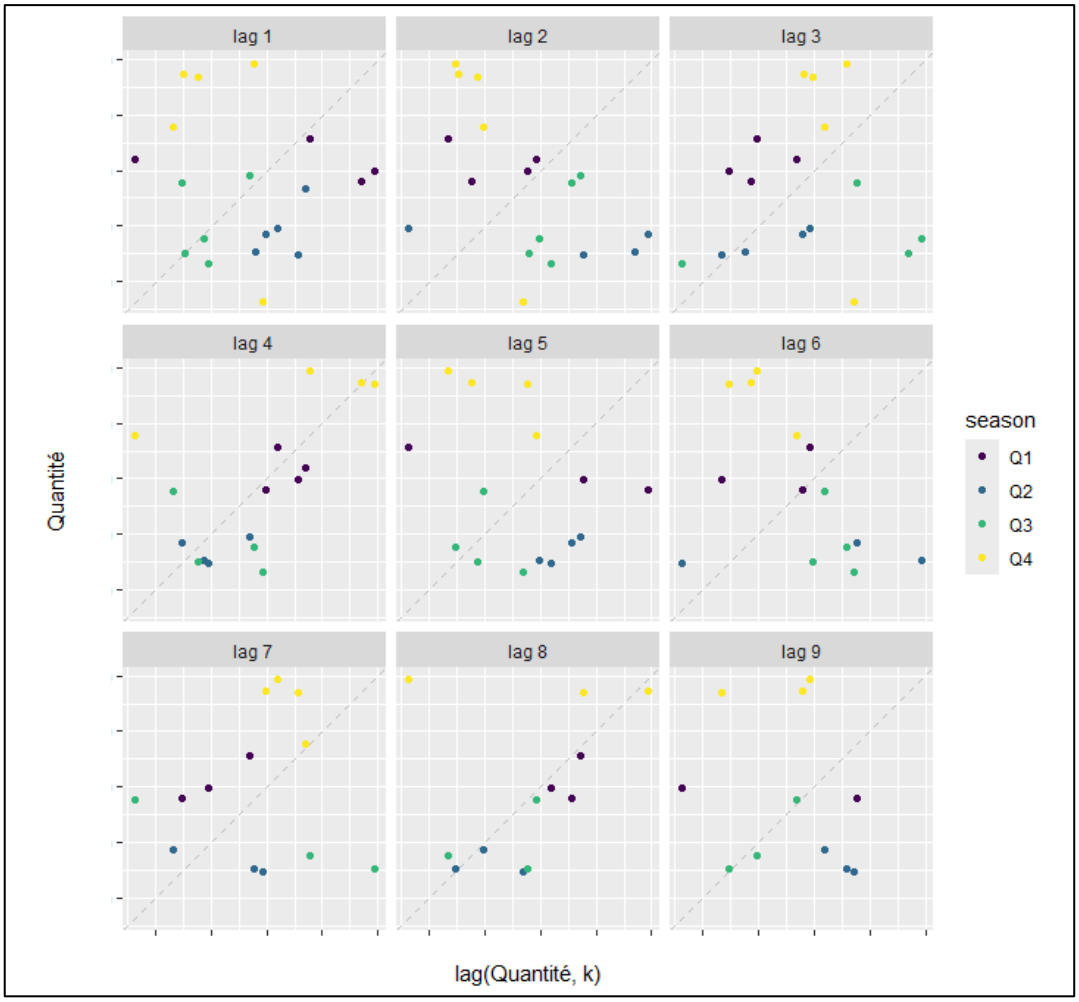


Figure 4.4 : Graphique de décalage

La figure 4.4 expose la relation entre le premier trimestre et les huit précédents (donc Q_t versus Q_{t-1} , Q_t versus Q_{t-2} et ainsi de suite). Bien qu'il semble y avoir une relation positive entre Q_t et Q_{t-4} de même qu'entre Q_t ainsi que Q_{t-8}, nous estimons que cela n'est pas suffisamment flagrant pour déclarer qu'il y a une forte saisonnalité à une échelle trimestrielle. Les visuels ci-dessous nous apporte quelques éclaircissements.

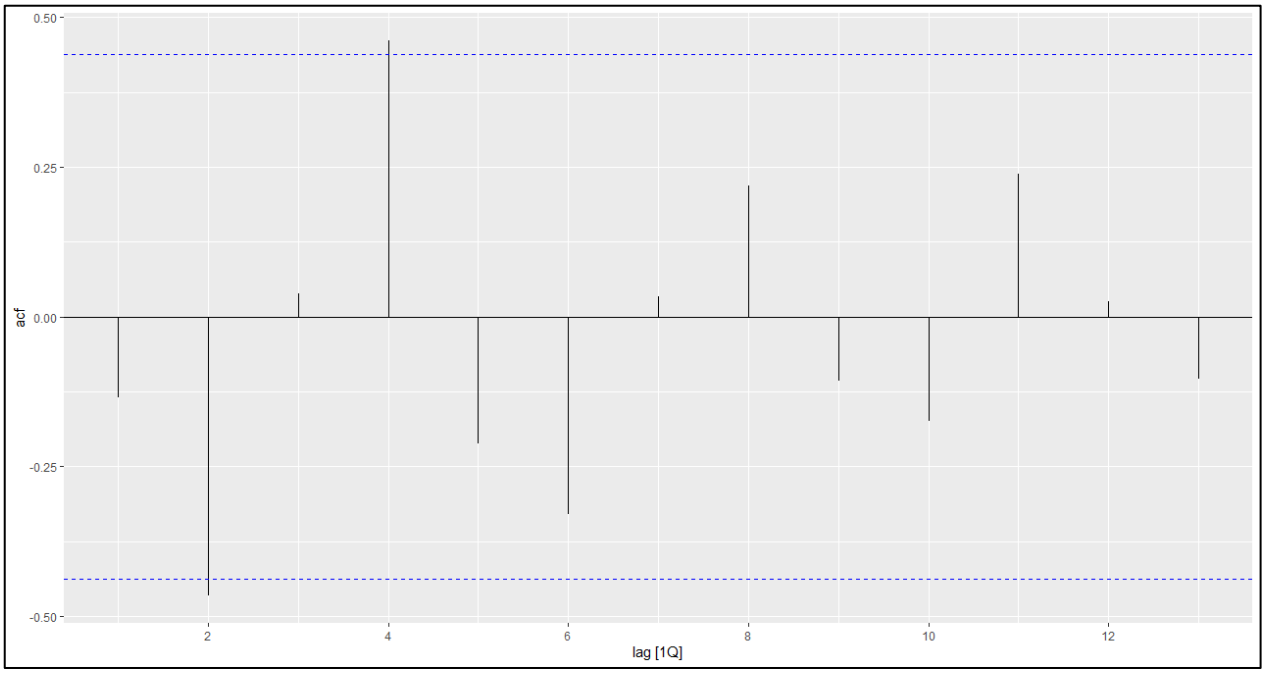


Figure 4.5 Corrélogramme des trimestres de la famille test entre 2013 et 2017

Le corrélogramme de la figure 4.5 laisse entrevoir qu'il y a une corrélation négative significative entre Q_t et Q_{t-2} ainsi qu'une corrélation positive significative entre Q_t et Q_{t-4}. Ces derniers résultats nous confirment la saisonnalité de notre série temporelle. En effet, il n'est pas surprenant pour une série présentant un schéma saisonnier d'exposer une forte corrélation entre un trimestre (1, 2, 3 ou 4) de l'année t et le même l'année d'avant (respectivement 1, 2, 3 ou 4). Nous nous attendons à ce que le schéma qui s'est produit l'année t-1 se reproduisent l'année suivante. De surcroît, il est attendu que le trimestre le plus éloigné (Q_{t-2}) soit négativement corrélé avec la période de référence (Q_t), car ils ne présentent pas le même schéma (un pic pour l'un et un creux pour l'autre). D'autre part, nous constatons que les corrélations tendent à diminuer à mesure que nous reculons dans le temps. Cela nous conduit à songer que le schéma varie au fil des années. Nous avions déjà remarqué en observant notre premier graphique que le comportement de la série temporelle semble évoluer, tout particulièrement les deux dernières années. Le corrélogramme à l'échelle mensuelle fourni ci-dessous (figure 4.6) nous emmène au même constat.

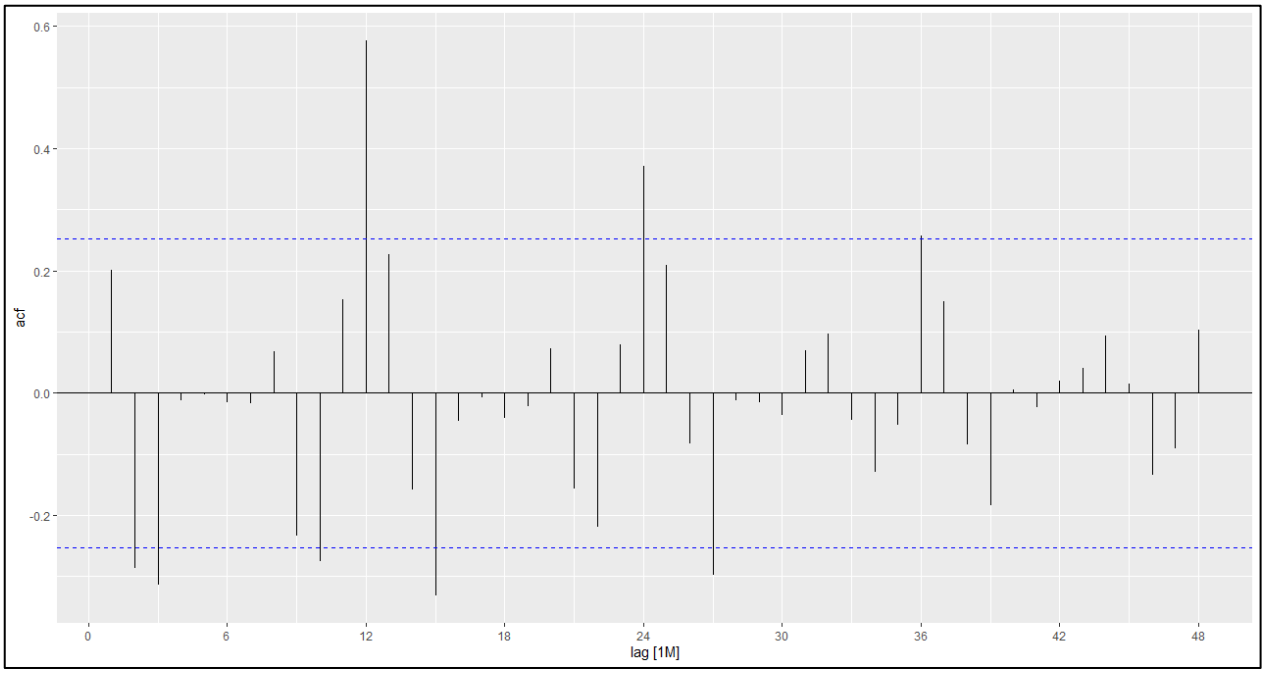


Figure 4.6 Corrélogramme des mois de la famille test entre 2013 et 2017

Afin de mieux visualiser l'évolution de notre schéma saisonnier, observons la variation trimestrielle et mensuelle de notre série temporelle au cours du temps.

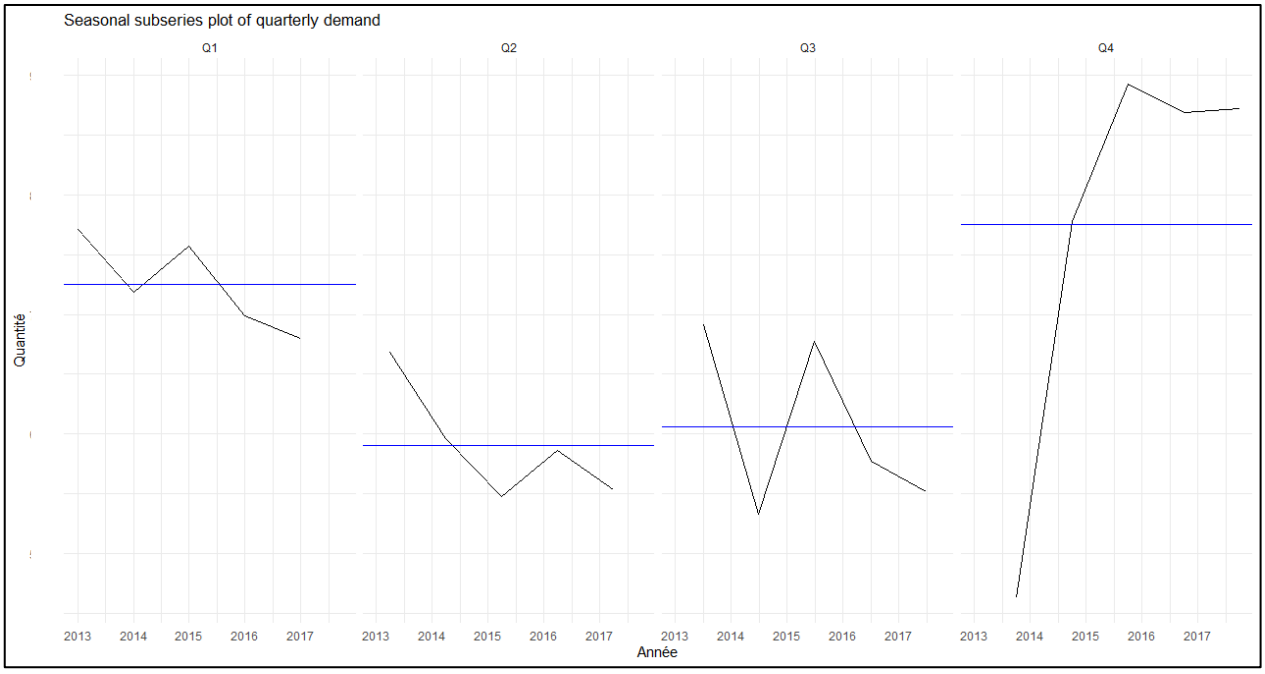


Figure 4.7 Graphique saisonnier par trimestre

Nous notons qu'en plus de varier au cours du temps, les quantités commandées par trimestre n'évoluent pas de la même manière. De fait, les trimestres 1 et 3 semblent alterner entre un pic et un creux. Bien que nous n'ayons pas un historique suffisant pour en avoir la certitude, les quantités commandées paraissent être sur une pente descendante. La diminution progressive de la demande au deuxième trimestre et l'augmentation drastique au quatrième sont plus évidentes et les oscillations plus discrètes.

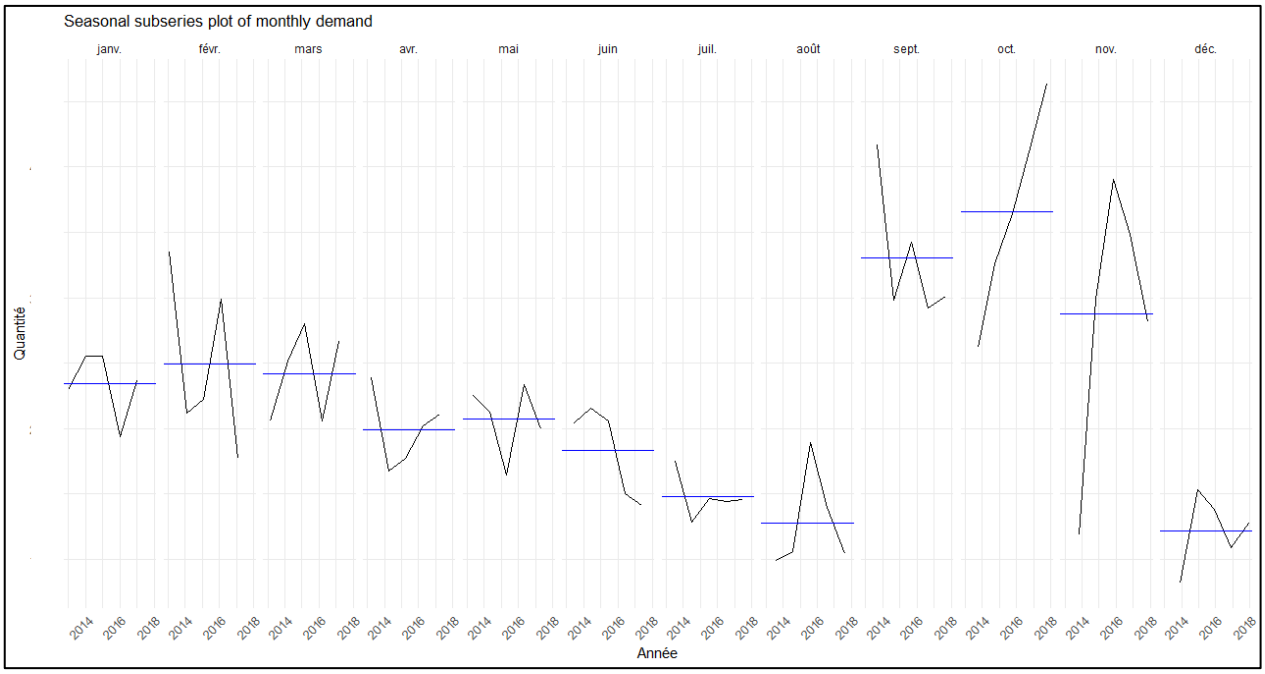


Figure 4.8 Graphique saisonnier par mois

Sans surprise, nous remarquons que pour un même mois, la demande fluctue au fil des ans. Nous observons un écart maximal d'environ 3700 unités entre 2013 et 2014 au mois de Novembre (ce qui représente une augmentation de plus de 100% entre ces deux années). Cela pourrait peut-être partiellement s'expliquer par la diminution des quantités commandées entre Mai et Août.

En étudiant l'intégralité de ces visuels, nous sommes tentés de croire en une évolution de la gestion des approvisionnements du client de l'entreprise dont nous ignorons le domaine d'activité. En effet, tel qu'observé dans les graphiques 4.2, 4.3, 4.7 et 4.8, il semblerait que les commandes importantes se soient déplacées du début d'année à la fin de l'année précédente, plus particulièrement en Novembre. La demande au cours du reste de l'année semble moins varier tandis que les commandes augmentent en fin de période. Bien qu'il serait possible de conduire une analyse plus rigoureuse en

discutant notamment avec le client de l'organisation, nous estimons que nos analyses graphiques sont suffisantes pour poursuivre notre mandat.

4.2.2 Décomposition

Nous réalisons ici 4 décompositions différentes sur python. Nous aurons plus tard l'occasion d'évaluer l'impact de cette étape sur la qualité des résultats que nous obtenons à l'aide des modèles bayésiens adaptés.

