



HAL
open science

Modélisation probabiliste de la dynamique de transmission de la fièvre typhoïde à Mayotte avec étude de risques épidémiques

Ibrahim Bouzalmat

► **To cite this version:**

Ibrahim Bouzalmat. Modélisation probabiliste de la dynamique de transmission de la fièvre typhoïde à Mayotte avec étude de risques épidémiques. Statistiques [math.ST]. Université de Montpellier, 2023. Français. NNT: . tel-04439679

HAL Id: tel-04439679

<https://hal.science/tel-04439679v1>

Submitted on 5 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biostatistiques

École doctorale : Information, Structures, Systèmes

Unité de recherche : Institut Montpelliérain Alexander Grothendieck

Modélisation probabiliste de la dynamique de transmission de la fièvre typhoïde à Mayotte avec étude de risques épidémiques

Présentée par Ibrahim BOUZALMAT
le 15/11/2023

Sous la direction de Benoîte DE SAPORTA
et Solym MANOU-ABI

Devant le jury composé de

Benoîte De Saporta

Marie-Pierre Etienne

Viet Chi Tran

Nathalie Peyrard

Benoît Liquet

Manon Costa

Bertrand Cloez

Solym Mawaki Manou-Abi

Professeur, Université de Montpellier

Maître de conférences, Institut Agro Rennes-Angers

Professeur, Université Gustave Eiffel

Directeur de recherche, INRAE

Professor, UPPA

Maître de conférences, Université Toulouse 3 - Paul Sabatier

Chargé de recherche, INRAE Montpellier

Maître de conférences, CUFR de Mayotte

Directrice de thèse

Rapporteur

Rapporteur

Présidente

Examineur

Examinatrice

Examineur

Co-encadrant de thèse



UNIVERSITÉ
DE MONTPELLIER

À l'âme de mon père.

"L'âme de mon père réside en moi, tissant un fil invisible qui me guide, m'inspire et m'accompagne tout au long de ma vie."

Remerciements

Je tiens à remercier toutes les personnes qui ont contribué peu ou prou au succès de cette thèse, et je tiens à saisir cette occasion unique pour les remercier.

En premier lieu, je tiens à remercier profondément ma directrice de thèse, Benoîte De Saporta. Son remarquable sens de l'organisation a été une source d'inspiration. Elle a guidé de près mon travail de recherche sur trois années, apportant constamment clarté et rigueur à nos échanges. Sa capacité à maintenir un haut niveau d'organisation dans notre collaboration a été essentielle pour le succès de ma thèse. De plus, je tiens à souligner que ma directrice est non seulement bien organisée, mais aussi d'une gentillesse exceptionnelle. Elle a toujours pris le temps de m'écouter, de répondre à mes questions et de m'apporter son soutien, peu importe la situation. Sa gentillesse naturelle a créé un environnement de travail agréable et motivant, favorisant ma progression dans mes recherches. Je voudrais également remercier mon co-encadrant de thèse, Solym Manou-Abi, qu'a fait preuve d'un dévouement remarquable tout au long de cette collaboration, en trouvant toujours du temps pour nos réunions malgré la distance et les contraintes qui se sont présentées. Malgré les barrières géographiques qui nous séparaient, mon co-encadrant a su maintenir un suivi régulier de notre travail en organisant des réunions à distance. Sa disponibilité et son engagement ont été essentiels pour le succès de notre projet de recherche.

Je tiens ensuite à remercier Marie-Pierre Etienne et Viet Chi Tran qui me font l'honneur de rapporter cette thèse ainsi que Nathalie Peyrard, Benoit Liquet, Manon Costa et Bertrand Cloez qui ont gracieusement accepté de faire partie du jury.

Je tiens également à exprimer ma gratitude envers toutes les personnes qui ont contribué à ce travail, avec une mention spéciale à Julien Balicchi et à l'Agence régionale de santé (ARS) de Mayotte pour leur précieuse collaboration dans la collecte des données. De plus, je souhaite souligner que ce projet a bénéficié d'un soutien financier partiel du LabEx NUMEV. Je suis reconnaissant d'avoir eu l'opportunité de participer aux journées scientifiques organisées par les membres de ce laboratoire d'excellence. J'ai également beaucoup apprécié les échanges constructifs avec le directeur, Andrea Parmeggiani de L2C, ainsi que Nabil Zemiti du LIRMM. Enfin, je tiens à exprimer ma gratitude envers Geneviève Carrière, la gestionnaire qui a veillé à la coordination de mes missions scientifiques. Toutes ces contributions ont été essentielles à la réussite de ce travail.

Intégrer l'équipe des chercheur/es et enseignant/es de l'IMAG à l'Université de Montpellier m'a ouvert la porte à de précieuses rencontres et à l'acquisition de connaissances. J'ai eu le privilège de faire la connaissance de personnes exceptionnelles, des amis et des collègues, chacun/une ayant sa place spéciale dans mon esprit. Je tiens à exprimer ma profonde gratitude à Thiziri Moulla, Florian Miralles, Marwa Salah, Pascal Azerad, Ali Gannoun, ainsi qu'à ceux qui ont depuis quitté le laboratoire, à savoir Ali Haidar,

Chayma Daayeb, Franshesko, Hanine Awada, Zaineb Smida et Antonia Jabbour. Un remerciement particulier revient à Sophie Cazanave, Nathalie Collain et Brigitte Labric, dont le dévouement exemplaire envers les membres de l'équipe a été remarquable.

Je ne saurais oublier de mentionner mes co-bureaux, Marien Hanno, notre maestro du laboratoire, un collègue hors pair et un ami toujours présent pour des discussions mathématiques, et Paul-Aimé-Latsouck Faye, un ami doté de nombreuses vertus et de valeurs admirables. Comme l'a si bien dit le chef du restaurant administratif, il est un véritable "prêtre sur terre".

Et bien sûr, je tiens également à exprimer ma gratitude envers ma famille : ma mère et ma sœur, qui m'ont apporté un soutien inestimable tout au long de cette aventure. Leur présence et leurs encouragements ont été une source de réconfort et de motivation, et je leur suis profondément reconnaissant pour leur soutien indéfectible.

Ibrahim BOUZALMAT

Résumé

Ce manuscrit de thèse vise à étudier la transmission de la fièvre typhoïde à Mayotte en utilisant des approches de modélisation mathématique. Nous introduisons tout d'abord le contexte de notre étude, les problématiques associées et les objectifs de la thèse. Un état de l'art sur la modélisation mathématique de la transmission de la fièvre typhoïde est présenté, mettant en évidence la spécificité de notre approche.

Nous proposons un modèle initial en deux versions, déterministe et stochastique, pour décrire la dynamique de transmission de la maladie à Mayotte. Nous explorons le comportement du modèle à travers des simulations numériques dans différents scénarios, en mettant en évidence les facteurs clés de la transmission.

Cependant, en raison des limitations du jeu de données disponibles, nous proposons un modèle stochastique simplifié et une méthode d'estimation paramétrique. Cette approche nous permet d'ajuster le modèle aux données disponibles et d'estimer les caractéristiques clés de la transmission de la fièvre typhoïde à Mayotte.

En enrichissant notre modèle, nous introduisons de nouvelles extensions. Nous incluons un compartiment pour les individus exposés, prenant en compte la durée d'incubation de la maladie. Les propriétés théoriques de ce modèle sont étudiées et illustrées par des simulations numériques. De plus, nous proposons une méthodologie d'estimation des paramètres adaptée à ce nouveau modèle, et des simulations numériques ont été réalisées pour évaluer la performance de notre approche d'estimation.

Nous examinons ensuite l'impact de la pluviométrie sur la transmission de la fièvre typhoïde à Mayotte, en utilisant des données publiques de précipitations. Nous identifions une saisonnalité de la pluie et estimons les paramètres du modèle sous différents régimes. Les résultats soulignent l'importance de cette variable météorologique dans la propagation de l'épidémie.

Ce manuscrit ouvre des perspectives de recherche, telles que l'extension du modèle à d'autres maladies infectieuses présentes à Mayotte et sa généralisation à d'autres territoires. Ces travaux contribueront à une meilleure compréhension et gestion des maladies infectieuses à Mayotte et dans d'autres régions similaires.

Abstract

The aim of this thesis manuscript is to study the transmission of typhoid fever in Mayotte using mathematical modelling approaches. We first introduce the context of our study, the associated issues, and the objectives of the thesis. A state-of-the-art review on mathematical modeling of typhoid fever transmission is presented, highlighting the specificity of our approach.

We propose an initial model in two versions, deterministic and stochastic, to describe the transmission dynamics of the disease in Mayotte. We explore the behavior of the model through numerical simulations in different scenarios, highlighting key factors of transmission.

However, due to the limitations of the available dataset, we propose a simplified stochastic model and a parametric estimation method. This approach enables us to fit the model to the available data and to estimate the key characteristics of typhoid fever transmission in Mayotte.

In enriching our model, we are introducing new extensions. We include a compartment for individuals exposed, taking into account the incubation period of the disease. The theoretical properties of this model are studied and illustrated by numerical simulations. In addition, we propose a parameter estimation methodology adapted to this new model, and numerical simulations have been carried out to evaluate the performance of our estimation approach.

We then examine the impact of rainfall on the transmission of typhoid fever in Mayotte, using publicly available precipitation data. We identify rainfall seasonality and estimate model parameters under different regimes. The results highlight the importance of this meteorological variable in the spread of the epidemic.

This manuscript opens up research perspectives, such as the extension of the model to other infectious diseases present in Mayotte and its generalisation to other territories. This work will contribute to a better understanding and management of infectious diseases in Mayotte and other similar regions.

Table des matières

1	Introduction générale	8
1.1	Introduction	9
1.2	Contexte	9
1.2.1	La fièvre typhoïde	9
1.2.2	Description de Mayotte	10
1.2.3	Accès à l'eau à Mayotte	12
1.2.4	La Typhoïde à Mayotte	13
1.3	Objectifs de la thèse	15
1.4	Outils mathématiques	15
1.4.1	Modèles à compartiments	15
1.4.2	Processus Markovien Déterministe par Morceaux	17
1.4.3	Processus de naissance et mort linéaire avec immigration	18
1.4.4	Modèle de Markov caché	24
1.5	État de l'art sur la modélisation mathématique de la fièvre typhoïde	29
1.6	Organisation de la thèse	30
2	Modèle à compartiments pour la transmission de la fièvre typhoïde à Mayotte	32
2.1	Introduction	33
2.2	Modélisation déterministe	33
2.2.1	Formulation du modèle	34
2.2.2	Calibration des paramètres	36
2.2.3	Discussion et simulations numériques	38
2.2.4	Le nombre de reproduction de base	45
2.3	Modélisation par un processus markovien déterministe par morceaux	48
2.3.1	Formulation du modèle	49
2.3.2	Discussion et simulations numériques	50
2.4	Conclusion	54
3	Estimation des paramètres d'un processus caché de naissance et mort linéaire avec immigration	55
3.1	Introduction	56
3.2	Modèle et schéma d'observations	57
3.2.1	Description du Modèle	57
3.2.2	Schéma d'observations	59
3.2.3	Stratégie d'estimation	60
3.3	Estimation paramétrique	60

3.3.1	Expression analytique des paramètres	61
3.3.2	Modèle de Markov caché	69
3.4	Étude numérique	74
3.4.1	Données synthétiques	74
3.4.2	Données de la typhoïde à Mayotte	79
3.4.3	Application à l'estimation de la probabilité d'atteindre un seuil épidémique dans une période de temps	80
3.5	Conclusion	84
4	Modélisation et estimation des paramètres d'un modèle multi-chaîne cachée	85
4.1	Introduction	86
4.2	Modèle et propriété	87
4.2.1	Définition du modèle	87
4.2.2	Propriétés mathématiques du modèle	88
4.3	Estimation des paramètres	92
4.3.1	Observations	93
4.3.2	Stratégie d'estimation	94
4.3.3	Expression analytique des paramètres	94
4.3.4	Modèle de Markov Multi-Chaîne Cachée	97
4.4	Simulation numérique	102
4.4.1	Observations complètes	102
4.4.2	Observations du cumul des nouveaux cas déclarés	103
4.5	Conclusion	107
5	Effet de la pluviométrie sur l'évolution de la fièvre typhoïde à Mayotte	108
5.1	Introduction	109
5.2	Description des données de précipitations	109
5.3	Détection de rupture	110
5.3.1	Méthode CUSUM	111
5.3.2	Application aux données de précipitations à Mayotte	111
5.4	Estimation des paramètres d'un modèle de Markov caché à deux régimes	114
5.5	Conclusion	118
6	Perspectives	119
Annexe		123
A	Formulaire de déclaration de la fièvre typhoïde	123
B	Matrice Jacobienne Dg	124
C	Log-vraisemblance $l(O M^n)$	127
D	Fonction génératrice des moments du processus $(E_t, I_t)_{t \geq 0}$	128
E	Séries chronologiques	130
F	Détection de rupture	135
Bibliographie		144

Chapitre 1

Introduction générale

Sommaire

1.1	Introduction	9
1.2	Contexte	9
1.2.1	La fièvre typhoïde	9
1.2.2	Description de Mayotte	10
1.2.3	Accès à l'eau à Mayotte	12
1.2.4	La Typhoïde à Mayotte	13
1.3	Objectifs de la thèse	15
1.4	Outils mathématiques	15
1.4.1	Modèles à compartiments	15
1.4.2	Processus Markovien Déterministe par Morceaux	17
1.4.3	Processus de naissance et mort linéaire avec immigration	18
1.4.4	Modèle de Markov caché	24
1.5	État de l'art sur la modélisation mathématique de la fièvre typhoïde	29
1.6	Organisation de la thèse	30

1.1 Introduction

L'épidémiologie est une science qui cherche à comprendre la dynamique régissant la propagation de maladies infectieuses, et étudier les facteurs de risques qui contribuent à cette propagation, leur prévalence et leurs conséquences afin d'établir des stratégies de prévention et d'intervention permettant de diminuer leur impact sur la santé publique. Le premier modèle utilisé en épidémiologie est celui qui décrit les infections par la variole (ou variole). Il a été publié par Daniel Bernoulli [1] en 1760. C'est la première fois que l'ensemble des observations épidémiologiques sur la variole ont été collectées, combinées avec les données démographiques de l'époque et insérées dans un modèle mathématique afin de tester les hypothèses et d'évaluer le rapport bénéfice-risque de la variolisation. Depuis lors et au début du XXe siècle, un grand nombre de modèles épidémiologiques ont été développés [2]- [3], avec en général l'un des objectifs principaux suivants :

- ◇ Traduire schématiquement la propagation de certaines maladies infectieuses, ce qui devrait permettre de mieux comprendre les phénomènes épidémiques et donc de mieux les contrôler.
- ◇ Décrire théoriquement et numériquement la dynamique épidémique des agents pathogènes d'intérêt.
- ◇ Quantifier l'impact et l'efficacité des différentes méthodes de prévention, évaluer également les risques associés à cette épidémie et aux interventions mise en place.
- ◇ Planifier, tester et optimiser des plans d'expérience, en complément de l'expérimentation par exemple.

Le but des modèles en épidémiologie est de décrire, comprendre, prévoir et/ou contrôler la dynamique des pathogènes. Dans ce contexte, la modélisation permet d'isoler les effets de certains facteurs et de sélectionner les interactions dont on tiendra compte. Par ailleurs, l'étude des effets de différents facteurs, ainsi que de leurs interactions est grandement enrichie par la possibilité de répliquer des mesures bien supérieures à ce que l'expérimentation pourrait apporter. De plus, par le jeu des changements des valeurs des paramètres, les modèles permettent d'étudier rapidement un grand nombre de situations. La littérature et le développement de l'épidémiologie mathématique sont bien documentés et peuvent être trouvés par exemple dans [4], [5] et [6].

Dans ce travail, nous nous intéressons à modéliser et étudier la dynamique de transmission de la maladie de la fièvre typhoïde à Mayotte dans le but de quantifier son influence et son impact sur la santé globale de la population.

1.2 Contexte

1.2.1 La fièvre typhoïde

La fièvre typhoïde (du grec tymphos, torpeur) ou typhus abdominal est une maladie infectieuse décrite en 1818 par Pierre Bretonneau, causée par une bactérie de la famille Entérobactérie, du genre des salmonelles, une bactérie qui se multiplie dans l'estomac et les intestins de personnes infectées, et dont les espèces responsables sont *Salmonella enterica* -Typhi ou Paratyphi A, B, C.

La principale voie de transmission de la fièvre typhoïde est la voie oro-fécale. Cela signifie qu'elle peut se propager directement à partir de porteurs infectés par le biais de matières

fécales contaminées, d'urines et de vomissements, ou indirectement par la consommation d'eau ou d'aliments contaminés (**Fig.1.1**).

Les symptômes de la fièvre typhoïde comprennent de la fièvre élevée, des maux de tête, une perte d'appétit, des douleurs abdominales et une faiblesse générale. Dans les cas graves, elle peut causer des complications sérieuses, y compris des saignements internes, une perforation intestinale et même la mort. Une fois diagnostiquée, la fièvre typhoïde peut être traitée avec des antibiotiques. Il est important de prévenir la fièvre typhoïde en adoptant des mesures d'hygiène alimentaire, en se vaccinant contre cette maladie et en traitant une infection présente dès que possible.

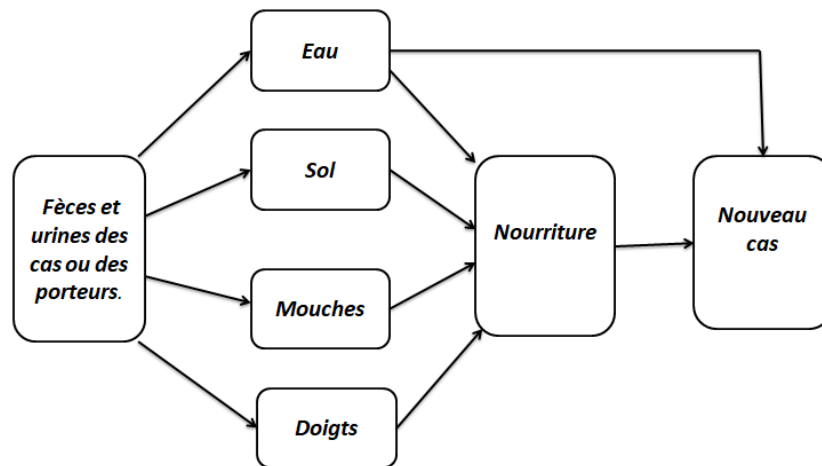


FIGURE 1.1 – Modes de transmission de la Typhoïde.

1.2.2 Description de Mayotte

Mayotte est une île de l'océan Indien, située dans l'archipel des Comores. C'est à la fois une région insulaire française et un département de France d'outre-mer qui sont administrés dans le cadre d'une collectivité territoriale unique dirigée par le conseil départemental de Mayotte [7]. Mayotte se compose de deux îles principales, Grande-Terre et Petite-Terre, entourées d'un lagon turquoise. Les reliefs sont formés de collines et de montagnes, couvertes de forêts tropicales luxuriantes. Elle est célèbre pour ses plages de sable blanc, ses eaux cristallines et sa faune marine diversifiée. Les principales activités économiques de l'île sont l'agriculture, la pêche et le tourisme.

Mayotte est le plus petit département d'outre-mer français, et le moins peuplé, cependant c'est aussi le territoire avec la plus forte croissance démographique, et la population de l'île devrait bientôt dépasser celle de la Guyane. Mayotte est cependant déjà largement en tête en terme de densité de population, avec 682 hab./km² (devant la Réunion, avec seulement 340 hab./km² [8]).

Elle est en particulier marquée par une expansion démographique spectaculaire : la population est passée de 11 000 habitants en 1911 à 256 518 en 2017 avec un taux de fécondité qui s'élève à 4,68 enfants par femme en 2019 [8], très au-dessus des autres départements français, et similaire à celui de l'Union des Comores voisine, l'hôpital de Mamoudzou est ainsi la première maternité d'Europe (**Fig.1.2**).

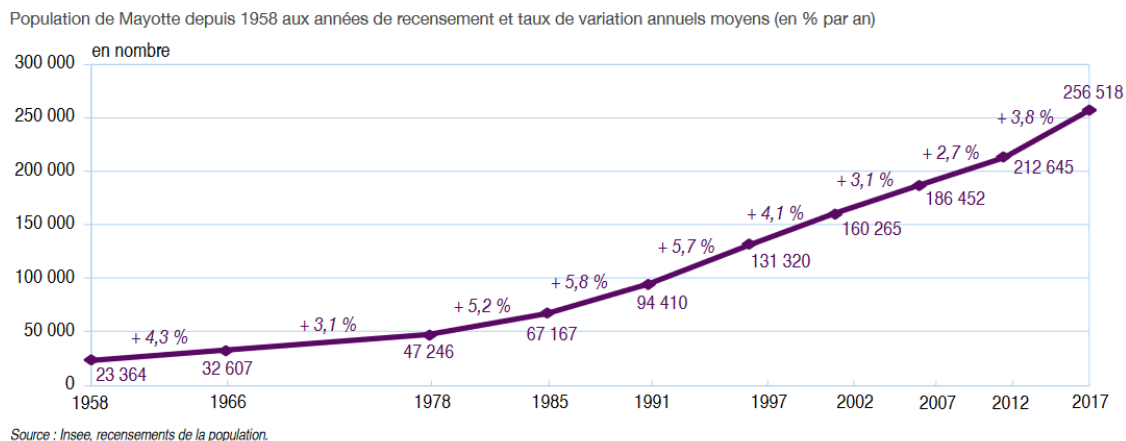


FIGURE 1.2 – Croissance démographique de la population de mayotte entre 1958 et 2017.

Sources : Insee, recensements de la population.

Entre 2012 et 2017, la répartition de la population de Mayotte par âge a très peu changé. L'âge moyen des résidents en 2017 était de 23 ans. Du fait de l'immigration et du taux de natalité élevé, le vieillissement de la population est encore limité. La population est encore beaucoup plus jeune que dans les autres régions de France : l'âge moyen est de 41 ans en métropole, de 35 ans à La Réunion et de 28 ans en Guyane [9].

En 2012, la moitié de la population avait moins de 18 ans et les trois dixièmes des Mahorais avaient moins de 10 ans. 11 000 personnes de plus de 60 ans ne représentent que 4% de la population totale, soit six fois moins qu'en France métropolitaine [9].

La pyramide des âges présente un creux marqué autour de 20 ans, du fait du déficit migratoire à ces âges (**Fig.1.3**).

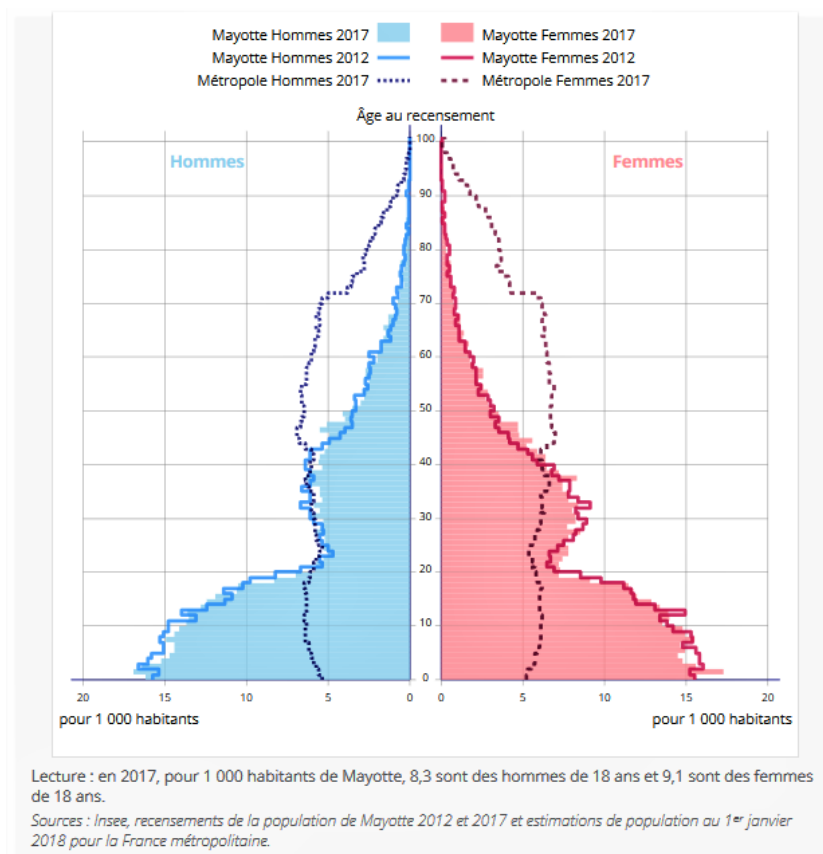


FIGURE 1.3 – Pyramide des âges à Mayotte (2012 et 2017) et en France métropolitaine (2017).

1.2.3 Accès à l'eau à Mayotte

Mayotte, en tant que département qui représente un contexte particulier de précarisation, présente une forte prévalence des maladies hydriques. Les difficultés pour accéder à l'eau courante et les conditions d'hygiène insuffisantes posent évidemment des problèmes sanitaires à Mayotte. Un habitant sur trois n'a pas accès à l'eau à l'intérieur de son logement et six logements sur dix ne proposent toujours pas le confort sanitaire de base (eau courante, toilettes, douche). A cette situation de très forte précarité s'ajoute l'absence d'un système d'assainissement de qualité.

Dans ce contexte, l'agence régionale de santé (ARS) est chargée d'organiser et de mettre en oeuvre le contrôle sanitaire des eaux destinées à la consommation. Outre cette mission, ses actions portent également sur l'amélioration de l'accès à l'eau potable pour tous puisque les habitants iraient s'alimenter chez un voisin, à une borne fontaine monétique (BFM), au puits ou encore à la rivière et peuvent recourir à un stockage de l'eau dans des conditions douteuses. Ainsi, l'installation de BFM (**Fig.1.4**) sur les quartiers d'habitat insalubre où résident les populations les plus démunies a été identifiée comme action de prévention prioritaire, depuis 2000, avec l'apparition de cas de choléra à Mayotte.