En réalisant une décomposition classique (fondée sur les moyennes mobiles), nous obtenons les résultats suivants :

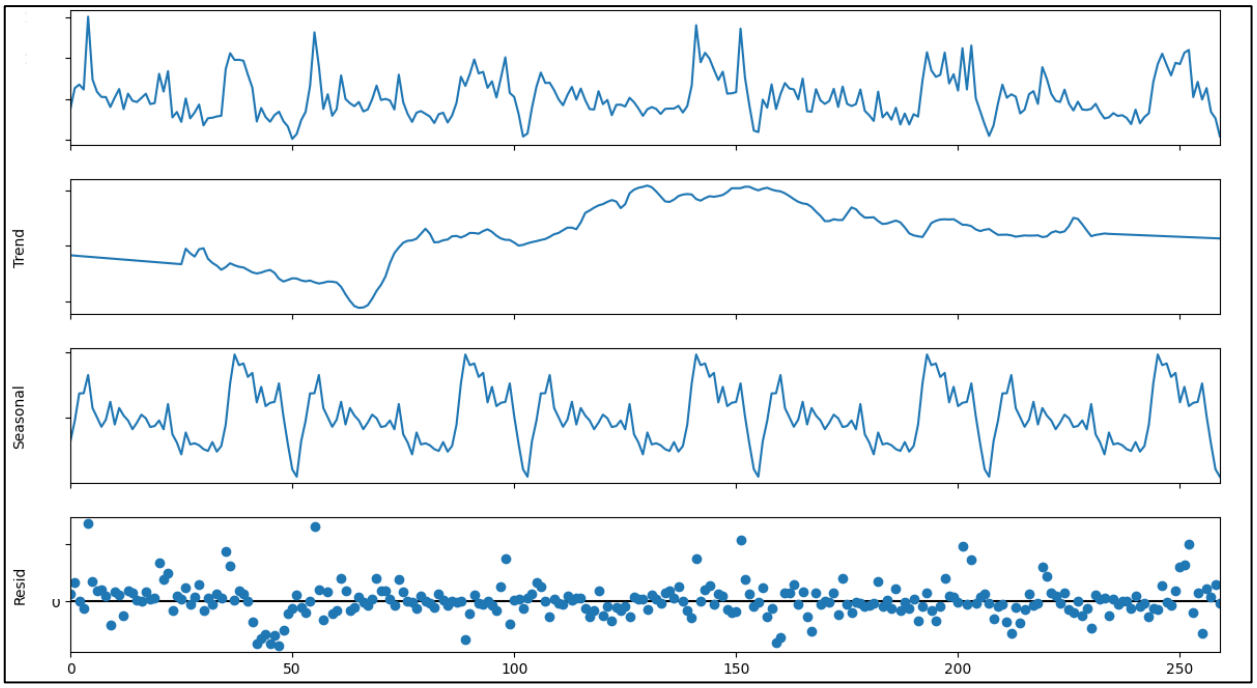


Figure 4.9 : Visuel de la décomposition additive obtenue sur python pour la demande hebdomadaire

Nous remarquons immédiatement la netteté du schéma saisonnier qui se répète chaque année. En revanche, nous observons que la tendance ne semble pas particulièrement maquée. La composante résiduelle semble osciller de manière aléatoire autour de 0, bien que certains points s'écartent grandement de la moyenne. Il se distingue cependant la répétition d'un même schéma au niveau de la composante résiduelle, de longueur plus au moins égale à celle du schéma saisonnier. Cela nous indique que les résidus ne sont pas parfaitement indépendants. Nous négligerons malgré tout cette information pour la suite de notre analyse. Notez finalement que les deux extrémités du signal de

la tendance sont extrapolées, car tel qu'exposé par Hyndman et al. (2021f), la décomposition classique ne nous permet pas d'obtenir une estimation de la composante tendance-cycle pour les premières et dernières observations.

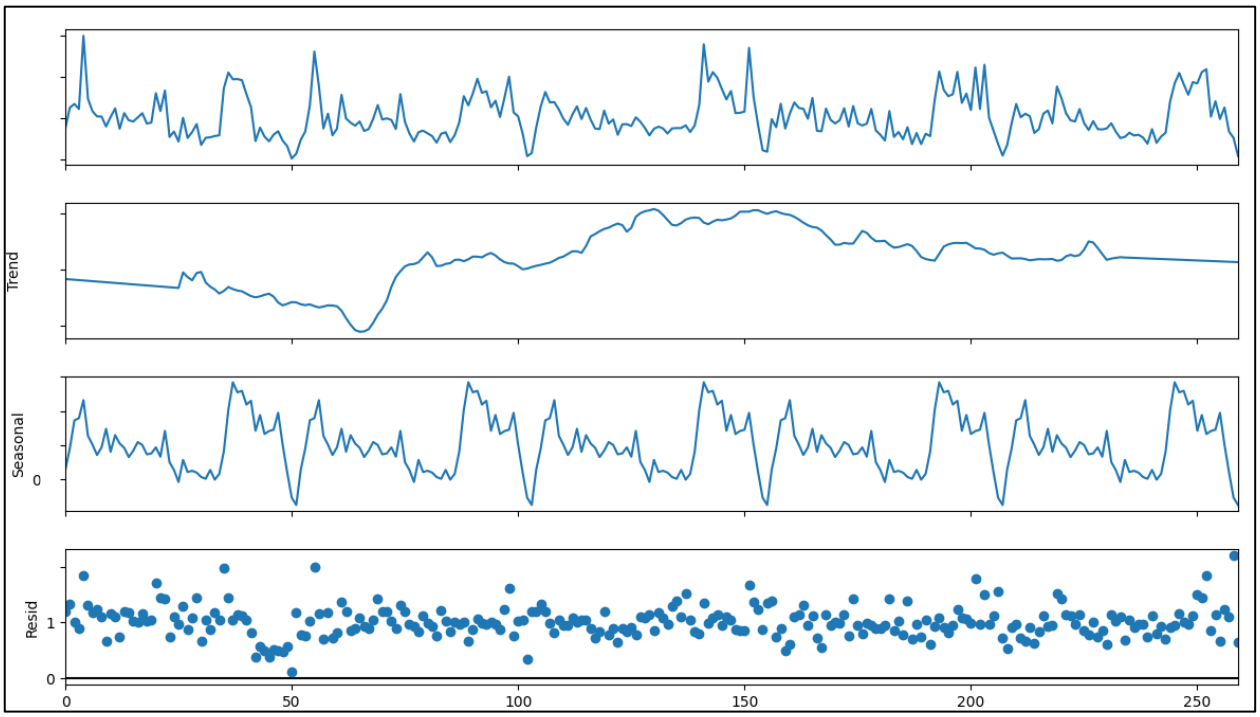


Figure 4.10 : Visuel de la décomposition multiplicative obtenue sur python pour la demande hebdomadaire

Ces résultats nous emmènent à des conclusions similaires. En revanche, la composante résiduelle semble osciller autour de 1. La décomposition STL nous conduit aux résultats suivants :

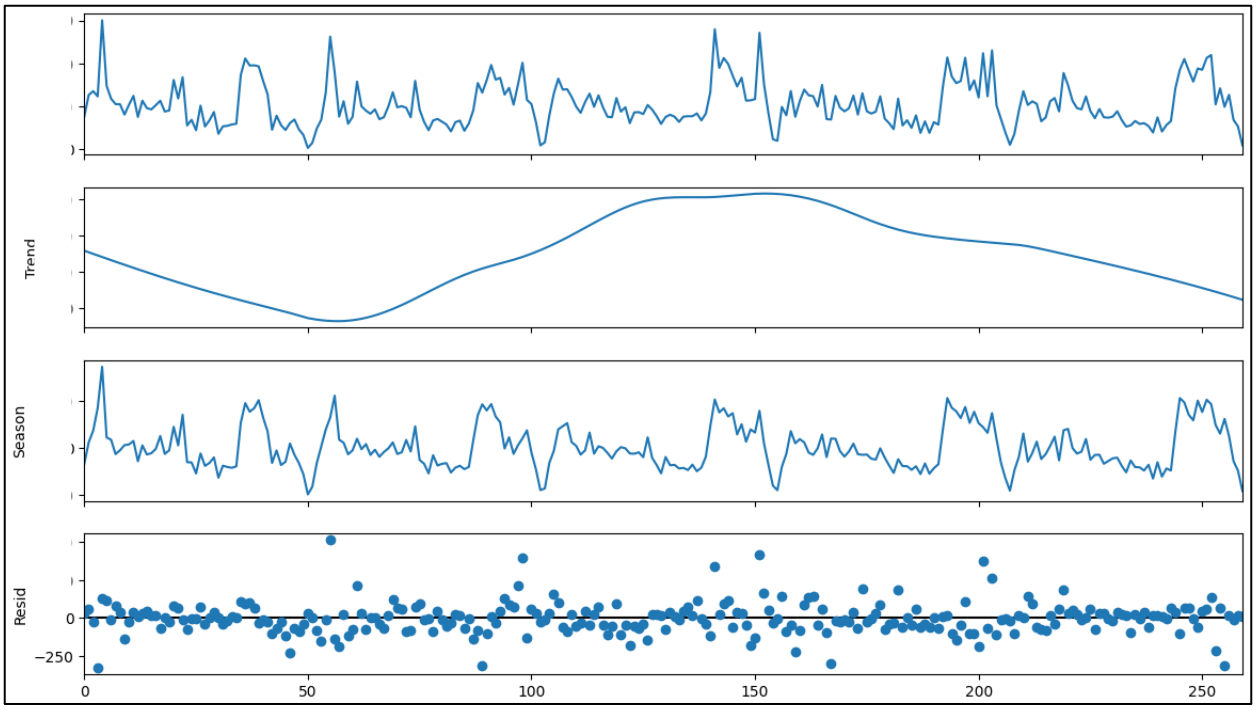


Figure 4.11 : Visuel de la décomposition STL obtenue sur python pour la demande hebdomadaire

Nous constatons en premier lieu que le schéma saisonnier, bien que clairement visible, semble différer quelque peu d'une année à l'autre. Cela n'a rien d'anormal, car la décomposition STL tient compte du fait que la composante saisonnière peut évoluer au cours du temps. La tendance semble quant à elle plus « régulière » que dans les graphiques précédents. C'est-à dire qu'elle semble plus lisse. Cela étant dit, elle ne semble toujours pas très imposante. Finalement, la composante résiduelle semble globalement un peu plus concentrée autour de 0.

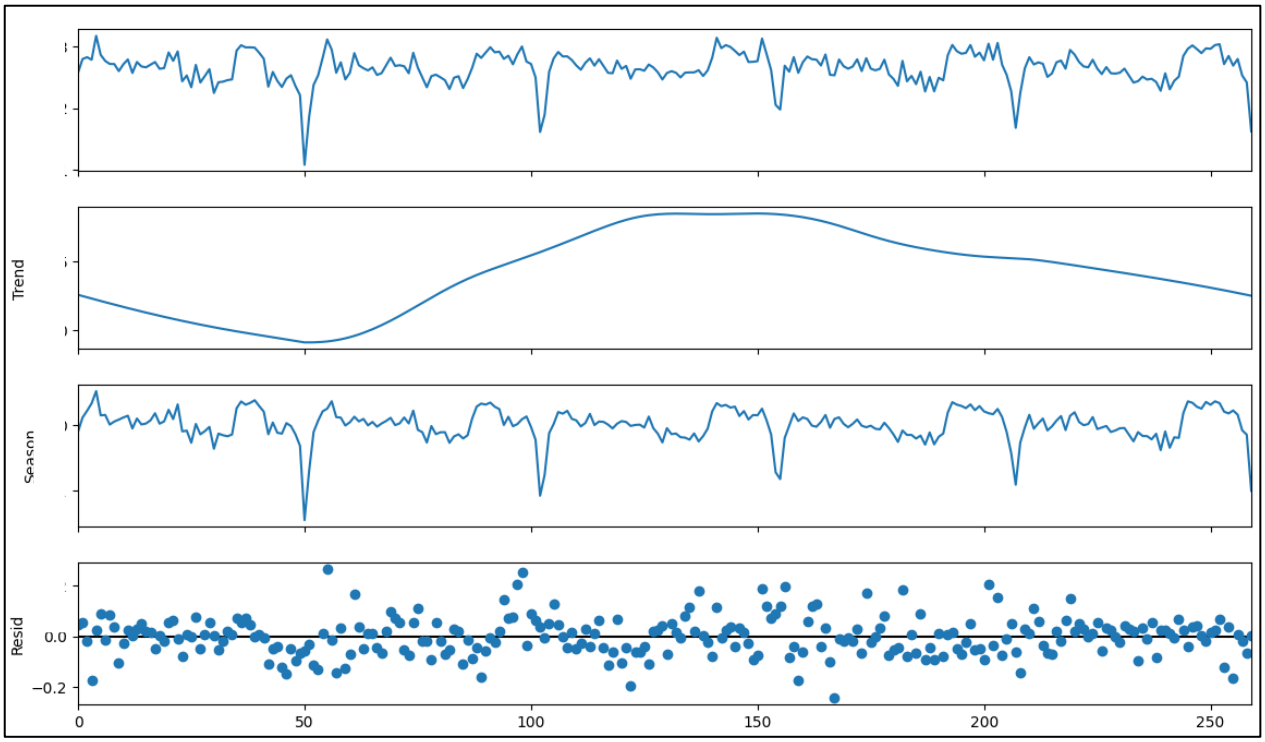


Figure 4.12 : Visuel de la décomposition STL du logarithme de la demande pour la demande hebdomadaire

Pour qu'elle soit multiplicative, la décomposition STL nécessite une transformation logarithmique préalable, tel que mentionné par Hyndman et al. (2021f). C'est la raison pour laquelle les valeurs présentées sur le graphique ci-dessus sont sans commune mesure avec les précédents. La décomposition accomplie, nous retransformons les données. Ces dernières ne seront cependant pas présentées graphiquement. Les conclusions sont ici semblables à celles que nous avons offertes pour la décomposition STL sans transformation logarithmique.

Nous avons d'autre part réalisé une décomposition STL sans transformation sur r (cette dernière nous permettant, sur R, d'obtenir des renseignements sur les caractéristiques de la série). Nous avons, à cette occasion, remarqué que les résultats obtenus diffèrent quelques peu, bien qu'il s'agisse pourtant de la même méthode que celle que nous avons employé sur python. Le graphique que nous obtenons ne nous permet cependant pas d'observer une si flagrante différence.

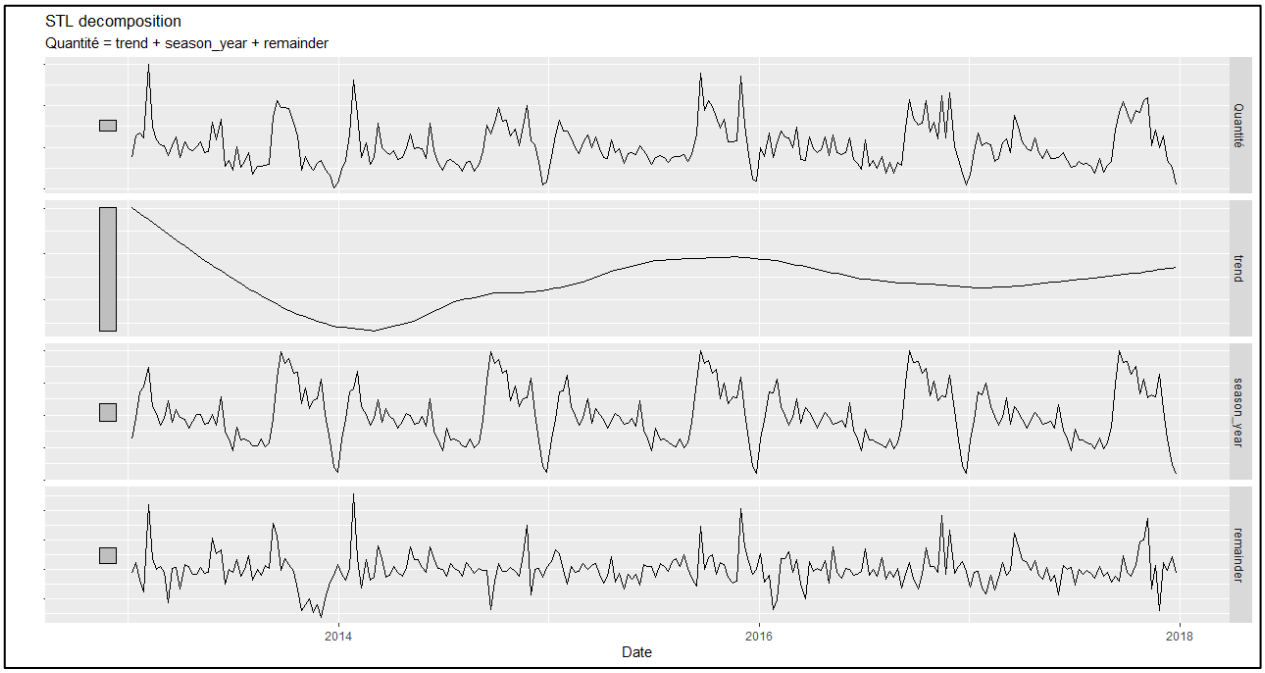


Figure 4.13 : Visuel de la décomposition STL réalisée sur R pour la demande hebdomadaire

Nous avons également réalisé l'ensemble de ces décompositions pour les données agrégées au mois. Les résultats sont disponibles en annexe A. Ces derniers ne sont pas commentés, car les conclusions sont comparables à celles que nous avons pour la demande hebdomadaire.

4.2.3 Évaluation des caractéristiques de la série

Les caractéristiques de notre série temporelle extraites, nous obtenons les résultats suivants :

Tableau 4.5 : Caractéristiques de la demande hebdomadaire

Entropie de Shannon	Force de la tendance	Force de la saison
0.868	0.054	0.694

Ces résultats nous permettent de constater en premier lieu que notre série ne présente pas une tendance très forte (force presque nulle). À l'inverse, la saisonnalité est relativement marquée (proche de 1). Une entropie proche de 1 laisse présager que la série risque de ne pas être évidente à prédire en raison de la présence d'un bruit important (Hyndman et al., 2021f). Comme annoncé antérieurement, par souci de confidentialité, nous avons décidé de ne pas présenter d'informations

relatives aux quartiles, au minimum, au maximum et à la médiane. Nous pouvons cependant affirmer que les quantités commandées nous semblent relativement dispersées.

Tableau 4.6 : Caractéristiques de la demande mensuelle

Entropie de Shannon	Force de la tendance	Force de la saison
0.554	0.101	0.771

Nous avons également décidé d'évaluer les caractéristiques de la demande agrégée au mois, pour évaluer la présence d'éventuelles différences. Nous remarquons en effet que le fait d'agréger la demande au mois semble avoir un impact sur les caractéristiques de la série. La force de la saison et de la tendance sont plus élevées, et l'entropie de Shannon plus faible. Cela nous laisse envisager que la prédiction mensuelle sera plus aisée que la prédiction hebdomadaire.

4.3 Spécification des modèles de prédiction

Les différents modèles présentés en 3.3 sont mis en œuvre ici.