Les BFM permettent de fournir de l'eau potable aux habitants des quartiers qui ne sont pas raccordés au réseau de distribution d'eau potable. Financées par l'ARS, installées par la Société Mahoraise des Eaux (SMAE), elles permettent à la population de récupérer le volume d'eau souhaité, à l'aide d'une carte (en vente à la SMAE). Depuis 2010, de

nouvelles implantations et des déplacements de BFM réalisés en faveur des quartiers de grande précarité portent le nombre de BFM en fonctionnement à 52, dont 33 se situent sur les communes de Mamoudzou et de Koungou. Au cours de la période 2015-2020, une demande de crédit de 1 million d'euros sur six ans a été déposée dans le cadre du contrat de plan État-région, ce qui équivaut à l'installation de 10 BFM/an pendant six ans en sachant qu'au regard des données du recensement de 2012, le nombre de logements sans accès à l'eau nécessiterait actuellement l'installation de 43 BFM (considérant qu'un BFM dessert environ 100 maisons) [10].



FIGURE 1.4 – Des fontaines d'eau potable à Mayotte [10].

Par conséquent, la procédure d'installation de BFM à Mayotte devrait permettre notamment de lutter contre la propagation des maladies liées à la consommation d'eau non conforme comme la fièvre typhoïde.

1.2.4 La Typhoïde à Mayotte

La fièvre typhoïde est une maladie endémique à Mayotte, et il est essentiel de souligner qu'elle est soumise à une déclaration obligatoire (voir la formulaire de déclaration en Annexe **A**). Au cours de l'année 2022, 100 cas de fièvre typhoïde confirmés par hémoculture et/ou coproculture ont été signalés par le laboratoire du Centre Hospitalier de Mayotte (CHM) à la Cellule de Veille, d'Alerte et de Gestion Sanitaire (CVAGS) de l'ARS. Le taux d'incidence de la fièvre typhoïde est en hausse au cours des trois dernières années, passant de 14 cas pour 100 000 habitants en 2020 à 35.7 cas pour 100 000 habitants en 2022. De 2016 à 2021, 39 cas de fièvre typhoïde par an, en moyenne, ont été signalés par l'ARS de Mayotte. [11] (**Fig.1.5**).

Parmi les 100 cas de fièvre typhoïde enregistrés en 2022, 71 cas ont été hospitalisés entre 1 et 7 jours (71.2 %) dont 9 dans le service de réanimation. 74 cas (73,7 %) sont survenus entre les semaines 17 et 25. Parmi les cas examinés, 53 cas avaient accès à l'eau courante ou à une borne-fontaine (BFM) à proximité et 9 cas ont déclaré ne pas avoir accès à l'eau courante et utiliser de l'eau de rivière ou de puits. Les cas de fièvre typhoïde sont survenus

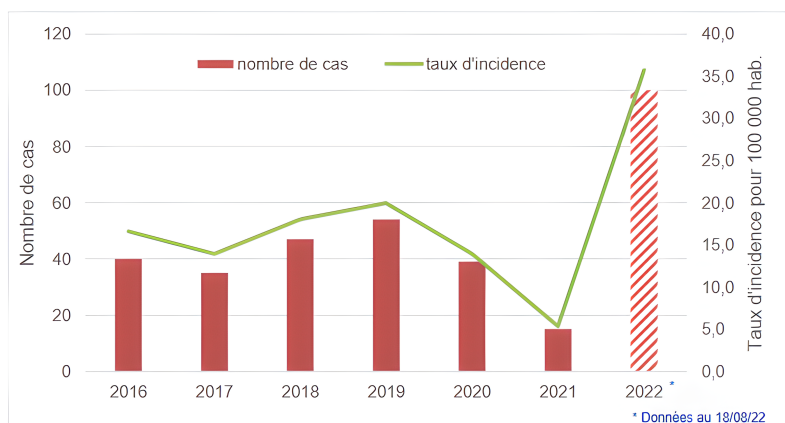


FIGURE 1.5 – Nombre de cas confirmés de fièvre typhoïde par année de prélèvement, janvier 2016-août 2022, Mayotte.

Source : enquêtes réalisées par la CVAGS de Mayotte, ARS.

tout au long de l'année à Mayotte, néanmoins, une recrudescence de cas a été observée entre les mois de février et juin de chaque année, avec des pics épidémiques variés d'une année à l'autre (**Fig.1.6**). Les raisons de ces variations peuvent être multiples, notamment des facteurs environnementaux, socio-économiques et comportementaux. Par exemple, des conditions météorologiques particulières, une densité de population élevée dans certaines zones, une insalubrité des installations sanitaires ou une faible hygiène personnelle peuvent favoriser la propagation de la bactérie *Salmonella Typhi*, qui est responsable de la fièvre typhoïde.

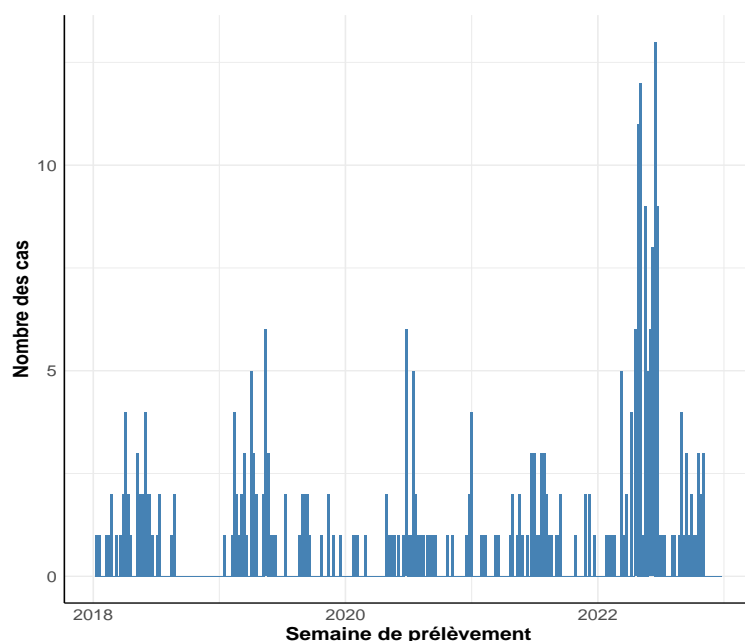


FIGURE 1.6 – Évolution hebdomadaire des nouveaux cas confirmés de fièvre typhoïde par semaine de prélèvement, janvier 2018-décembre 2022.

1.3 Objectifs de la thèse

Cette thèse vise à proposer et étudier de nouveaux modèles de propagation de la fièvre typhoïde adaptés à Mayotte. L'objectif principal est de prendre en compte l'accès à l'eau potable des populations, ainsi que la durée d'incubation de la maladie et l'impact de la pluviométrie sur la dynamique de transmission. Plus précisément, nous visons à :

- ▶ Définir un modèle global de propagation de la fièvre typhoïde pour Mayotte, en tenant compte des spécificités de l'île comme l'accès ou non à l'eau potable.
- ▶ Étudier théoriquement et numériquement le comportement des modèles proposés.
- ▶ Estimer les paramètres des modèles à partir des données fournies par l'Agence Régionale de Santé de Mayotte correspondant au nombre cumulé de nouveaux cas déclarés par jour.
- ▶ Prendre en compte la durée d'incubation de la maladie dans notre modèle afin de mieux comprendre la dynamique de cette maladie infectieuse.
- ▶ Étudier l'impact de la pluviométrie sur la transmission de la fièvre typhoïde à Mayotte en utilisant des données de précipitations et en identifiant des ruptures saisonnières.
- ▶ Explorer les perspectives de recherche futures dans ce domaine.

La thèse contribuera également à développer l'axe III du volet recherche du projet d'établissement du Centre Universitaire de Formation et de Recherche de Mayotte, à savoir : *Modélisations et analyse épistémologique des systèmes complexes pour comprendre le fonctionnement et simuler le devenir des systèmes insulaires : développer des outils de modélisation statistique, descriptive ou dynamique dans une approche interdisciplinaire.* Cette démarche correspond aux priorités de recherche identifiées par la commission scientifique du Centre Universitaire de Mayotte.

1.4 Outils mathématiques

Afin de garantir la rigueur et l'analyse nécessaires pour atteindre les objectifs de cette thèse, il est essentiel de s'appuyer sur des concepts et des outils mathématiques appropriés. Cette section fournit un aperçu détaillé des outils mathématiques clés qui sont utilisés tout au long de cette thèse.

1.4.1 Modèles à compartiments

Les modèles compartimentaux, également connus sous le nom de modèles à compartiments, sont des outils mathématiques puissants pour décrire la dynamique d'une population où les individus traversent successivement plusieurs états, représentés par des compartiments distincts. Dans ces modèles, la population est répartie dans les différents compartiments en fonction de leur statut infectieux par exemple. Chaque compartiment est alors associé à une dynamique propre et des échanges avec les autres compartiments.

Les modèles compartimentaux sont largement utilisés dans le domaine de l'épidémiologie pour étudier la propagation des maladies infectieuses au sein d'une population. Ils permettent de modéliser comment les individus passent d'un compartiment à un autre en fonction de leur statut infectieux et des interactions qu'ils ont les uns avec les autres. Par exemple, les modèles SI, SIS, SIR, SIRS, etc., sont couramment utilisés pour représenter différentes classes épidémiologiques, tels que les Susceptibles, les Infectés, les Retirés (morts ou immunisés), etc. La dynamique de ces modèles, dans le cas déterministe, est généralement décrite par un ensemble d'équations différentielles [2, 12]. Elles peuvent être basées sur diverses hypothèses, telles que la transmission de la maladie à travers des contacts directs ou indirects entre individus, la distribution de la population dans différents groupes d'âge ou de susceptibilité, et la variation des taux d'infection, de guérison et de mortalité. Ces modèles mathématiques permettent de prédire l'évolution temporelle de l'épidémie, d'estimer les paramètres épidémiologiques clés, et d'évaluer l'impact des interventions de santé publique sur la propagation de la maladie.

La **Fig.1.7** représente un schéma d'un modèle à compartiment de type **SIR** dont les retirés ne redeviennent jamais susceptibles, où les bacs représentent les compartiments dans lesquels sont regroupés les individus de même statut épidémiologique (susceptibles, infectés, retirés) et les robinets représentent les échanges entre ces compartiments.

Dans le chapitre 2 (2.2), nous utiliserons un modèle à compartiments pour étudier la dynamique de propagation de la fièvre typhoïde et analyser l'interaction entre les différentes populations impliquées dans le processus épidémique.

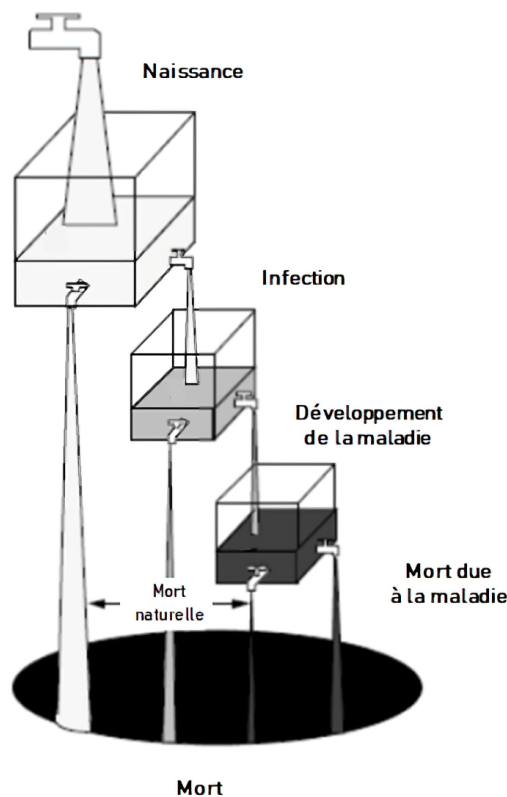


FIGURE 1.7 – Représentation schématique d'un modèle à compartiment.

1.4.2 Processus Markovien Déterministe par Morceaux

Selon une définition de Davis [13], la classe des Processus Markoviens Déterministes par Morceaux (en anglais : Piecewise Deterministic Markov Process (**PDMP**)) est une famille générale de modèles stochastiques hybrides couvrant pratiquement toutes les applications de non diffusion.

Un PDMP $(X_t) = (M_t, Y_t)$ est un processus de Markov en temps continu à valeurs dans un espace d'état $\mathbb{E} = M \times \bigcup_{m \in M} F_m$, où M est un ensemble fini ou dénombrable et F_m est un ouvert de \mathbb{R}^d pour chaque $m \in M$. Il est constitué de deux variables. La première M_t , appelée *régime* ou *mode*, est discrète et permet de caractériser l'environnement et/ou le régime dans lequel le processus évolue. La seconde variable, notée Y_t , est appelée *variable euclidienne* et décrit les grandeurs physiques du modèle.

Un exemple concret pourrait être un modèle météorologique où la variable discrète *mode* pourrait représenter les différents types de temps (ensoleillé, nuageux, pluvieux, etc.), tandis que la variable continue *variable euclidienne* pourrait représenter des grandeurs physiques telles que la température, la pression atmosphérique, l'humidité, etc. Le modèle pourrait alors utiliser la variable discrète pour déterminer le régime météorologique dans lequel il se trouve, et la variable continue pour représenter les grandeurs physiques associées à ce régime.

Les PDMP sont une grande famille de différents processus stochastiques qui comprend les processus de Markov en temps discret, les chaînes de Markov en temps continu, les processus déterministes avec sauts, les systèmes dynamiques avec commutation aléatoire et certains processus ponctuels. Ce type de modèle peut être utilisé dans de nombreux domaines, tels que la physique, la biologie, la dynamique des populations, la fiabilité des systèmes, etc. [14, 15, 16, 17].

L'évolution d'un PDMP est déterminée par trois caractéristiques locales (dépendant du mode $m \in M$) :

- Le flot $\phi_m : \mathbb{R}^d \times \mathbb{R}^+ \rightarrow \mathbb{R}^d$ qui décrit la trajectoire déterministe du processus entre deux sauts. Il peut s'interpréter comme une solution au temps t d'un système d'équations différentielles partant de y au temps 0. Il vérifie également la propriété de semi-groupe,

$$\forall s, t \geq 0, \quad \phi_m(\cdot, t + s) = \phi_m(\phi_m(\cdot, s), t).$$

On appelle $t^*(m, y)$ le temps déterministe d'atteinte de la frontière de l'espace F_m en partant du point $x = (m, y)$:

$$t^*(m, y) = \inf \{t > 0 : \phi_m(y, t) \in \partial F\}.$$

Si $t^*(m, y) = \infty$, cela signifie que le processus n'atteint pas la frontière de F en temps fini.

- L'intensité de saut dans le mode m , $\lambda_m : \bar{F} \rightarrow \mathbb{R}^+$ une fonction mesurable. Elle vérifie la propriété d'intégrabilité locale suivante :

$$\forall x = (m, y) \in \mathbb{E}, \quad \exists \varepsilon > 0, \quad \int_0^\varepsilon \lambda_m(\phi_m(y, t)) dt < +\infty.$$

On suppose que le temps de saut aléatoire T est distribué comme suit :

$$\mathbb{P}_{(m, y)}(T > t) = e^{-\int_0^t \lambda_m(\phi_m(y, s)) ds}.$$

Ainsi, le premier temps du saut T_1 est défini par :

$$T_1 = \min\{t^*(m, y), T\}.$$

C'est le minimum entre le temps déterministe d'atteinte de la frontière et le temps de saut aléatoire, sa loi est donnée par :

$$\mathbb{P}_{(m,y)}(T_1 > t) = e^{-\int_0^t \lambda_m(\phi_m(y,s))ds} \mathbb{1}_{\{t < t^*(m,y)\}}. \quad (1.1)$$

• Le noyau de transition $Q_m : \bar{F} \times \mathcal{B}(\bar{F}) \rightarrow [0, 1]$, qui permet de déterminer la position après le saut. Le nouveau mode et/ou position (M_{T_1}, Y_{T_1}) sélectionnés au temps T_1 par le noyau markovien Q_m :

$$\mathbb{P}_{(m,y)}((M_{T_1}, Y_{T_1}) \in A \mid T_1) = \int_A Q_m(\phi_m(y, T_1), dx).$$

La dynamique d'un PDMP peut être définie de manière itérative suivante : le processus démarre d'un point $x_0 = (m, y)$ dans l'espace d'état \mathbb{E} . Le premier temps de saut $T_1 = S_1$ est déterminé par sa loi **1.1**, il peut se produire soit aléatoirement avec un taux λ_m ou bien lorsqu'on touche la frontière ∂F . Sur l'intervalle $[0, T_1[$, la variable euclidienne Y_t suit le flot déterministe $t \rightarrow \phi_m(y, t)$. La position du processus à l'instant de saut T_1 est sélectionnée selon la distribution $Q_m(\phi_m(y, T_1), \cdot)$, et le processus repart à partir d'une nouvelle position $X_{T_1} = (m_1, Y_1)$. Ensuite, et de même manière, un temps intersaut $S_2 = T_2 - T_1$ est déterminée à partir de la fonction de survie **1.1**, et sur l'intervalle $[T_1, T_2[$, le processus X_t suit le flot déterministe $\phi_{m_1}(Y_1, t - T_1)$. La nouvelle position après le saut T_2 est X_{T_2} et ainsi de suite. L'exemple d'un PDMP représenté jusqu'à son deuxième temps de saut est illustré dans la **Fig.1.8**.

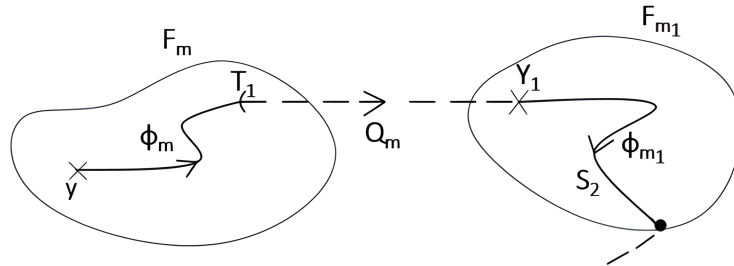


FIGURE 1.8 – Exemple d'une trajectoire d'un PDMP.

Dans le chapitre 2 (**2.3**), nous utiliserons ce type de modèle afin de proposer une modélisation stochastique de la dynamique de transmission de la fièvre typhoïde à Mayotte.

1.4.3 Processus de naissance et mort linéaire avec immigration

Le chapitre **3** de ce manuscrit se concentre sur l'estimation des paramètres d'un processus de Markov de saut pur appartenant à une famille particulière. Cette famille représente un cas particulier de PDMP et une généralisation des processus de comptage, dans lesquels la population peut augmenter ou diminuer d'une unité à chaque saut. Dans cette

partie, nous présentons les notions de base pour caractériser ce processus et décrire ses propriétés mathématiques.

Un processus de naissance et de mort linéaire avec immigration (en anglais : linear Birth-Death Process with Immigration (**LBDPI**)), noté $(X_t)_{t \geq 0}$, est un processus de Markov à temps continu à valeurs dans \mathbb{N} , comptant le nombre d'individus dans une population à l'instant t , son générateur infinitésimal Q est donné par :

$$\begin{aligned} b_i = Q_{i,i+1} &= \lambda i + \nu && \text{pour } i \geq 0, \\ d_i = Q_{i,i-1} &= \mu i && \text{pour } i \geq 1, \\ Q_{i,i} &= -(\lambda + \mu)i - \nu && \text{pour } i \geq 0, \end{aligned}$$

où λ, μ et ν sont positifs.

$(X_t)_{t \geq 0}$ est un processus qui décrit une population qui reçoit des immigrants au taux ν et dans laquelle chaque individu donne naissance à un autre individu au taux λ et meurt au taux μ , indépendamment des autres.

Les fonctions b_i et d_i sont respectivement appelées taux de naissance (birth) et taux de mort (death) du processus $(X_t)_{t \geq 0}$ en présence de i individus. Le paramètre λ s'interprète comme le taux de *naissance individuel*, et μ est comme le taux de *mort (retrait) individuel*.

La dynamique du processus X_t est défini comme suit : une fois à l'état $i \geq 0$, on y reste un temps de loi $Exp(b_i + d_i)$ avant d'aller soit à l'état $i + 1$ avec probabilité $b_i/(b_i + d_i)$, soit à l'état $i - 1$ avec probabilité $d_i/(b_i + d_i)$ (voir **Fig.1.9** et **Fig.1.10**).

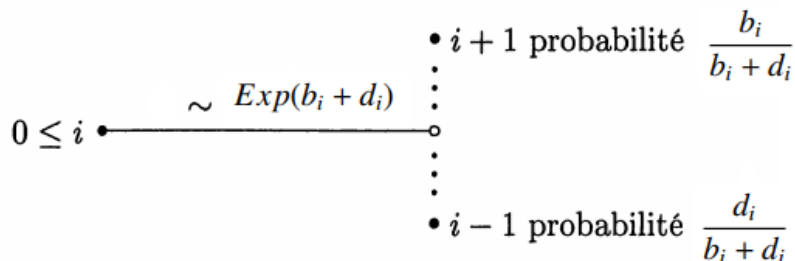


FIGURE 1.9 – Processus de naissance et de mort.

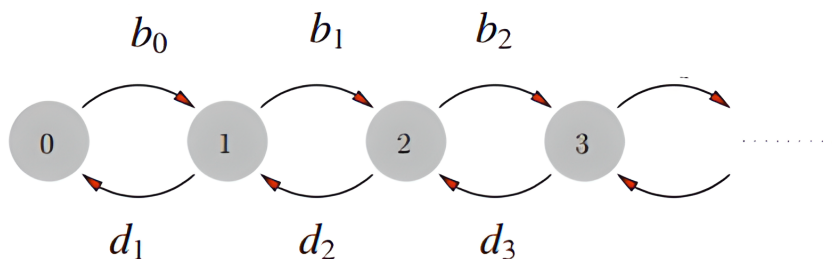


FIGURE 1.10 – Diagramme de Markov du processus $(X_t)_{t \geq 0}$ d'intensités de naissance b_i et de mort d_i , si $b_i > 0 \ \forall i \geq 0$ et $d_i > 0 \ \forall i \geq 1$.

Le semi-groupe de transition défini par $(P(t) = (p_{i,j}(t), i, j \in \mathbb{N}))_{t \geq 0}$, où

$$p_{i,j}(t) = \mathbb{P}(X_t = j | X_0 = i), i, j \in \mathbb{N}, t \geq 0,$$

satisfait l'équation de Kolmogorov progressive suivante :

$$\frac{dP(t)}{dt} = P(t)Q, \quad (1.2)$$

pour $t > 0$, qui peut être réécrite pour $i, j \in \mathbb{N}$ avec $j \geq 1$ comme suit

$$\begin{aligned} \frac{dp_{i,0}(t)}{dt} &= \mu p_{i,1}(t) - \nu p_{i,0}(t), \\ \frac{dp_{i,j}(t)}{dt} &= (\lambda(j-1) + \nu) p_{i,j-1}(t) + \mu(j+1) p_{i,j+1}(t) - ((\lambda + \mu)j + \nu) p_{i,j}(t), \end{aligned}$$

et avec la condition initiale $p_{i,i}(0) = 1$ et $p_{i,j}(0) = 0$ pour $j \neq i$. L'équation de Kolmogorov correspond à la relation infinitésimale : pour $h > 0$,

$$\begin{aligned} \mathbb{P}(X_{t+h} = j | X_t = i) &= Q_{i,j}h + o(h) && \text{si } j \neq i, \\ \mathbb{P}(X_{t+h} = i | X_t = i) &= 1 + Q_{i,i}h + o(h), \end{aligned}$$

uniformément en t . Les trajectoires de $(X_t)_{t \geq 0}$ sont constantes par morceaux et ponctuées de sauts d'amplitude $+1$ ou -1 .

Par la suite, nous présentons certaines propriétés théoriques bien connues du processus X_t (voir par exemple [18]), qui permettent de décrire son comportement moyen et asymptotique. Nous complétons ces résultats par des simulations numériques qui viennent illustrer ces propriétés.

Théorème 1.1. (*Comportement moyen de X_t*). Pour tout $i \geq 0$ et $t \geq 0$,

$$\mathbb{E}_i[X_t] = \begin{cases} \nu t + i & \text{si } \lambda = \mu, \\ i e^{(\lambda-\mu)t} + \frac{\nu}{\lambda-\mu} (e^{(\lambda-\mu)t} - 1) & \text{si } \lambda \neq \mu. \end{cases}$$

Preuve. En utilisant l'équation de Kolmogorov progressive 1.2, on obtient

$$\begin{aligned} \frac{d}{dt} \mathbb{E}_i[X_t] &= \sum_{j \geq 0} j p'_{i,j}(t) \\ &= \sum_{j \geq 0} j \sum_{k \geq 0} p_{i,k}(t) Q_{k,j} \\ &= \sum_{k \geq 0} (\lambda - \mu) k p_{i,k}(t) + \nu. \end{aligned}$$

On a alors

$$\frac{d}{dt} \mathbb{E}_i[X_t] = (\lambda - \mu) \mathbb{E}_i[X_t] + \nu,$$

avec la condition initiale $\mathbb{E}_i[X_0] = i$.

On résout cette équation différentielle par la méthode de variation de la constante, on trouve

$$\mathbb{E}_i[X_t] = \begin{cases} \nu t + i & \text{si } \lambda = \mu, \\ i e^{(\lambda-\mu)t} + \frac{\nu}{\lambda-\mu} (e^{(\lambda-\mu)t} - 1) & \text{si } \lambda \neq \mu. \end{cases}$$

□

Corollaire 1.1. *Asymptotiquement et conditionnellement à $X_0 = i$, si $\lambda \geq \mu$, le processus moyen tend vers l'infini ($\lim_{t \rightarrow \infty} \mathbb{E}_i[X_t] = \infty$) de manière linéaire lorsque $\lambda = \mu$ (**Fig.1.12**), et de manière exponentielle lorsque $\lambda > \mu$ (**Fig.1.11**). En revanche, si $\lambda < \mu$, l'espérance de X_t se stabilise et dans ce cas $\lim_{t \rightarrow \infty} \mathbb{E}_i[X_t] = \nu/(\mu - \lambda)$ (**Fig.1.13**).*

Théorème 1.2. *(Comportement asymptotique de X_t). Le processus $(X_t)_{t \geq 0}$ admet une unique probabilité invariante si et seulement $\mu > \lambda$.*

De plus, si on note $\pi = (\pi_k, k \in \mathbb{N})$ cette probabilité invariante, alors pour tout $k \geq 0$:

$$\pi_k = \binom{\frac{\nu}{\lambda} + k - 1}{k} \left(\frac{\lambda}{\mu}\right)^k \left(1 - \frac{\lambda}{\mu}\right)^{\nu/\lambda}, \quad (1.3)$$

où

$$\binom{\frac{\nu}{\lambda} + k - 1}{k} = \frac{(\frac{\nu}{\lambda} + k - 1)(\frac{\nu}{\lambda} + k - 2) \cdots \frac{\nu}{\lambda}}{k!}.$$

Preuve. $(X_t)_{t \geq 0}$ est un processus de naissance et mort linéaire, tel que :

$$\forall i \neq 0, \quad b_i \neq 0, d_i \neq 0 \text{ et } b_0 > 0.$$

Donc, il est irréductible.

• *Existence de la loi invariante :*

D'après [**19, Théorème 1.4**], le processus $(X_t)_{t \geq 0}$ admet une loi invariante si et seulement si :

$$C = \sum_{k \geq 1} \frac{b_0 \cdots b_{k-1}}{d_1 \cdots d_k} = \sum_{k \geq 1} \frac{\nu(\lambda + \nu) \cdots ((k-1)\lambda + \nu)}{\mu 2\mu \cdots k\mu} < +\infty.$$

Ainsi, $(X_t)_{t \geq 0}$ admet une loi invariante $(\pi_k)_{k \in \mathbb{N}}$ si et seulement si $\mu > \lambda$.

• *Expression de $(\pi_k)_{k \in \mathbb{N}}$:*

D'après [**19, Théorème 1.4**], l'unique loi invariante est donnée par :

$$\pi_0 = \frac{1}{1 + C}, \quad \pi_k = \frac{b_0 \cdots b_{k-1}}{d_1 \cdots d_k} \pi_0 \quad \forall k \geq 1.$$

Nous commençons par calculer π_0 . On a :

$$C = \sum_{k \geq 1} \frac{(\frac{\nu}{\lambda} + k - 1)(\frac{\nu}{\lambda} + k - 2) \cdots \frac{\nu}{\lambda} \lambda^k}{k! \mu^k},$$

donc :

$$1 + C = \sum_{k \geq 0} \binom{\frac{\nu}{\lambda} + k - 1}{k} \left(\frac{\lambda}{\mu}\right)^k.$$

Or, on sait que : pour tout $x \in [0, 1[$ et pour tout $n \in \mathbb{R}_+$, on a :

$$\frac{1}{(1-x)^n} = \sum_{k=0}^{\infty} \binom{n+k-1}{k} x^k.$$

Ainsi :

$$\pi_0 = \left(1 - \frac{\lambda}{\mu}\right)^{\nu/\lambda}.$$

Nous déduisons maintenant l'expression de π_k pour tout $k \geq 1$,

$$\begin{aligned}\pi_k &= \frac{\nu(\lambda + \nu) \dots ((k-1)\lambda + \nu)}{\mu 2\mu \dots k\mu} \left(1 - \frac{\lambda}{\mu}\right)^{\nu/\lambda} \\ &= \frac{(\frac{\nu}{\lambda} + k - 1)(\frac{\nu}{\lambda} + k - 2) \dots \frac{\nu}{\lambda}}{k!} \frac{\lambda^k}{\mu^k} \left(1 - \frac{\lambda}{\mu}\right)^{\nu/\lambda}.\end{aligned}$$

Par conséquent, pour tout $k \geq 1$:

$$\pi_k = \binom{\frac{\nu}{\lambda} + k - 1}{k} \left(\frac{\lambda}{\mu}\right)^k \left(1 - \frac{\lambda}{\mu}\right)^{\nu/\lambda}.$$

□

Théorème 1.3. *Le LBDPI $(X_t)_{t \geq 0}$ (avec $\nu > 0$) a trois régimes possibles de comportement à long terme : croissance exponentielle (transitoire) si $\lambda > \mu$ (**Fig.1.11**). Tous les états sont visités infiniment souvent avec un temps de retour moyen infini (réurrence nulle) si $\lambda = \mu$ (**Fig.1.12**). Tous les états sont visités infiniment souvent avec un temps de retour moyen fini (réurrence positive) si $\lambda < \mu$ (**Fig.1.13**). Dans ce dernier cas, il existe une unique loi invariante donnée par **1.3**.*

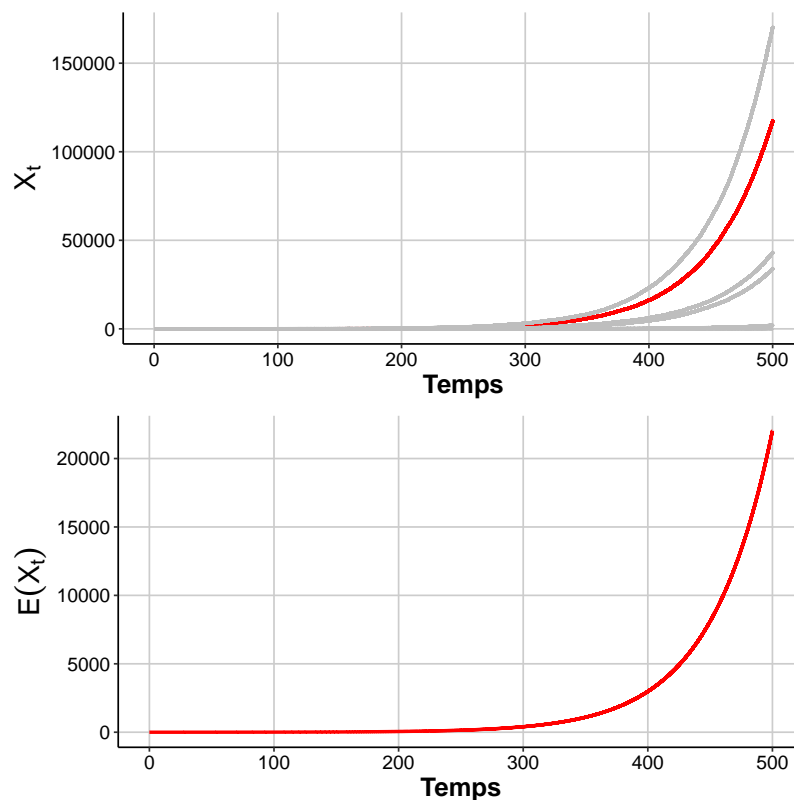


FIGURE 1.11 – Exemple des trajectoires de X_t et son espérance $\mathbb{E}[X_t]$ pour $\lambda = 0.1$, $\mu = 0.08$ et $\nu = 0.02$ (régime transitoire).

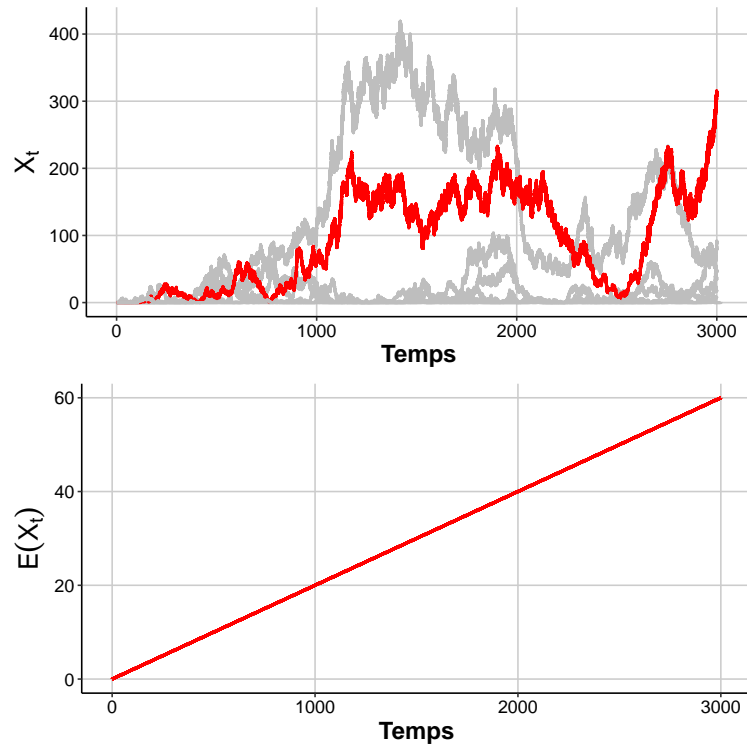


FIGURE 1.12 – Exemple des trajectoires de X_t et son espérance $\mathbb{E}[X_t]$ pour $\lambda = \mu = 0.1$ et $\nu = 0.02$ (régime récurrent nul).

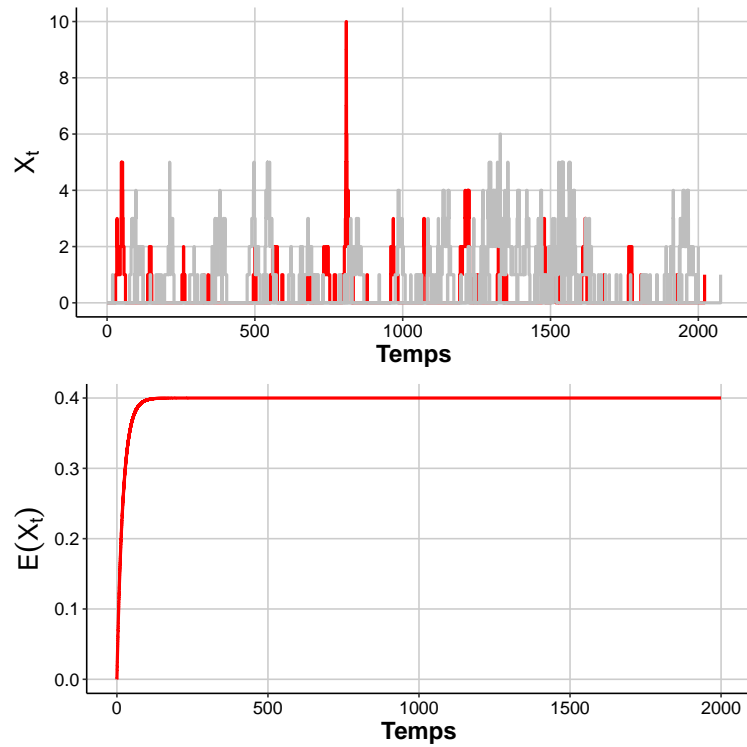


FIGURE 1.13 – Exemple des trajectoire de X_t et son espérance $\mathbb{E}[X_t]$ pour $\lambda = 0.1$, $\mu = 0.15$ et $\nu = 0.02$ (régime récurrent positif).

1.4.4 Modèle de Markov caché

Cette section introductive de la thèse est dédiée aux problèmes d'information cachée, qui constituent une grande partie des problèmes abordés dans ce travail. Nous y présentons les principes fondamentaux, les notations clés et les algorithmes principaux qui seront développés et adaptés pour notre contexte.

Dans la pratique, de nombreux processus suivent un modèle de Markov, mais ne sont pas observés directement. Dans ce cas, on parle de chaînes de Markov cachées, où seule une partie de l'état du processus est observée. L'objectif est alors d'obtenir le maximum d'informations sur la partie cachée de la chaîne à partir des observations partielles disponibles.

Les modèles de Markov cachés (en anglais : Hidden Markov Model (**HMM**)) sont largement utilisés pour modéliser des données séquentielles dans de nombreux domaines, tels que la reconnaissance de la parole [20], la reconnaissance d'écriture manuscrite [21], la bioinformatique [22], la finance et la surveillance [23]. Les HMM sont des modèles statistiques qui permettent de représenter des données observées qui sont la résultante de processus stochastiques non observables, également appelés processus cachés. L'objectif de cette modélisation est de découvrir les états cachés à partir des données observées, et d'inférer les paramètres du modèle à partir des données observées. Les HMM sont particulièrement utiles pour les problèmes de classification et de prédiction de séquences, et constituent un outil puissant pour l'analyse de données séquentielles. Nous allons présenter les principes et les notions de base des HMM.

Caractéristiques des HMM à temps discret

On considère une chaîne de Markov homogène $X := X_{0:T} = (X_0, X_1, \dots, X_T)$ à valeurs dans un espace d'états fini ou dénombrable \mathcal{S} . La chaîne X n'est pas directement observable, elle est dite cachée. Considérons une autre suite des observations $O := O_{0:T} = (O_0, O_1, \dots, O_T)$ prenant des valeurs dans un ensemble \mathcal{O} appelé l'espace d'observation. Cet espace peut être au plus dénombrable. Nous supposons que les variables aléatoires $O_{0:T}$ sont indépendantes conditionnellement à la suite d'états cachée $X_{0:T}$ et que chaque observation O_k ne dépend que de l'état X_k (l'hypothèse du canal d'observation sans mémoire : HMM standard). C'est-à-dire :

$$\mathbb{P}(O_{0:T} = o_{0:T} \mid X_{0:T} = x_{0:T}) = \prod_{t=0}^T \mathbb{P}(O_t = o_t \mid X_t = x_t), \quad x_{0:T} \in \mathcal{S}^{T+1}, o_{0:T} \in \mathcal{O}^{T+1}.$$

Un HMM (X, O) à espace d'observation dénombrable est caractérisé par :

- Une matrice de transition P du processus d'état caché X :

$$p_{x,x'} = \mathbb{P}(X_{t+1} = x' \mid X_t = x), \quad x, x' \in \mathcal{S}.$$

- Une matrice d'émission ψ définie par :

$$\psi_x(o) = \mathbb{P}(O_t = o \mid X_t = x), \quad x \in \mathcal{S}, o \in \mathcal{O}.$$

- Une loi initiale ρ du processus X :

$$\rho_x = \mathbb{P}(X_0 = x), \quad x \in \mathcal{S}.$$

On note par le triplet $M = (P, \psi, \rho)$ les paramètres du HMM.

La probabilité conjointe des états cachés et des observations se factorise comme suit en utilisant les propriétés de Markov et d'indépendance conditionnelle du HMM :

$$\mathbb{P}(X_{0:T} = x_{0:T}, O_{0:T} = o_{0:T}) = \rho_{x_0} \psi_{x_0}(o_0) \prod_{t=1}^T \psi_{x_t}(o_t) p_{x_{t-1}, x_t}.$$

La figure **1.14** illustre une représentation graphique du HMM standard.

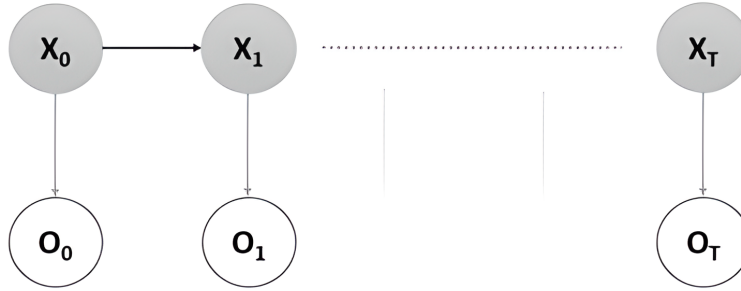


FIGURE 1.14 – Schéma d'émission du HMM standard.

Algorithme Forward-Backward de Baum [20, 24]

Dans les problèmes du HMM, il est essentiel de pouvoir évaluer efficacement la probabilité $\mathbb{P}_M(O_{0:T})$ d'observer la suite d'observations $O_{0:T}$, étant donné le modèle M (la vraisemblance des observations).

Il est important de remarquer que cette probabilité peut être exprimée sous la forme suivante :

$$\begin{aligned} \mathbb{P}_M(O_{0:T} = o_{0:T}) &= \sum_{x_{0:T} \in \mathcal{S}^{T+1}} \mathbb{P}_M(X_{0:T} = x_{0:T}, O_{0:T} = o_{0:T}) \\ &= \sum_{x_{0:T} \in \mathcal{S}^{T+1}} \rho_{x_0} \psi_{x_0}(o_0) \prod_{t=1}^T \psi_{x_t}(o_t) p_{x_{t-1}, x_t}, \end{aligned} \quad (1.4)$$

pour tout $o_{0:T} \in \mathcal{O}^{T+1}$.

Il convient de noter que le calcul direct de la probabilité $\mathbb{P}_M(O_{0:T})$ est possible, mais requiert des calculs d'ordre exponentiel ($T \times N^T$, où N représente le nombre d'états cachés dans le modèle), ce qui est très coûteux en temps de calcul. C'est pourquoi, une approche récursive (Algorithme Forward-Backward) plus astucieuse et plus rapide a été développée pour réduire le temps de calcul nécessaire à l'évaluation de cette probabilité.

• Algorithme Forward :

On s'intéresse à évaluer la probabilité conjointe des t premières observations $O_{0:t}$ et $X_t = x \in \mathcal{S}$, étant donné le modèle M :

$$\alpha_x(t) = \mathbb{P}_M(O_{0:t} = o_{0:t}, X_t = x), \quad \text{pour } t \geq 0.$$

Cette probabilité satisfait une formule récursive donnée par le théorème suivant.

Théorème 1.4. Pour tout $1 \leq t \leq T$, $x \in \mathcal{S}$,

$$\begin{cases} \alpha_x(0) = \rho_x \psi_x(o_0), \\ \alpha_x(t) = \psi_x(o_t) \sum_{x' \in \mathcal{S}} p_{x',x} \alpha_{x'}(t-1). \end{cases} \quad (1.5)$$

Preuve. Soit $x \in \mathcal{S}$, d'abord, pour $t = 0$, on a :

$$\alpha_x(0) = \mathbb{P}_M(O_0 = o_0, X_0 = x) = \rho_x \psi_x(o_0).$$

Ensuite, on peut remarquer que : pour tout $t \geq 2$,

$$\begin{aligned} \alpha_x(t-1) &= \sum_{x_{0:t-2} \in \mathcal{S}^{t-1}} \mathbb{P}_M(O_{0:t} = o_{0:t}, X_{0:t-2} = x_{0:t-2}, X_{t-1} = x) \\ &= \sum_{x_{0:t-2} \in \mathcal{S}^{t-1}} \rho_{x_0} \psi_{x_0}(o_0) \prod_{l=1}^{t-1} \psi_{x_l}(o_l) p_{x_{l-1}, x_l}. \end{aligned}$$

Ainsi,

$$\begin{aligned} \alpha_x(t) &= \sum_{x_{0:t-1} \in \mathcal{S}^t} \rho_{x_0} \psi_{x_0}(o_0) \prod_{l=1}^t \psi_{x_l}(o_l) p_{x_{l-1}, x_l} \\ &= \psi_{x_t}(o_t) \sum_{x_{t-1} \in \mathcal{S}} p_{x_{t-1}, x_t} \sum_{x_{0:t-2} \in \mathcal{S}^{t-1}} \rho_{x_0} \psi_{x_0}(o_0) \prod_{l=1}^{t-1} \psi_{x_l}(o_l) p_{x_{l-1}, x_l} \\ &= \psi_{x_t}(o_t) \sum_{x_{t-1} \in \mathcal{S}} p_{x_{t-1}, x_t} \alpha_x(t-1). \end{aligned}$$

Le résultat souhaité est donc obtenu. □

Remarque 1.1. La fonction de vraisemblance $\mathbb{P}_M(O_{0:T})$ peut être calculée à partir de l'algorithme Forward en utilisant la formule suivante :

$$\mathbb{P}_M(O_{0:T}) = \sum_{x \in \mathcal{S}} \alpha_x(T).$$

• **Algorithme Backward :**

Maintenant, nous définissons la probabilité Backward présentant la probabilité d'observer la suite partielle ultérieure $O_{t+1:T}$, étant donné le modèle M et $X_t = x \in \mathcal{S}$:

$$\beta_x(t) = \mathbb{P}_M(O_{t+1:T} = o_{t+1:T} \mid X_t = x), \quad \text{pour } 0 \leq t < T.$$

Contrairement à la méthode Forward, la valeur de $\beta_x(T)$ est arbitrairement choisie :

$$\beta_x(T) = 1.$$

De même, nous pouvons calculer la probabilité Backward à partir d'une formule de récurrence, présentée dans le théorème suivant.

Théorème 1.5. Pour tout $0 \leq t < T$, $x \in \mathcal{S}$,

$$\begin{cases} \beta_x(T) = 1, \\ \beta_x(t) = \sum_{x' \in \mathcal{S}} \psi_{x'}(o_{t+1}) p_{x,x'} \beta_{x'}(t+1). \end{cases} \quad (1.6)$$

Preuve. Soit $x \in \mathcal{S}$, nous commençons par $t = T - 1$, on a

$$\begin{aligned}
\beta_x(T-1) &= \mathbb{P}_M(O_T = o_T \mid X_{T-1} = x) \\
&= \sum_{x_T \in \mathcal{S}} \mathbb{P}_M(O_T = o_T, X_T = x_T \mid X_{T-1} = x) \\
&= \sum_{x_T \in \mathcal{S}} \mathbb{P}_M(O_T = o_T \mid X_T = x_T) \mathbb{P}_M(X_T = x_T \mid X_{T-1} = x) \\
&= \sum_{x_T \in \mathcal{S}} \psi_{x_T}(o_T) p_{x, x_T}.
\end{aligned}$$

Ensuite, pour tout $0 \leq t < T - 1$,

$$\begin{aligned}
\beta_x(t) &= \mathbb{P}_M(O_{t+1:T} = o_{t+1:T} \mid X_t = x) \\
&= \sum_{x_{t+1:T} \in \mathcal{S}^{T-t}} \mathbb{P}_M(O_{t+1:T} = o_{t+1:T}, X_{t+1:T} = x_{t+1:T} \mid X_t = x) \\
&= \sum_{x_{t+1:T} \in \mathcal{S}^{T-t}} \mathbb{P}_M(O_{t+1:T} = o_{t+1:T} \mid X_{t+1:T} = x_{t+1:T}) \mathbb{P}_M(X_{t+1:T} = x_{t+1:T} \mid X_t = x) \\
&= \sum_{x_{t+1:T} \in \mathcal{S}^{T-t}} \prod_{l=t+1}^T \psi_{x_l}(o_l) p_{x, x_{t+1}} \prod_{l=t+2}^T p_{x_{l-1}, x_l} \\
&= \sum_{x_{t+1} \in \mathcal{S}} \psi_{x_{t+1}}(o_{t+1}) p_{x, x_{t+1}} \sum_{x_{t+2:T} \in \mathcal{S}^{T-t-1}} \prod_{l=t+2}^T \psi_{x_l}(o_l) p_{x_{l-1}, x_l} \\
&= \sum_{x_{t+1} \in \mathcal{S}} \psi_{x_{t+1}}(o_{t+1}) p_{x, x_{t+1}} \sum_{x_{t+2:T} \in \mathcal{S}^{T-t-1}} \mathbb{P}_M(O_{t+2:T} = o_{t+2:T} \mid X_{t+2:T} = x_{t+2:T}) \\
&\quad \mathbb{P}_M(X_{t+2:T} = x_{t+2:T} \mid X_{t+1} = x_{t+1}) \\
&= \sum_{x_{t+1} \in \mathcal{S}} \psi_{x_{t+1}}(o_{t+1}) p_{x, x_{t+1}} \mathbb{P}_M(O_{t+2:T} = o_{t+2:T} \mid X_{t+1} = x_{t+1}) \\
&= \sum_{x_{t+1} \in \mathcal{S}} \psi_{x_{t+1}}(o_{t+1}) p_{x, x_{t+1}} \beta_{x_{t+1}}(t+1)
\end{aligned}$$

Ainsi, la preuve est complète. \square

Proposition 1.1. *La fonction de vraisemblance des observations $\mathbb{P}_M(O_{0:T})$ est donnée par :*

$$\mathbb{P}_M(O_{0:T}) = \sum_{x \in \mathcal{S}} \alpha_x(t) \beta_x(t),$$

où $\alpha_x(t)$ et $\beta_x(t)$ sont les probabilités Forward et Backward calculées à partir de **1.5** et **1.6**.

Preuve. On a :

$$\mathbb{P}_M(O_{0:T}) = \sum_{x \in \mathcal{S}} \mathbb{P}_M(O_{0:T}, X_t = x),$$

où, pour tout $x \in \mathcal{S}$,

$$\begin{aligned}
\mathbb{P}_M(O_{0:T}, X_t = x) &= \sum_{x_{0:t-1} \in \mathcal{S}^t} \sum_{x_{t+1:T} \in \mathcal{S}^{T-t}} \mathbb{P}_M(O_{0:T}, X_{0:T} = x_{0:T}) \\
&= \sum_{x_{0:t-1} \in \mathcal{S}^t} \sum_{x_{t+1:T} \in \mathcal{S}^{T-t}} \rho_{x_0} \psi_{x_0}(o_0) \prod_{t=1}^T \psi_{x_t}(o_t) p_{x_{t-1}, x_t} \\
&= \sum_{x_{0:t-1} \in \mathcal{S}^t} \rho_{x_0} \psi_{x_0}(o_0) \prod_{l=1}^t \psi_{x_l}(o_l) p_{x_{l-1}, x_l} \sum_{x_{t+1:T} \in \mathcal{S}^{T-t}} \prod_{l=t+1}^T \psi_{x_l}(o_l) p_{x_{l-1}, x_l} \\
&= \alpha_x(t) \beta_x(t).
\end{aligned} \tag{1.7}$$

Par conséquent,

$$\mathbb{P}_M(O_{0:T}) = \sum_{x \in \mathcal{S}} \alpha_x(t) \beta_x(t).$$

□

Remarque 1.2. *Le schéma Forward-Backward permet de calculer l'évaluation de la probabilité $\mathbb{P}_M(O_{0:T})$ de manière plus efficace que la méthode naïve donnée par 1.4, en réduisant l'ordre de calcul de $T \times N^T$ à $T \times N^2$. Cette amélioration est obtenue grâce à la combinaison des probabilités Forward et Backward (Pour $N = 4$ et $T = 100$, l'évaluation de $\mathbb{P}(O_{0:100})$ grâce à la procédure forward-backward passe de l'ordre de calcul $T \times N^T = 100 \times 4^{100}$ à $T \times N^2 = 100 \times 4^2 = 1600$. Cela représente une réduction massive du temps de calcul nécessaire pour évaluer la vraisemblance de la séquence d'observations pour ce modèle de HMM).*

Ré-estimation des paramètres du HMM

Une fois que nous avons déterminé la vraisemblance de notre modèle HMM pour une séquence d'observations donnée, nous pouvons ré-estimer les paramètres du modèle pour améliorer sa performance. Cette étape est appelée la ré-estimation des paramètres du HMM et se base sur l'algorithme de Baum-Welch [25].

L'idée générale de l'algorithme de Baum-Welch est d'ajuster les paramètres du modèle de telle sorte que la probabilité d'observer la séquence d'observations soit maximisée. Il s'agit d'un processus itératif qui commence par une initialisation des paramètres, puis ré-estime à chaque itération les paramètres en utilisant l'algorithme Expectation-Maximization (EM).

La première étape de l'algorithme EM consiste à calculer une fonction objectif appelée fonction de vraisemblance attendue définie par :

$$Q(M, M^{(n)}) = \sum_{x_{0:T}} P(X_{0:T} = x_{0:T} \mid O_{0:T}, M^{(n)}) \log P(O_{0:T}, X_{0:T} = x_{0:T} \mid M), \quad (1.8)$$

où M représente les paramètres du HMM, $M^{(n)}$ est une estimation courante de ces paramètres.