4.3.1 Définition des modèles bayésiens

Tel que présenté dans le tableau B.1 présenté en annexe, nous avons sélectionné trois modèles différents pour réaliser 7 essais. Le premier modèle auquel nous avons songé est celui qui suppose que les quantités commandées (entières et positives) peuvent être modélisées par une loi de Poisson. Nous avons alors choisi pour la moyenne une *a priori* exponentielle, ce paramètre étant nécessairement positif et probablement un nombre décimal. L'*a posteriori* obtenue analytiquement est une binomiale négative. Les résultats que nous avons obtenus avec ce modèle ont été insatisfaisants, en raison de la présence d'une importante saisonnalité incorrectement traitée. Nos prédictions étaient ainsi toujours incorrectes. Nous avons éliminé ce modèle dont l'évaluation de la performance ne sera pas abordée, celui-ci n'étant pas un candidat adapté à notre cas d'étude. Pour paramétrer l'*a priori*, nous nous sommes servis de la relation suivante :

$$Si\ \lambda \sim \mathrm{Exp}(\gamma), \text{alors } E(\lambda) = \frac{1}{\gamma}$$

Nous utilisons des données historiques de 2013 à 2015 pour paramétrer l'*a priori*. Nous le faisons parce-que nous n'avons pas accès aux connaissances de l'expert de l'entreprise. D'autre part, toutes les familles de produits sont "anonymisées", ce qui fait que nous ne savons pas avec quel genre de produits nous travaillons, et surtout pour quel client. Si nous nous servions par exemple d'une autre famille de produits (donc d'autres vêtements) pour paramétrer l'*a priori*, nous aurions de grandes chances de prendre une famille de produits destinée à un client radicalement différent, pour des raisons tout aussi singulières. Nous n'aurions alors pas un juste paramétrage. Nous avons donc estimé que notre seul moyen d'approcher l'avis d'un expert était de nous servir des données historiques de la famille de produits étudiée, bien que cela puisse nuire à la rigueur de notre analyse. Nous estimons ainsi γ par l'inverse de la moyenne des données historiques.

Pour le deuxième modèle, nous avons décidé de modéliser la vraisemblance par une loi normale, avec pour *a priori* pour la moyenne et la variance respectivement une exponentielle et une inverse gamma. Cependant, le modèle est inadéquat lorsque nous utilisons les données résiduelles (données auxquelles ont été soustraites la tendance et la saison) obtenues par les décompositions additives et STL régulière. Les données résiduelles étant des nombres décimaux pouvant être positifs comme négatifs, nous pouvons les modéliser par une loi normale. En revanche, leur moyenne étant susceptible d'être négative, elle ne peut être modélisée par une loi exponentielle. Ce problème ne se pose pas lorsque nous utilisons les données résiduelles obtenues par décomposition multiplicative, car elles sont positives. Nous n'utiliserons donc ce modèle que sur certaines données résiduelles et pas d'autres. Le paramètre de l'*a priori* de la moyenne, γ, est estimé de la même manière que pour l'*a priori* du modèle précédent. Pour estimer les paramètres α et β de l'*a priori* de la variance, nous nous servons des relations suivantes :

$$\forall\, \alpha > 1, \quad E(X) = \frac{\beta}{\alpha - 1}$$

$$\forall\, \alpha > 2, \quad V(X) = \frac{\beta^2}{(\alpha - 1)^2 \times (\alpha - 2)}$$

Après réarrangement, nous obtenons les relations suivantes :

$$\forall\ V(X) > 0, \quad \alpha = \frac{E(X)^2}{V(X)} + 2$$

$$\forall\ V(X) > 0, \quad \beta = E(X) \times \left(\frac{E(X)^2}{V(X)} + 1 \right)$$

Nous faisons donc usage de la moyenne et de la variance des variances des différentes années historiques (2013, 2014 et 2015) dont nous disposons pour estimer α et β.

Pour le dernier modèle, nous avons choisi de modéliser la vraisemblance par une loi normale, avec pour *a priori* pour la moyenne et la variance respectivement une normale et une inverse gamma. Cette modélisation est adaptée aux différentes données résiduelles que nous obtenons avec les différentes méthodes de décomposition, ces dernières étant des nombres décimaux pouvant être positifs comme négatifs. La moyenne étant probablement également un réel, une loi normale nous semble indiquée. Finalement, nous avons choisi une inverse gamma comme *a priori* pour la variance car nous nous attendons à ce que cette dernière soit un nombre réel positif. Les paramètres sont estimés de la même manière que pour les autres modèles.

4.3.2 Définition des modèles de référence

Débutons par les modèles d'espace-état, sur R :

En nous inspirant des codes que nous offrent Hyndman et al. (2021f), nous obtenons pour la demande hebdomadaire un résultat infructueux. La fonction utilisée ne fonctionne que pour des périodes inférieures ou égales à 24. Or, nos données étant hebdomadaires et la saisonnalité annuelle, notre période est de 52 et nous ne pouvons obtenir de résultats.

En revanche, nous obtenons des résultats pour la demande mensuelle. En tenant compte des paramètres obtenus et disponibles en annexe, les composantes des modèles sont les suivantes :

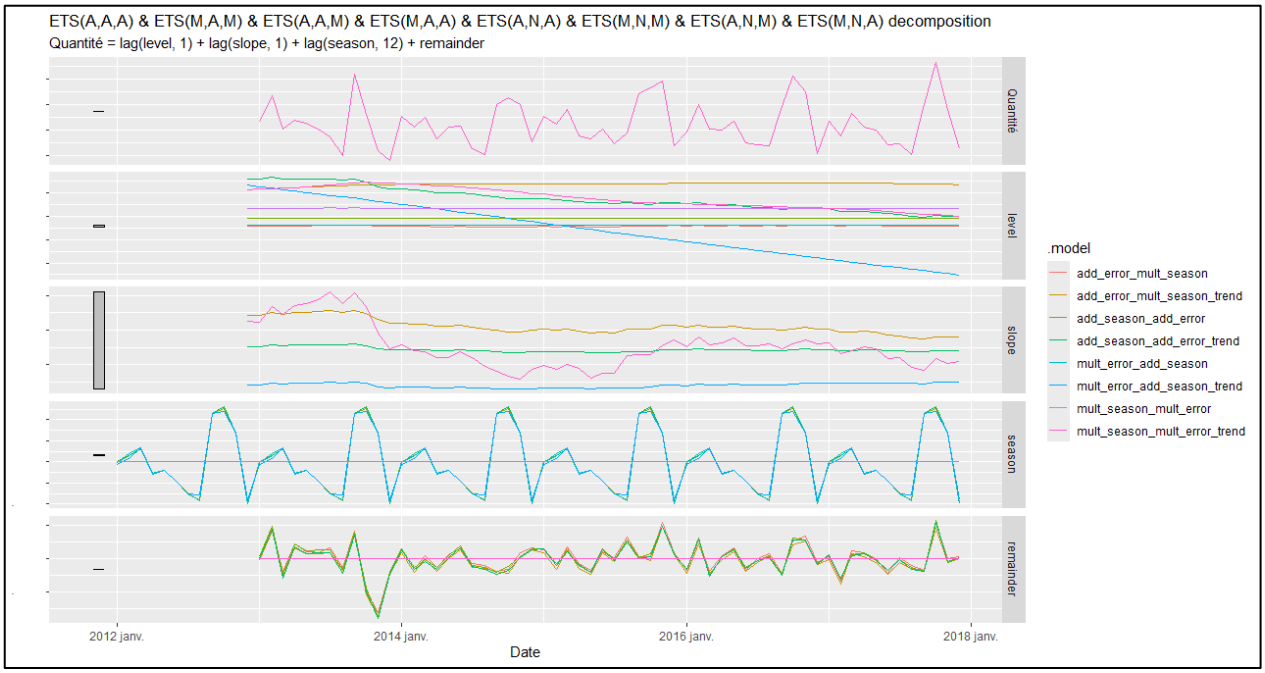


Figure 4.14 : Composantes des modèles d'espace-état

Le tableau 4.7 explique les correspondances entre la légende affichée sur le graphique et les modèles.

Tableau 4.7 : Correspondances entre la légende affichée sur le graphique et les modèles

Nom présenté dans la légende	Modèle
add_error_mult_season	ETS(A,N,M)
add_error_mult_season_trend	ETS(A,A,M)
add_season_add_error	ETS(A,N,A)
add_season_add_error_trend	ETS(A,A,A)
mult_error_add_season	ETS(M,N,A)
mult_error_add_season_trend	ETS(M,A,A)
mult_season_mult_error	ETS(M,N,M)
mult_season_mult_error_trend	ETS(M,A,M)

Poursuivons avec les modèles SARIMA :

Pour la demande hebdomadaire et à l'aide de la librairie fable, la fonction automatique ARIMA() nous propose les modèles suivants :

```
# A mable: 1 x 2
                    stepwise                      search
                     <model>                     <model>
1 <ARIMA(2,0,2)(1,1,1)[52]> <ARIMA(1,0,1)(0,1,1)[52]>
```

Figure 4.15 : Modèles proposés pour la demande hebdomadaire

Les paramètres des modèles sont disponibles en annexe.

Pour la demande mensuelle nous obtenons plutôt l'extrant suivant :

```
# A mable: 1 x 2
                    stepwise                      search
                     <model>                     <model>
1 <ARIMA(0,0,0)(0,1,0)[12]> <ARIMA(0,0,0)(0,1,0)[12]>
```

Figure 4.16 : Modèle proposé pour la demande mensuelle

Ici, nos deux tentatives nous ont conduit à un seul et même modèle. Les paramètres ne sont pas présentés en annexe, car l'ordre autorégressif et l'ordre des moyennes mobiles sont nuls.

4.4 Évaluation des modèles

4.4.1 Analyse des résidus

Modèles espace-état :

Avant de procéder à l'analyse des résidus, nous avons comparé nos 8 modèles pour ne conserver que le meilleur d'entre eux. Nous nous sommes servis pour ce faire du critère de l'AICc. À l'aide de la fonction report(), nous avons obtenu les résultats suivants :

```
# A tibble: 8 × 9
  .model                  sigma2 log_lik   AIC  AICc   BIC    MSE   AMSE     MAE
  <chr>                    <dbl>   <dbl> <dbl> <dbl> <dbl>  <dbl>  <dbl>   <dbl>
1 add_season_add_error_… 2.94e+5   -491. 1017. 1031. 1052. 2.16e5 2.15e5 340.
2 mult_season_mult_erro… 4.64e-2   -482.  998. 1013. 1034. 2.21e5 2.20e5   0.145
3 add_error_mult_season… 2.87e+5   -491. 1015. 1030. 1051. 2.11e5 2.10e5 339.
4 mult_error_add_season… 5.16e-2   -484. 1002. 1017. 1038. 2.27e5 2.26e5   0.153
5 add_season_add_error   2.76e+5   -491. 1011. 1022. 1043. 2.12e5 2.10e5 340.
6 mult_season_mult_error 4.44e-2   -482.  993. 1004. 1025. 2.10e5 2.10e5   0.148
7 add_error_mult_season  2.70e+5   -490. 1010. 1021. 1042. 2.07e5 2.06e5 337.
8 mult_error_add_season  4.71e-2   -483.  995. 1006. 1027. 2.14e5 2.13e5   0.152
```

Figure 4.17 : Extrant de la fonction report() pour les modèles d'espace-état

En privilégiant l'AICc, nous remarquons que le modèle le plus indiqué semble être le sixième. Ainsi, nous ne conserverons que ce dernier pour la suite de notre analyse.

Analysons les résidus obtenus pour le modèle ETS(M,N,M) :

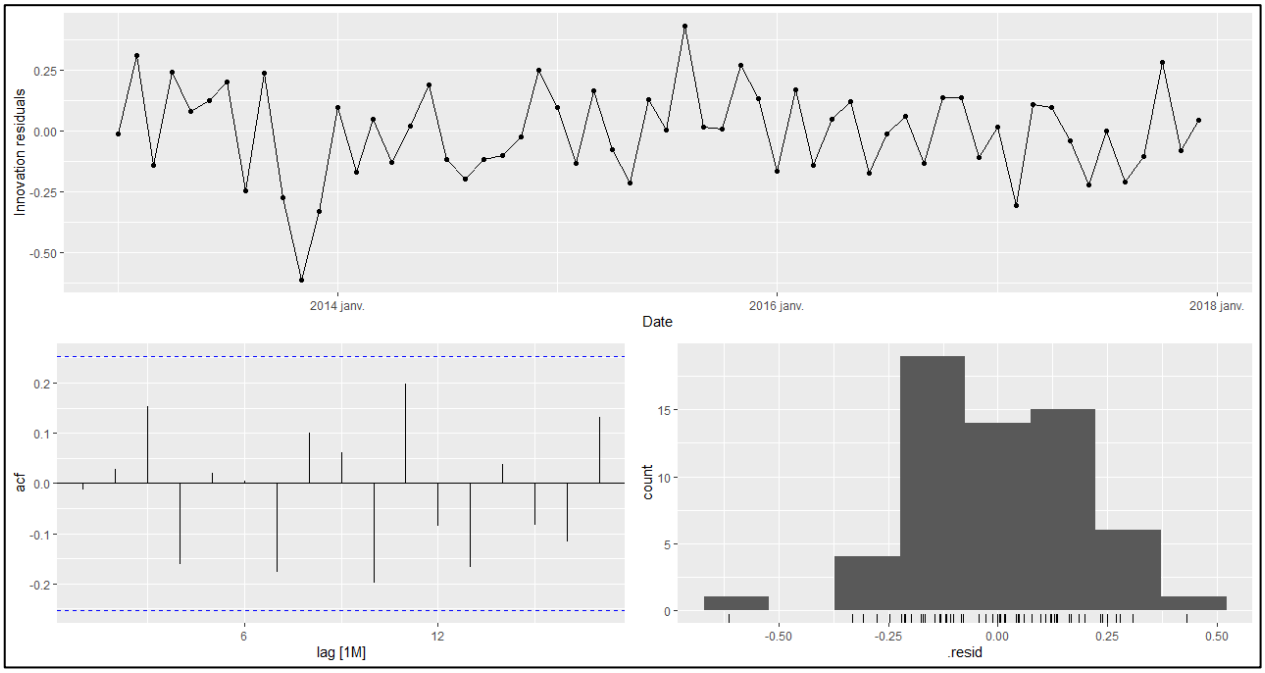


Figure 4.18 : Graphique obtenu avec gg_tsresiduals() pour ETS(M,N,M)

Nous remarquons que les résidus présentent une certaine variabilité autour de 0. La condition d'homoscédasticité ne semble pas tout à fait respectée. D'autre part, les résidus semblent non corrélés entre eux. Leur distribution n'a pas l'air normale.

En réalisant le test de Ljung-Box, nous obtenons le résultat suivant :

```
# A tibble: 8 × 3
  .model                       lb_stat lb_pvalue
  <chr>                          <dbl>     <dbl>
1 add_error_mult_season           22.8    0.534
2 add_error_mult_season_trend     22.4    0.558
3 add_season_add_error            21.1    0.634
4 add_season_add_error_trend      21.2    0.627
5 mult_error_add_season           32.1    0.124
6 mult_error_add_season_trend     27.0    0.304
7 mult_season_mult_error          33.5    0.0934
8 mult_season_mult_error_trend    33.5    0.0948
```

Figure 4.19 : Résultats du test de Ljung-Box

Nous déduisons de ces résultats que les résidus ne sont pas significativement différents d'un bruit blanc. Autrement dit, ils ne présentent pas d'autocorrélation.

Nous pouvons conclure de l'ensemble de ces résultats que les prédictions ne seront probablement pas significativement biaisées (les résidus sont plutôt centrés en 0). Également, n'étant pas auto-corrélés, les résidus nous permettent de conclure que le modèle a exploité de façon satisfaisante les informations qui étaient à disposition. Les conditions de normalité et d'homoscédasticité n'étant pas exactement respectées, nous pouvons nous attendre à avoir des résultats insatisfaisants en matière de prédiction.

Modèles SARIMA :

Nous suivons pour SARIMA les mêmes étapes que dans le cas des modèles d'espace-état.

Ainsi, nous produisons à l'aide de la fonction report() pour la demande hebdomadaire le résultat suivant :

```
# A tibble: 2 × 8
  .model   sigma2 log_lik   AIC  AICc   BIC ar_roots   ma_roots
  <chr>     <dbl>   <dbl> <dbl> <dbl> <dbl> <list>     <list>
1 stepwise 38013.  −1396. 2807. 2808. 2830. <cpl [54]> <cpl [54]>
2 search   35250.  −1393. 2795. 2795. 2808. <cpl [1]>  <cpl [53]>
```

Figure 4.20 : Extrant de la fonction report() pour les modèles SARIMA obtenus pour la demande hebdomadaire

Nous remarquons ici que search nous propose le meilleur modèle, à savoir ARIMA(1,0,1)(0,1,1)[52].

Nous procédons donc à l'analyse des résidus pour ARIMA(1,0,1)(0,1,1)[52] Nous obtenons le graphique suivant :

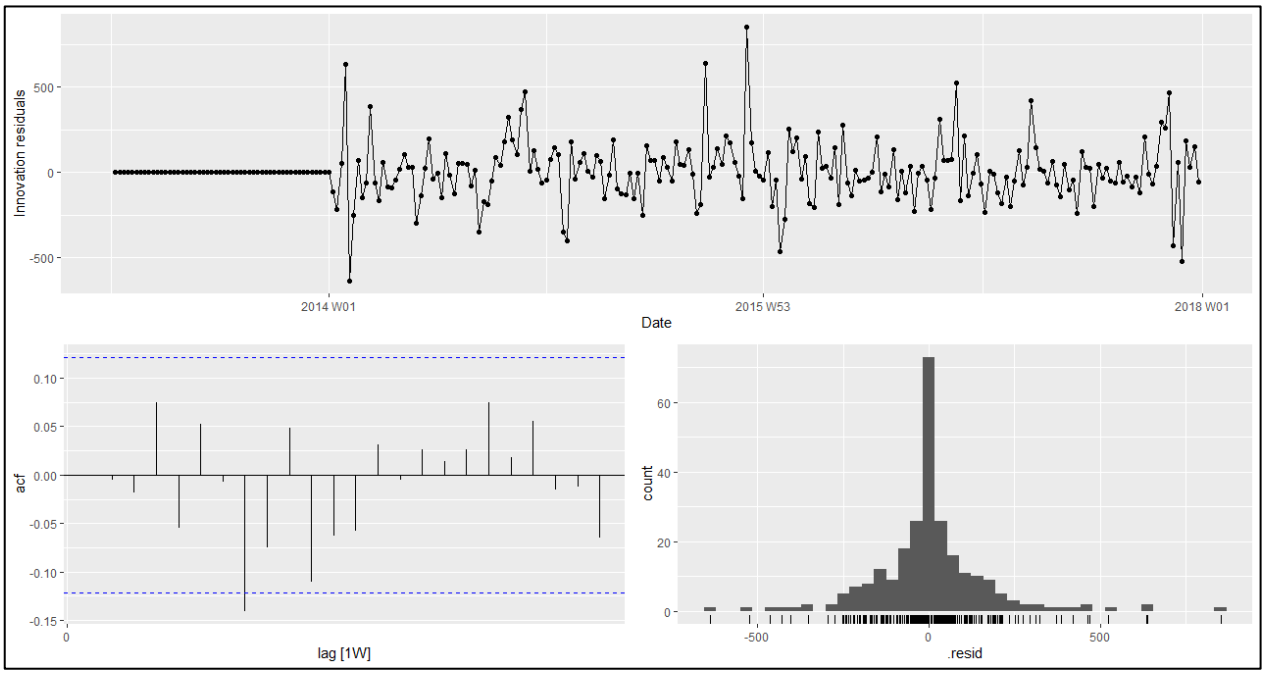


Figure 4.21 : Graphique obtenu avec gg_tsresiduals() pour le modèle ARIMA(1,0,1)(0,1,1)[52]

Nous observons en premier lieu que les 52 premières valeurs semblent nulles. Cela vient de la différenciation saisonnière qui a été effectuée par la fonction pour rendre la série stationnaire. En dehors de cela, les résidus semblent assez variables, et la distribution ressemble davantage à un pic qu'à une cloche. Les résidus ne semblent pas significativement auto-corrélés, bien que nous observions cependant une forme de schéma saisonnier se dessiner.