La deuxième étape consiste à maximiser cette fonction de vraisemblance attendue par rapport aux paramètres du HMM M , ce qui donne la nouvelle estimation $M^{(n+1)}$:

$$M^{(n+1)} = \arg \max_M Q(M, M^{(n)}).$$

Ces deux étapes sont répétées jusqu'à convergence, c'est-à-dire lorsque la variation entre les estimations successives est suffisamment petite. Ainsi, l'algorithme EM permet de trouver les paramètres optimaux du HMM qui maximisent la vraisemblance des observations.

Pour calculer explicitement les ré-estimations de Baum-Welch, nous définissons les deux probabilités suivantes :

$$\xi_{x,x'}(t) = \mathbb{P}_M(X_{t+1} = x', X_t = x \mid O_{0:T} = o_{0:T}),$$

et

$$\gamma_x(t) = \sum_{x' \in \mathcal{S}} \xi_{x,x'}(t) = \mathbb{P}_M(X_t = x \mid O_{0:T} = o_{0:T}).$$

En utilisant les probabilités calculées par l'algorithme Forward-Backward $\alpha_x(t)$ et $\beta_x(t)$, Ces deux probabilité peuvent être exprimées de la manière suivante :

$$\begin{aligned}\xi_{x,x'}(t) &= \frac{\mathbb{P}_M(X_{t+1} = x', X_t = x, O_{0:T} = o_{0:T})}{\mathbb{P}_M(O_{0:T} = o_{0:T})} \\ &= \frac{\alpha_x(t)p_{x,x'}\psi_{x'}(o_{t+1})\beta_{x'}(t+1)}{\sum_{x,x' \in \mathcal{S}} \alpha_x(t)p_{x,x'}\psi_{x'}(o_{t+1})\beta_{x'}(t+1)}.\end{aligned}$$

Pour la deuxième probabilité, d'après **1.7**, on obtient :

$$\begin{aligned}\gamma_x(t) &= \frac{\mathbb{P}_M(X_t = x, O_{0:T} = o_{0:T})}{\mathbb{P}_M(O_{0:T} = o_{0:T})} \\ &= \frac{\alpha_x(t)\beta_x(t)}{\sum_{x \in \mathcal{S}} \alpha_x(t)\beta_x(t)}.\end{aligned}$$

En sommant les coefficients $\xi_{x,x'}(t)$ et $\gamma_x(t)$ sur les instants $t = 0, \dots, T-1$, on obtient des estimateurs du nombre de transitions de l'état x à l'état x' et du nombre de transitions de l'état x dans la suite $X_{0:T}$, respectivement. Ces estimateurs sont basés sur les observations de la suite $O_{0:T} = o_{0:T}$.

Par conséquent, les ré-estimateurs de Baum-Welch qui maximise la fonction **1.8**, étant donné le modèle $M^{(n)}$ sont donnés par

$$\begin{aligned}p_{x,x'}^{(n+1)} &= \frac{\sum_{t=0}^{T-1} \xi_{x,x'}^{(n)}(t)}{\sum_{t=0}^{T-1} \gamma_x^{(n)}(t)}, \\ \psi_x^{(n+1)}(o) &= \frac{\sum_{t=0}^T \mathbb{1}_{o_t=o} \gamma_x^{(n)}(t)}{\sum_{t=0}^T \gamma_x^{(n)}(t)}, \\ \rho_x^{(n+1)} &= \gamma_x^{(n)}(0).\end{aligned}$$

L'algorithme de Baum-Welch se termine lorsque la vraisemblance des observations converge vers un maximum local, ce qui indique que les paramètres ont été ajustés de manière optimale pour la suite d'observations donnée. La qualité de la solution obtenue dépend de la qualité de l'initialisation des paramètres et de la qualité de la suite d'observations utilisée pour l'entraînement.

Cet algorithme sera étendu et appliqué à différents cas d'étude dans les chapitres 3 et 4.

1.5 État de l'art sur la modélisation mathématique de la fièvre typhoïde

La modélisation de la transmission de la fièvre typhoïde est un sujet de recherche important, abordé par de nombreux chercheurs et mathématiciens tels que Lauria et al.[**26**], Kalajdziewska et al. [**27**], Mushayabasa et al.[**28**], Abboubakar et al. [**29**] et Pitzer et al.[**30**]. Cependant, malgré les nombreux travaux existants, l'influence de la qualité de l'eau sur la transmission de la maladie n'a pas été suffisamment considérée [**28, 29, 30**] et aucun modèle ne prend en compte deux populations en interaction avec accès ou non à l'eau potable. De plus, les modèles classiques utilisés sont souvent des modèles compartimentaux déterministes (**1.4.1**) du type S, I, R et C (le compartiment des porteurs

chroniques), qui supposent que toutes les variables d'entrée sont des fonctions déterministes du temps. Ces modèles ne permettent pas de prendre en compte les variations aléatoires de certains facteurs qui peuvent influencer la dynamique de transmission de la maladie.

Plusieurs travaux ont été réalisés dans ce cadre. Par exemple, dans [27], les auteurs ont développé un modèle mathématique pour évaluer les effets des porteurs chronique (porteurs de germes, demeurant asymptomatique, pendant une très longue période, parfois jusqu'à la mort) sur la dynamique de transmission des maladies infectieuses telles que la typhoïde. Ils ont conclu que les porteurs jouent un rôle important dans la transmission des maladies infectieuses. **Mushayabasa et al.** [28] ont étudié un modèle épidémiologique pour la transmission directe (personne à personne) et indirecte (par l'eau) de la fièvre typhoïde. Ils ont montré à partir d'une analyse de sensibilité que la transmission indirecte de la typhoïde a plus d'impact sur la détermination de la prévalence de la maladie que la transmission directe. Ils ont également souligné l'importance de prendre en compte la qualité de l'eau dans les modèles de transmission de la maladie. **Abboubakar et al.**[29] présentent un modèle mathématique déterministe qui prend en compte l'effet de l'environnement sur la transmission de la fièvre typhoïde, et une stratégie de contrôle optimale pour réduire la propagation de la maladie. Les auteurs ont utilisé des données épidémiologiques pour paramétrer le modèle et une analyse de sensibilité pour identifier les paramètres les plus importants pour la transmission de la maladie. Ils ont également établi un problème de contrôle optimal pour étudier l'impact des différentes stratégies de vaccination sur la transmission de la maladie et ont conclu que la vaccination généralisée était plus efficace pour réduire la prévalence de la maladie que la vaccination ciblée sur des groupes à risque élevé.

Pitzer et al.[30] ont pris en compte l'impact de la qualité de l'eau sur la propagation de la typhoïde en élaborant un modèle parcimonieux structuré par âge qui décrit l'histoire naturelle et l'immunité à l'infection typhoïde. Ils ont montré que l'amélioration de la qualité de l'eau peut avoir un impact significatif sur la prévention de la transmission de la typhoïde et ont évalué l'efficacité de différentes stratégies de vaccination en milieu scolaire utilisant des vaccins vivants oraux, des vaccins polysaccharidiques et des vaccins conjugués.

Aucun de ces modèles présentés ci-dessus ne peut décrire de manière adéquate la propagation de la typhoïde dans le contexte particulier de Mayotte, vu qu'ils ne considèrent pas les spécificités de Mayotte, comme l'accès aux BFM limité à certaines zones géographiques ainsi que de la pression migratoire que subit l'île. Cependant, ces modèles restent pertinents pour servir de références permettant d'identifier les ordres de grandeur de chaque paramètre de notre modèle décrit dans la section 2.2.

1.6 Organisation de la thèse

Dans cette section, nous allons présenter l'organisation des chapitres de cette thèse.

Le chapitre 2 propose un modèle initial à compartiments pour caractériser et décrire la dynamique de transmission de la fièvre typhoïde à Mayotte, en prenant en compte le contexte spécifique de cette maladie hydrique ainsi que les particularités de la population mahoraise. Le modèle est présenté dans ses versions déterministe et stochastique (PDMP), et est exploré tant sur le plan théorique que numérique, avec des simulations

de trajectoires dans différents scénarios, ainsi que l’analyse de leur comportement qualitatif. En outre, nous évaluons la virulence de la maladie en utilisant le paramètre clé R_0 représentant le nombre de reproduction de base.

Le jeu de données fourni par l’ARS de Mayotte est restreint et ne permet pas une calibration précise de notre modèle initial, ni une quantification précise de la propagation épidémique à Mayotte. Dans le chapitre **3**, nous proposons donc un modèle stochastique simplifié à un compartiment de type naissance et mort avec immigration **1.4.3**, qui permet de décrire les effectifs de la population infectée. Nous utilisons cette approche afin d’obtenir un modèle ajusté aux données disponibles et d’estimer les paramètres clés de la dynamique de la fièvre typhoïde à Mayotte en adaptant un modèle de Markov caché **1.4.4** à notre contexte particulier. La procédure d’estimation et les principales propriétés des estimateurs sont établies théoriquement et illustrées numériquement sur des données synthétiques ainsi que sur des données réelles provenant de Mayotte.

Dans les chapitres **4** et **5**, nous enrichissons notre modèle décrit dans le chapitre **3** en proposant deux extensions. Tout d’abord, dans le chapitre **4**, nous introduisons un nouveau compartiment dans le modèle, qui correspond aux individus exposés à la bactérie Salmonella Typhi. Cette nouvelle extension nous permet de prendre en compte la durée d’incubation de la maladie. Les propriétés théoriques de ce nouveau modèle, appelé modèle infecté-exposé, sont étudiées et illustrées par des simulations numériques. Nous présentons également une nouvelle méthodologie d’estimation des paramètres du modèle bivarié en utilisant un modèle de Markov multi-chaîne cachée. Les résultats obtenus sont illustrés par des études numériques.

La deuxième extension de notre modèle présentée dans le chapitre **5** vise à étudier l’impact de la pluviométrie sur la transmission de la maladie à Mayotte, en exploitant des données publiques de précipitations. Pour cela, nous avons utilisé une méthode de détection de rupture afin de sélectionner et caractériser la saisonnalité de la pluie. Les résultats d’estimation des paramètres de notre modèle sous les deux régimes mettent en évidence l’importance de cette variable météorologique sur la propagation de l’épidémie, soulignant la nécessité de prendre en compte cet effet dans la planification des interventions de santé publique pour lutter contre la fièvre typhoïde à Mayotte.

Le chapitre **6** conclut ce manuscrit en explorant les pistes de recherche futures.

Chapitre 2

Modèle à compartiments pour la transmission de la fièvre typhoïde à Mayotte

Sommaire

2.1	Introduction	33
2.2	Modélisation déterministe	33
2.2.1	Formulation du modèle	34
2.2.2	Calibration des paramètres	36
2.2.3	Discussion et simulations numériques	38
2.2.4	Le nombre de reproduction de base	45
2.3	Modélisation par un processus markovien déterministe par morceaux	48
2.3.1	Formulation du modèle	49
2.3.2	Discussion et simulations numériques	50
2.4	Conclusion	54

2.1 Introduction

La fièvre typhoïde est une maladie infectieuse d'origine bactérienne, causée par la *Salmonella typhi*, qui se propage principalement par la consommation d'eau ou d'aliments contaminés. À Mayotte, territoire français d'outre-mer, la fièvre typhoïde constitue un problème de santé publique majeur, classant l'île parmi les foyers endémiques de la maladie en France. La population mahoraise, en particulier les individus les plus vulnérables, sont confrontés à des défis importants liés à l'accès limité à une eau potable de qualité et à des systèmes d'assainissement efficaces **1.2.3**.

Ce chapitre constitue une première étape importante dans notre travail de modélisation de la fièvre typhoïde à Mayotte, visant à mieux comprendre les mécanismes sous-jacents de propagation de la maladie dans ce contexte spécifique. Dans ce chapitre, nous proposons un nouveau modèle à compartiments qui prend en compte l'impact de la qualité de l'eau sur la dynamique de la fièvre typhoïde à Mayotte.

Le modèle que nous présentons comprend deux populations distinctes, chacune ayant un accès à une eau de qualité différente. Nous explorons les variations de l'évolution de la maladie dans ces deux populations, en tenant compte des caractéristiques sanitaires spécifiques à Mayotte. Nous examinons le comportement du modèle dans deux versions : déterministe et stochastique, afin d'analyser la dynamique de la maladie dans différents scénarios.

L'organisation du chapitre est comme suit : dans la section **2.2**, nous présentons la version déterministe de notre modèle et décrivons la formulation du modèle à compartiments. Nous effectuons ensuite des simulations pour caractériser le comportement qualitatif du modèle dans différents scénarios. Nous évaluons également la virulence de la maladie en utilisant le paramètre du nombre de reproduction de base, qui nous permet de quantifier la capacité de la maladie à se propager dans la population.

Dans la section **2.3**, nous présentons la version stochastique du modèle, basée sur un processus markovien déterministe par morceaux (PDMP), qui permet de prendre en compte la variabilité intrinsèque des systèmes biologiques et d'étudier la dynamique de la maladie dans des conditions plus réalistes. Nous effectuons ensuite des simulations des trajectoires du PDMP pour différents valeurs des paramètres du modèle.

Enfin, dans la section **2.4**, nous concluons ce chapitre en résumant les résultats obtenus.

2.2 Modélisation déterministe

Comme évoqué dans l'état de l'art sur la modélisation mathématique de la fièvre typhoïde **1.5**, de nombreux modèles déterministes ont été proposés pour décrire la dynamique de propagation de cette maladie hydrique (voir par exemple [28]). Cependant, très peu de ces modèles prennent explicitement en compte la qualité de l'eau [30], et aucun ne considère spécifiquement deux populations interagissant avec un accès à des eaux de qualité différente.

Notre modèle déterministe, développé dans cette section, tient compte des particularités épidémiologiques de la fièvre typhoïde à Mayotte, notamment de deux populations présentant des caractéristiques sanitaires distinctes et ayant accès à des eaux de qualité différente, en adéquation avec les caractéristiques propres à la population mahoraise.

2.2.1 Formulation du modèle

Dans cette section, un modèle mathématique déterministe à compartiments pour décrire la dynamique de transmission de la fièvre typhoïde est formulé de manière à étendre et à compléter ceux qui existent dans la littérature [4]-[6], [26]-[30]. Le modèle subdivise la population humaine en quatre types de compartiments : S susceptible, I infecté (personnes infectieuses présentant des signes cliniques), C porteur infecté (porteurs chroniques infectieux sans signes cliniques) et R retiré.

Le modèle (**Fig.2.1**) prend en compte une transmission directe de la typhoïde des individus infectés aux individus susceptibles. Cependant, la typhoïde est en grande partie transmise par des bactéries environnementales par l'intermédiaire d'eau, d'aliments, de boissons contaminées. Pour intégrer ce phénomène biologique, nous considérons un compartiment supplémentaire W , qui représente la présence des bactéries dans l'eau.

Nous subdivisons notre population susceptible en deux sous-groupes :

S1 : Population ayant un accès à l'eau de bonne qualité où la contamination se produit à un taux $\beta_p(I + \omega C)$, c'est-à-dire que la transmission de la maladie se fait d'une manière directe (Personne à Personne avec un taux de transmission spécifique β_p) entre les individus infectés ou chroniques et les individus susceptibles.

S2 : Population n'ayant pas l'accès à l'eau de bonne qualité où la contamination se produit à un taux $\beta_p(I + \omega C) + \beta_w W$ où $\beta_w W$ est la force d'infection véhiculée par l'eau, c'est-à-dire que la maladie est transmise par des bactéries environnementales par l'intermédiaire d'eau en plus du contact direct de personne à personne.

Nous supposons que les échanges démographiques entre ces deux parties de la population sont négligeables mais physiquement elles sont complètement mélangées et donc en contact direct l'une avec l'autre à Mayotte, ce qui explique aussi le choix de ne pas séparer les I, C, R en deux catégories.

Les individus de la classe infectée restent contagieux pendant une période de $1/\delta$. Nous supposons que les individus retirés sont immunisés à vie. Une petite fraction ($\theta \in [0, 1]$) deviennent des porteurs chroniques (à vie) et asymptomatiques de la typhoïde (C) qui contribuent à la transmission à un taux relatif $\omega < 1$ par rapport aux infectieux. Le reste des individus connaissent une mortalité induite par la maladie ($\alpha \in [0, 1]$). Les porteurs chroniques et les personnes infectées excrètent tous deux des bactéries dans l'environnement. Toutefois, le taux d'excrétion γ par le groupe infectieux est supérieur à celui du groupe de chroniques $\gamma\omega$, cela s'explique par le fait que les porteurs infectieux ne présentent aucun signe d'infection. Les bactéries infectieuses perdent leur viabilité (c'est-à-dire qu'elles se dégradent dans l'eau) à un taux ξ .

Nous considérons que les individus naissent (à un taux b) susceptibles d'être infectés par la typhoïde avec une probabilité $p \in [0, 1]$ d'être dans le compartiment $S1$ et une probabilité $1 - p$ d'être en $S2$. Le taux de mortalité naturelle est noté par μ . Notons par $N := S1 + S2 + I + C + R$ la taille de la population totale.

À partir des hypothèses ci-dessus, nous formulons le système d'équations différentielles suivant :

$$\begin{cases} \frac{dS_1}{dt} = pbN - \mu S_1 - \beta_p (I + \omega C) S_1 \\ \frac{dS_2}{dt} = (1-p)bN - \mu S_2 - \left(\beta_p (I + \omega C) + \beta_w W \right) S_2 \\ \frac{dI}{dt} = -(\mu + \delta) I + \beta_p (I + \omega C) (S_1 + S_2) + \beta_w W S_2 \\ \frac{dC}{dt} = -\mu C + \delta \theta I \\ \frac{dR}{dt} = -\mu R + \delta (1 - \alpha - \theta) I \\ \frac{dW}{dt} = -\xi W + \gamma (I + \omega C) \end{cases} \quad (2.1)$$

où les variables et les paramètres sont résumés dans les **tables 2.1** et **2.2**. Le diagramme en compartiments correspondant à ce système est donné par la **figure 2.1**.

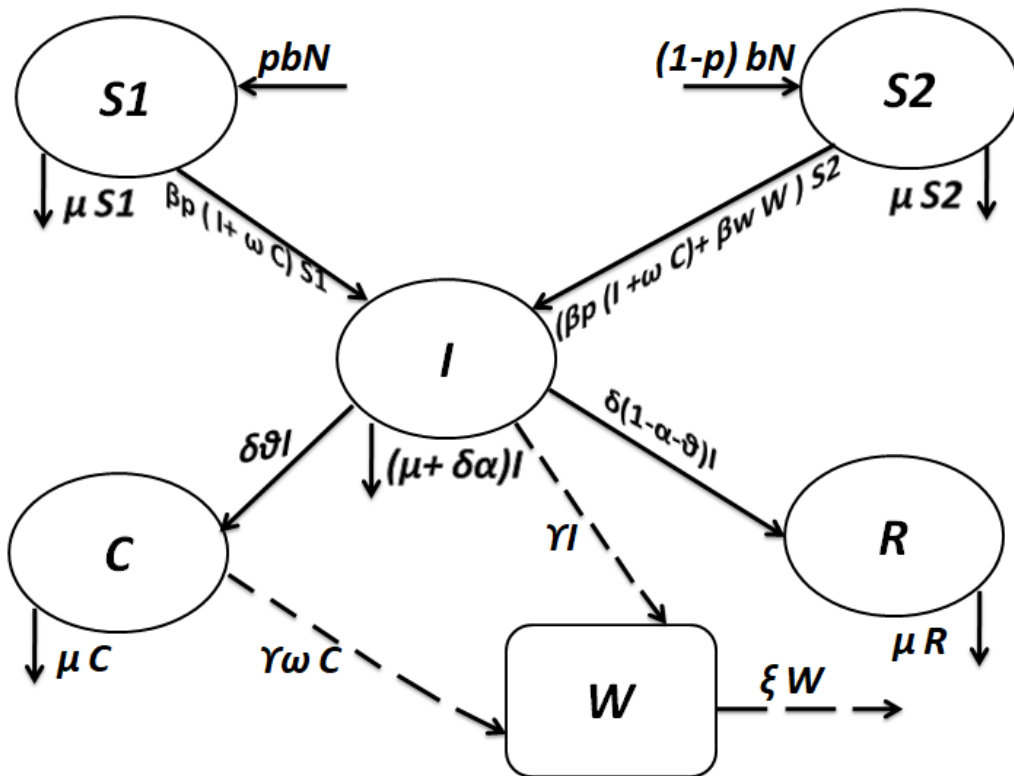


FIGURE 2.1 – Diagramme de la structure du modèle.

TABLE 2.1 – *Description des variables d'état du modèle épidémique.*

Variable d'Etat	Description
S_1	Nombre de personnes susceptibles ayant un accès à l'eau de bonne qualité
S_2	Nombre de personnes susceptibles n'ayant pas un accès à l'eau de bonne qualité
I	Nombre de personnes infectieuses présentant des signes cliniques
C	Nombre de personnes infectieuses sans signes cliniques (porteurs chroniques)
R	Nombre de personnes ayant une immunité permanente
N	Taille de la population totale
W	Concentration de bactéries dans l'eau en UFC/ml [31]

2.2.2 Calibration des paramètres

La calibration des paramètres d'un modèle est une étape importante pour obtenir des résultats de simulation qui correspondent au mieux aux données réelles. Pour notre modèle décrivant la propagation de la fièvre typhoïde à Mayotte (voir **Fig.2.1**),

► Le taux de natalité b et le taux de mortalité naturelle μ sont souvent les plus faciles à calibrer car ils sont généralement bien documentés. À Mayotte, ils sont respectivement de 38.9 et 2.9 pour 1000 habitants [9].

Les autres paramètres du modèle présentés dans la table **2.2** ne sont pas identifiés dans la littérature disponible à Mayotte.

► La durée de la contagiosité peut varier en fonction de facteurs tels que l'âge, la souche de la bactérie *Salmonella Typhi*, la gravité de l'infection et le traitement antibiotique administré. Les valeurs peuvent donc varier en fonction de la population étudiée. Elle a été étudiée dans plusieurs pays d'Afrique et d'Asie, et les valeurs rapportées varient dans une plage de 1 à 5 semaines [30, 32, 33].

► Les taux de transmission directe et indirecte de la fièvre typhoïde varient considérablement selon les contextes épidémiologiques et les populations étudiées. Dans la littérature [34, 35], les valeurs de β_p peuvent varier entre 0.1 et 2.3 par personne et par jour, tandis que les valeurs de β_w peuvent varier entre 0.01 et 0.6 par personne et par jour. Cependant, il est important de noter que ces valeurs sont souvent spécifiques à chaque contexte et doivent être ajustées en fonction des caractéristiques de la population étudiée.

► Le taux d'excrétion dans l'approvisionnement en eau (γ) et le taux de dégradation des particules infectieuses dans l'eau (ξ) sont des paramètres qui dépendent de plusieurs facteurs tels que la qualité de l'eau, la présence d'autres microorganismes dans l'eau, etc. Il peut donc être plus difficile de trouver des données précises pour les calibrer. Pour donner un ordre de grandeur, certaines études ont rapporté des valeurs pour γ allant de 0.01 à 1.5 par jour [30, 36], tandis que pour ξ , les valeurs varient entre 0.01 et 1.5 par jour

TABLE 2.2 – *Description des paramètres du modèle.*

Les Paramètres	Description
b	Taux de natalité
p	Probabilité que les nouveau-nés soient en S1
μ	Taux de mortalité naturelle
$1/\delta$	Durée de la contagiosité
β_p	Taux de transmission directe (de personne à personne)
β_w	Taux de transmission indirecte (transmise par l'eau)
α	Fraction des infectés qui connaissent une mortalité induite par la maladie
θ	Fraction des infectés qui deviennent porteurs chroniques
ω	Infectiosité relative des malades chroniques (Contribution des porteurs chroniques dans la propagation)
γ	Taux d'excrétion dans l'approvisionnement en eau
ξ	Taux de dégradation des particules infectieuses dans l'eau

[36, 37]. Cependant, ces valeurs doivent être considérées avec prudence car elles peuvent varier considérablement selon les conditions locales.

► les paramètres α , θ et ω sont des paramètres spécifiques au modèle et à la maladie étudiée. Les valeurs de ces paramètres peuvent varier selon les différentes populations étudiées et les caractéristiques de la maladie.

En général, la fraction des infectés qui connaissent une mortalité induite par la maladie (α) peut varier en fonction de la gravité de la maladie et de l'accès aux soins médicaux. Pour la fièvre typhoïde, la mortalité peut varier entre 0.1% et 10% en fonction des conditions sanitaires et de l'accès aux soins [38].

La fraction des infectés qui deviennent porteurs chroniques (θ) peut également varier selon la population étudiée et les caractéristiques de la maladie. Pour la fièvre typhoïde, la proportion de porteurs chroniques peut varier de 1% à 5% [39].

Enfin, l'infectiosité relative des malades chroniques (ω) peut être difficile à quantifier, car elle dépend également de facteurs tels que la dose infectieuse et la durée de l'infection chronique. Les valeurs rapportées dans la littérature pour ce paramètre peuvent varier de 0.5% à 40% [40, 30].

Il convient donc de bien définir ces paramètres et de trouver des données spécifiques à la population étudiée pour pouvoir les calibrer correctement. Ainsi, des études futures seront nécessaires pour estimer ces paramètres afin de calibrer notre modèle général pour la propagation de la fièvre typhoïde à Mayotte. Toutefois, il est possible de réaliser des simulations numériques sous différents scénarios pour caractériser qualitativement le comportement de la transmission de cette maladie hydrique.

2.2.3 Discussion et simulations numériques

En nous basant sur la partie précédente portant sur la calibration des paramètres, nous effectuons des simulations numériques pour illustrer le comportement du système (2.1) sous trois scénarios différents. Les simulations ont été effectuées en utilisant le logiciel R en implémentant un schéma d'Euler, dont le pseudo-code (2.2.3) est présenté dans l'algorithme ci-dessous :

Algorithme 1 : Algorithme d'Euler pour le système (2.1)

entrée : paramètres du modèles : $b, \mu, p, \beta_p, \beta_w, \omega, \alpha, \delta, \theta, \gamma, \xi$.

valeurs initiales des variables : $s_1^0, s_2^0, i^0, c^0, r^0, w^0$.

horizon de simulation H , pas de discrétisation en temps Δt .

sortie : Trajectoire $(X_t = (S1(t), S2(t), I(t), C(t), R(t), W(t)))_{0 \leq t \leq H}$ par un pas de Δt

1 début

2 Initialisation

3 $(S1(0), S2(0), I(0), C(0), R(0), W(0)) \leftarrow (s_1^0, s_2^0, i^0, c^0, r^0, w^0)$

4 $N(0) \leftarrow S1(0) + S2(0) + I(0) + C(0) + R(0)$

5 pour h de 0 à H par pas de Δt faire

6 $S1(h + \Delta t) \leftarrow S1(h) + \Delta t (pbN(h) - \mu S1(h) - \beta_p(I(h) + \omega C(h)) S1(h))$

7 $S2(h + \Delta t) \leftarrow S2(h) + \Delta t ((1 - p)bN(h) - \mu S2(h) - (\beta_p(I(h) + \omega C(h)) + \beta_w W(h)) S2(h))$

8 $I(h + \Delta t) \leftarrow I(h) + \Delta t (- (\mu + \delta)I(h) + (\beta_p(I(h) + \omega C(h))) (S1(h) + S2(h)) + \beta_w W(h) S2(h))$

9 $C(h + \Delta t) \leftarrow C(h) + \Delta t (- \mu C(h) + \delta \theta I(h))$

10 $R(h + \Delta t) \leftarrow R(h) + \Delta t (- \mu R(h) + \delta (1 - \alpha - \theta) I(h))$

11 $W(h + \Delta t) \leftarrow W(h) + \Delta t (- \xi W(h) + \gamma (I(h) + \omega C(h)))$

12 $N(h + \Delta t) \leftarrow S1(h + \Delta t) + S2(h + \Delta t) + I(h + \Delta t) + C(h + \Delta t) + R(h + \Delta t)$

13 fin

14 renvoyer : $(X_t = (S1(t), S2(t), I(t), C(t), R(t), W(t)))_{0 \leq t \leq H}$

15 fin

Premier jeu des paramètres : évolution de l'épidémie sans prise en compte de la démographie.

Dans un premier temps, nous considérons que la démographie de la population est négligée par rapport à l'évolution de l'épidémie, ce qui signifie que l'impact de l'expansion démographique sur la propagation de la typhoïde est ignoré ($b = \mu = 0$). Cette simplification permet de se concentrer sur les dynamiques de transmission de la maladie sans être influencé par des facteurs externes.

En ce qui concerne les compartiments, la **figure 2.2** représente la population comme étant divisée en plusieurs catégories, chacune étant caractérisée par son état de santé. Les robinets de naissance et de mort sont supposés être fermés, ce qui signifie que la taille de la population reste constante tout au long de la simulation. Initialement, tous les individus sont sains, ce qui est représenté par le premier compartiment en haut $S1 + S2$.

Lorsque les robinets $S1/S2$ et I sont ouverts, l'ensemble de la population s'écoule vers les compartiments C et R , tandis qu'une petite fraction des individus infectés s'écoule également vers le compartiment D (Morts), ce qui reflète la mortalité associée à la maladie. Il est important de noter que, tant que $C \neq 0$ ou $W \neq 0$, le robinet $S1/S2$ restera ouvert, même si $I = 0$. Cependant, la propagation de l'épidémie ainsi que sa durée dépendent essentiellement du flux d'entrée/sortie des individus infectés. En effet, plus le flux de sortie

est faible par rapport à l'entrée, plus l'épidémie sera aiguë et sa disparition prendra plus de temps.

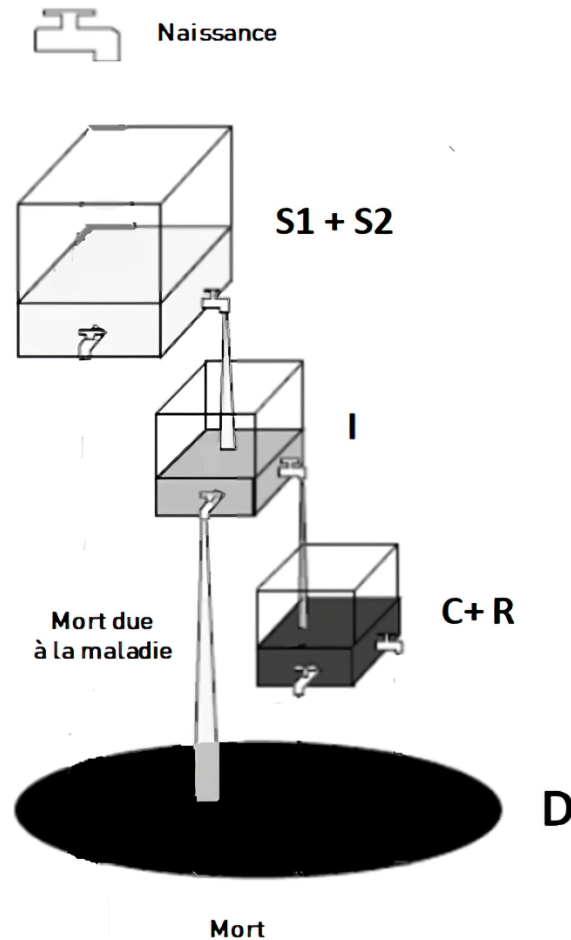


FIGURE 2.2 – Représentation schématique d'un modèle à compartiment sans démographie.

Dans le cadre de cette simulation, la population initiale est composée de 300000 individus (ordre de grandeur de la population mahoraise) qui sont susceptibles d'être infectés à l'instant $t = 0$. Parmi ces individus, 30% (soit 90000 personnes) n'ont pas accès à de l'eau de bonne qualité, ce qui augmente leur risque d'infection. De plus, nous avons introduit 5 porteurs chroniques dans la population et avons fixé la concentration de bactéries dans l'eau à 20 UFC/ml, et le nombre initial de personnes infectées et retirées à 0. Tout au long des simulations, nous avons utilisé un pas de discrétisation Δt de 0.001. Les autres paramètres du modèle sont choisis et fixés selon 2.2.2 et sont répertoriés dans la table 2.3.

TABLE 2.3 – *Premier jeu de paramètres.*

Les paramètres	β_p	β_w	ω	δ	θ	α	ξ	γ
Valeurs numériques (1/ semaine)	0.8	1.8	0.07	0.25	0.03	0.001	0.8	2

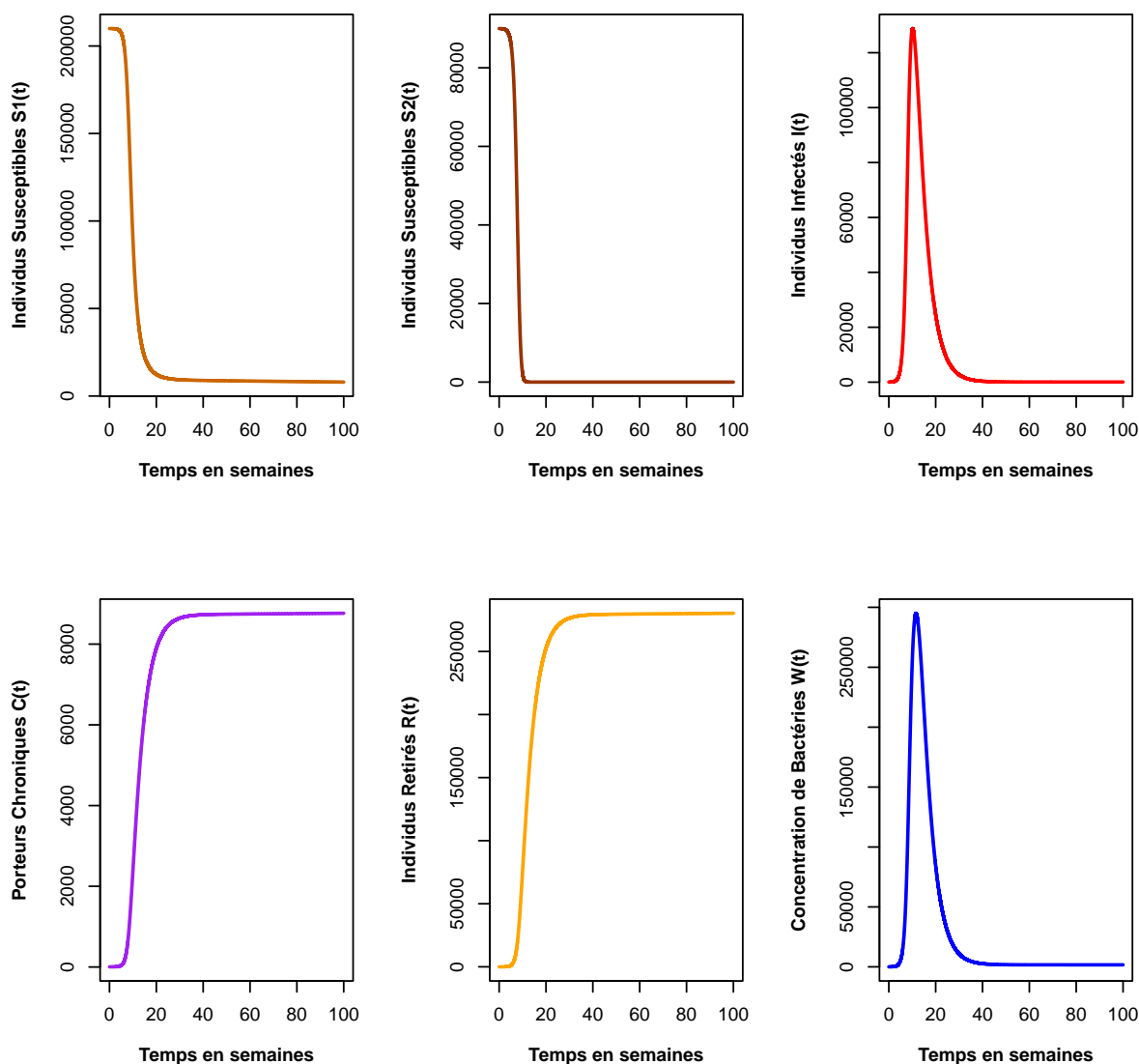


FIGURE 2.3 – **Évolution des populations $S1, S2, I, C, R$ et la concentration de bactéries dans l'eau W sur un horizon de 100 semaines.**

Au fil du temps, le nombre d'individus infectés augmente rapidement jusqu'à atteindre un pic de 130000 personnes (soit 43.3% de la population totale) lors de la 18ème semaine, puis diminue graduellement jusqu'à atteindre zéro en 50 semaines. Au cours de cette période, le nombre de personnes saines tend vers zéro dans les deux sous-ensembles $S1$ et $S2$. À la fin de l'épidémie, la population se trouve dans deux états distincts : les porteurs chroniques (représentant 3% de la population) et les retirés (représentant 97% de

la population). En outre, la concentration de bactéries dans l'eau augmente rapidement, atteignant un maximum de 300000 UFC/ml après 18 semaines, avant de diminuer et de se stabiliser à un seuil de 4479 UFC/ml. Malgré la propagation rapide de la maladie, le nombre de décès liés à la typhoïde reste faible, avec seulement 200 décès à la fin de la propagation (**Fig.2.3**).

Il convient de noter que l'épidémie a pu se propager dans la population même en l'absence d'individus infectés à $t = 0$, en raison de deux sources d'infectiosité : les porteurs chroniques et la concentration de bactéries dans l'eau.

Enfin, il est important de souligner que la population S_2 décroît plus rapidement que la population S_1 en raison de l'effet supplémentaire de la qualité de l'eau (représentée par le paramètre W). Cette observation met en évidence l'importance de la qualité de l'eau dans la propagation de l'épidémie de typhoïde et souligne la nécessité de mettre en place des mesures efficaces pour prévenir la contamination de l'eau.

Deuxième jeu de paramètres

Nous maintenons les valeurs des paramètres ξ et γ , qui décrivent la propagation de la fièvre typhoïde par l'eau, et augmentons les autres paramètres, à l'exception de δ qui est réduit afin d'obtenir une durée de contagiosité assez longue. Les valeurs modifiées des paramètres sont répertoriées dans la **table 2.4** ci-dessous.

TABLE 2.4 – *Deuxième jeu de paramètres.*

Les paramètres	β_p	β_w	ω	δ	θ	α	ξ	γ
Valeurs numériques (1/ semaine)	1.2	2	0.4	0.15	0.05	0.01	0.8	2

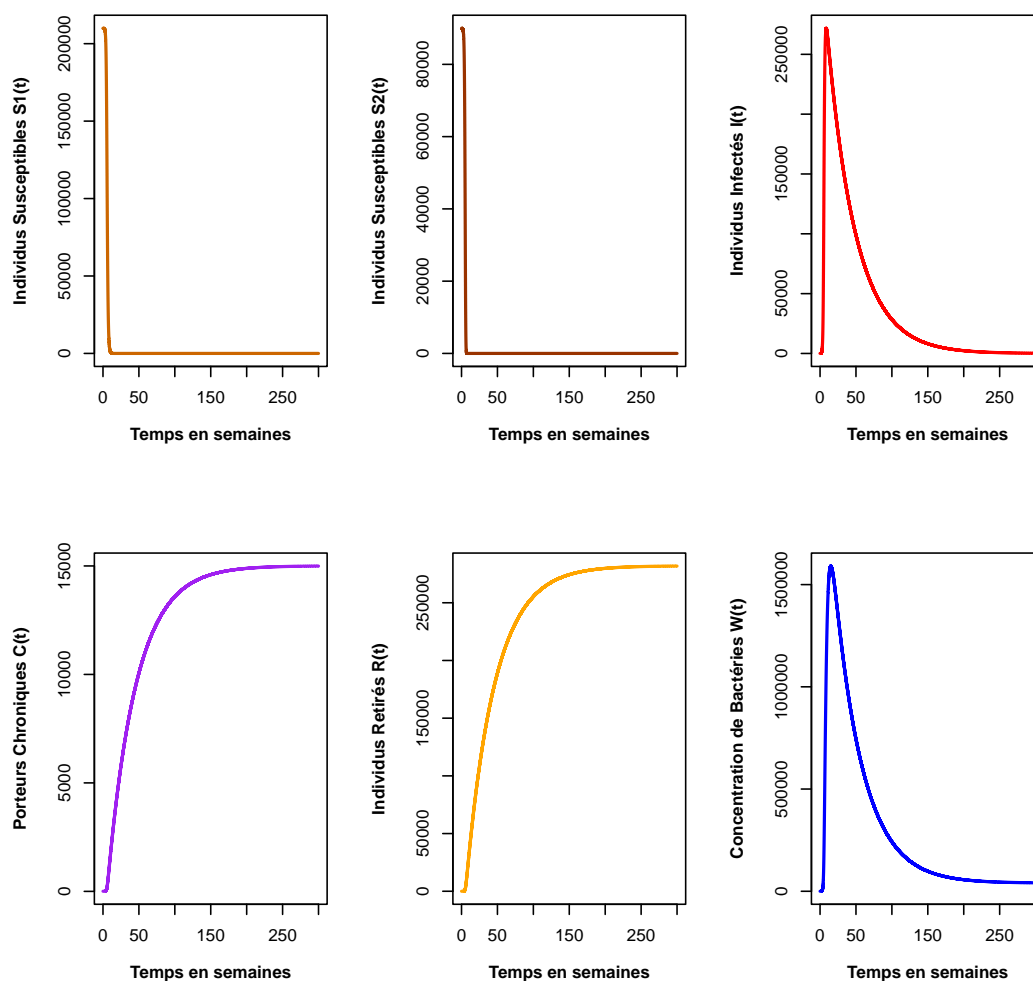


FIGURE 2.4 – Évolution des populations S_1, S_2, I, C, R et la concentration de bactéries dans l'eau W sur un horizon de 300 semaines.

En partant des mêmes conditions initiales, l'analyse des résultats montre que l'épidémie se propage de manière similaire à l'exemple précédent, conformément aux attentes (Fig.2.4). Cependant, la durée de la maladie est considérablement plus longue, avec un nombre d'individus infectés qui tend vers zéro après près de 250 semaines. De plus, la décroissance est beaucoup plus lente, tandis que le pic épidémique est plus élevé et se produit beaucoup plus rapidement, avec un nombre maximal de 272000 individus infectés (soit 90.6% de la population). Au fil du temps, la population évolue progressivement vers les deux états : les porteurs chroniques (5%) et les retirés (93.9%). Toutefois, un petit pourcentage d'individus subit une mortalité induite par la typhoïde, avec un total de 3000 décès enregistrés au cours de 300 semaines.

Troisième jeu de paramètres : La dimension démographique de la pandémie

Lorsque la démographie de la population est prise en compte, avec un taux de natalité élevé, le comportement asymptotique de l'épidémie est considérablement altérée (Fig.2.5). Tout d'abord, le nombre d'individus infectés et la concentration de bactéries dans l'eau augmentent rapidement, puis diminuent, ce qui est similaire au modèle sans

démographie. Cependant, l'expansion rapide de la population a un impact significatif sur la propagation de l'épidémie, car les naissances font réaugmenter rapidement les individus infectés, maintenant ainsi les compartiments $S1$ et $S2$ remplis d'un côté, et faisant exploser le nombre d'individus infectés (et la concentration de bactéries) de l'autre côté. Les porteurs chroniques, les individus guéris et la taille de la population totale augmentent de manière exponentielle.

Les valeurs des paramètres utilisés dans ce scénario sont indiquées dans le tableau suivant **Table.2.5**. Il convient de noter que nous avons fixé le pourcentage de partage des nouveau-nés entre les deux populations $S1$ et $S2$ à 50% ($p = 0.5$).

TABLE 2.5 – *Troisième jeu de paramètres.*

Les paramètres	b	μ	β_p	β_w	ω	δ	θ	α	ξ	γ
Valeurs numériques (1/ semaine)	$38.9 \cdot 10^{-3}/52$	$2.9 \cdot 10^{-3}/52$	0.8	1.8	0.07	0.15	0.03	0.001	0.8	2

En d'autres termes, ces modèles mettent en évidence l'interaction complexe entre la dynamique de la population et la propagation de la maladie, soulignant ainsi l'importance de la démographie et de la croissance de la population dans la propagation de la fièvre typhoïde. L'expansion rapide de la population peut accélérer la propagation de l'épidémie, tandis que un taux de natalité plus faible peut atténuer la gravité et la vitesse de transmission.

Cependant, quel que soit le scénario démographique, la population évolue dans une seule direction, car il n'y a pas de possibilité de revenir en arrière et la population restera dans les états infectés, porteurs chroniques ou rétablis.

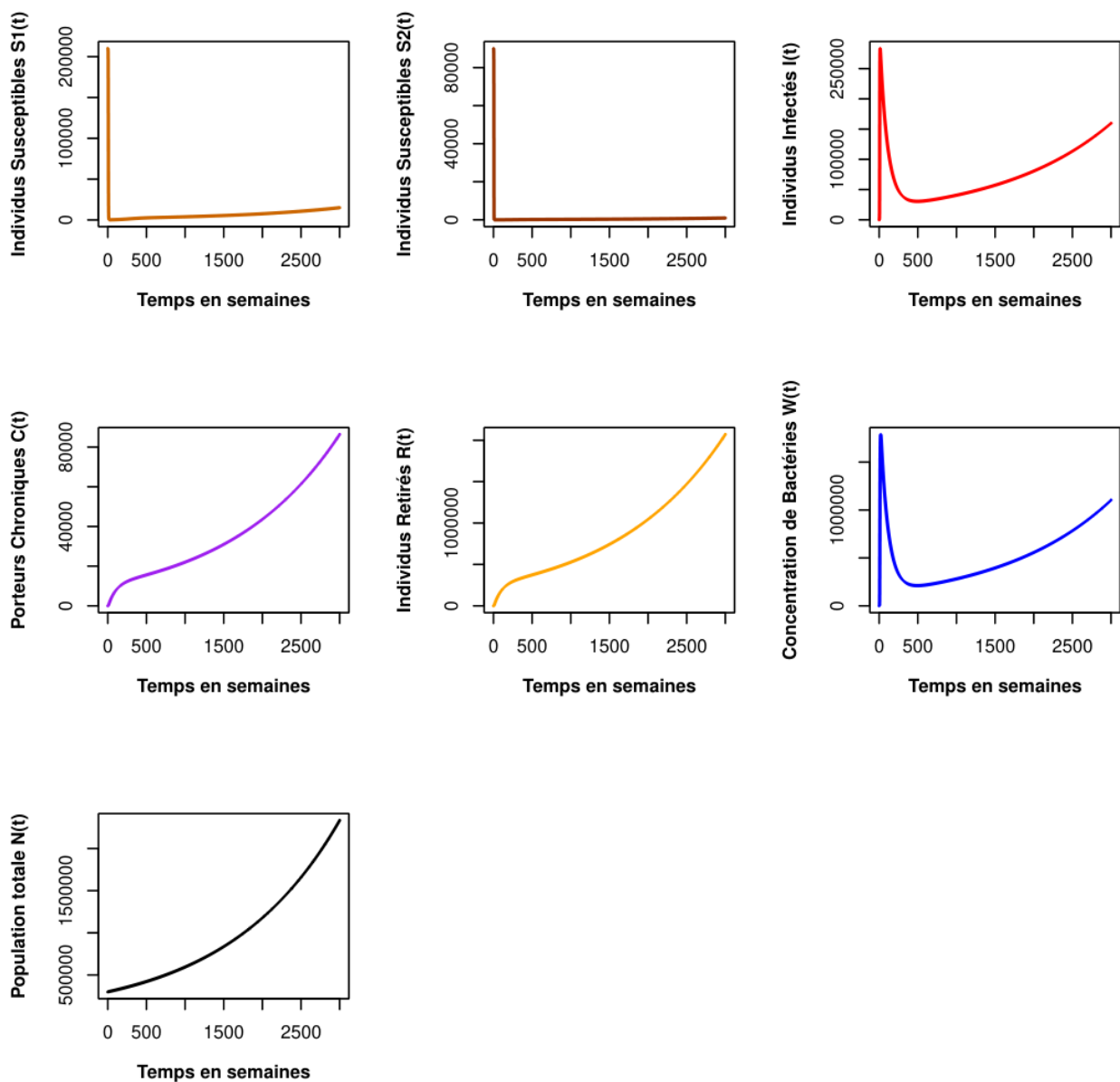


FIGURE 2.5 – Évolution des populations S_1, S_2, I, C, R, N et W des bactéries sur un horizon de 3000 semaines.

Ces simulations ont permis de mieux comprendre la dynamique complexe de la propagation de la fièvre typhoïde au sein d'une population composée de deux compartiments sains présentant des caractéristiques sanitaires distinctes. Elles ont permis d'explorer différents scénarios et d'analyser l'impact de la démographie, de la transmission directe et indirecte, ainsi que d'autres facteurs sur la propagation de la maladie.

Dans la section suivante, nous allons définir le nombre de reproduction de base, un indicateur clé en épidémiologie permettant d'estimer la dynamique de l'épidémie de la fièvre typhoïde et de projeter son évolution.

2.2.4 Le nombre de reproduction de base

Le nombre de reproduction de base, souvent noté R_0 , est un concept clé en épidémiologie. Il représente le nombre moyen de personnes qu'un individu infecté peut contaminer dans une population susceptible, en l'absence de toute mesure de prévention ou de contrôle de l'épidémie [41]. Ce paramètre est très important dans l'étude de la propagation des épidémies, étant donné qu'il permet de quantifier la virulence initiale d'une maladie infectieuse, et donc de prévoir l'évolution de l'épidémie à court terme. Si $R_0 < 1$, cela signifie que chaque individu infectieux produira en moyenne moins d'une infection secondaire. Et si $R_0 > 1$, chaque individu infectieux produira en moyenne plus d'une infection secondaire et donc que la maladie se répandra dans la population. Ce paramètre R_0 est calculé en utilisant l'approche développée par van den Driessche et Watmough [42].

Théorème 2.1. *Si $b = \mu \neq 0$ (le taux de natalité est égal au taux de mortalité naturelle) alors, le nombre de reproduction de base pour le modèle (2.1) est donné par :*

$$R_0 = \frac{1}{(\mu + \delta)p} \left(\beta_p + \frac{\gamma\beta_w(1-p)}{\xi} \right) \left(1 + \frac{\delta\theta\omega}{\mu} \right).$$

Preuve. Notons par $X = (S1, S2, I, C, R, W)$ et $Y = (I, C, W)$. Nous calculons le nombre de reproduction de base en utilisant l'approche *next generation* [42]. Nous définissons le taux de changement des compartiments infectieux à l'équilibre sans maladie (les compartiments infectés sont I, C et W) :

$$\frac{dY}{dt} = \begin{pmatrix} \frac{dI}{dt} \\ \frac{dC}{dt} \\ \frac{dW}{dt} \end{pmatrix} = \mathcal{F}(X) - \mathcal{V}(X).$$

La partie $\mathcal{F}(X)$ représente les taux d'individus nouvellement infectés rejoignant chaque compartiment. Ces taux dépendent des interactions entre les compartiments et d'autres facteurs influençant la propagation de la maladie.

La partie $\mathcal{V}(X)$ correspond à la différence entre les taux d'individus sortant et entrant dans chaque compartiment par d'autres moyens que l'infection. Cela peut inclure des processus tels que la guérison, la mortalité ou d'autres mécanismes spécifiques à la fièvre typhoïde.

$$\mathcal{F}(X) = \begin{pmatrix} \beta_p(I + \omega C)(S1 + S2) + \beta_w W S2 \\ 0 \\ 0 \end{pmatrix} \quad \text{et} \quad \mathcal{V}(X) = \begin{pmatrix} (\mu + \delta)I \\ \mu C - \delta\theta I \\ \xi W - \gamma(I + \omega C) \end{pmatrix}.$$

Le point d'équilibre sans maladie du système (2.1) $X^* = (S1^*, S2^*, 0, 0, 0, 0)$ est obtenu en résolvant :

$$\frac{dS1}{dt} = \frac{dS2}{dt} = \frac{dI}{dt} = \frac{dC}{dt} = \frac{dR}{dt} = \frac{dW}{dt} = 0,$$

sous la contrainte de l'absence de maladie, c'est-à-dire :

$$I = C = W = 0.$$

Ainsi, $S1^*$ et $S2^*$ sont toute solution strictement positive de :

$$\begin{cases} pb(S1^* + S2^*) = \mu S1^*, \\ (1-p)b(S1^* + S2^*) = \mu S2^*. \end{cases}$$

Ce système admet une solution si et seulement si $b = \mu$ et dans ce cas, il a une infinité de solutions :

$$S2^* = \frac{1-p}{p} S1^*.$$

Ainsi, $X^* = (S1^*, \frac{1-p}{p} S1^*, 0, 0, 0, 0)$.