Le test de Ljung-Box nous offre les résultats suivants :

```
# A tibble: 2 × 3
  .model   lb_stat lb_pvalue
  <chr>      <dbl>     <dbl>
1 search      131.  0.0389
2 stepwise    161.  0.000279
```

Figure 4.22 : Résultats du test de Ljung-Box pour les modèles SARIMA de la demande hebdomadaire

En évaluant ce dernier résultat, nous pouvons déduire que les résidus sont significativement corrélés, et qu'ils sont ainsi significativement différents d'un bruit blanc.

Nous pouvons conclure de l'ensemble de ces informations que les prédictions ne seront probablement pas significativement biaisées (les résidus sont plutôt centrés en 0). D'autre part, les résidus étant auto-corrélés, le modèle n'a vraisemblablement pas exploité de façon satisfaisante les informations qui étaient à disposition. Il a donc le potentiel d'être amélioré. Les conditions de normalité et d'homoscédasticité n'étant pas exactement respectées, nous pouvons nous attendre à avoir des résultats insatisfaisants en matière de prédiction. Pour ce qui est de la demande mensuelle, inutile de faire usage de la fonction report(), puisque la fonction ARIMA ne nous propose qu'un modèle. Nous passons ainsi directement à l'analyse des résidus.

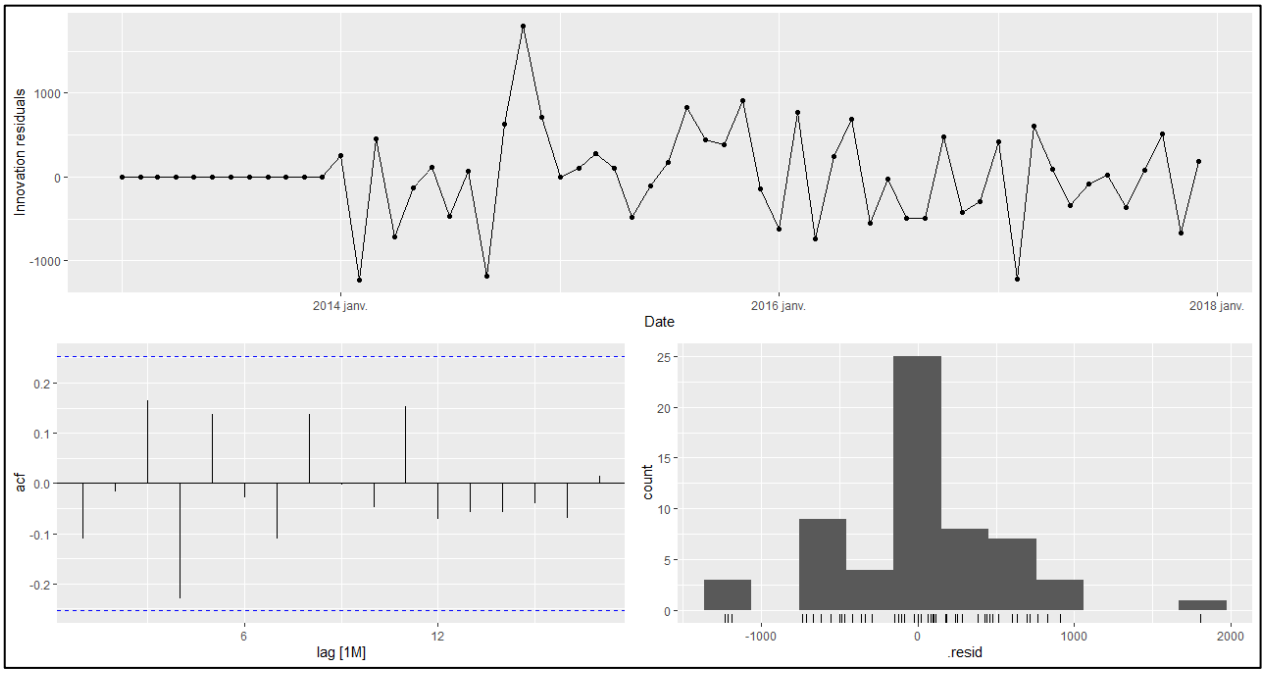


Figure 4.23 : Graphique obtenu avec gg_tsresiduals() pour le modèle ARIMA(0,0,0)(0,1,0)[12]

Plus ou moins centrés en 0, les résidus sont variables. La condition d'homoscédasticité n'est pas respectée. La distribution n'est pas normale. Les résidus semblent non-auto corrélés.

Appliquons le test de Ljung-Box :

```
# A tibble: 2 × 3
  .model   lb_stat lb_pvalue
  <chr>      <dbl>     <dbl>
1 search      22.7     0.537
2 stepwise    22.7     0.537
```

Figure 4.24 : Résultats du test de Ljung-Box pour les modèles SARIMA de la demande mensuelle

Le test de Ljung-box nous permet de conclure que les résidus sont effectivement non-auto corrélés, et non significativement différent d'un bruit blanc.

Nous pouvons conclure de l'ensemble de ces résultats que les prédictions ne seront probablement pas significativement biaisées (les résidus sont plutôt centrés en 0). Également, n'étant pas auto-corrélés, les résidus nous permettent de conclure que le modèle a exploité de façon satisfaisante les informations qui étaient à disposition. Les conditions de normalité et d'homoscédasticité n'étant pas respectées, nous pouvons encore une fois nous attendre à avoir des résultats insatisfaisants en matière de prédiction.

4.4.2 Évaluation de la qualité des simulations

À l'issue de nos simulations, nous obtenons les résultats suivants :

Tableau 4.8 : Statistiques mesurant la qualité de la simulation appliquée aux données résiduelles hebdomadaires

Modèle	Paramètre	SD	$\hat{R}$	ESS_bulk	ESS_tail	MCSE_mean	MCSE_mean/SD	MCSE_sd	MCSE_sd/SD
NEI-mul	μ	0.028	1	11571	12359	0	0	0	0
NEI-mul	σ^2	0.011	1	11425	11963	0	0	0	0
NEI-mul	σ	0.019	1	11425	11963	0	0	0	0
NEI-STL-log	μ	0.018	1	11840	13479	0	0	0	0
NEI-STL-log	σ^2	0.005	1	12464	12670	0	0	0	0
NEI-STL-log	σ	0.012	1	12464	12670	0	0	0	0
NNI-add	μ	0.879	1	12572	13765	0.008	0.009101251	0.006	0.0068259
NNI-add	σ^2	2792.47	1	11831	12959	25.817	0.00924522	18.305	0.0065551
NNI-add	σ	9.513	1	11831	12959	0.088	0.009250499	0.062	0.0065174
NNI-mul	μ	0.01	1	12357	13709	0	0	0	0
NNI-mul	σ^2	0.011	1	11289	11681	0	0	0	0
NNI-mul	σ	0.019	1	11289	11681	0	0	0	0
NNI-STL	μ	4.675	1	13341	14038	0.04	0.00855615	0.031	0.006631
NNI-STL	σ^2	1293.89	1	11837	12372	11.852	0.009159975	8.381	0.0064774
NNI-STL	σ	6.424	1	11837	12372	0.059	0.009184309	0.042	0.006538
NNI-STL-log	μ	0.013	1	11645	12211	0	0	0	0
NNI-STL-log	σ^2	0.005	1	12056	12436	0	0	0	0
NNI-STL-log	σ	0.012	1	12056	12436	0	0	0	0

Notez que « SD » fait référence à la déviation standard, « ESS_bulk » et « ESS_tail » respectivement au nombre estimatif d'échantillons effectifs au centre et aux extrémités de la distribution. « MCSE_mean » se réfère à l'erreur standard de Monte Carlo pour l'estimation de la moyenne du paramètre étudié. « MCSE_sd » représente quant à lui l'erreur standard de Monte Carlo pour l'estimation de la déviation standard du paramètre. Pour plus de renseignement sur ces deux derniers éléments, le lecteur est invité à consulter (Kenney et al., 1951).

Pour comprendre la correspondance entre les noms des modèles qui figurent sur la colonne la plus à gauche et ce qui les caractérise, le lecteur est invité à consulter le tableau B.1.

En analysant cet extrant, nous constatons en premier lieu que le nombre estimatif d'échantillons efficaces est supérieur à 10 000 pour l'ensemble des simulations. Nous en déduisons que nous devrions obtenir un aperçu suffisamment clair de la distribution *a posteriori* pour les différents paramètres. La taille de l'ESS est pour l'ensemble des essais équivalente à environ la moitié du nombre d'échantillons simulés (hors rodage). Nous estimons que, bien que nous pourrions être plus efficients, ce résultat n'est pas pour autant insatisfaisant. D'autre part, nous constatons que la statistique de Gelman-Rubin est de 1 pour l'ensemble des expérimentations. Cela nous laisse supposer que nos chaînes ont convergé. La déviation standard est généralement inférieure à 1, en dehors des simulations de certains paramètres pour les modèles NNI-add et NNI-STL. Pour ces derniers, pour les paramètres concernés, cela signifie qu'il plane une incertitude plus importante dans leur approximation. Cela étant dit, le rapport entre le MCSE et la déviation standard étant très petit (inférieur à 0.01), nous en déduisons que les chaînes sont plutôt stables, et que l'estimation des paramètres varie peu.

Tableau 4.9 : Évaluation qualitative des tracés de la simulation appliquée aux données résiduelles hebdomadaires

Modèle	Tracé de la chaîne	Graphique de la densité a posteriori
NEI-mul	Aucune anomalie	Densité sensiblement normale, symétrique, fluctuations modérées
NEI-mul	Aucune anomalie	Densité sensiblement positivement asymétrique, légères fluctuations
NEI-mul	Aucune anomalie	Densité sensiblement positivement asymétrique, légères fluctuations
NEI-STL-log	Aucune anomalie	Densité sensiblement normale, symétrique, fluctuations modérées
NEI-STL-log	Aucune anomalie	Densité sensiblement positivement asymétrique, fluctuations modérées
NEI-STL-log	Aucune anomalie	Densité sensiblement positivement asymétrique, fluctuations modérées
NNI-add	Aucune anomalie	Densité sensiblement normale, symétrique, légères fluctuations
NNI-add	Aucune anomalie	Densité sensiblement positivement asymétrique, flagrantes fluctuations
NNI-add	Aucune anomalie	Densité sensiblement positivement asymétrique, flagrantes fluctuations
NNI-mul	Aucune anomalie	Densité sensiblement normale, symétrique, légères fluctuations
NNI-mul	Aucune anomalie	Densité sensiblement positivement asymétrique, légères fluctuations
NNI-mul	Aucune anomalie	Densité sensiblement positivement asymétrique, légères fluctuations
NNI-STL	Aucune anomalie	Densité sensiblement normale, symétrique, légères fluctuations
NNI-STL	Aucune anomalie	Densité sensiblement positivement asymétrique, légères fluctuations
NNI-STL	Aucune anomalie	Densité sensiblement positivement asymétrique, légères fluctuations
NNI-STL-log	Aucune anomalie	Densité sensiblement normale, symétrique, fluctuations modérées
NNI-STL-log	Aucune anomalie	Densité sensiblement positivement asymétrique, légères fluctuations
NNI-STL-log	Aucune anomalie	Densité sensiblement positivement asymétrique, légères fluctuations

Tel qu'énoncé dans ce tableau, la plupart des tracés ne présentaient aucune anomalie particulière. Les itérations des différentes chaînes semblent plutôt superposées et présentent l'apparence recherchée (celle d'une grosse chenille). Les graphiques de densité sont plutôt semblables d'une simulation à l'autre, présentant parfois quelques fluctuations. Les courbes ne sont pas parfaitement superposées, mais pas très éloignées non plus, ce qui ne nous permet pas de conclure que la représentativité des échantillons est tout à fait satisfaisante. La moyenne est représentée par une densité qui semble normale, tandis que la variance ressemble à une inverse-gamma. La déviation standard semble quant à elle se trouver entre les deux. Elle n'est pas parfaitement symétrique, mais elle l'est davantage que la courbe qui représente la variance. Cela n'est pas surprenant, car l'application de la racine carrée aux échantillons simulés pour la variance fait en sorte que les valeurs qui étaient inférieures à 1 sont plus dispersées, tandis que les valeurs plus élevées se compressent. La « queue » de la distribution de droite rapetisse alors un peu, tandis que celle de gauche se développe. C'est d'ailleurs la raison pour laquelle dans le tableau, nous observons parfois que la déviation standard associée à σ est un peu plus importante que celle associées à σ^2. Un exemple des tracés obtenus est disponible en annexe.

Tableau 4.10 : Extrants obtenus en appliquant la méthode PSIS pour la simulation appliquée aux données hebdomadaires

Modèle	elpd_loo	p_loo	k
NEI-mul	-19.1	3.17	≤ 0.5
NEI-STL-log	27.09	2.7	≤ 0.5
NNI-add	-665.58	1.69	≤ 0.5
NNI-mul	-18.44	2.14	≤ 0.5
NNI-STL	-626.47	2.18	≤ 0.5
NNI-STL-log	27.56	2.19	≤ 0.5

Nous remarquons que, pour l'ensemble des modèles, k est inférieur à 0.5. Cela nous laisse supposer que l'approximation est probablement plutôt fiable. Par ailleurs, en dehors du modèle NEI-mul, nous avons p_loo qui est à la fois inférieur au nombre total d'observations et au nombre de paramètres (3). Nous pouvons en conclure que la majorité des modèles semblent disposer d'une capacité prédictive satisfaisante, *a priori*. Finalement, la colonne « elpd_loo » laisse entrevoir que parmi les modèles, les plus précis seront probablement NNI-STL-log et NEI-STL-log. Nous remarquons en particulier que les modèles avec décomposition additive affichent un « elpd_loo » largement plus petit, et donc qu'ils risquent d'être bien plus imprécis que les autres.

Les tableaux présentés ci-dessous nous ont mené à des conclusions hautement semblables, pour les données résiduelles mensuelles.

Tableau 4.11 : Statistiques mesurant la qualité de la simulation appliquée aux données résiduelles mensuelles

Modèle	Paramètre	SD	$\hat{R}$	ESS_bulk	ESS_tail	MCSE_mean	MCSE_mean/SD	MCSE_sd	MCSE_sd/SD
NEI-mul	μ	0.034	1	12033	12067	0	0	0	0
NEI-mul	σ^2	0.007	1	11408	12339	0	0	0	0
NEI-mul	σ	0.021	1	11408	12339	0	0	0	0
NEI-STL-log	μ	0.027	1	11336	12063	0	0	0	0
NEI-STL-log	σ^2	0.003	1	10891	13148	0	0	0	0
NEI-STL-log	σ	0.012	1	10891	13148	0	0	0	0
NNI-add	μ	13.416	1	12208	12731	0.122	0.00909362	0.087	0.0064848
NNI-add	σ^2	43572.534	1	12163	12730	399.569	0.009170203	284.317	0.0065251
NNI-add	σ	51.958	1	12163	12730	0.475	0.009141999	0.337	0.006486
NNI-mul	μ	0.011	1	11535	11728	0	0	0	0
NNI-mul	σ^2	0.007	1	11971	12842	0	0	0	0
NNI-mul	σ	0.021	1	11971	12842	0	0	0	0
NNI-STL	μ	22.745	1	13143	14830	0.199	0.008749176	0.147	0.006463
NNI-STL	σ^2	10122.849	1	11741	11891	93.604	0.009246804	66.23	0.0065426
NNI-STL	σ	18.035	1	11741	11891	0.167	0.009259773	0.118	0.0065428
NNI-STL-log	μ	0.016	1	11536	12698	0	0	0	0
NNI-STL-log	σ^2	0.003	1	12396	13007	0	0	0	0
NNI-STL-log	σ	0.012	1	12396	13007	0	0	0	0

Tableau 4.12 : Évaluation qualitative des tracés de la simulation appliquée aux données résiduelles mensuelles

Modèle	Tracé de la chaîne	Graphique de la densité a posteriori
NEI-mul	Aucune anomalie	Densité sensiblement normale, symétrique, flagrantes fluctuations
NEI-mul	Aucune anomalie	Densité sensiblement positivement asymétrique, légères fluctuations
NEI-mul	Aucune anomalie	Densité sensiblement positivement asymétrique, légères fluctuations
NEI-STL-log	Aucune anomalie	Densité sensiblement normale, symétrique, fluctuations modérées
NEI-STL-log	Aucune anomalie	Densité sensiblement positivement asymétrique, flagrantes fluctuations
NEI-STL-log	Aucune anomalie	Densité sensiblement positivement asymétrique, flagrantes fluctuations
NNI-add	Aucune anomalie	Densité sensiblement normale, symétrique, flagrantes fluctuations
NNI-add	Aucune anomalie	Densité sensiblement positivement asymétrique, flagrantes fluctuations
NNI-add	Aucune anomalie	Densité sensiblement positivement asymétrique, flagrantes fluctuations
NNI-mul	Aucune anomalie	Densité sensiblement normale, symétrique, légères fluctuations
NNI-mul	Aucune anomalie	Densité sensiblement positivement asymétrique, légères fluctuations
NNI-mul	Aucune anomalie	Densité sensiblement positivement asymétrique, légères fluctuations
NNI-STL	Aucune anomalie	Densité sensiblement normale, symétrique, fluctuations modérées
NNI-STL	Aucune anomalie	Densité sensiblement positivement asymétrique, fluctuations modérées
NNI-STL	Aucune anomalie	Densité sensiblement positivement asymétrique, fluctuations modérées
NNI-STL-log	Aucune anomalie	Densité sensiblement normale, symétrique, flagrantes fluctuations
NNI-STL-log	Aucune anomalie	Densité sensiblement positivement asymétrique, légères fluctuations
NNI-STL-log	Aucune anomalie	Densité sensiblement positivement asymétrique, légères fluctuations

Tableau 4.13 : Extrants obtenus en appliquant la méthode PSIS pour la simulation appliquée aux données mensuelles

Modèle	elpd_loo	p_loo	k
NEI-mul	10.23	1.23	≤ 0.5
NEI-STL-log	16.95	0.99	≤ 0.5
NNI-add	-177.74	0.88	≤ 0.5
NNI-mul	10.24	0.48	≤ 0.5
NNI-STL	-166.81	0.28	≤ 0.5
NNI-STL-log	17.5	0.51	≤ 0.5

Nous remarquons ici qu'à la différence des simulations réalisées pour les données résiduelles hebdomadaires, le nombre effectif de paramètres est ici inférieur au nombre total d'observations et au nombre de paramètres des modèles pour l'ensemble des expériences. Cela signifie qu'*a priori*, aucun d'entre eux n'a une capacité prédictive particulièrement faible. Encore une fois, les modèles étant identifiés comme ayant le meilleur potentiel en matière de précision sont NNI-STL-log et NEI-STL-log.