Ensuite, soit F et V les matrices de 3×3 (nombre des compartiments infectés) définies par :

$$F = \left[\frac{\partial \mathcal{F}_i}{\partial Y_j}(X^*) \right] \quad \text{et} \quad V = \left[\frac{\partial \mathcal{V}_i}{\partial Y_j}(X^*) \right] \quad \text{avec} \quad 1 \leq i, j \leq 3.$$

L'indice de reproduction de base R_0 est égal à le rayon spectral de la matrice de *next generation*, FV^{-1} .

Dans notre cas, on a :

$$F = \begin{pmatrix} \frac{\beta_p S1^*}{p} & \frac{\beta_p \omega S1^*}{p} & \frac{\beta_w (1-p) S1^*}{p} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{et} \quad V = \begin{pmatrix} \delta + \mu & 0 & 0 \\ -\delta\theta & \mu & 0 \\ -\gamma & -\omega\gamma & \xi \end{pmatrix},$$

par conséquent

$$V^{-1} = \frac{1}{\xi\mu(\delta + \mu)} \begin{pmatrix} \xi\mu & 0 & 0 \\ \xi\delta\theta & \xi(\delta + \mu) & 0 \\ \gamma(\mu + \delta\theta\omega) & \omega\gamma(\delta + \mu) & \mu(\delta + \mu) \end{pmatrix} \quad \text{et} \quad FV^{-1} = \begin{pmatrix} R_0 & A & B \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

où

$$\begin{aligned} R_0 &= \frac{S1^*}{(\delta + \mu)} \left(\frac{\beta_p}{p} + \frac{\gamma\beta_w(1-p)}{p\xi} \right) \left(1 + \frac{\delta\theta\omega}{\mu} \right), \\ A &= \frac{\omega S1^*}{\mu p} \left(\beta_p + \frac{\beta_w(1-p)\gamma}{\xi} \right), \\ B &= \frac{\beta_w(1-p) S1^*}{p\xi}. \end{aligned}$$

Sans perte de généralité, nous pouvons supposer que $S1^* = 1$, ainsi :

$$R_0 = \frac{1}{(\delta + \mu)} \left(\frac{\beta_p}{p} + \frac{\gamma\beta_w(1-p)}{p\xi} \right) \left(1 + \frac{\delta\theta\omega}{\mu} \right).$$

□

Cette expression de R_0 peut être décomposée pour montrer les contributions de la transmission directe et indirecte. Il est possible de réécrire $R_0 = R_{0,p} + R_{0,w}$, où

$$R_{0,p} = \frac{\beta_p}{p(\delta + \mu)} \left(1 + \frac{\delta\theta\omega}{\mu} \right) \quad \text{et} \quad R_{0,w} = \frac{\gamma\beta_w(1-p)}{p(\delta + \mu)\xi} \left(1 + \frac{\delta\theta\omega}{\mu} \right).$$

Le terme $R_{0,p}$ représente la contribution de la transmission directe de personne à personne. Il dépend du taux de transmission directe (β_p), du taux de guérison (δ), du taux de mortalité naturelle (μ), de la fraction des infectés qui deviennent porteurs chroniques (θ) et de l'infectiosité relative des malades chroniques (ω). Plus le taux de transmission directe est élevé et plus la probabilité de devenir porteur chronique est importante, plus

la contribution de la transmission directe sera élevée.

Le terme $R_{0,w}$ représente la contribution de la transmission indirecte par l'eau. Il dépend du taux d'excrétion dans l'approvisionnement en eau (γ), du taux de transmission indirecte par l'eau (β_w), de la fraction des nouveaux-nés qui partagent l'eau contaminée ($1 - p$), du taux de dégradation des particules infectieuses dans l'eau (ξ), du taux de guérison (δ), du taux de mortalité naturelle (μ), de la fraction des infectés qui deviennent porteurs chroniques (θ) et de l'infectiosité relative des malades chroniques (ω). Si le taux d'excrétion dans l'eau est élevé, le taux de transmission indirecte par l'eau est important et la contribution de la transmission indirecte sera élevée.

À titre d'illustration, la **figure 2.6** met en évidence l'évolution du R_0 en fonction du taux de transmission par l'eau β_w et de l'infectiosité relative des porteurs chroniques ω , tandis que les autres paramètres sont maintenus constants (**Table 2.6**).

Dans le cas de la fièvre typhoïde, qui est transmise principalement par l'eau contaminée, la valeur de R_0 dépend de plusieurs facteurs, tels que le taux de transmission par l'eau, l'infectiosité relative des porteurs chroniques et la qualité de l'eau (**Fig.2.6**). Plus ces facteurs sont élevés, plus le R_0 dépasse la valeur seuil de 1, indiquant ainsi que la maladie se répandra massivement dans la population. En revanche, en l'absence de ces deux facteurs d'infectiosité ($\beta_w = \omega = 0$), le R_0 reste inférieur à 1, signifiant que chaque individu infectieux produira en moyenne moins d'une infection secondaire.

Il est donc crucial de prendre en compte ces facteurs dans la mise en place de mesures de prévention et de contrôle de la fièvre typhoïde. Des actions telles que l'amélioration de la qualité de l'eau, l'identification et le traitement des porteurs chroniques, ainsi que la sensibilisation de la population à l'importance de l'hygiène et de la propreté, peuvent contribuer à réduire la valeur de R_0 et donc à ralentir la propagation de la maladie.

TABLE 2.6 – *Les valeurs utilisées pour le calcul de R_0 .*

Les paramètres	β_p	μ	δ	θ	ξ	γ
Valeurs numériques	0.8	$2.9 \cdot 10^{-3} / 52$	0.25	0.03	0.8	2

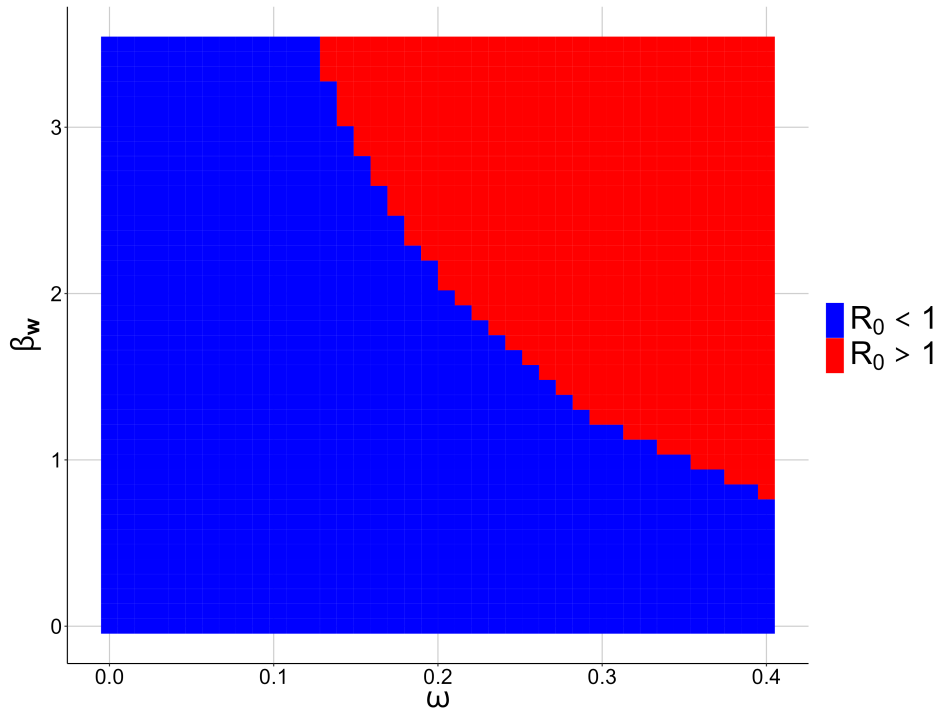


FIGURE 2.6 – Évolution du R_0 en fonction du taux de transmission par l'eau β_w et de l'infectiosité relative des malades chroniques ω .

En conclusion, les simulations réalisées avec différents scénarios de propagation de la maladie de typhoïde ont permis d'identifier les comportements qualitatifs de l'épidémie. Cependant, le taux de reproduction de base R_0 est un outil intrinsèque pour évaluer la virulence initiale de la maladie en fonction des différents facteurs épidémiologiques. Tandis que les modèles déterministes permettent de calculer facilement le R_0 et de prévoir les conditions de transmission de la maladie. Cependant, le profil déterministe proposé ne colle pas du tout à Mayotte, où la taille de la population est petite, et le nombre des cas déclarés reste faible, fluctue et présente une extinction et une réapparition de la maladie (voir **Fig.1.6**). Les modèles stochastiques seront donc mieux adaptés à ce type de contexte. Ils permettent de prendre en compte les variations aléatoires et les effets de petite taille de la population, ce qui peut mieux correspondre à la réalité observée.

Ces résultats mettent en évidence l'importance de l'utilisation de simulations stochastiques pour mieux comprendre et prévoir la propagation de maladies infectieuses dans des contextes épidémiologiques complexes comme celui de Mayotte.

2.3 Modélisation par un processus markovien déterministe par morceaux

Les modèles stochastiques sont capables de capturer les fluctuations des populations de petite taille contrairement aux modèles déterministes. Ces modèles prennent en compte l'aspect aléatoire des événements impliqués dans la propagation et le développement de maladies infectieuses, et permettent de travailler sur des petits effectifs souvent rencontrés dans les études portant sur des ensembles de populations locales interconnectées d'une

même espèce.

Comme mentionné dans la section **1.2.4** et dans la conclusion précédente, l'évolution de la fièvre typhoïde à Mayotte est caractérisée par une fluctuation entre des valeurs faibles, nécessitant une étude probabiliste de cette dynamique. Dans cette partie, nous introduisons une modélisation hybride de la transmission de la typhoïde en utilisant un processus markovien déterministe par morceaux (**1.4.2**). Cette méthode permet de prendre en compte simultanément des variables continues et des variables discrètes.

2.3.1 Formulation du modèle

Nous proposons une approximation hybride originale $(I_t, X_t)_{t \geq 0}$ du modèle déterministe (**2.1**) pour formuler notre PDMP. La première composante I_t est discrète et représente le nombre d'individus infectés à l'instant t , prenant des valeurs dans \mathbb{N} . La deuxième composante $X_t = (S1_t, S2_t, C_t, R_t, W_t)$ est continue à valeurs dans \mathbb{R}_+^5 . Elle décrit l'évolution des autres variables d'état définies dans la **Table 2.1**.

La **Table 2.7** décrit les sauts possibles du processus I_t partant de l'état i , avec $i \in \mathbb{N}$, conditionnellement à $X_t = (s1, s2, c, r, w)$.

TABLE 2.7 – *Sauts possibles du processus I_t .*

Événements	Transitions	Taux
Une nouvelle infection	$i \rightarrow i + 1$	$\beta_p (i + \omega c) (s1 + s2) + \beta_w w s2$
Une guérison ou passage à l'état chronique ou mort	$i \rightarrow i - 1$	$(\mu + \delta) i$

Entre deux sauts de la partie discrète $(I_t)_{t \geq 0}$, la trajectoire de $(X_t)_{t \geq 0}$ déterministe, est solution du système d'équations différentielles (**2.2**) qui dépend de l'état discret i . Lors d'un saut de I_t , de i vers $i + 1$ ou bien $i - 1$, la variable continue X_t ne bouge pas.

$$\left\{ \begin{array}{l} \frac{dS1}{dt} = pbN_t - \mu S1_t - \beta_p (i + \omega C_t) S1_t \\ \frac{dS2}{dt} = (1 - p)bN_t - \mu S2_t - \left(\beta_p (i + \omega C_t) + \beta_w W_t \right) S2_t \\ \frac{dC}{dt} = -\mu C_t + \delta \theta i \\ \frac{dR}{dt} = -\mu R_t + \delta (1 - \alpha - \theta) i \\ \frac{dW}{dt} = -\xi W_t + \gamma (i + \omega C_t) \\ N_t = S1_t + S2_t + i + C_t + R_t. \end{array} \right. \quad (2.2)$$

Le flot du PDMP est déterminé par les équations différentielles du système (**2.2**) reliant les variables continues $S1_t, S2_t, C_t, R_t, W_t$ en fonction du temps t , et ce, entre deux sauts du processus I_t . Le flot est donné par ces équations différentielles.

L'intensité du PDMP est définie par les taux de transition des sauts du processus I_t décrits dans la Table **2.7**. Ces taux de transition dépendent de l'état discret i et des valeurs des variables continues $s1, s2, c, r, w$. Lorsque le processus fait un saut de l'état i à $i + 1$, le taux de transition est donné par $\beta_p (i + \omega c) (s1 + s2) + \beta_w w s2$. Lorsque le processus fait un saut de l'état i à $i - 1$, le taux de transition est $(\mu + \delta) i$.

Le noyau du PDMP est défini comme la probabilité de transition d'un état à un autre

entre les sauts du processus discret, conditionnellement à la trajectoire continue entre ces sauts. Dans le cas présent, le noyau du PDMP correspond à la probabilité conditionnelle de transition de l'état i à $i + 1$ ou de l'état i à $i - 1$, étant donné les valeurs continues de X_t .

En utilisant les taux de transition mentionnés dans la Table **2.7**, la probabilité conditionnelle de transition de i à $i + 1$ entre les sauts est donnée par :

$$p(i \rightarrow i + 1 \mid s1, s2, c, r, w) = \frac{\beta_p(i + \omega c)(s1 + s2) + \beta_w w s2}{\beta_p(i + \omega c)(s1 + s2) + \beta_w w s2 + (\mu + \delta)i}.$$

De même, la probabilité conditionnelle de transition de i à $i - 1$ entre les sauts est donnée par :

$$p(i \rightarrow i - 1 \mid s1, s2, c, r, w) = \frac{(\mu + \delta)i}{\beta_p(i + \omega c)(s1 + s2) + \beta_w w s2 + (\mu + \delta)i}.$$

En grande population pour I_0 , le modèle déterministe **(2.1)** présenté dans la partie **2.2** est une approximation du PDMP. En effet, dans ce cas, les fluctuations dues à la petite taille de la population sont négligeables et le modèle déterministe suffit à décrire l'évolution de la maladie. Cependant, dans le cas d'une population infectée de petite taille, les fluctuations peuvent avoir un impact significatif sur la dynamique de la maladie, ce qui justifie l'utilisation d'un PDMP pour capturer l'effet stochastique de la propagation de la maladie.

2.3.2 Discussion et simulations numériques

Nous présentons quelques exemples de trajectoires simulées de notre PDMP en fonction des paramètres décrits dans la Table **2.3**, en utilisant les mêmes conditions initiales pour chaque simulation. le pseudo-code de R est présenté dans l'algorithme **2.3.2**.

1^{er} cas : l'évolution de l'épidémie sans démographie (Fig.2.7)

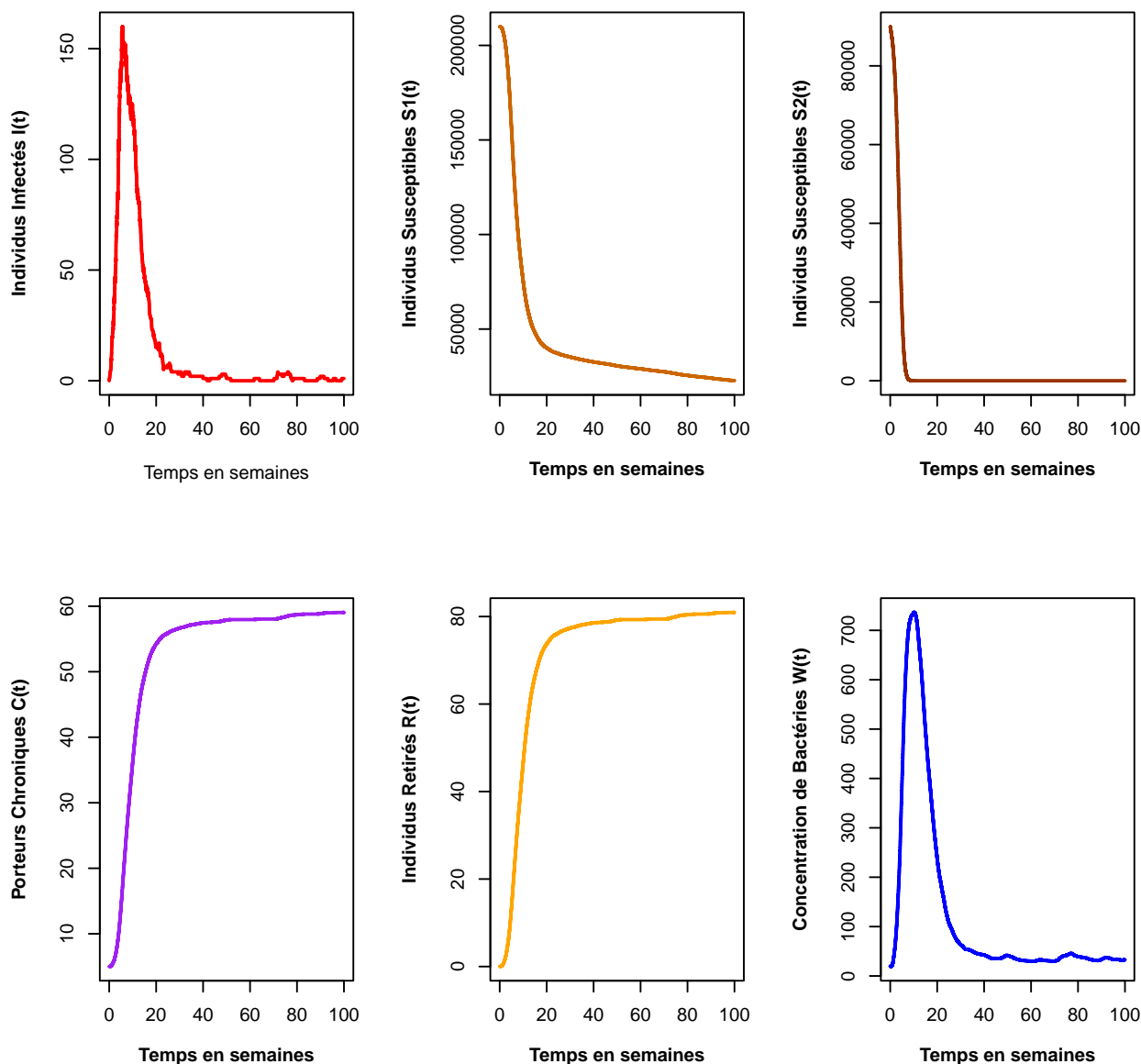


FIGURE 2.7 – Évolution des populations $S1$, $S2$, I , C , R et la concentration de bactéries dans l'eau W sur un horizon de 100 semaines.

La figure 2.7 illustre l'évolution de la population selon deux mécanismes distincts. Au début, la dynamique de l'épidémie suit un comportement très similaire à celui du modèle déterministe présenté dans la partie 2.2, avec une croissance rapide de l'infection et un pic épidémique important. Cette croissance rapide est rendue possible par le grand nombre de susceptibles au début de l'épidémie. Cependant, en dessous d'un certain seuil de population $S1 + S2$, la dynamique de la typhoïde ralentit considérablement. Le nombre d'infectés présente des fluctuations avec de petites valeurs, et même s'il atteint zéro et disparaît complètement de la population ($I=0$), il peut réapparaître en raison de la présence des porteurs chroniques et des bactéries dans l'eau, qui constituent les deux sources d'infectiosité dans notre population. Cette propagation lente est caractérisée par des retours fréquents à l'état 0, ce qui signifie que l'épidémie est s'éteint, mais qu'elle

reprend si les conditions sont à nouveau favorables (à cause de C ici).

Algorithme 2 : Algorithme de simulation pour un PDMP

entrée : paramètres du modèles : $b, \mu, p, \beta_p, \beta_w, \omega, \alpha, \delta, \theta, \gamma, \xi$.

valeurs initiales des variables : $s1^0, s2^0, i^0, c^0, r^0, w^0$.

un majorant global M de l'intensité :

horizon de simulation H , pas de discrétisation en temps Δt

sortie : Trajectoire $(I_t, X_t = (S1(t), S2(t), C(t), R(t), W(t)))_{0 \leq t \leq H}$ par pas de Δt

1 début

2 Initialisation

3 $(S1(0), S2(0), i, C(0), R(0), W(0)) \leftarrow (s1^0, s2^0, i^0, c^0, r^0, w^0)$

4 $t \leftarrow 0$

5 tant que $t < H$ faire

6 Tirer le prochain saut virtuel : Tirer T selon la loi $\text{Exp}(M)$

7 Ne pas dépasser l'horizon : $T \leftarrow \min\{T, H - t\}$

8 Faire avancer le flot :

9 pour h de 0 à T , par pas de Δt faire :

10 $S1(h + \Delta t) \leftarrow S1(h) + \Delta t \left(pbN(h) - \mu S1(h) - \beta_p(i + \omega C(h)) S1(h) \right)$

11 $S2(h + \Delta t) \leftarrow S2(h) + \Delta t \left((1 - p)bN(h) - \mu S2(h) - (\beta_p(i + \omega C(h)) + \beta_w W(h)) S2(h) \right)$

12 $C(h + \Delta t) \leftarrow C(h) + \Delta t \left(-\mu C(h) + \delta \theta i \right)$

13 $R(h + \Delta t) \leftarrow R(h) + \Delta t \left(-\mu R(h) + \delta(1 - \alpha - \theta) i \right)$

14 $W(h + \Delta t) \leftarrow W(h) + \Delta t \left(-\xi W(h) + \gamma(i + \omega C(h)) \right)$

15 $N(h + \Delta t) \leftarrow S1(h + \Delta t) + S2(h + \Delta t) + i + C(h + \Delta t) + R(h + \Delta t)$

16 fin

17 Mettre à jour le temps : $t \leftarrow t + T$

18 Déterminer la nature du saut si on n'a pas atteint l'horizon :

19 Si $t < H$ alors

20 Tirer U selon la loi $\mathcal{U}[0, 1]$

21 Si $U < \frac{\beta_p(i + \omega C(T))(S1(T) + S2(T)) + \beta_w W(T) S2(T)}{M}$ alors

22 naissance : $i \leftarrow i + 1$

23 fin

24 Sinon Si $U < \frac{\beta_p(i + \omega C(T))(S1(T) + S2(T)) + \beta_w W(T) S2(T) + (\delta + \mu)i}{M}$ alors

25 mort : $i \leftarrow i - 1$

26 fin

27 Sinon ce n'est pas un saut : ne rien faire

28 fin

29 fin

30 renvoyer : $(I_t, X_t)_{0 \leq t \leq H}$

30 fin

2^{eme} cas : l'évolution de l'épidémie avec démographie (Fig.2.8)

Dans ce scénario avec démographie, nous simulons notre PDMP en utilisant les mêmes valeurs des paramètres que celles utilisées dans le cas déterministe (**Tableau 2.5**).

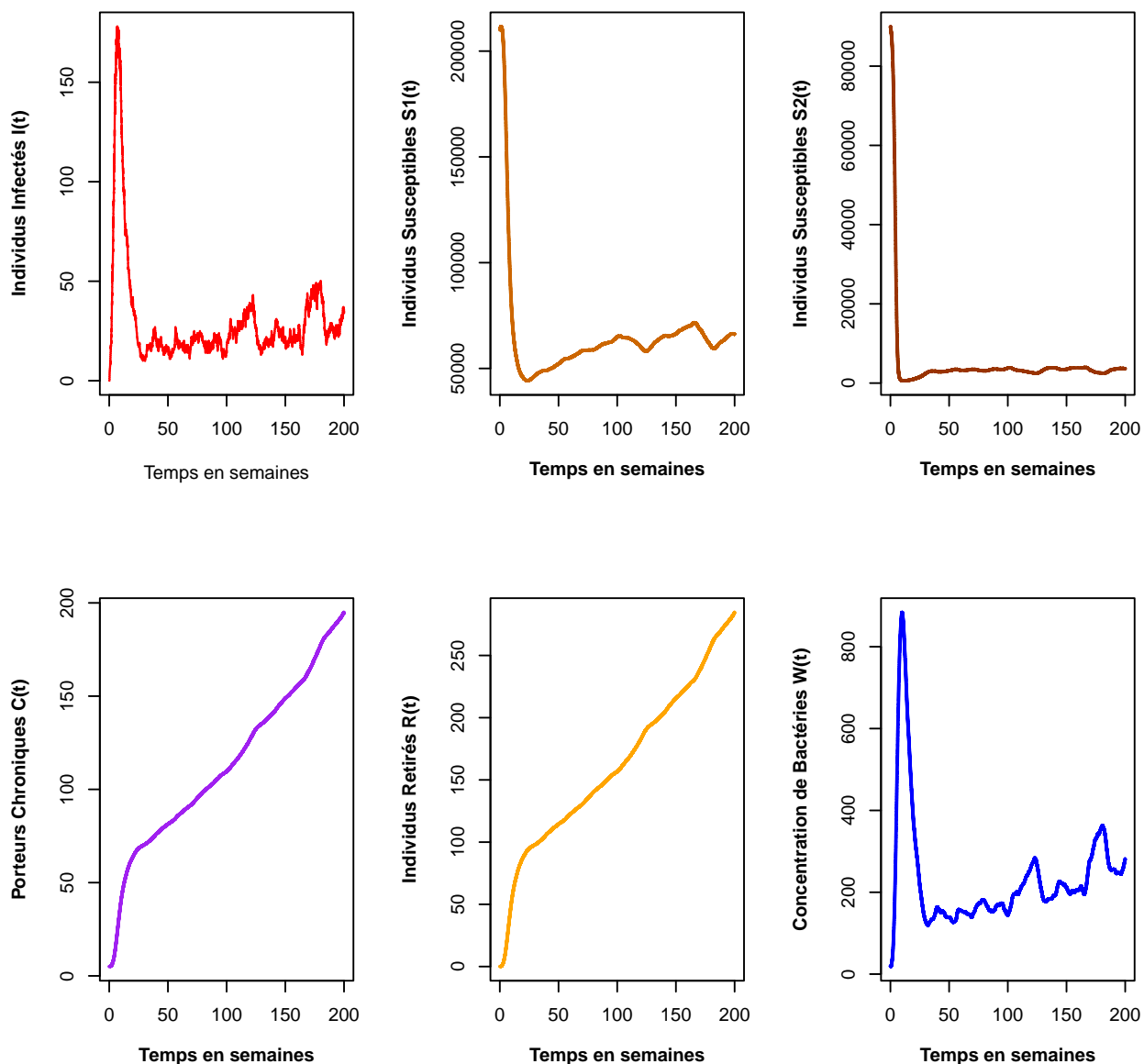


FIGURE 2.8 – Evolution des populations $S1$, $S2$, I , C , R et W des bactéries sur un horizon de 200 semaines.

Dans le cas de la simulation présentée **2.8**, la population suit un mécanisme similaire à celui décrit précédemment, mais avec des variations notables. Tout d’abord, après environ 200 semaines, on observe un changement dans la population : au début de la maladie, nous remarquons une diminution rapide des sous-groupes $S1$ et $S2$. Cependant, ceux-ci augmentent à nouveau par la suite en raison de la démographie de la population, notamment les naissances qui contribuent à l’augmentation du nombre d’individus susceptibles d’être infectés. En outre, le nombre d’individus retirés et de porteurs chroniques augmente de manière significative. Les individus infectés présentent également des fluctuations plus marquées, avec plusieurs pics épidémiques plus forts que dans l’exemple précédent **2.7**. Ces résultats suggèrent une transition de la dynamique de la maladie, passant d’un ré-

gime de propagation rapide à un régime de fluctuations modérées et de persistance de la maladie.

Nous avons réussi à simuler notre PDMP pour différents scénarios, mais il est important de noter que nous n'avons pas encore calibré notre modèle sur les données réelles de Mayotte (**Fig.1.6**) afin de mieux refléter la réalité locale. En effet, notre modèle contient de nombreux paramètres à calibrer spécifiquement pour Mayotte, et nous disposons seulement des cumuls journaliers des nouveaux cas déclarés, ce qui rend impossible l'ajustement précis du modèle à ces observations. Par conséquent, nous proposons de simplifier ce modèle probabiliste dans le but de réaliser des estimations des paramètres à partir des données disponibles.

2.4 Conclusion

Dans ce chapitre, nous avons proposé deux modèles pour étudier théoriquement et numériquement la dynamique de la propagation de la fièvre typhoïde à Mayotte. Nous avons commencé par développer un nouveau modèle déterministe qui prend en compte l'effet de la qualité de l'eau sur la transmission de la typhoïde et son impact sur deux populations de caractéristiques sanitaires différentes ayant accès à des eaux de qualité différente. Nous avons pu simuler des trajectoires pour différents scénarios et interpréter leur comportement qualitatif. Ensuite, nous avons proposé une méthode pour calculer le nombre de reproduction de base, qui est crucial pour évaluer la virulence d'une épidémie. Cependant, ce type de modèle déterministe n'est pas adapté à la réalité de l'évolution de la typhoïde à Mayotte, car sa dynamique qualitative ne correspond pas aux données observées. Nous avons donc choisi de passer à une modélisation stochastique par un PDMP, où nous avons considéré une évolution aléatoire des individus infectés et déterministe pour les autres variables d'état. Nous avons pu simuler certaines trajectoires possibles du PDMP pour différentes valeurs des paramètres. Néanmoins, il n'a pas été possible de calibrer le modèle de manière précise pour Mayotte en raison du grand nombre de paramètres impliqués et du manque de données disponibles.

Pour toutes ces raisons, nous proposons dans le chapitre suivant de simplifier notre PDMP en utilisant un modèle de type naissance et mort avec immigration (**1.4.3**), afin d'essayer de mettre en place une procédure de calibration et d'estimation des paramètres à partir des données fournies par l'ARS de Mayotte. Le but de cette simplification est de créer un modèle avec le minimum de paramètres possible, afin qu'il soit identifiable et puisse être ajusté aux données de manière plus efficace. Nous nous concentrerons sur l'estimation des paramètres clés qui déterminent le comportement du modèle, ce qui nous permettra de mieux quantifier la propagation de la fièvre typhoïde à Mayotte.

Chapitre 3

Estimation des paramètres d'un processus caché de naissance et mort linéaire avec immigration

Sommaire

3.1	Introduction	56
3.2	Modèle et schéma d'observations	57
3.2.1	Description du Modèle	57
3.2.2	Schéma d'observations	59
3.2.3	Stratégie d'estimation	60
3.3	Estimation paramétrique	60
3.3.1	Expression analytique des paramètres	61
3.3.2	Modèle de Markov caché	69
3.4	Étude numérique	74
3.4.1	Données synthétiques	74
3.4.2	Données de la typhoïde à Mayotte	79
3.4.3	Application à l'estimation de la probabilité d'atteindre un seuil épidémique dans une période de temps	80
3.5	Conclusion	84

3.1 Introduction

Le premier modèle proposé dans le chapitre précédent **2.1**, qui comporte de nombreux paramètres, s'est avéré trop complexe pour être calibré avec précision à l'aide des données fournies par l'agence régionale de santé. En effet, on ne dispose d'aucune information sur les éventuels malades chroniques ni aucune mesure sur les bactéries dans l'eau. Les seules données disponibles sont les cumuls journaliers des nouveaux cas déclarés, qui sont considérés comme isolés selon l'ARS de Mayotte [11]. Afin d'obtenir un modèle ajusté aux données disponibles et d'estimer les paramètres clés de la dynamique de la fièvre typhoïde à Mayotte, nous proposons un modèle stochastique simplifié à un compartiment (**Fig.3.1**), qui ne prend en compte que les effectifs de la population infectée I . Dans ce modèle simplifié, nous négligeons les compartiments $S1$ et $S2$ qui sont supposés être infinis par rapport à I , et nous n'incluons pas le suivi des porteurs chroniques C ni des individus rétablis R , puisque nous n'observons que les cumuls de cas isolés. Ce modèle stochastique est un processus de *naissance et de mort linéaire avec immigration* (LBDPI)(**1.4.3**). L'*immigration* correspond ici à une contamination exogène, ou contamination de l'environnement, la *naissance* correspond à une nouvelle contamination de personne à personne, et la *mort* correspond à l'isolement. Ces processus appartiennent à la famille des processus de Markov de sauts pures et sont utilisés dans un grand nombre de disciplines quantitatives, telles que la biologie évolutive, l'épidémiologie, l'écologie, la génétique des populations, la spéciation et l'extinction, ou la théorie des files d'attente [43, 44, 45, 46, 47].

Ce modèle de LBDPI peut être décrit avec seulement trois paramètres : les taux d'*immigration* (contamination exogène ou environnementale), de *naissance* (contamination de personne à personne) et de *mort* (isolement). Son comportement en temps long est bien connu (voir le **Th.1.2**), de même que les estimateurs du maximum de vraisemblance basés sur des observations discrètes ou continues dès que les observations disponibles incluent les temps de saut et les positions après saut du processus [48, 49, 50, 51].

Toutefois, dans la pratique épidémiologique, il est rare que l'on dispose de comptages continus ou périodiques des individus infectés. Pour les maladies à déclaration obligatoire, les données les plus couramment collectées sont les cumuls des nouveaux infectés sur des périodes données, généralement des intervalles quotidiens ou hebdomadaires. L'inférence des paramètres pour le processus de LBDPI dans ce cadre d'observation est difficile pour deux raisons principales :

Tout d'abord, lorsque les observations sont discrètes, il est nécessaire d'extraire les probabilités de transition des infectés pour une période de temps donné. La composante *immigration* invalide la propriété de branchement, de sorte que l'obtention d'une forme analytique pour les probabilités de transition n'est pas simple, et l'obtention d'une forme numérique peut être sujette à des instabilités numériques [52]. Plusieurs chercheurs ont proposé des approches pour estimer les paramètres des processus de naissance et mort généraux observés discrètement. Crawford et Suchard et Crawford et al [47, 53] proposent d'utiliser la théorie des fractions continues pour trouver l'expression explicite de la transformée de Laplace des probabilités de transition afin de considérer le problème d'inférence avec des observations discrètes comme un problème de données manquantes en utilisant l'algorithme d'expectation-maximisation (EM) pour trouver des estimations du maximum de vraisemblance. Xu et al [54] utilisent la même méthode d'algorithme EM pour proposer une approximation de processus de branchement multi-types pour birth-death shift processes. Dans [55], les auteurs proposent d'utiliser deux approches pour estimer les

taux de *naissance* et de *mort* à partir d'un processus de naissance et mort linéaire observé discrètement de type Galton-Watson process et en maximisant la vraisemblance par une méthode d'approximation du point-selle. Malheureusement, aucune de ces approches ne peut être utilisée dans notre cas.

Deuxièmement, dans tous les articles cités précédemment, les observations sont discrètes mais correspondent à des comptages de population infectée. L'extraction du nombre de personnes infectées à partir des cumuls de nouveaux cas déclarés correspond à un modèle de Markov caché (HMM) non standard. En effet, en raison de la dynamique continue sous-jacente, les nouveaux cas déclarés cumulés sur une période donnée dépendent à la fois du nombre de personnes infectées au début et à la fin de la période. Par conséquent, la distribution des émissions du HMM dépend de l'état actuel et de l'état passé de la chaîne cachée. L'algorithme standard de Baum-Welch [25] ne peut donc pas être appliqué directement et doit être adapté à ce cas spécifique.

Ce chapitre est adapté d'un preprint d'article [56], aborde ces deux défis. Tout d'abord, nous décrivons le modèle LBDPI stochastique et notre schéma d'observation spécifique dans la section 3.2. Ensuite, dans la section 3.3, nous présentons la procédure d'estimation des paramètres, où nous établissons une formule analytique pour les paramètres d'un LBDPI en termes de probabilités de transition à pas discret. Nous discutons également les principales propriétés des estimateurs. Nous adaptons ensuite l'algorithme de Baum-Welch à notre forme spéciale de distribution des émissions pour estimer les probabilités de transition de la chaîne cachée à partir des données de cumuls de nouveaux cas déclarés. La section 3.4 est dédiée à l'étude des propriétés numériques des estimateurs, à la fois sur des données synthétiques et réelles. Elle met également en évidence l'application pratique de ces estimations pour évaluer la probabilité d'atteindre un seuil épidémique des infectés à Mayotte. Enfin, la section 3.5 présente quelques remarques finales.

3.2 Modèle et schéma d'observations

Dans cette section, nous présentons d'abord le modèle stochastique que nous utilisons pour quantifier le nombre d'individus infectés dans la population, noté I (3.2.1). Nous définissons ensuite notre schéma d'observation et énonçons notre problème d'estimation des paramètres (3.2.2). Enfin, nous expliquons notre stratégie pour obtenir des estimateurs (3.2.3).

3.2.1 Description du Modèle

Notre nouveau modèle repose sur plusieurs hypothèses simplificatrices par rapport au modèle initial décrit dans 2.1. Tout d'abord, nous supposons que les taux de natalité et de mortalité naturelle sont négligeables, ce qui signifie que l'effet de la démographie de la population est négligeable. De plus, nous considérons que les individus susceptibles (S_1, S_2) et la concentration des bactéries dans l'eau (W) sont constantes, puisque les individus infectés sont peu nombreux par rapport aux individus susceptibles.

Dans notre modèle simplifié (Fig.3.1), nous nous concentrons principalement sur la dynamique des individus infectés (I), en considérant une seule source d'infectiosité (W) sans transmission de la typhoïde par les porteurs chroniques (C). Les individus retirés (R) ne sont pas pris en compte, car les données disponibles ne représentent que des cumuls des cas isolés qui sont considérés guéris après une période de temps inconnue. Cette approche

nous permet de simplifier notre modèle pour être en mesure d'inférer la dynamique. Comme nous l'avons souligné dans la section **1.2.4**, la prévalence de la fièvre typhoïde à Mayotte est caractérisée par des niveaux relativement faibles et fluctuants, et des résurgences de la maladie après des plages sans nouvelles infections. Un modèle stochastique de type processus de comptage est donc bien adapté pour modéliser ces données. Parmi ces modèles, l'un des plus simples et pertinent, avec peu de paramètres est le *processus de naissance et mort linéaire avec immigration*. Un tel processus $(I_t)_{t \geq 0}$ comptabilise le nombre d'individus infectés. Chaque individu contamine un nouvel individu à un taux (de *naissance*) $\lambda > 0$ et devient isolé avec un taux (de *mort*) $\mu > 0$ indépendamment des autres individus, ce qui génère des taux de naissance et mort linéaires pour la population globale. La contamination via l'environnement génère également de nouveaux infectés via un processus de Poisson de taux (d'*immigration*) $\nu > 0$.

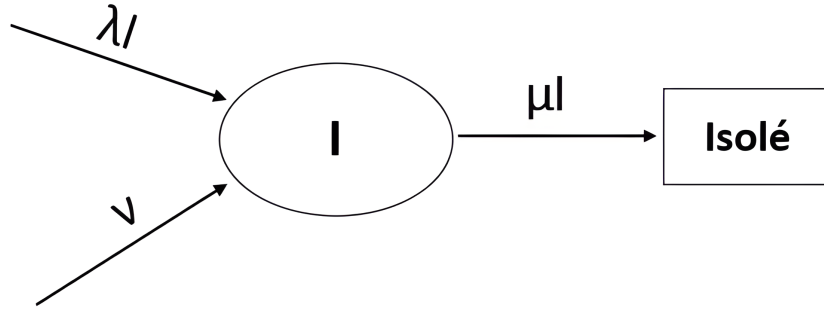


FIGURE 3.1 – **Modèle à un seul compartiment : Infecté.**

La loi de I est caractérisée par son générateur infinitésimal Q donné par :

$$\begin{aligned} Q(i, i+1) &= \lambda i + \nu && \text{pour } i \geq 0, \\ Q(i, i-1) &= \mu i && \text{pour } i \geq 1, \\ Q(i, i) &= -(\lambda + \mu)i - \nu && \text{pour } i \geq 0. \end{aligned}$$

Le semi-groupe de transition défini par $(P(t) = (p_{i,j}(t), i, j \in \mathbb{N}))_{t \geq 0}$, où

$$p_{i,j}(t) = \mathbb{P}(I_t = j | I_0 = i), \quad i, j \in \mathbb{N}, t \geq 0,$$

satisfait l'équation de Kolmogorov progressive suivante : pour $t > 0$ et $i, j \in \mathbb{N}$ avec $j \geq 1$

$$\begin{aligned} \frac{dp_{i,0}(t)}{dt} &= \mu p_{i,1}(t) - \nu p_{i,0}(t), \\ \frac{dp_{i,j}(t)}{dt} &= (\lambda(j-1) + \nu) p_{i,j-1}(t) + \mu(j+1) p_{i,j+1}(t) - ((\lambda + \mu)j + \nu) p_{i,j}(t), \end{aligned} \tag{3.1}$$

et avec la condition initiale $p_{i,i}(0) = 1$ et $p_{i,j}(0) = 0$ pour $j \neq i$.

Dans toute la suite, nous supposons que le processus LBDPI est en régime récurrent positif, c'est-à-dire $\nu > 0$ et $\lambda < \mu$. Notre objectif est d'estimer le triple (λ, μ, ν) .

3.2.2 Schéma d'observations

La principale difficulté de notre étude provient du schéma d'observation. En effet, le nombre d'individus infectés à un moment donné I_t n'est pas directement observé dans notre jeu de données. Au lieu de cela, nous avons seulement accès aux cumuls de nouveaux cas isolés par jour. Cette formulation est mathématiquement complexe et originale, mais elle est très courante dans les données de santé publique, notamment pour les maladies à déclaration obligatoire.

Notons par $O_{]s,t]}$ le cumul des nouveaux déclarés correspondant au processus I sur l'intervalle de temps $]s, t]$. Ainsi, $O_{]s,t]}$ correspond au nombre de sauts d'amplitude -1 de I sur l'intervalle de temps $]s, t]$: $O_{]s,t]}$ est le cardinal de l'ensemble $\{t \in]s, t]; I_t - I_{t-} = -1\}$, où $I_{t-} = \lim_{r < t, r \rightarrow t} I_r$.

Le processus joint $(I_t, O_{]0,t]}$ est encore un processus de Markov de saut pur et sa distribution est caractérisée par les probabilités de transition suivantes : pour $t > 0$

$$p_{i,(j,o)}(t) := \mathbb{P}(I_t = j, O_{]0,t]} = o \mid I_0 = i), \quad i, j, o \in \mathbb{N}.$$

Elles satisfont également une équation de Kolmogorov (rétrograde) donnée par le lemme suivant.

Lemme 3.1. *Pour $t > 0$ et $i, j, o \in \mathbb{N}$, on a*

$$\frac{dp_{i,(j,o)}(t)}{dt} = -((\lambda + \mu)i + \nu)p_{i,(j,o)}(t) + (\lambda i + \nu)p_{i+1,(j,o)}(t) + (\mu i)p_{i-1,(j,o-1)}(t), \quad (3.2)$$

pour $i \leq j + o$, avec $p_{i,(j,o)}(t) = 0$ quand $i > j + o$ et avec les conditions initiales $p_{i,(j,o)}(0) = \delta_{i=j}\delta_{o=0}$.

Preuve. Nous utilisons la caractérisation infinitésimale de la distribution du processus joint.

Pour $h > 0$ et $i \leq j + o$, on a

$$\begin{aligned} p_{i,(j,o)}(t+h) &= \mathbb{P}(I_{t+h} = j, O_{]0,t+h]} = o \mid I_0 = i) \\ &= \mathbb{P}(I_h = i, O_{]0,h]} = 0, I_{t+h} = j, O_{]h,t+h]} = o \mid I_0 = i) \\ &\quad + \mathbb{P}(I_h = i+1, O_{]0,h]} = 0, I_{t+h} = j, O_{]h,t+h]} = o \mid I_0 = i) \\ &\quad + \mathbb{P}(I_h = i-1, O_{]0,h]} = 1, I_{t+h} = j, O_{]h,t+h]} = o-1 \mid I_0 = i) + 0(h) \\ &= \mathbb{P}(I_h = i, O_{]0,h]} = 0 \mid I_0 = i) \mathbb{P}(I_{t+h} = j, O_{]h,t+h]} = o \mid I_h = i) \\ &\quad + \mathbb{P}(I_h = i+1, O_{]0,h]} = 0 \mid I_0 = i) \mathbb{P}(I_{t+h} = j, O_{]h,t+h]} = o \mid I_h = i+1) \\ &\quad + \mathbb{P}(I_h = i-1, O_{]0,h]} = 1 \mid I_0 = i) \mathbb{P}(I_{t+h} = j, O_{]h,t+h]} = o-1 \mid I_h = i-1) + 0(h) \end{aligned}$$

Puisque $(I_t)_{t \geq 0}$ est un processus de Markov homogène, on obtient :

$$\begin{aligned} p_{i,(j,o)}(t+h) &= (1 - ((\lambda + \mu)i + \nu)h)p_{i,(j,o)}(t) + (\lambda i + \nu)hp_{i+1,(j,o)}(t) \\ &\quad + (\mu i)hp_{i-1,(j,o-1)}(t) + 0(h). \end{aligned}$$

Par conséquent

$$\frac{dp_{i,(j,o)}(t)}{dt} = -((\lambda + \mu)i + \nu)p_{i,(j,o)}(t) + (\lambda i + \nu)p_{i+1,(j,o)}(t) + (\mu i)p_{i-1,(j,o-1)}(t).$$

□

Nous définissons Δt comme étant un pas de temps fixe, généralement d'une durée d'un jour ou d'une semaine. Notre objectif principal est d'estimer les trois paramètres (λ, μ, ν) à partir des observations disponibles, qui sont représentées par la suite $(O_{]n\Delta t, (n+1)\Delta t])_{n \in \mathbb{N}}$.

3.2.3 Stratégie d'estimation

Notre problème d'estimation fait partie de la classe des problèmes à information cachée. En effet, une suite donnée d'observations $(O_{]n\Delta t, (n+1)\Delta t])_{n \in \mathbb{N}}$ peut correspondre à plusieurs trajectoires différentes possibles de I . Toutefois, notre objectif n'est pas de reconstruire la trajectoire la plus probable, mais simplement d'estimer les paramètres (λ, μ, ν) .

Malheureusement, nous n'avons pas pu obtenir une forme analytique des probabilités de transitions $p_{i,(j,o)}(t)$ à partir de l'équation (3.2). Ainsi, il n'est pas possible de mettre en œuvre une méthode d'estimation directe par maximum de vraisemblance avec ce schéma d'observations. Pour pallier ce problème, nous procédons en deux étapes :

La première étape **3.3.1** consiste à étudier la chaîne de Markov à pas discret $(I_{n\Delta t})_{n \in \mathbb{N}}$. Nous obtenons une formule analytique pour sa matrice de transition en fonction des paramètres inconnus (λ, μ, ν) . Cette formule peut être inversée pour obtenir une expression des paramètres (λ, μ, ν) en termes de probabilités de transition bien choisies. Nous obtenons ensuite des estimateurs par plug-in à partir de n'importe quel estimateur des probabilités de transition, et nous établissons que la consistance et la normalité asymptotique des estimateurs des probabilités de transition se transfèrent aux estimateurs de (λ, μ, ν) .

La deuxième étape **3.3.2** tient compte du cadre à information manquante en interprétant la chaîne de Markov à temps discret $(I_{n\Delta t}, O_{]n\Delta t, (n+1)\Delta t])_{n \in \mathbb{N}}$ comme un modèle de Markov caché, où $(I_{n\Delta t})$ est la chaîne cachée et $(O_{]n\Delta t, (n+1)\Delta t])$ les observations. Cependant, dans cette structure, le processus d'émission est non standard car le nombre cumulé de nouveaux cas déclarés $O_{]n\Delta t, (n+1)\Delta t]}$ dépend de l'ensemble de la trajectoire en temps continu de I sur l'intervalle $]n\Delta t, (n+1)\Delta t]$ et pas seulement de sa valeur finale $I_{(n+1)\Delta t}$. En utilisant la propriété de Markov, on peut obtenir les probabilités d'émission de $O_{]n\Delta t, (n+1)\Delta t]}$ conditionnellement à $I_{n\Delta t}$ et $I_{(n+1)\Delta t}$ seulement. Notre première idée a été d'exécuter la procédure HMM standard du package R HMM sur la chaîne bidimensionnelle $Z_n = (I_{n-1}, I_n)$ pour estimer les probabilités de transition de la chaîne cachée (Z_n) . Malheureusement, la méthode était numériquement instable et entachée d'erreurs dues aux nombreux zéros dans la matrice de transition de (Z_n) . Au lieu de cela, nous avons complètement réécrit les probabilités forward et backward dans notre contexte particulier et adapté l'algorithme de Baum-Welsh en conséquence. Cette deuxième étape nous permet d'obtenir des estimations des probabilités de transition cachées de la chaîne de Markov $(I_{n\Delta t})_{n \in \mathbb{N}}$ sur la base des observations $(O_{]n\Delta t, (n+1)\Delta t])$ uniquement. Nous utilisons ensuite les formules explicites de la première étape pour obtenir des estimateurs du triple (λ, μ, ν) .

3.3 Estimation paramétrique

Dans cette section, nous détaillons notre méthodologie d'estimation paramétrique. Tout d'abord, nous établissons une formule analytique pour les paramètres d'un LBDPI en termes de probabilités de transition dans la partie **3.3.1**. Nous adaptons ensuite l'algorithme de Baum-Welsh à notre forme particulière de loi d'émission pour estimer les

probabilités de transition de la chaîne cachée à partir des cumuls de nouveaux cas déclarés dans **3.3.2**. Enfin, nous obtenons des estimations des paramètres originaux par plug-in.

3.3.1 Expression analytique des paramètres

La première étape de notre procédure d'estimation consiste à établir une formule analytique pour les paramètres (λ, μ, ν) en tant que fonctions des probabilités de transition en pas discret de la chaîne de Markov $(I_{n\Delta t})_{n \in \mathbb{N}}$.

Nous présentons d'abord les lemmes suivants qui seront utiles pour établir une solution analytique de l'équation de Kolmogorov **(3.1)**.

Lemme 3.2. *Soit $G_i(s, t)$ une fonction différentiable sur $\mathbb{R} \times [0, +\infty[$, pour tout $i \in \mathbb{N}$, la solution de l'équation aux dérivées partielles linéaire*

$$\begin{cases} \frac{\partial G_i(s, t)}{\partial t} = (\mu - \lambda s)(1 - s) \frac{\partial G_i(s, t)}{\partial s} + \nu(s - 1)G_i(s, t), \\ G_i(s, 0) = s^i, \end{cases} \quad (3.3)$$

est donnée par :

$$G_i(s, t) = \begin{cases} \frac{(\mu - \lambda)^{\frac{\nu}{\lambda}} \left(s(\mu e^{(\lambda - \mu)t} - \lambda) - \mu(e^{(\lambda - \mu)t} - 1) \right)^i}{\left(\mu - \lambda e^{(\lambda - \mu)t} + \lambda s(e^{(\lambda - \mu)t} - 1) \right)^{\frac{\nu}{\lambda} + i}} & \lambda \neq \mu, \\ \left(\frac{\lambda t + s(1 - \lambda t)}{1 + \lambda t - \lambda t s} \right)^i (1 + \lambda t - \lambda t s)^{-\frac{\nu}{\lambda}} & \lambda = \mu. \end{cases}$$

Preuve. Pour résoudre l'équation **(3.3)**, nous utilisons la méthode des caractéristiques (voir par exemple ([57], chapitre 1)).

Étape 1. Détermination des caractéristiques. Considérons l'équation caractéristique suivante :

$$\frac{dt}{ds} = -\frac{1}{(\mu - \lambda s)(1 - s)} = \begin{cases} \frac{1}{\lambda - \mu} \left(\frac{1}{1 - s} - \frac{\lambda}{\mu - \lambda s} \right) & \text{if } \lambda \neq \mu \\ -\frac{1}{\lambda(1 - s)^2} & \text{if } \lambda = \mu. \end{cases}$$

En intégrant par rapport à s , nous avons

$$t = k - \begin{cases} \frac{1}{\lambda - \mu} \ln \left(\frac{1 - s}{\mu - \lambda s} \right) & \text{si } \lambda \neq \mu \\ \frac{1}{\lambda(1 - s)} & \text{si } \lambda = \mu. \end{cases} \quad (3.4)$$

L'équation **(3.4)** définit une famille de caractéristiques paramétrées par la constante d'intégration k .