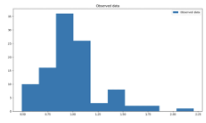
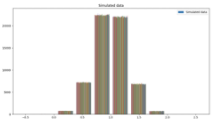
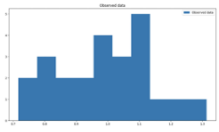
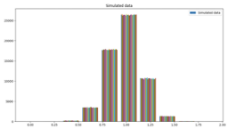
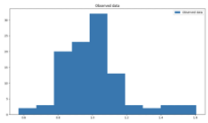
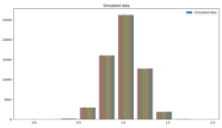
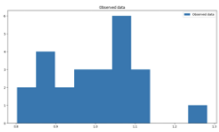
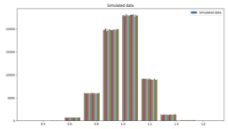
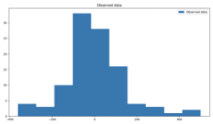
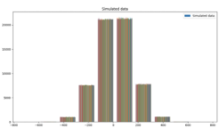
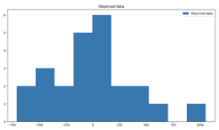
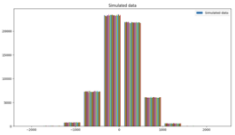
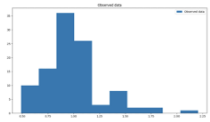
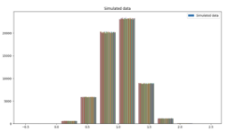
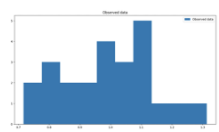
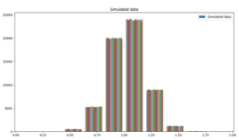
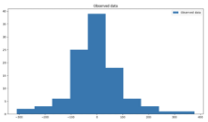
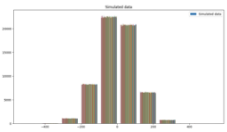
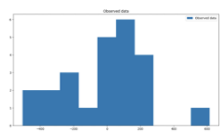
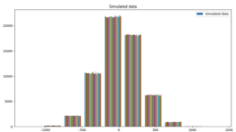
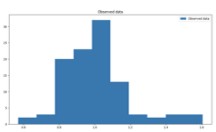
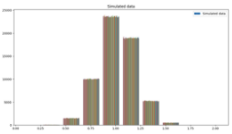
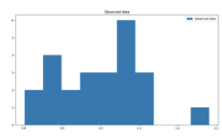
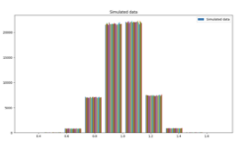
4.4.3 Analyse graphique des données observées et simulées

Comparons à présent les données observées avec les données simulées (tableau 4.14). En premier lieu, notez pour les graphiques des données simulées qu'il n'y a en réalité pas de « trous » comme nous le laissent penser les espaces blancs. Cet effet est strictement lié au type de graphique généré.

Nous remarquons que les données simulées semblent se rapprocher d'une distribution normale, tant pour les simulations réalisées à partir des données résiduelles hebdomadaires que mensuelles. Les données observées ne s'en approchent pas particulièrement. Pour les résultats hebdomadaires, à gauche, nous évaluons que les données résiduelles sont distribuées de façon plutôt positivement asymétrique, avec des queues relativement étendues. Pour les résultats mensuels, à droite, nous constatons plutôt que les données résiduelles ont une distribution généralement bimodale. Pour l'ensemble des modèles, les données simulées n'englobent pas toujours toutes les valeurs de données résiduelles observées. Pour certains, les données simulées englobent des valeurs qui n'ont pas été observées (plus grandes ou plus petites). Nous en déduisons que nos modèles ne reflètent pas correctement le comportement de nos données. Le type de décomposition que nous avons employé ne nous semble visuellement pas avoir un grand impact sur le comportement global des données résiduelles. Notez par ailleurs que bien que les résultats obtenus pour les deux catégories de modèles (Normale-Normale-Inverse Gamma et Normale-Exponentielle-Inverse Gamma) aient

l'air similaires, ils diffèrent quelque peu. Cela n'est pas visible sur le tableau car les figures sont petites, mais les données simulées issues des modèles « NEI » sont davantage décalées vers la gauche que celles qui sont obtenues à l'aide des modèles « NNI ». Cela est lié au fait que l'*a priori* de la moyenne (l'exponentielle) « tire » les données vers la gauche (les plus faibles valeurs positives étant associées à une plus grande probabilité). Les différences entre les deux modèles ne sont pratiquement pas perceptibles pour les données hebdomadaires car, étant plus nombreuses (104 observations plutôt que 24), la vraisemblance domine davantage les *a priori*.

Tableau 4.14 : Données observées versus données simulées

	Hebdomadaire		Mensuel	
	Données observées	**Données simulées**	**Données observées**	**Données simulées**
NEI-mul				
NEI-STL-log				
NNI-add				
NNI-mul				
NNI-STL				
NNI-STL-log				

Nous pouvons conclure de l'ensemble de ces résultats que la façon dont nous avons choisi de modéliser nos données est probablement sujette à amélioration. Les données simulées ne sont pas tout à fait représentatives des données réelles, comme nous avons pu le constater à l'aide des

différents graphiques analysés. Cependant, en dépit d'une modélisation imparfaite, les statistiques que nous avons évaluées nous laissent entendre que les simulations se sont réalisées sans véritable encombre. Leurs efficience, stabilité et précision respectives semblent décentes. La méthode PSIS nous a offert des résultats qui nous permettent d'envisager une certaine capacité de prédiction. Nous nous attendons à des résultats moyennement satisfaisants, particulièrement pour les données plus extrêmes qui ne semblent pas avoir été correctement prises en compte par nos modèles.

4.5 Prédiction/Estimation

4.5.1 Modèles d'espace-état :

Bien que les conditions d'homoscédasticité et de normalité ne soient pas respectées, nous avons tenté de réaliser des prédictions pour les 12 prochains mois. Nous obtenons le résultat ci-dessous :

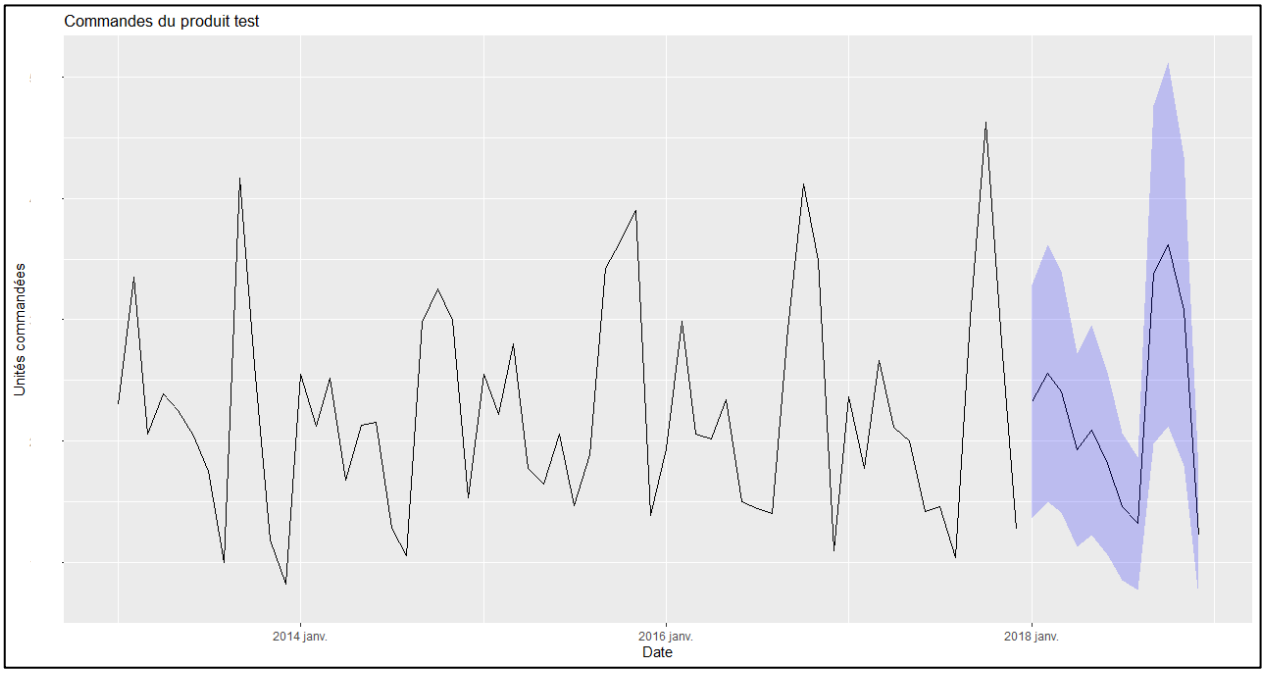


Figure 4.25 : Prédictions réalisées à l'aide du modèle espace-état pour les 12 prochains mois

La plage bleue représente l'intervalle de prédiction pour une probabilité de couverture de 95%.

La fonction forecast() nous renvoie une distribution de probabilité pour chaque mois que nous souhaitons prédire, avec pour prédiction ponctuelle la moyenne de ladite distribution. Par souci de

confidentialité, nous ne présenterons pas les résultats numériques obtenus. Il en sera de même pour les résultats obtenus avec les autres méthodes.

4.5.2 SARIMA :

En réalisant des prédictions pour les 12 prochaines semaines à l'aide du modèle ARIMA(1,0,1)(0,1,1)[52], nous obtenons les résultats ci-dessous.

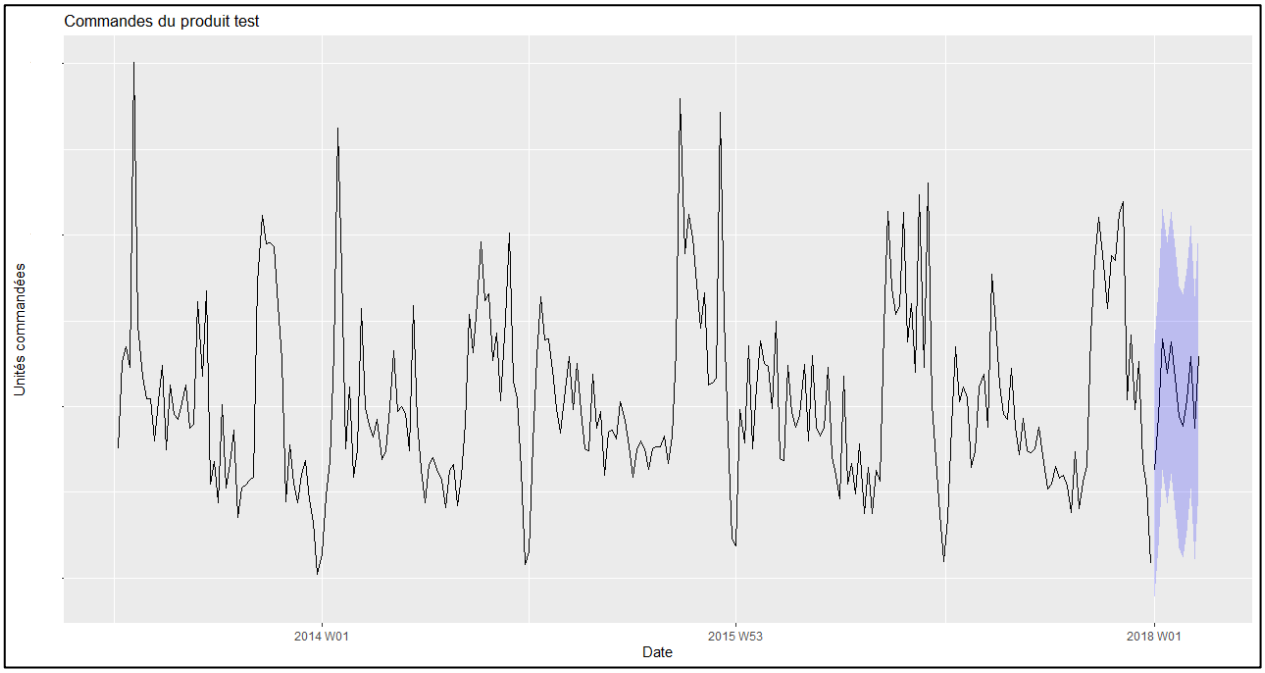


Figure 4.26 : Prédictions réalisées à l'aide du modèle ARIMA(1,0,1)(0,1,1)[52] pour les 12 prochaines semaines

En réalisant des prédictions pour les 12 prochains mois à l'aide du modèle ARIMA(0,0,0)(0,1,0)[12], nous obtenons les résultats ci-dessous.

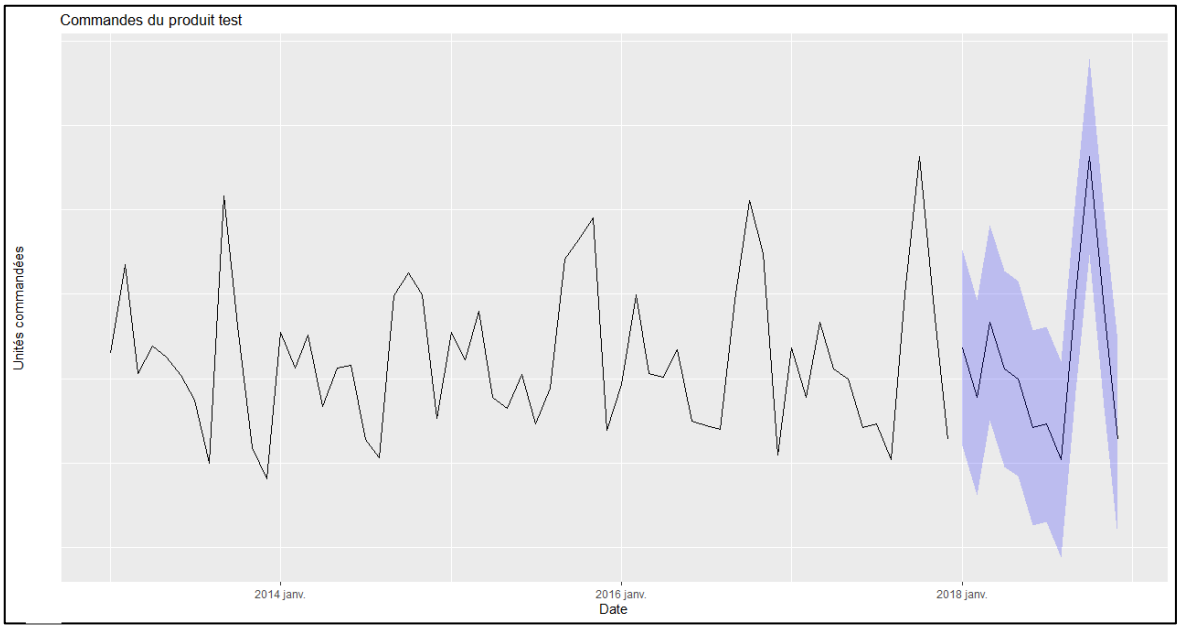


Figure 4.27 : Prédictions réalisées à l'aide du modèle ARIMA(0,0,0)(0,1,0)[12] pour les 12 prochains mois

4.5.3 Modèles bayésiens :

En réalisant nos prédictions, nous obtenons tant pour la demande hebdomadaire que mensuelle et pour chaque période prédite une distribution comme celle que nous présentons ci-dessous.

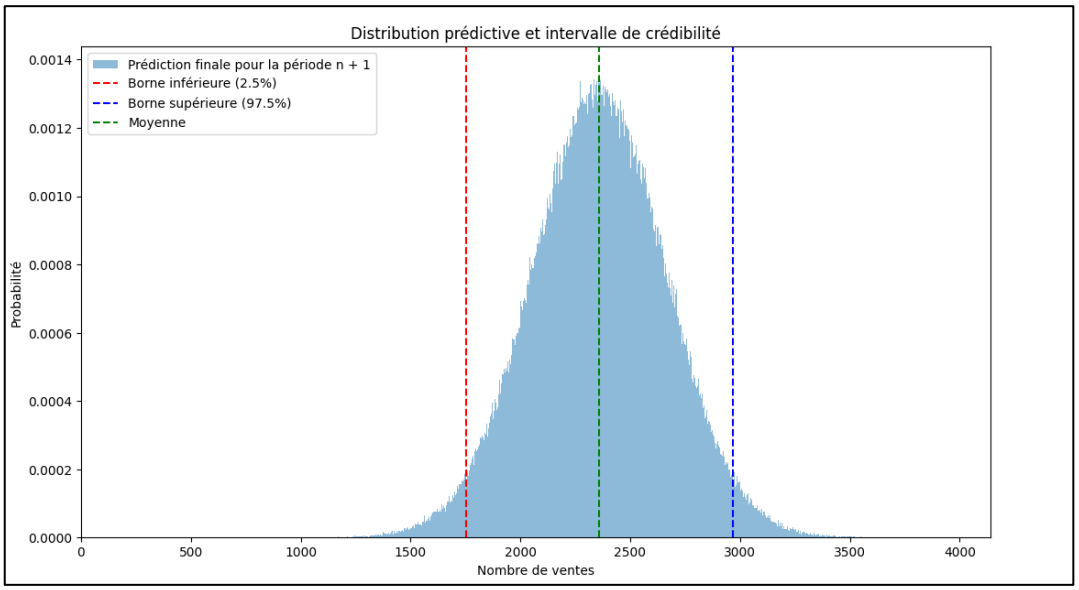


Figure 4.28 : Exemple de distribution prédictive obtenue à l'aide des modèles bayésiens adaptés

La distribution obtenue est semblable pour l'ensemble des modèles.

4.6 Évaluation de la performance

Nous avons, pour les modèles SARIMA, Holt-Winters et bayésiens calculé le MAE, le MAPE et le score quantile moyen. La moyenne a été réalisée pour les 12 périodes prédites. Les calculs ont été effectués séparément pour les prévisions hebdomadaires et mensuelles. Nous avons utilisé la moyenne des échantillons de prédictions comme estimation de la valeur prédite.

Tableau 4.15 : Classement des méthodes en matière de performance pour les prédictions hebdomadaires

MAE	Outil	MAPE (%)	Outil	Score Quantile	Outil
198.25	SARIMA	32.89	SARIMA	77.1	SARIMA
224.35	NEI-STL-log-ext	36.17	NEI-STL-log-ext	116.96	NNI-add-ext
224.63	NNI-STL-ext	36.2	NNI-STL-ext	118.32	NNI-add
225.34	NNI-STL-log-ext	36.52	NNI-STL-log-ext	119.46	NNI-mult-ext
232.26	NEI-mul-ext	39.46	NEI-mul	121.66	NNI-mul
232.35	NEI-mul	39.5	NEI-STL-log	123.92	NEI-mul-ext
232.59	NEI-STL-log	39.57	NEI-mul-ext	126.09	NEI-mul
233.28	NNI-STL	39.87	NNI-STL-log	157.95	NNI-STL
233.62	NNI-STL-log	40.23	NNI-STL	164.14	NNI-STL-log
235.72	NNI-add-ext	41.09	NNI-mul	165.1	NEI-STL-log
235.87	NNI-add	41.19	NNI-mul-ext	173.39	NNI-STL-ext
236.76	NNI-mul-ext	41.2	NNI-add	182.64	NNI-STL-log-ext
236.86	NNI-mult	41.3	NNI-add-ext	183.55	NEI-STL-log-ext

La présence de « -ext » dans le nom de certaines de nos expérimentations évoquent l'extrapolation de la tendance. Son absence évoque l'estimation de la tendance par une moyenne.

Nous observons pour la prédiction hebdomadaire que le modèle le plus performant est SARIMA, avec un pourcentage d'erreur absolue moyenne de 32.89%. Le modèle NEI-STL-log-ext occupe la seconde place avec un pourcentage d'erreur absolue moyenne de 3.28% plus élevé que SARIMA. Nous estimons qu'il ne s'agit pas d'une différence très importante. La performance reste cependant plutôt insatisfaisante selon le MAPE que nous jugeons trop élevé. L'écart entre SARIMA et NEI-STL-log semble un peu plus conséquent lorsque nous observons les deux autres métriques. Le score quantile est ici calculé pour $p = 0.975$. Il ne nous semble pas étonnant que nous ayons un score quantile significativement plus élevé que notre méthode de référence SARIMA. Cela car

nous avions remarqué précédemment que nos modélisations ne rendaient pas toujours compte adéquatement des valeurs résiduelles extrêmes. Il est donc tout à fait possible que la valeur quantile associée au 97.5ème centile de nos distributions prédictives soit plus faible que ce que nous aurions voulu, et donc que nous soyons davantage pénalisés. Nous considérons que la méthode de Holt-Winters telle que nous avons voulu l'utiliser occupe la dernière place, puisqu'elle n'a produit aucune prédiction.

Tableau 4.16 : Classement des méthodes en matière de performance pour les prédictions mensuelles

MAE	Outil	MAPE (%)	Outil	Score Quantile	Outil
280.58	NNI-STL-log-ext	12.40	NEI-STL-log-ext	54.20	Holt-Winters
283.63	NEI-STL-log-ext	12.46	NNI-STL-log-ext	62.26	NNI-mul
284.58	NNI-STL-ext	13.10	NNI-STL-ext	62.56	SARIMA
289.90	NEI-mul-ext	13.35	NEI-mul-ext	65.39	NNI-mul-ext
312.86	NNI-add-ext	14.93	NNI-add-ext	68.48	NNI-add
324.53	NNI-mul-ext	15.95	NNI-mul-ext	70.17	NNI-add-ext
327.15	NEI-mul	15.96	SARIMA	74.34	NEI-mul
330.25	SARIMA	16.10	NEI-mul	77.37	NEI-mul-ext
332.56	NEI-STL-log	16.93	NEI-STL-log	88.63	NNI-STL-log
340.60	NNI-STL-log	17.54	NNI-STL-log	89.93	NEI-STL-log
351.84	NNI-STL	18.10	NNI-STL	100.65	NNI-STL
365.75	NNI-add	19.02	NNI-add	116.41	NNI-STL-ext
366.07	NNI-mul	19.04	NNI-mul	117.25	NNI-STL-log-ext
393.83	Holt-Winters	19.81	Holt-Winters	119.85	NEI-STL-log-ext

Nous observons pour la prédiction mensuelle que le modèle le plus performant selon le MAE est NNI-STL-log-ext, avec une erreur absolue moyenne de 280.58 unités. Ce dernier est suivi de près par le modèle NEI-STL-log-ext, qui affiche une performance plus faible de 3.05 unités. Selon le MAPE, les deux modèles s'échangent leur place et affichent une performance de 12.40% (pour NEI-STL-log-ext) et 12.46% respectivement. Nous remarquons encore une fois que l'écart entre les deux méthodes est très faible. Pour le score quantile, nos deux outils de références semblent ici aussi plus performants. SARIMA occupe pour le MAE la 8ème position avec un écart de performance avec NNI-STL-log-ext de 49.67 unités. Pour le MAPE, la méthode se trouve en 7ème place et s'écarte de la première de 3.56% (ce qui demeure plutôt faible). Nos modèles d'espace-

états se situent en dernier pour le MAE et le MAPE, tandis qu'ils sont en première position pour le score quantile.

Nous remarquons que globalement, les expérimentations qui sont les plus réussies sont celles qui impliquent l'usage des méthodes de décomposition STL-log avec extrapolation de la tendance. Cela nous semble peu surprenant, considérant que notre évaluation préliminaire des modèles nous indiquait que les modèles avec décomposition STL-log étaient les plus à-même de fournir de bonnes prédictions. D'autre part, les méthodes de décomposition STL permettent au schéma saisonnier de varier au cours du temps, ce qui rend les données décomposées plus proche de la réalité. Nous n'avons aussi pas de perte d'informations au début et à la fin de la période décomposée selon ladite méthode de décomposition. L'extrapolation de la tendance nous permet également davantage de saisir l'évolution de la demande, ce qui contribue à une meilleure performance. Le type de décomposition employée semble donc avoir un impact sur la qualité de nos prédictions.

Nous constatons que la différence entre la performance hebdomadaire et mensuelle est très importante. Le meilleur score associé aux prédictions hebdomadaires est pour le MAPE environ 2.65 fois plus élevé que pour les prévisions mensuelles. Le score quantile est également plus conséquent (1.42 fois). Le MAE nous semble meilleur pour les anticipations hebdomadaires, ce qui ne repose en réalité que sur une différence d'échelle. Les prédictions mensuelles sont réalisées à partir des données agrégées. Une erreur de 100 unités pour les prévisions du prochain mois est donc moindre comparativement à une erreur de 100 unités pour les prévisions de la semaine suivante. Ainsi, globalement, nous sommes meilleur pour prédire les quantités commandées à l'échelle mensuelle. Ce résultat est conforme à nos attentes. L'analyse des caractéristiques de notre série temporelle allait en ce sens. Nous estimons que cela est lié au fait que les quantités commandées varient certainement davantage à l'échelle hebdomadaire. Lorsque nous les agrégeons au mois, nous remarquons que la demande nous semble plus stable dans le temps, ce qui peut faciliter la prédiction. Tout cela repose selon nous sur une problématique que nous avions évoquée au chapitre précédent, à savoir la minimisation de l'incertitude associée à la saison et à la tendance. Cela serait cohérent avec le fait que l'usage d'une méthode de décomposition qui tient compte du caractère variable du schéma saisonnier nous conduise à une meilleure performance.

CHAPITRE 5 DISCUSSION

Nous avons au chapitre précédent exposé les résultats que nous avons obtenus à l'aide des modèles que nous avons construits. Nous les avons comparés avec les résultats que nous obtenons avec deux méthodes de référence. Les modèles ont été appliqués sur une famille de produit. Notre objectif était d'obtenir une distribution prédictive de la demande hebdomadaire et mensuelle par famille de produits en nous servant des données historiques de 2013 à 2018. Nous l'avons partiellement atteint. Nous avons obtenu des distributions prédictives pour la demande hebdomadaire et mensuelle, mais pour une seule famille de produits. Nous avons adapté notre façon de procéder de sorte à pouvoir traiter une demande qui était marquée par une saisonnalité relativement importante. Cependant, la méthode pourrait ne pas convenir à d'autres catégories de produits dont le comportement pourrait être plus chaotique, intermittent et/ou marqué par des bris structurels. La décomposition pourrait alors être plus difficile, voire inadaptée. Il sera donc nécessaire dans les études subséquentes d'évaluer la compatibilité de notre méthodologie avec les autres regroupements de produits. Si le comportement de la famille de produits que nous avons étudiée a changé depuis 2018 (ce qui serait tout à fait envisageable), il est même possible que notre méthode ne soit plus adaptée pour prédire sa demande.

D'autre part, bien que nos meilleures méthodes n'affichent pas une performance très éloignée de celle de SARIMA pour la demande hebdomadaire en termes de MAPE et de MAE, les prédictions sont insatisfaisantes. L'erreur de prévision nous semble en effet non négligeable. Nous n'avons cependant pas observé l'évolution des erreurs. C'est-à-dire que nous n'avons analysé que des erreurs moyennes. Il est possible que les prédictions pour un horizon n+1 soient meilleures que pour un horizon n+12, pour tous les modèles. Il serait intéressant de l'observer pour les prochaines analyses.

Si nous avons une idée des origines du manque d'efficacité de nos méthodes, nous n'avons pas quantifié combien chacune d'entre elles impactent nos résultats, isolément. Parmi les causes potentielles, nous avons identifié la minimisation de l'incertitude des composantes tendance-cycle et saisonnière ainsi qu'une modélisation inadéquate. Nous n'avons pas analysé combien le fait de traiter le bruit de nos données comme une variable aléatoire (et donc d'en caractériser l'incertitude) nous rapproche d'une meilleure performance. En d'autres termes, nous ignorons si nos résultats

actuels sont significativement différents des résultats que nous aurions obtenu à l'aide d'une méthode naïve saisonnière, ou non. Peut-être que d'approximer la demande au temps n+1 par la demande de l'année d'avant au même moment nous aurait offert des résultats peu différents de ceux que nous avons obtenus avec nos méthodes. Il est possible que la part la plus importante de l'incertitude de la demande réside dans les composantes saisonnière et tendancielle plutôt que dans la composante résiduelle. La tendance étant peu marquée, nous supposons cependant que l'incertitude serait davantage concentrée dans la composante saisonnière. Cependant, l'étude que nous avons menée ne nous permet pas de le savoir. Nous n'avons pas caractérisé l'incertitude relative aux composantes saisonnières et tendancielles et ignorons ainsi son impact sur nos résultats.

Aussi, nous n'avons pas tenté de modéliser la vraisemblance autrement que par une loi normale. Nous ignorons ainsi combien l'usage d'une loi de probabilité différente nous permettrait d'obtenir des échantillons plus représentatifs des données que nous pouvons observer (en particulier les plus extrêmes).

Par ailleurs, bien que nous ayons une idée de la qualité des distributions obtenues à l'aide du score quantile, cette métrique ne nous permet pas de les évaluer dans leur ensemble, ni d'évaluer la qualité des intervalles. De plus, nous n'avons calculé le score que pour un seul centile. Selon l'usage précis que souhaite faire l'entreprise des distributions et les aspects de la performance qu'elle souhaite analyser, il est possible que les métriques sélectionnées soient insuffisantes. L'évaluation de la performance de nos résultats pourrait donc à l'avenir faire l'objet d'améliorations.

Également, bien que nous ayons comparé nos méthodes à deux outils largement utilisés dans le domaine, nous estimons que d'autres méthodes pourraient être évaluées dans des travaux futurs. Il nous semble que cela est trop peu pour les situer par rapport à la multitude d'outils qui existent pour réaliser des prédictions sur des séries temporelles. Comparer leur performance à davantage de méthodes pourrait être envisagé dans des études subséquentes.

Finalement, bien que la qualité de nos simulations nous ait semblée suffisante pour poursuivre notre analyse, il est possible que l'usage d'une autre méthode que Métropolis aurait amélioré nos résultats. Nous n'avons pas considéré l'efficience des simulations dans nos critères de performance, alors qu'il s'agit d'un détail qui revêt une certaine importance en industrie.

CHAPITRE 6 CONCLUSION ET RECOMMANDATIONS

Logistik Unicorp, entreprise spécialisée dans la production de produits textiles, souhaite améliorer la gestion de ses approvisionnements. Cette dernière nous a donné pour mandat d'obtenir des distributions prédictives des demandes hebdomadaire et mensuelle de familles de produits pour un horizon de taille 12 à l'aide de modèles bayésiens. Après réception des données de l'entreprise, nous les avons nettoyées et organisées, de sorte à obtenir pour chaque famille de produits la demande hebdomadaire et mensuelle séparément. Nous avons sélectionné une famille de produits et l'avons statistiquement analysée entre les années 2013 et 2017. Nous avons constaté qu'elle était marquée par une saisonnalité relativement importante et une tendance plutôt discrète. Devant la volonté de l'entreprise de faire usage d'un modèle bayésien, nous avons entrepris de modéliser la demande. Cette dernière étant dépendante du temps, nous avons estimé qu'il nous était impossible de considérer les observations comme étant indépendantes les unes des autres dans notre modélisation descriptive. Pour obtenir des données indépendantes entre elles, nous avons décidé de retirer le cycle saisonnier de la demande et de modéliser les données résiduelles. Nous avons ainsi eu l'idée de prédire le bruit de la demande sur un horizon de taille 12, auquel nous incorporerions par la suite une estimation des composantes saisonnières et tendancielles associées. Ces composantes ont été estimées de deux façons différentes, la première étant en réalisant une moyenne de la saison et de la tendance des années précédentes, et l'autre en extrapolant la tendance. En parallèle et pour comparer notre méthode à celles qui existent, nous avons réalisé des prédictions sur un horizon de taille 12 avec un modèle SARIMA et un modèle d'espace état (méthode Holt-Winters). Nous avons décidé d'évaluer la performance des différentes méthodes à l'aide du MAE, du MAPE et du score quantile, pour un centile de 97.5%. En comparant nos meilleurs modèles avec ceux qui nous servent de références, nous avons remarqué que nous n'obtenions pas de résultats très différents. Tantôt un peu moins, tantôt un peu plus performants, nos modèles se sont en revanche montrés moins aptes à la prédiction lorsqu'évalués à l'aide du score quantile. Nos prédictions mensuelles sont plus précises que nos prédictions hebdomadaires, dont la demande affiche un comportement saisonnier plus incertain. Le manque de précision de nos résultats peut selon nous principalement s'expliquer par la minimisation de l'incertitude liée à la composante saisonnière et à une modélisation sous-optimale.

Pour améliorer nos prédictions, nous recommandons de modéliser la saison par une variable aléatoire, en exploitant les statistiques circulaires. Tel qu'exposé par Veatch et al. (2019), les distributions circulaires comme celle de Von Mises permettent de modéliser les données qui présentent un schéma qui se répète de façon périodique (comme des saisons). Il est possible qu'il faille utiliser un mélange de Von Mises, cette loi étant unimodale et n'étant pas tout à fait indiquée lorsque les données affichent plusieurs modes (comme c'est le cas pour notre schéma saisonnier). En réalisant une telle modélisation, il serait possible de mesurer l'incertitude associée à la composante saisonnière.

Par la suite, nous recommandons pour les données résiduelles hebdomadaires d'essayer de modéliser la vraisemblance par une loi de Student, pour accorder une probabilité plus importante aux valeurs plus éloignées de la moyenne. Cela pourrait mieux représenter les données observées, qui semblaient comporter des ailes/queues plus étendues de part et d'autre. Les données observées présentant une certaine asymétrie, il serait également peut-être envisageable de modéliser la vraisemblance par une loi Normale asymétrique. Pour ce qui est des données mensuelles, un mélange de loi normale pourrait peut-être offrir une meilleure représentation des données, qui semblaient bidimensionnelles.

Aussi, nous recommandons d'évaluer la possibilité de traiter la composante résiduelle de telle sorte à ce que les résidus soient parfaitement indépendants et identiquement distribués.

Finalement, nous recommandons de mener une analyse de la performance de nos résultats en faisant usage de métriques différentes, pour l'évaluer sous d'autres angles.

RÉFÉRENCES

Adjengue, L. (2016). *Méthodes statistiques : Concepts, applications et exercices*. Presses Internationales Polytechnique. (530p)

Alhasawi, E., Hajli, M., & Dennehy, D. (2023). A Review of Artificial Intelligence (AI) and Machine Learning (ML) for Supply Chain Resilience: Preliminary Findings. *2023 IEEE International Symposium on Technology and Society (ISTAS)*, 1-8.

arviz.loo — ArviZ 0.20.0 documentation. (2018). Arviz.org. https://python.arviz.org/en/stable/api/generated/arviz.loo.html

Babai, M.Z., Boylan, J.E., & Tabar, B.R. (2021). Demand forecasting in supply chains: a review of aggregation and hierarchical approaches. *International Journal of Production Research, 60*, 324 - 348.

Benziane, B.A., Lardeux, B., Jridi, M., & Mcharek, A. (2023). Supply Chain Forecasting in a Fast Moving Global Economy: Review, Limits and Future Directions. *2023 28th International Conference on Automation and Computing (ICAC)*, 1-6.

Calatayud, A., Mangan, J. and Christopher, M. (2019), "The self-thinking supply chain", *SupplyChain Managemen*t, Vol. 24 No. 1, pp. 22-38. https://doi.org/10.1108/SCM-03-2018-0136

Champagne, S. (2022). Sociétés les mieux gérées: Logistik Unicorp : des uniformes d'ici…partout dans le monde. Parut le 10 Mai 2022 dans *La Presse*. https://www.lapresse.ca/affaires/portfolio/2022-05-10/societes-les-mieux-gerees/logistik-unicorp-des-uniformes-d-ici-partout-dans-le-monde.php

Davidson-Pilon, C. (2016). *Bayesian methods for hackers : probabilistic programming and bayesian inference*. Addison-Wesley. (225p)

Department of Economic and Social Affairs, Population Division. (2022). *World Population Prospects 2022: Summary of Results.* United Nations. https://www.un.org/development/desa/pd/content/World-Population-Prospects-2022

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., & Rubin, D.B. (2021). *Bayesian Data Analysis* (3e éd.). CRC Press. (667p)

Guéron, A. (2019a). *Deep Learning avec Keras et TensorFlow : mise en œuvre et cas concret* (2e éd). (553p)

Guéron, A. (2019b). *Machine learning avec Scikit-Learn : mise en œuvre et cas concret* (2e éd.). Dunod. (297p)

Ghouati, S., El Amri, A., & Oulfarsi, S. (2022). The impact of Artificial Intelligence on Supply Chain: literature review and conceptual framework. *2022 14th International Colloquium of Logistics and Supply Chain Management (LOGISTIQUA)*, 1-6.

Hammarberg, R., Highfield, L., Walton, G., Wermuth, P., & Bowman, Ann. (2023). "Hot cities" and their experiences and planning for rapid growth: An analysis of strategic planning documents. *IET Smart Cities*. Volume 5, no 3, pages 220-229. https://doi.org/10.1049/smc2.12061

Hyndman, R. J., & Athanasopoulos, G. (2021a). *ARIMA Modelling in fable* [Vidéo]. OTexts. https://otexts.com/fpp3/arima-r.html

Hyndman, R. J., & Athanasopoulos, G. (2021b). *Autoregressive models* [Vidéo]. OTexts. https://otexts.com/fpp3/AR.html

Hyndman, R. J., & Athanasopoulos, G. (2021c). *Distributional forecasts and prediction intervals* [Vidéo]. OTexts. https://otexts.com/fpp3/prediction-intervals.html

Hyndman, R. J., & Athanasopoulos, G. (2021d). *Evaluating distributional forecast accuracy* [Vidéo]. OTexts. https://otexts.com/fpp3/distaccuracy.html

Hyndman, R. J., & Athanasopoulos, G. (2021e). *Evaluating point forecast accuracy* [Vidéo].OTexts. https://otexts.com/fpp3/accuracy.html

Hyndman, R. J., & Athanasopoulos, G. (2021f). *Forecasting Principles and Practice* (3e éd.). OTexts. (440p)

Hyndman, R. J., & Athanasopoulos, G. (2021g). *Forecasting with ETS models* [Vidéo]. OTexts.https://otexts.com/fpp3/ets-forecasting.html

Hyndman, R. J., & Athanasopoulos, G. (2021h). *Innovations state space models for exponentialSmoothing* [Vidéo]. OTexts. https://otexts.com/fpp3/ets.html

Hyndman, R. J., & Athanasopoulos, G. (2021i). *Residual diagnostics for forecasting models*[Vidéo]. OTexts. https://otexts.com/fpp3/diagnostics.html

Hyndman, R. J., & Athanasopoulos, G. (2021j). *Seasonal ARIMA model* [Vidéo]. OTexts. https://otexts.com/fpp3/seasonal-arima.html

Ingle, C., Bakliwal, D., Jain, J., Singh, P., Kale, P.A., & Chhajed, V. (2021). Demand Forecasting: Literature Review On Various Methodologies. *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1-7.

Jalbert, J. (2023). [Notes de cours]. Moodle@Polytechnique Montréal. https://moodle.polymtl.ca/

Johnson, A., Ott, M.Q., & Dogucu, M. (2021). *Bayes rules! – An Introduction to Applied Bayesian Modeling*. CRC Press. https://www.bayesrulesbook.com

Kenney, J.F., & Keeping, E.S. (1951). *Mathematics of statistics* (2e éd.). Volume 2. Van Nostrand Company. (429p)

Koutsandreas, D., Spiliotis, E., Petropoulos, F., & Assimakopoulos, V. (2021). On the selection of forecasting accuracy measures. *Journal of the Operational Research Society*, *73*(5), 937–954. https://doi.org/10.1080/01605682.2021.1892464

Kruschke, J. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. *Academic Press.*

Logistik Unicorp. (2024). *Clients et Divisions*. Logistik Unicorps. Consulté le 01 Juin 2024 https://www.logistikunicorp.com/fr/default.asp?Lang=fr&PageName=clients

Makridakis, S., Hyndman, R.J., & Petropoulos, F. (2020). Forecasting in social settings: The state of the art. *International Journal of Forecasting*. Volume 36, Issue 1, pages 15 à 28.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE, 13*.

Makridakis, S., Spiliotis, E., Assimakopoulos, V., Semenoglou, A. A., Mulder, G., & Nikolopoulos, K. (2022). Statistical, machine learning and deep learning forecasting

methods: Comparisons and ways forward. *Journal of the Operational Research Society*, *74*(3), 840–859. https://doi.org/10.1080/01605682.2022.2118629

McElreath, R., (2020). *Statistical Rethinking : A Bayesian Course with Examples in R and Stan* (2e éd.). CRC Press. (593p)

Petropoulos, F., et al. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, Volume 38, Issue 3, Pages 705-871, ISSN 0169-2070, https://doi.org/10.1016/j.ijforecast.2021.11.001.

Ross, K., (2022). *An Introduction to Bayesian Reasoning and Methods*. https://bookdown.org/kevin_davisross/bayesian-reasoning-and-methods/

SAS Institute Inc. (2017). *SAS/STAT® 14.3 User's Guide: The MCMC Procedure* (Version 14.3). SAS Institute Inc. https://support.sas.com/documentation/onlinedoc/stat/143/mcmc.pdf

Stan Development Team. (2025). *Glossary: p_loo (effective number of parameters)*. Stan Project. Consulté le 12 Février 2025. https://mc-stan.org/loo/reference/loo-glossary.html#:~:text=p_loo%20(effective%20number%20of%20parameters)&text=In%20well%20behaving%20cases%20p_loo,indicate%20a%20severe%20model%20misspecification.

Tyralis, H., Papacharalampous, G. (2024) A review of predictive uncertainty estimation with machine learning. *Artif Intell Rev* **57**, 94. https://doi.org/10.1007/s10462-023-10698-8

Veatch, W., Villarini, G. (2020) Modeling the seasonality of extreme coastal water levels with mixturesof circular probability density functions. *Theor Appl Climatol* **140**, 1199–1206. https://doi.org/10.1007/s00704-020-03143-1

Vehtari, A., Gelman, A., and Gabry, J. (2017a). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*. Volume 27, pages 1413-1432. https://doi.org/10.1007/s11222-016-9696-4

Vehtari, A., Simpson, D., Gelman, A., Yao, Y., & Gabry, J. (2024). Pareto Smoothed Importance Sampling. *Journal of Machine Learning Research*, Volume 25, pages 1-58, https://doi.org/10.48550/arXiv.1507.02646.

Zhongxu, Z., Ying, P., Jing, Z., Junxi, W., & Ran, Z. (2022). The Impact of Urbanization on the Delivery of Public Service–Related SDGs in China, Sustainable Cities and Society (Volume 80). https://doi.org/10.1016/j.scs.2022.103776.

ANNEXE A – VISUELS ET CODES

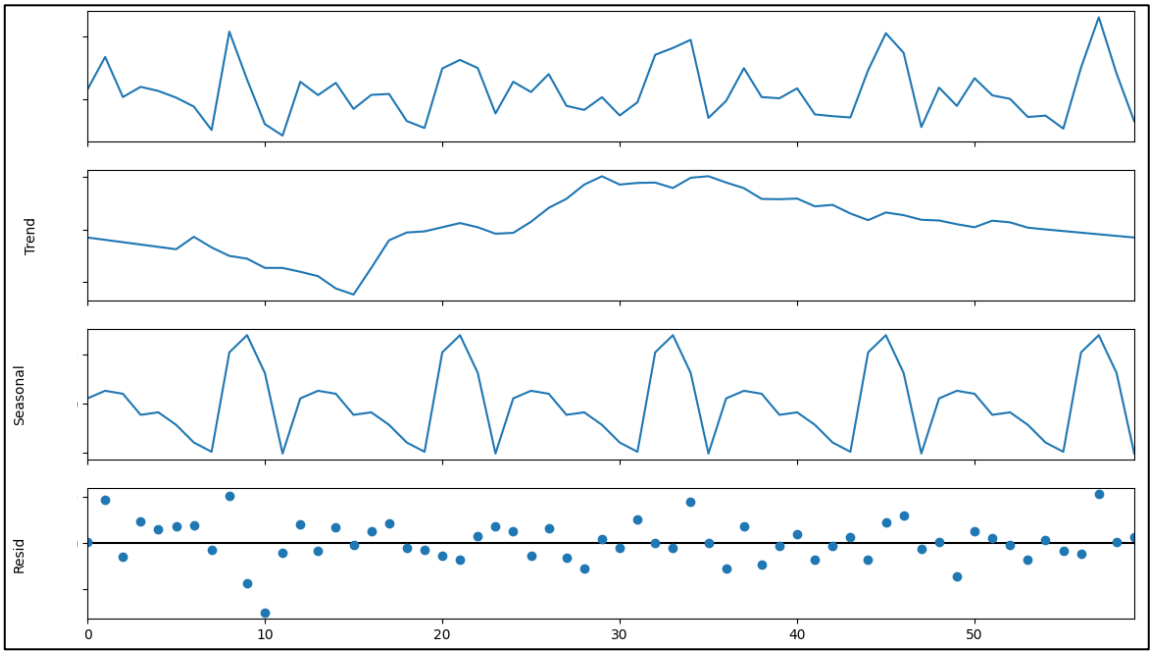


Figure A.1 : Visuel de la décomposition additive obtenue sur python pour la demande mensuelle

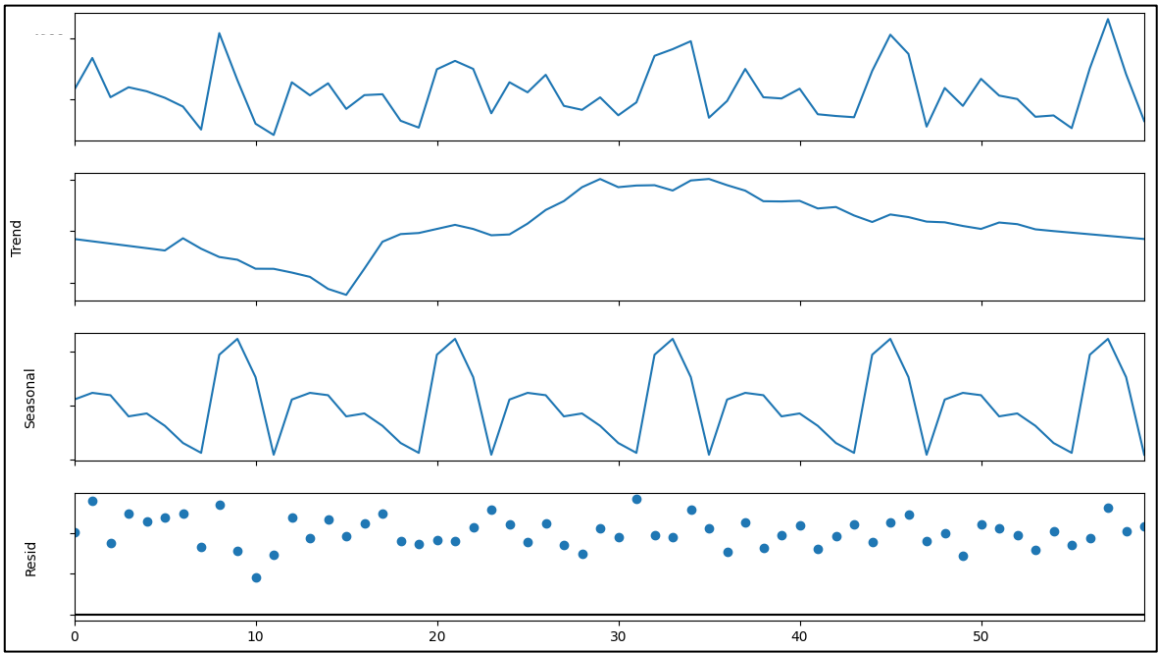


Figure A.2 : Visuel de la décomposition multiplicative obtenue sur python pour la demande mensuelle

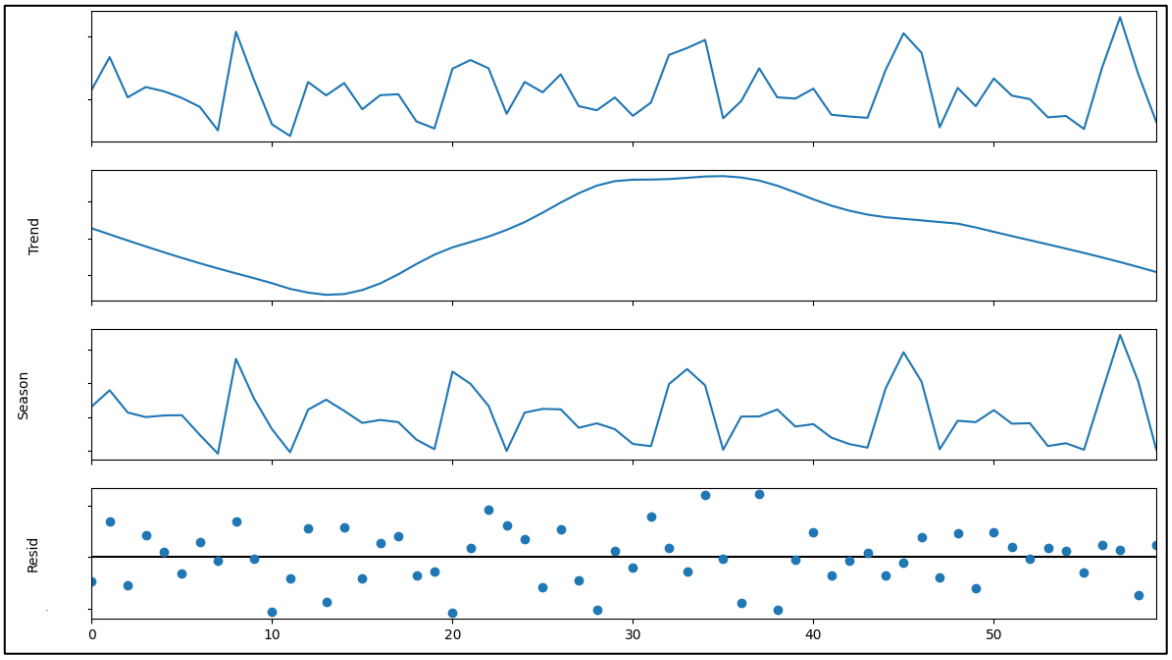


Figure A.3 : Visuel de la décomposition STL obtenue sur python pour la demande mensuelle

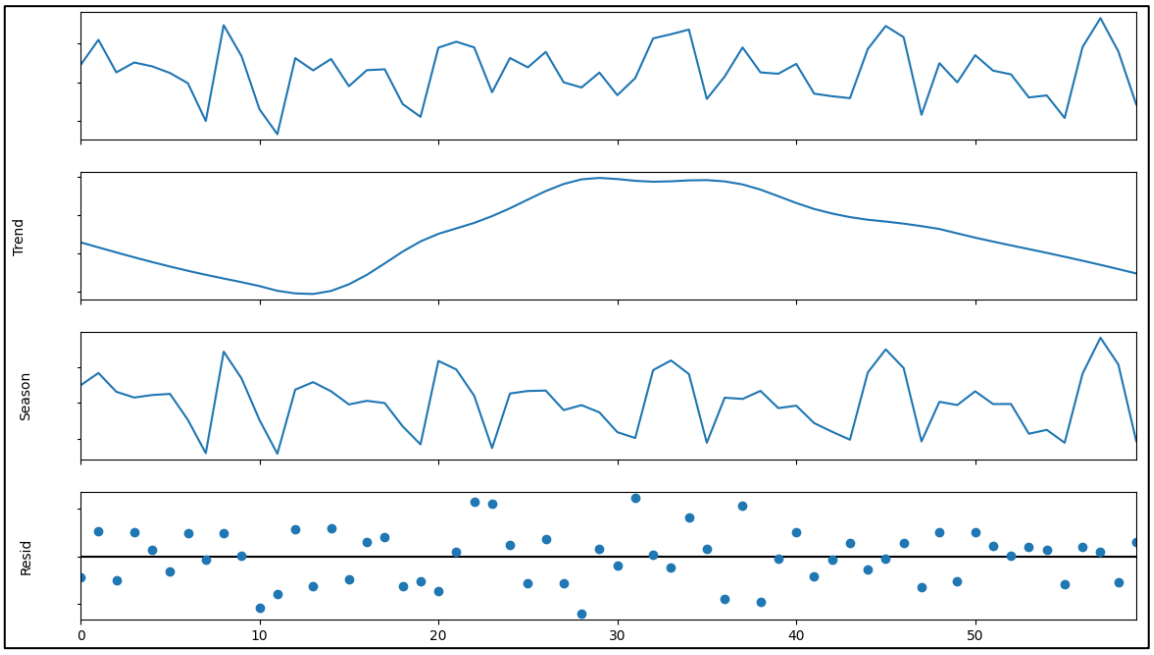


Figure A.4 : Visuel de la décomposition STL du logarithme de la demande pour la demande mensuelle

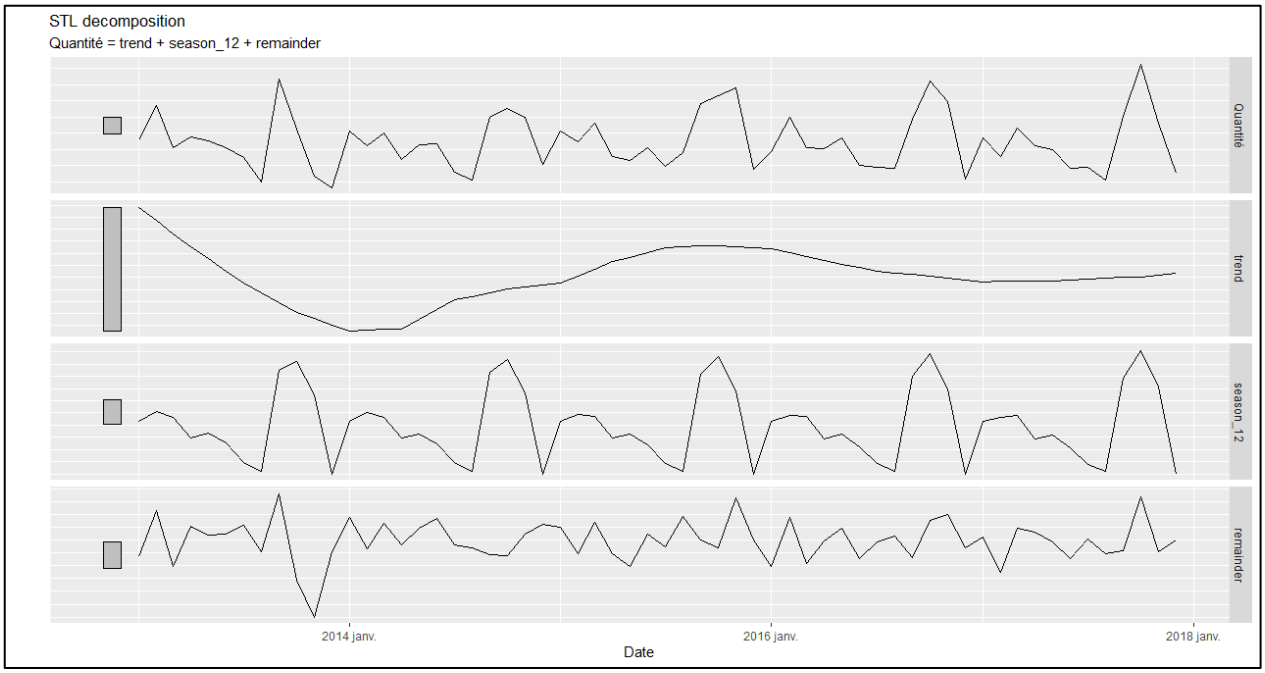


Figure A.5 : Visuel de la décomposition STL réalisée sur R pour la demande mensuelle

```
#Modèles d'espace-état usant de la méthode Holt-Winters

library(fpp3)
library(readr)

data <- read_csv("chemin/d/accès/fichier-demande-mensuelle-famille-test.csv")
tsib <- as_tsibble(data, index = Date)
tsib <- tsib %>%
  mutate(Date = yearmonth(Date)) %>%
  as_tsibble(index = Date)

#Ajustement
Qte_fit <- tsib %>%
  model(
    add_season_add_error_trend = ETS(Quantité ~ error("A") + trend("A") + season("A", period = "1 year")),
    mult_season_mult_error_trend = ETS(Quantité ~ error("M") + trend("A") + season("M", period = "1 year")),
    add_error_mult_season_trend = ETS(Quantité ~ error("A") + trend("A") + season("M", period = "1 year")),
    mult_error_add_season_trend = ETS(Quantité ~ error("M") + trend("A") + season("A", period = "1 year")),
    add_season_add_error = ETS(Quantité ~ error("A") + trend("N") + season("A", period = "1 year")),
    mult_season_mult_error = ETS(Quantité ~ error("M") + trend("N") + season("M", period = "1 year")),
    add_error_mult_season = ETS(Quantité ~ error("A") + trend("N") + season("M", period = "1 year")),
    mult_error_add_season = ETS(Quantité ~ error("M") + trend("N") + season("A", period = "1 year"))
    )

tidy(Qte_fit)
components(Qte_fit) %>%
  autoplot()

#Analyse des résidus
print(augment(Qte_fit), n = 500)
gg_tsresiduals(Qte_fit %>% select(mult_season_mult_error))
augment(Qte_fit) |>
  features(.innov, ljung_box, lag = 24)

#Prédiction
Qte_forecast <- tsib %>%
  model(
    mult_season_mult_error = ETS(Quantité ~ error("M") + trend("N") + season("M", period = "1 year"))
    ) %>%
  forecast(h = "12 months") %>%
  hilo(95)%>%
  mutate(lower=`95%`$lower, upper=`95%`$upper)
  autoplot(Qte_forecast) + labs(y = "Unités commandées", title = "Commandes du produit test")
print(Qte_forecast, n = 1000)
autoplot(Qte_forecast) +
geom_ribbon(aes(ymin = lower, ymax = upper), fill = "blue", alpha = 0.2) +
labs(y = "Unités commandées", title = "Commandes du produit test")

autoplot(Qte_forecast) +
  geom_line(data = tsib, aes(x = Date, y = Quantité), color = "black", size = 0.5) + # Add historical data
  geom_ribbon(aes(ymin = lower, ymax = upper), fill = "blue", alpha = 0.2) + # Forecast intervals
  labs(
    y = "Unités commandées",
    title = "Commandes du produit test",
    x = "Date"
  ) +
  guides(colour = guide_legend(title = "Forecast"))
```

Figure A.6 : Code des modèles d'espaces-état

```
#Sarima pour la demande hebdomadaire

library(fpp3)
library(readr)

data <- read_csv("chemin/d/acc`s/fichier-demande-hebdomadaire-famille-test.csv")
tsib <- as_tsibble(data, index = Date)
tsib <- tsib %>%
  mutate(Date = yearweek(Date)) %>%
  as_tsibble(index = Date)

#Ajustement
Qte_fit <- tsib %>%
  model(stepwise = ARIMA(Quantité),
        search = ARIMA(Quantité, stepwise=FALSE, approx = FALSE))
tidy(Qte_fit)

#Analyse des résidus
gg_tsresiduals(Qte_fit %>% select(search))
augment(Qte_fit) |>
  features(.innov, ljung_box, lag = 104)

#Prédiction
Qte_forecast <- tsib %>%
  model(search = ARIMA(Quantité, stepwise=FALSE, approx = FALSE)) %>%
  forecast(h = "12 weeks")%>%
  hilo(95)%>%
  mutate(lower=`95%`$lower, upper=`95%`$upper)
  autoplot(Qte_forecast) + labs(y = "Unités commandées", title = "Commandes du produit test")
print(Qte_forecast, n = 1000)
autoplot(Qte_forecast) +
geom_ribbon(aes(ymin = lower, ymax = upper), fill = "blue", alpha = 0.2) +
labs(y = "Unités commandées", title = "Commandes du produit test")

autoplot(Qte_forecast) +
  geom_line(data = tsib, aes(x = Date, y = Quantité), color = "black", size = 0.5) + # Add historical data
  geom_ribbon(aes(ymin = lower, ymax = upper), fill = "blue", alpha = 0.2) + # Forecast intervals
  labs(
    y = "Unités commandées",
    title = "Commandes du produit test",
    x = "Date"
  ) +
  guides(colour = guide_legend(title = "Forecast"))
```

Figure A.7 : Code de SARIMA pour la demande hebdomadaire

```
# A tibble: 9 × 6
  .model   term  estimate std.error statistic  p.value
  <chr>    <chr>    <dbl>     <dbl>     <dbl>    <dbl>
1 stepwise ar1     0.268      0.121     2.23  2.70e- 2
2 stepwise ar2    -0.786      0.133    -5.93  1.26e- 8
3 stepwise ma1    -0.176      0.105    -1.67  9.65e- 2
4 stepwise ma2     0.862      0.107     8.08  5.10e-14
5 stepwise sar1   -0.0688     0.210    -0.327 7.44e- 1
6 stepwise sma1   -0.447      0.223    -2.00  4.66e- 2
7 search   ar1     0.657      0.159     4.14  5.00e- 5
8 search   ma1    -0.446      0.187    -2.39  1.78e- 2
9 search   sma1   -0.588      0.113    -5.20  4.83e- 7
```

Figure A.8 : Paramètres des modèles ARIMA proposés pour la demande hebdomadaire

```
import pymc3 as pm
import matplotlib.pyplot as plt
import numpy as np
import arviz as az
import pandas as pd

df = pd.read_csv('données-désaisonnalisées.csv')
df['Date'] = pd.to_datetime(df['Date'])
df_2013_m = df[df['Date'] < '2014-01-06']
df_2014_m = df[(df['Date'] >= '2014-01-06') & (df['Date'] <= '2014-12-29')]
df_2015_m = df[(df['Date'] > '2014-12-29') & (df['Date'] <= '2015-12-28')]
df_2016_2017_m = df[df['Date'] > '2015-12-28']
demand_data = df_2016_2017_m['Résidu'].tolist()
print(len(demand_data))

mean_ref_data = np.mean(df_2013_m['Résidu'])
var_ref_data = np.var(df_2013_m['Résidu'])
print("2013 mean : ", mean_ref_data)
print("2013 variance : ", var_ref_data)

mean_ref_data1 = np.mean(df_2014_m['Résidu'])
var_ref_data1 = np.var(df_2014_m['Résidu'])
print("2014 mean : ", mean_ref_data1)
print("2014 variance : ", var_ref_data1)

mean_ref_data2 = np.mean(df_2015_m['Résidu'])
var_ref_data2 = np.var(df_2015_m['Résidu'])
print("2015 mean : ", mean_ref_data2)
print("2015 variance : ", var_ref_data2)

#mu parameters
data_ref = [mean_ref_data, mean_ref_data1, mean_ref_data2]
mean_ref_final = np.mean(data_ref)
print("Nu : ", mean_ref_final)
TauSquare_M1 = np.var(data_ref)
print("Tau : ", TauSquare_M1)

#sigSquare_M1 parameters
data_ref1 = [var_ref_data, var_ref_data1, var_ref_data2]
k = np.mean(data_ref1)
print("k : ", k)
q = np.var(data_ref1)
print("q : ", q)
alphaP = (((k**2)/q)+2)
print("Alpha parameter for InvG : ", alphaP)
BetaP = k*(alphaP-1)
print("Beta parameter for InvG : ", BetaP)

if __name__ == '__main__':
    with pm.Model() as model:
        nu_M1 = np.mean(data_ref)
        TauSquare_M1 = np.var(data_ref)
        mu_M1 = pm.Normal("mu_M1", nu_M1, sigma = TauSquare_M1**0.5)
        sigSquare_M1 = pm.InverseGamma("sigSquare_M1", alpha = alphaP, beta = BetaP)
        sigma = pm.Deterministic("sigma", pm.math.sqrt(sigSquare_M1))
        likelihood = pm.Normal("Likelihood", mu_M1, sigma = sigma, observed = demand_data)
        step = pm.Metropolis() #based on the posterior likelihood*prior...
        trace = pm.sample(20000, tune=10000, step=step, return_inferencedata=False, cores=3, chains=3)

        print(pm.summary(trace))
        az.plot_trace(trace)
        plt.show()
        loo_result = az.loo(trace)
        print(loo_result)
        print(loo_result.pareto_k)

        mu_M1_samples = trace['mu_M1']
        sigSquare_M1_samples = trace['sigSquare_M1']
        sigma_samples = trace['sigma']

    ############ Predictive Distribution processing #############
    with model:
        post_pred = pm.sample_posterior_predictive(trace)
```

Figure A.9 : Code des simulations pour les modèles Normale-Normale-Inverse Gamma

```
import pymc3 as pm
import matplotlib.pyplot as plt
import numpy as np
import arviz as az
import pandas as pd

df = pd.read_csv('données_désaisonnalisées.csv')
df['Date'] = pd.to_datetime(df['Date'])
df_2013_m = df[df['Date'] < '2014-01-06']
df_2014_m = df[(df['Date'] >= '2014-01-06') & (df['Date'] <= '2014-12-29')]
df_2015_m = df[(df['Date'] > '2014-12-29') & (df['Date'] <= '2015-12-28')]
df_2016_2017_m = df[df['Date'] > '2015-12-28']
demand_data = df_2016_2017_m['Résidu'].tolist()

mean_ref_data = np.mean(df_2013_m['Résidu'])
var_ref_data = np.var(df_2013_m['Résidu'])
print("2013 mean : ", mean_ref_data)
print("2013 variance : ", var_ref_data)

mean_ref_data1 = np.mean(df_2014_m['Résidu'])
var_ref_data1 = np.var(df_2014_m['Résidu'])
print("2014 mean : ", mean_ref_data1)
print("2014 variance : ", var_ref_data1)

mean_ref_data2 = np.mean(df_2015_m['Résidu'])
var_ref_data2 = np.var(df_2015_m['Résidu'])
print("2015 mean : ", mean_ref_data2)
print("2015 variance : ", var_ref_data2)

#mu parameters
data_ref = [mean_ref_data, mean_ref_data1, mean_ref_data2]
mean_ref_final = np.mean(data_ref)
print("Nu : ", mean_ref_final)

#sigSquare_M1 parameters
data_ref1 = [var_ref_data, var_ref_data1, var_ref_data2]
k = np.mean(data_ref1)
print("k : ", k)
q = np.var(data_ref1)
print("q : ", q)
alphaP = (((k**2)/q)+2)
print("Alpha parameter for InvG : ", alphaP)
BetaP = k*(alphaP-1)
print("Beta parameter for InvG : ", BetaP)

if __name__ == '__main__':
    with pm.Model() as model:
        gamma_M1 = 1/mean_ref_final
        mu_M1 = pm.Exponential("mu_M1", gamma_M1)
        sigSquare_M1 = pm.InverseGamma("sigSquare_M1", alpha = alphaP, beta = BetaP)
        sigma = pm.Deterministic("sigma", pm.math.sqrt(sigSquare_M1))
        likelihood = pm.Normal("Likelihood", mu = mu_M1, sigma = sigma, observed = demand_data)
        step = pm.Metropolis() #based on the posterior likelihood*prior...
        trace = pm.sample(20000, tune=10000, step=step, return_inferencedata=False, cores=3, chains=3)

        print(pm.summary(trace))
        az.plot_trace(trace)
        plt.show()
        loo_result = az.loo(trace)
        print(loo_result)
        print(loo_result.pareto_k)

    mu_M1_samples = trace['mu_M1']
    sigSquare_M1_samples = trace['sigSquare_M1']
    sigma_samples = trace['sigma']

############ Predictive Distribution processing #############
    with model:
        post_pred = pm.sample_posterior_predictive(trace)
```

Figure A.10 : Code des simulations pour les modèles Normale-Exponentielle-Inverse Gamma

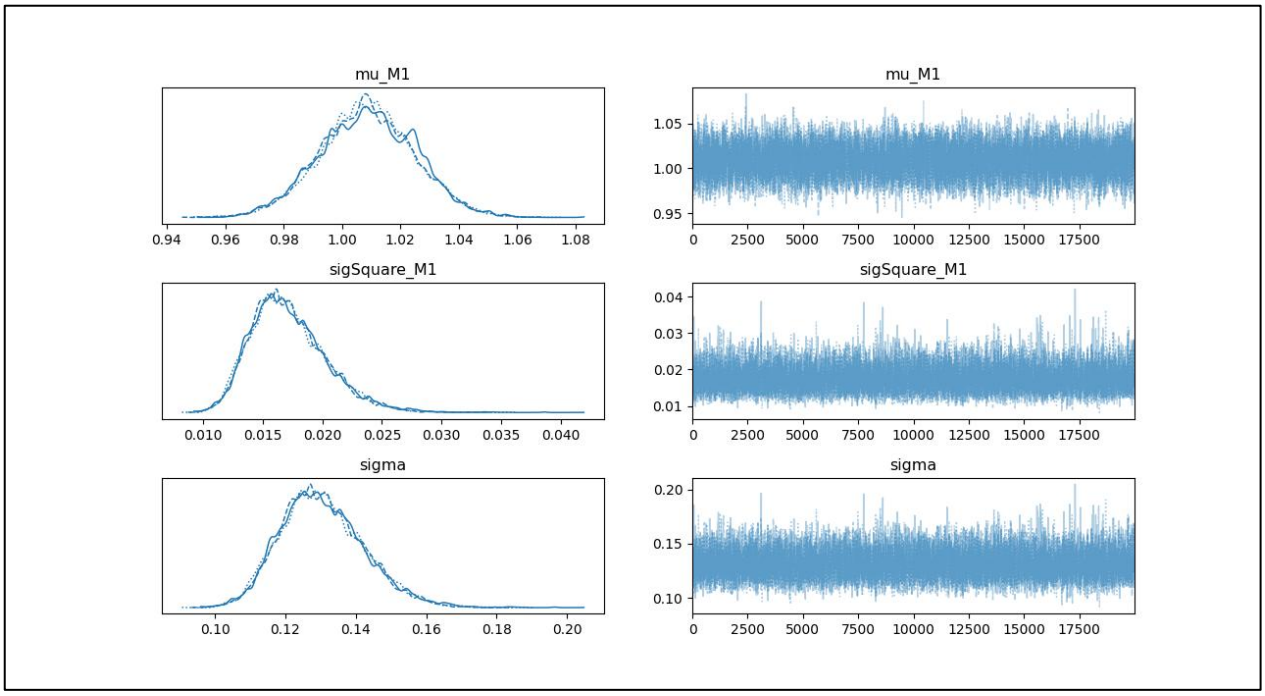


Figure A.11 : Exemple de tracés de chaînes de Markov

ANNEXE B – CARACTÉRISTIQUES DES DIFFÉRENTS MODÈLES

Tableau B.1 : Caractéristiques des différents modèles

Expérience	Vraisemblance	*A priori*	Loi *a posteriori*	Distribution prédictive	Données historiques	Décomposition
PE	$f(Y)_{(y\|\lambda)} = \prod_{i=1}^{n} Poi(y_i\|\lambda)$	$f(\lambda)_{\lambda} = Exp(\lambda\|\gamma)$	$f(\lambda)_{\lambda\|Y=y} = Gamma(n\bar{y}+1, n+\gamma)$	$f(y_{n+1})_{y_{n+1}\|(Y=y,\ \lambda)} = \mathfrak{BN}\left(n\bar{y}+1, \frac{n+\gamma}{n+\gamma+1}\right)$.	2013-2017	-
NNI-add	$f(Y)_{(y\|\mu,\sigma^2)} = \prod_{i=1}^{n} N(y_i\|\mu,\sigma^2)$	$f(\mu,\sigma^2)_{\mu,\sigma^2} = N(\mu\|\upsilon,\sigma^2) \times IG(\sigma^2\|\alpha,\beta)$	Approximation par MCMC	Approximation par MCMC	2013-2017	Additive classique
NNI-mul	$f(Y)_{(y\|\mu,\sigma^2)} = \prod_{i=1}^{n} N(y_i\|\mu,\sigma^2)$	$f(\mu,\sigma^2)_{\mu,\sigma^2} = N(\mu\|\upsilon,\sigma^2) \times IG(\sigma^2\|\alpha,\beta)$	Approximation par MCMC	Approximation par MCMC	2013-2017	Multiplicative classique
NNI-STL-log	$f(Y)_{(y\|\mu,\sigma^2)} = \prod_{i=1}^{n} N(y_i\|\mu,\sigma^2)$	$f(\mu,\sigma^2)_{\mu,\sigma^2} = N(\mu\|\upsilon,\sigma^2) \times IG(\sigma^2\|\alpha,\beta)$	Approximation par MCMC	Approximation par MCMC	2013-2017	STL-log
NNI-STL	$f(Y)_{(y\|\mu,\sigma^2)} = \prod_{i=1}^{n} N(y_i\|\mu,\sigma^2)$	$f(\mu,\sigma^2)_{\mu,\sigma^2} = N(\mu\|\upsilon,\sigma^2) \times IG(\sigma^2\|\alpha,\beta)$	Approximation par MCMC	Approximation par MCMC	2013-2017	STL
NEI-mul	$f(Y)_{(y\|\mu,\sigma^2)} = \prod_{i=1}^{n} N(y_i\|\mu,\sigma^2)$	$f(\mu,\sigma^2)_{\mu,\sigma^2} = Exp(\mu\|\gamma) \times IG(\sigma^2\|\alpha,\beta)$	Approximation par MCMC	Approximation par MCMC	2013-2017	Multiplicative classique
NEI-STL-log	$f(Y)_{(y\|\mu,\sigma^2)} = \prod_{i=1}^{n} N(y_i\|\mu,\sigma^2)$	$f(\mu,\sigma^2)_{\mu,\sigma^2} = Exp(\mu\|\gamma) \times IG(\sigma^2\|\alpha,\beta)$	Approximation par MCMC	Approximation par MCMC	2013-2017	STL-log

```
# A tibble: 17 × 9
   term   add_error_mult_season add_error_mult_season_trend add_season_add_error add_season_add_error_trend mult_error_add_season
   <chr>                  <dbl>                        <dbl>                <dbl>                       <dbl>                  <dbl>
 1 alpha                0.00102                     0.000566             0.000100                     0.00569               0.000100
 2 b[0]                NA                           0.855               NA                           -0.936                NA
 3 beta                NA                           0.000260            NA                            0.000100             NA
 4 gamma                0.000212                    0.000101             0.000104                     0.000552              0.000101
 5 l[0]              2228.                       2307.                2245.                        2330.                 2231.
 6 s[-1]                1.27                        1.27                 696.                         674.                  714.
 7 s[-10]               1.10                        1.10                 144.                         147.                  191.
 8 s[-11]               1.03                        1.03                  16.4                        -19.5                  -3.73
 9 s[-2]                1.61                        1.60                1262.                        1317.                 1187.
10 s[-3]                1.51                        1.45                1153.                        1157.                 1158.
11 s[-4]                0.569                       0.585               -911.                        -887.                 -903.
12 s[-5]                0.647                       0.679               -747.                        -754.                 -754.
13 s[-6]                0.836                       0.818               -460.                        -468.                 -450.
14 s[-7]                0.907                       0.922               -191.                        -197.                 -197.
15 s[-8]                0.872                       0.893               -286.                        -292.                 -275.
16 s[-9]                1.10                        1.09                 317.                        320.                   324.
17 s[0]                 0.553                       0.556               -993.                        -996.                 -991.

   mult_error_add_season_trend mult_season_mult_error mult_season_mult_error_trend
                         <dbl>                  <dbl>                       <dbl>
 1                    0.000127               0.000103                    0.00117
 2                   -3.16                   NA                          0.517
 3                    0.000100               NA                          0.00108
 4                    0.000118               0.000106                    0.000105
 5                 2316.                  2267.                       2307.
 6                  678.                     1.36                        1.40
 7                   87.1                    1.13                        1.08
 8                  -68.2                    1.03                        1.01
 9                 1195.                     1.60                        1.62
10                 1158.                     1.49                        1.44
11                 -772.                     0.583                       0.591
12                 -754.                     0.645                       0.636
13                 -467.                     0.803                       0.815
14                 -199.                     0.922                       0.919
15                 -292.                     0.850                       0.858
16                  330.                     1.06                        1.09
17                 -895.                     0.541                       0.539
```

Figure B.1 : Paramètres estimés pour les modèles d’espace-état