Étape 2. Tout au long des caractéristiques de **(3.4)**, G_i satisfait

$$\frac{dG_i}{ds} = \frac{\nu G_i}{\mu - \lambda s}.$$

L'intégration par rapport à s donne

$$G_i = C(k) (\mu - \lambda s)^{-\frac{\nu}{\lambda}},$$

où $C(k)$ est une autre constante d'intégration, fonction du paramètre k . En utilisant l'équation (3.4), nous obtenons

$$G_i(s, t) = \begin{cases} C \left(t + \frac{1}{\lambda - \mu} \ln \left(\frac{1-s}{\mu - \lambda s} \right) \right) (\mu - \lambda s)^{-\frac{\nu}{\lambda}} & \text{si } \lambda \neq \mu \\ C \left(t + \frac{1}{\lambda(1-s)} \right) (\lambda(1-s))^{-\frac{\nu}{\lambda}} & \text{si } \lambda = \mu. \end{cases} \quad (3.5)$$

Nous utilisons maintenant la condition initiale $G_i(s, 0) = s^i$ pour calculer $C(k)$. En fixant $t = 0$ dans l'équation (3.5), nous trouvons

$$s^i = \begin{cases} C \left(\frac{1}{\lambda - \mu} \ln \left(\frac{1-s}{\mu - \lambda s} \right) \right) (\mu - \lambda s)^{-\frac{\nu}{\lambda}} & \text{si } \lambda \neq \mu \\ C \left(\frac{1}{\lambda(1-s)} \right) (\lambda(1-s))^{-\frac{\nu}{\lambda}} & \text{si } \lambda = \mu. \end{cases} \quad (3.6)$$

Posons

$$w = \begin{cases} \frac{1}{\lambda - \mu} \ln \left(\frac{1-s}{\mu - \lambda s} \right) & \text{si } \lambda \neq \mu \\ \frac{1}{\lambda(1-s)} & \text{si } \lambda = \mu. \end{cases}$$

En mettant

$$s = \begin{cases} \frac{e^{(\lambda-\mu)w} \mu - 1}{e^{(\lambda-\mu)w} \lambda - 1} & \text{si } \lambda \neq \mu \\ 1 - \frac{1}{\lambda w} & \text{si } \lambda = \mu, \end{cases}$$

dans (3.6) pour obtenir la fonction $C(w)$ requise,

$$C(w) = \begin{cases} \left(\frac{e^{(\lambda-\mu)w} \mu - 1}{e^{(\lambda-\mu)w} \lambda - 1} \right)^i \left(\mu - \lambda \frac{e^{(\lambda-\mu)w} \mu - 1}{e^{(\lambda-\mu)w} \lambda - 1} \right)^{\frac{\nu}{\lambda}} & \text{si } \lambda \neq \mu \\ \left(1 - \frac{1}{\lambda w} \right)^i w^{-\frac{\nu}{\lambda}} & \text{si } \lambda = \mu. \end{cases} \quad (3.7)$$

En combinant les équations (3.5) et (3.7), nous pouvons maintenant écrire la forme complète de $G_i(s, t)$, pour $\lambda \neq \mu$,

$$G_i(s, t) = \frac{(\mu - \lambda)^{\frac{\nu}{\lambda}} \left(s \mu e^{(\lambda-\mu)t} - \lambda - \mu (e^{(\lambda-\mu)t} - 1) \right)^i}{\left(\mu - \lambda e^{(\lambda-\mu)t} + \lambda s (e^{(\lambda-\mu)t} - 1) \right)^{\frac{\nu}{\lambda} + i}},$$

et pour $\lambda = \mu$

$$G_i(s, t) = \left(\frac{\lambda t + s(1 - \lambda t)}{1 + \lambda t - \lambda t s} \right)^i (1 + \lambda t - \lambda t s)^{-\frac{\nu}{\lambda}},$$

□

Lemme 3.3. Pour $i, j, n \geq 0$, on a

$$\sum_{k=0}^n \binom{i+k}{i} \binom{j+n-k}{j} = \binom{i+j+n+1}{n}.$$

Preuve. Pour tout $x \in]-1, 1[$ et $\alpha \in \mathbb{R}$, on a

$$(1-x)^{-\alpha-1} = \sum_{k=0}^{\infty} \binom{\alpha+k}{\alpha} x^k.$$

Ainsi, d'une part, nous avons :

$$(1-x)^{-(i+j+1)-1} = \sum_{n=0}^{\infty} \binom{i+j+n+1}{n} x^n. \quad (3.8)$$

et d'autre part :

$$(1-x)^{-(i+j+1)-1} = (1-x)^{-i-1} (1-x)^{-j-1},$$

où

$$\begin{aligned} (1-x)^{-i-1} &= \sum_{k=0}^{\infty} \binom{i+k}{i} x^k, \\ (1-x)^{-j-1} &= \sum_{k=0}^{\infty} \binom{j+k}{j} x^k. \end{aligned}$$

En utilisant le produit de Cauchy de deux séries infinies, on trouve :

$$(1-x)^{-(i+j+1)-1} = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n \binom{i+k}{i} \binom{j+n-k}{j} \right) x^n. \quad (3.9)$$

En identifiant les coefficients de (3.8) et (3.9), nous obtenons le résultat souhaité. \square

Nous présentons maintenant l'expression explicite des probabilités de transition du processus I_t . Bien que nous ayons supposé que $\lambda < \mu$, nous donnons des formules générales qui incluent également le cas où $\lambda \geq \mu$.

Théorème 3.1. Pour $t \geq 0$, les probabilités de transition de la chaîne de Markov à temps continu I_t sont données pour $i, j \in \mathbb{N}$ par :

$$p_{i,j}(t) = q(t)^r \sum_{l=0}^{\min(i,j)} \binom{i}{l} \binom{r+i+j-l-1}{j-l} \left(\frac{\mu}{\lambda}\right)^{i-l} (1-q(t))^{i+j-l} \left(\frac{1 - (\frac{\mu}{\lambda} + 1)(1-q(t))}{1-q(t)}\right)^l, \quad (3.10)$$

où

$$r = \frac{\nu}{\lambda} \quad \text{et} \quad q(t) = \begin{cases} \frac{\mu - \lambda}{\mu - \lambda e^{(\lambda - \mu)t}} & \lambda \neq \mu, \\ \frac{1}{1 + \lambda t} & \lambda = \mu. \end{cases}$$

Preuve. La première étape consiste à transformer l'équation de Kolmogorov 3.1 en une équation différentielle partielle linéaire standard du premier ordre équivalente. Dans ce

but, nous introduisons la fonction génératrice des moments du processus I_t : pour tout $(s, t) \in \times[0, +\infty]^2$, et $i \in \mathbb{N}$

$$G_i(s, t) = \sum_{j=0}^{+\infty} s^j p_{i,j}(t).$$

En utilisant l'équation de Kolmogorov progressive et après quelques réarrangements, nous trouvons :

$$\begin{aligned} \frac{\partial G_i(s, t)}{\partial t} &= \sum_{j=0}^{\infty} s^j \frac{dp_{i,j}(t)}{dt} \\ &= \left(\lambda s^2 - (\lambda + \mu) s + \mu \right) \sum_{j=1}^{\infty} j s^{j-1} p_{i,j}(t) + \nu(s-1) \sum_{j=0}^{\infty} s^j p_{i,j}(t). \end{aligned}$$

Ainsi, la fonction génératrice des moments satisfait l'équation différentielle partielle linéaire du premier ordre suivante :

$$\frac{\partial G_i(s, t)}{\partial t} = (\mu - \lambda s)(1 - s) \frac{\partial G_i(s, t)}{\partial s} + \nu(s-1)G_i(s, t),$$

avec la condition initiale $G_i(s, 0) = s^i$.

Maintenant, en utilisant le lemme **3.2**, nous avons : pour $\lambda \neq \mu$,

$$G_i(s, t) = \frac{(\mu - \lambda)^{\frac{\nu}{\lambda}} \left(s(\mu e^{(\lambda-\mu)t} - \lambda) - \mu (e^{(\lambda-\mu)t} - 1) \right)^i}{\left(\mu - \lambda e^{(\lambda-\mu)t} + \lambda s (e^{(\lambda-\mu)t} - 1) \right)^{\frac{\nu}{\lambda} + i}}$$

et pour $\lambda = \mu$

$$G_i(s, t) = \left(\frac{\lambda t + s(1 - \lambda t)}{1 + \lambda t - \lambda t s} \right)^i (1 + \lambda t - \lambda t s)^{-\frac{\nu}{\lambda}}.$$

Le cas $i = 0$ donne la fonction génératrice suivante pour les probabilités de transition $p_{0,j}(t)$, $t \geq 0$:

$$G_0(s, t) = \begin{cases} \left(\frac{\mu - \lambda}{\mu - \lambda e^{(\lambda-\mu)t}} \right)^{\frac{\nu}{\lambda}} \left(1 - \frac{\lambda(1 - e^{(\lambda-\mu)t})}{\mu - \lambda e^{(\lambda-\mu)t}} s \right)^{-\frac{\nu}{\lambda}} & \lambda \neq \mu, \\ \left(\frac{1}{1 + \lambda t} \right)^{\frac{\nu}{\lambda}} \left(1 - \frac{\lambda t}{1 + \lambda t} s \right)^{-\frac{\nu}{\lambda}} & \lambda = \mu. \end{cases}$$

Il s'agit de la fonction génératrice de la loi binomiale négative généralisée avec des paramètres

$$r = \frac{\nu}{\lambda} \quad \text{et} \quad q(t) = \begin{cases} \frac{\mu - \lambda}{\mu - \lambda e^{(\lambda-\mu)t}} & \lambda \neq \mu, \\ \frac{1}{1 + \lambda t} & \lambda = \mu. \end{cases}$$

Par conséquent, pour $j \in \mathbb{N}$ et $t \geq 0$, on obtient

$$p_{0,j}(t) = \binom{j+r-1}{r-1} q(t)^r (1 - q(t))^j. \quad (3.11)$$

Ensuite, la forme générale de la fonction G_i pour $i \neq 0$ peut être écrite comme suit :

$$G_i(s, t) = G_0(s, t)N_i(s, t), \quad (3.12)$$

où

$$N_i(s, t) = \begin{cases} \left(\frac{\mu - \lambda s - \mu(1-s)e^{(\lambda-\mu)t}}{\mu - \lambda s - \lambda(1-s)e^{(\lambda-\mu)t}} \right)^i & \lambda \neq \mu, \\ \left(\frac{1 + (\lambda t - 1)(1-s)}{1 + \lambda t(1-s)} \right)^i & \lambda = \mu, \end{cases}$$

est la fonction génératrice des moments d'un processus de naissance et mort linéaire sans immigration. En principe, l'équation (3.12) reflète l'indépendance entre le processus de contamination par l'environnement (G_0) et l'auto-contamination de la population (N). Le processus de naissance et mort linéaire sans immigration possède la propriété de branchement qui se traduit par la forme puissance de sa fonction génératrice des moments, voir par exemple [58] ou [59]. On peut facilement retrouver ses probabilités de transition, notées $p_{i,j}^{(N)}(t), t \geq 0$ à partir du développement en série de $N_i(s, t)$:

$$p_{i,j}^{(N)}(t) = \sum_{l=0}^{\min(i,j)} \binom{i}{l} \binom{i+j-l-1}{i-1} \alpha(t)^{i-l} \beta(t)^{j-l} (1 - \alpha(t) - \beta(t))^l, \quad (3.13)$$

où $\alpha(t) = \frac{\mu}{\lambda}(1 - q(t))$ et $\beta(t) = 1 - q(t)$.

Ainsi, les probabilités $p_{i,j}(t)$ sont données par la convolution suivante de l'équation (3.12) :

$$p_{i,j}(t) = \sum_{k=0}^j p_{0,k}(t) p_{i,j-k}^{(N)}(t),$$

et les équations (3.11), (3.13) donnent finalement

$$\begin{aligned} p_{i,j}(t) &= q(t)^r \sum_{l=0}^{\min(i,j)} \binom{i}{l} \left(\frac{\mu}{\lambda} \right)^{i-l} (1 - q(t))^{i+j-l} \left(\frac{1 - (\frac{\mu}{\lambda} + 1)(1 - q(t))}{1 - q(t)} \right)^l \\ &\quad \times \sum_{k=0}^{j-l} \binom{r-1+k}{r-1} \binom{i-1+j-l-k}{i-1}, \end{aligned}$$

En utilisant le lemme 3.3, nous obtenons le résultat souhaité

$$\begin{aligned} p_{i,j}(t) &= q(t)^r \sum_{l=0}^{\min(i,j)} \binom{i}{l} \binom{r+i+j-l-1}{j-l} \left(\frac{\mu}{\lambda} \right)^{i-l} (1 - q(t))^{i+j-l} \\ &\quad \times \left(\frac{1 - (\frac{\mu}{\lambda} + 1)(1 - q(t))}{1 - q(t)} \right)^l. \end{aligned}$$

□

Corollaire 3.1. *En dérivant la fonction génératrice des moments par rapport à s et en prenant $s = 1$, nous pouvons retrouver l'expression bien connue de l'espérance du LBDPI, $m(t) = \mathbb{E}_i[I_t]$:*

$$m(t) = \mathbb{E}[I_t | I_0 = i] = \begin{cases} \frac{\nu}{\lambda - \mu} (e^{(\lambda - \mu)t} - 1) + i e^{(\lambda - \mu)t} & \lambda \neq \mu, \\ \nu t + i & \lambda = \mu. \end{cases}$$

Remarque 3.1. *Le résultat ci-dessus sur l'espérance est cohérent avec celui obtenu dans [60] en utilisant une autre méthode qui consiste à dériver une équation pour $m(t+h)$, en conditionnant sur I_t , puis en résolvant directement une équation différentielle sur $m(t)$. De plus, nous constatons que lorsque $\lambda \geq \mu$ et $\nu > 0$, L'espérance tend vers l'infini. Lorsque $\lambda > \mu$, l'espérance augmente exponentiellement vite. Lorsque $\lambda = \mu$, la moyenne conditionnelle de I_t augmente linéairement. Cependant, dans le cas où $\lambda < \mu$ et $\nu > 0$, l'espérance se stabilise et converge vers une constante $\nu/(\mu - \lambda)$.*

Il est bien connu que la matrice de transition de la chaîne $(I_{n\Delta t})_{n \in \mathbb{N}}$ est $P(\Delta t)$, nous avons donc une forme analytique pour les probabilités de transition en temps discret. Nous voulons maintenant inverser l'équation (3.10) afin d'obtenir une expression des paramètres d'intérêt (λ, μ, ν) en fonction des probabilités de transition. Comme il n'y a que trois paramètres inconnus, il devrait suffire de sélectionner trois probabilités de transition spécifiques. Au vu des données réelles décrites dans la section 3.4.2, nous choisissons d'utiliser $p_{0,0}(\Delta t)$, $p_{0,1}(\Delta t)$ et $p_{1,0}(\Delta t)$ car elles correspondent aux transitions les plus fréquemment observées et ont les expressions analytiques les plus simples. Pour simplifier les notations, dans la suite du texte, nous noterons $p_{0,0}$, $p_{0,1}$ et $p_{1,0}$ au lieu de $p_{0,0}(\Delta t)$, $p_{0,1}(\Delta t)$ et $p_{1,0}(\Delta t)$, et de manière générale $p_{i,j}$ au lieu de $p_{i,j}(\Delta t)$.

Nous présentons maintenant le théorème suivant qui donne les estimateurs ponctuels de λ , μ et ν du processus I_t dans le cas où $\lambda \neq \mu$.

Théorème 3.2. *Les paramètres λ , μ et ν sont donnés par les relations suivantes $(\lambda, \mu, \nu) = g(p_{0,0}, p_{0,1}, p_{1,0})$ avec*

$$\begin{aligned}\lambda &= g_1(p_{0,0}, p_{0,1}, p_{1,0}) = \frac{\ln\left(\frac{u}{q}\right)(q-1)}{\Delta t(q-u)} \\ \mu &= g_2(p_{0,0}, p_{0,1}, p_{1,0}) = \frac{\ln\left(\frac{u}{q}\right)(u-1)}{\Delta t(q-u)} \\ \nu &= g_3(p_{0,0}, p_{0,1}, p_{1,0}) = \frac{\ln(p_{0,0}) \ln\left(\frac{u}{q}\right)(q-1)}{\ln(q) \Delta t(q-u)},\end{aligned}$$

où

$$\begin{aligned}q &= \frac{p_{0,1}}{p_{0,0} \ln(p_{0,0})} W \left(\frac{\left(\frac{p_{0,0}}{p_{0,1}} + 1\right) \ln(p_{0,0})}{p_{0,1}} \right), \\ u &= 1 - \frac{p_{1,0}}{p_{0,0}},\end{aligned}$$

et W est la fonction W de Lambert (parfois aussi appelée la fonction Oméga [61, 62]).

Preuve. Nous utilisons l'équation (3.10) pour $(i, j) \in \{(0, 0), (0, 1), (1, 0)\}$ et $t = \Delta t$ pour obtenir

$$\begin{aligned}p_{0,0} &= q^r \\ p_{0,1} &= q^r r (1 - q) \\ p_{1,0} &= q^r \left(\frac{\mu}{\lambda}\right) (1 - q),\end{aligned}$$

avec $r = \frac{\nu}{\lambda}$ et $q = q(\Delta t) = \frac{\mu - \lambda}{\mu - \lambda e^{(\lambda - \mu)\Delta t}}$.

Les deux premières équations ci-dessus sont équivalentes au système

$$\begin{cases} \ln(p_{0,0}) &= r \ln(q) \\ \frac{p_{0,1}}{p_{0,0}} &= r(1 - q), \end{cases} \quad (3.14)$$

ce qui donne

$$\frac{p_{0,0}}{p_{0,1}} \ln(p_{0,0}) q e^{\frac{p_{0,0}}{p_{0,1}} \ln(p_{0,0}) q} = \frac{\left(\frac{p_{0,0}}{p_{0,1}} + 1\right) \ln(p_{0,0})}{p_{0,1}},$$

on a donc,

$$q = \frac{p_{0,1}}{p_{0,0} \ln(p_{0,0})} W \left(\frac{\left(\frac{p_{0,0}}{p_{0,1}} + 1\right) \ln(p_{0,0})}{p_{0,1}} \right), \quad (3.15)$$

où W est la fonction de Lambert. A partir des équations (3.14) et (3.15), nous pouvons facilement trouver que

$$r = \frac{\ln(p_{0,0})}{\ln(q)}.$$

Définissons maintenant

$$u = q e^{(\lambda - \mu)\Delta t} = 1 - \frac{p_{1,0}}{p_{0,0}}.$$

L'équation (3.10) pour $(i, j) \in \{(0, 0), (0, 1), (1, 0)\}$ peut maintenant être réécrite comme suit

$$\begin{cases} r &= \frac{\nu}{\lambda} \\ q &= \frac{\mu - \lambda}{\mu - \lambda e^{(\lambda - \mu)\Delta t}} \\ u &= \frac{(\mu - \lambda) e^{(\lambda - \mu)\Delta t}}{\mu - \lambda e^{(\lambda - \mu)\Delta t}}. \end{cases} \quad (3.16)$$

Étudions maintenant l'identifiabilité du système 3.16.

Soit $\lambda, \lambda', \mu, \mu', \nu, \nu' > 0$ tel que $\lambda \neq \mu$ et $\lambda' \neq \mu'$. Supposons que

$$\begin{cases} \frac{\nu}{\lambda} &= \frac{\nu'}{\lambda'} \\ \frac{\mu - \lambda}{\mu - \lambda e^{(\lambda - \mu)\Delta t}} &= \frac{\mu' - \lambda'}{\mu' - \lambda' e^{(\lambda' - \mu')\Delta t}} \\ \frac{(\mu - \lambda) e^{(\lambda - \mu)\Delta t}}{\mu - \lambda e^{(\lambda - \mu)\Delta t}} &= \frac{(\mu' - \lambda') e^{(\lambda' - \mu')\Delta t}}{\mu' - \lambda' e^{(\lambda' - \mu')\Delta t}}. \end{cases} \quad (3.17)$$

En prenant le rapport des deux dernières équations, on obtient

$$\mu = \lambda - \lambda' + \mu'.$$

En substituant l'expression de μ dans la deuxième équation de (3.17), on obtient

$$\lambda \left(1 - e^{(\lambda' - \mu')\Delta t}\right) = \lambda' \left(1 - e^{(\lambda' - \mu')\Delta t}\right),$$

puisque $\lambda' \neq \mu'$, nous déduisons que $\lambda = \lambda'$, donc $\mu = \mu'$.

A partir de la première équation de (3.17), nous obtenons également que $\nu = \nu'$.

Par conséquent, le système d'équations 3.16 est identifiable.

Enfin, la résolution du système (3.16) donne lieu aux expressions de λ , μ et ν en fonction des probabilités de transition $p_{0,0}$, $p_{0,1}$ et $p_{1,0}$ données dans le théorème 3.2. \square

Corollaire 3.2. *Soit $\hat{p}_{0,0}^n$, $\hat{p}_{0,1}^n$ et $\hat{p}_{1,0}^n$ des estimateurs consistants et asymptotiquement normaux de $p_{0,0}$, $p_{0,1}$ et $p_{1,0}$ avec une matrice de variance-covariance asymptotique Σ' . Alors, les estimateurs $(\hat{\lambda}^n, \hat{\mu}^n, \hat{\nu}^n) = g(\hat{p}_{0,0}^n, \hat{p}_{0,1}^n, \hat{p}_{1,0}^n)$ obtenus par plug-in à partir des formules du théorème 3.2 sont également consistants et asymptotiquement normaux avec une matrice de variance-covariance asymptotique donnée par $\Sigma = Dg\Sigma'Dg^T$, où Dg est la matrice jacobienne de l'application g et Dg^T est sa transposée.*

Preuve. Les $\hat{p}_{0,0}^n$, $\hat{p}_{0,1}^n$ et $\hat{p}_{1,0}^n$ sont des estimateurs consistants de $p_{0,0}$, $p_{0,1}$ et $p_{1,0}$, c'est-à-dire, quand $n \rightarrow \infty$

$$\begin{aligned}\hat{p}_{0,0}^n &\xrightarrow{p.s.} p_{0,0} \\ \hat{p}_{0,1}^n &\xrightarrow{p.s.} p_{0,1} \\ \hat{p}_{1,0}^n &\xrightarrow{p.s.} p_{1,0}.\end{aligned}$$

Étant donné que g est une fonction continue (elle est définie par la composition et le produit de fonctions continues), nous déduisons que

$$\begin{aligned}\hat{\lambda}^n &\xrightarrow{p.s.} \lambda \\ \hat{\mu}^n &\xrightarrow{p.s.} \mu \\ \hat{\nu}^n &\xrightarrow{p.s.} \nu.\end{aligned}$$

Maintenant, pour la propriété de la normalité asymptotique, les $\hat{p}_{0,0}^n$, $\hat{p}_{0,1}^n$ et $\hat{p}_{1,0}^n$ sont des estimateurs asymptotiquement normaux de $p_{0,0}$, $p_{0,1}$ et $p_{1,0}$ avec une matrice de variance-covariance asymptotique Σ' , c'est-à-dire, quand $n \rightarrow \infty$

$$\sqrt{n} \begin{pmatrix} \hat{p}_{0,0}^n - p_{0,0} \\ \hat{p}_{0,1}^n - p_{0,1} \\ \hat{p}_{1,0}^n - p_{1,0} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_3(0, \Sigma'). \quad (3.18)$$

Puisque g est une fonction différentiable, la δ -méthode à 3.18, donne, quand $n \rightarrow \infty$

$$\sqrt{n} \left(g \left(\hat{p}_{0,0}^n, \hat{p}_{0,1}^n, \hat{p}_{1,0}^n \right) - g \left(p_{0,0}, p_{0,1}, p_{1,0} \right) \right) \xrightarrow{\mathcal{L}} \mathcal{N}_3(0, Dg\Sigma'Dg^T),$$

i.e :

$$\sqrt{n} \begin{pmatrix} \hat{\lambda}^n - \lambda \\ \hat{\mu}^n - \mu \\ \hat{\nu}^n - \nu \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_3(0, \Sigma).$$

Le calcul de la matrice jacobienne de g est donné dans l'annexe B. \square

Application : Considérons par exemple $\hat{p}_{i,j}^n$, les estimateurs du maximum de vraisemblance de $p_{i,j}$ pour les observations $(I_{k\Delta t}, 0 \leq k \leq n)$ et pour $(i, j) \in \{(0, 0), (0, 1), (1, 0)\}$. Ces estimateurs correspondent au rapport entre le nombre de transitions de i à j et le nombre total de transitions de i dans la suite $(I_{k\Delta t}, 0 \leq k \leq n)$, voir 3.4.1. Sous l'hypothèse de récurrence positive ($\lambda < \mu$ et $\nu > 0$) de la chaîne $(I_{k\Delta t})$, ils sont consistants et

asymptotiquement normaux avec une matrice de variance-covariance asymptotique [63, **Théorème 2.1**] donnée par

$$\Sigma' = \begin{pmatrix} \frac{p_{0,0}(1-p_{0,0})}{\pi_0} & -\frac{p_{0,0}p_{0,1}}{\pi_0} & 0 \\ -\frac{p_{0,0}p_{0,1}}{\pi_0} & \frac{p_{0,1}(1-p_{0,1})}{\pi_0} & 0 \\ 0 & 0 & \frac{p_{1,0}(1-p_{1,0})}{\pi_1} \end{pmatrix},$$

où π est la loi invariante donnée par l'équation **1.3**. Plus précisément, on a, quand $n \rightarrow \infty$,

$$\sqrt{n} \begin{pmatrix} \hat{p}_{0,0}^n - p_{0,0} \\ \hat{p}_{0,1}^n - p_{0,1} \\ \hat{p}_{1,0}^n - p_{1,0} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_3(0, \Sigma'),$$

Ainsi, ces estimateurs satisfont aux hypothèses du corollaire **3.2**. Cependant, ils ne peuvent pas être utilisés dans notre contexte car les $(I_{k\Delta t}, 0 \leq k \leq n)$ sont cachés. Nous nous tournons donc vers une autre procédure d'estimation pour $p_{0,0}$, $p_{0,1}$ et $p_{1,0}$ via un modèle de Markov caché.

3.3.2 Modèle de Markov caché

Nous passons maintenant à l'estimation des probabilités de transition $p_{0,0}$, $p_{0,1}$ et $p_{1,0}$ de la chaîne cachée $(I_{n\Delta t})_{n \in \mathbb{N}}$ à partir des observations disponibles $(O_{\lfloor n\Delta t, (n+1)\Delta t \rfloor})_{n \in \mathbb{N}}$. Pour simplifier les notations, nous notons dans cette section I_n au lieu de $I_{n\Delta t}$ et O_n au lieu de $O_{\lfloor n-1 \rfloor \Delta t, n \Delta t \rfloor}$.

L'objectif de cette section est d'adapter les algorithmes HMM standard pour obtenir des estimations des probabilités de transition de la chaîne cachée (I_n) à partir des observations (O_n) . Cependant, comme expliqué dans la section précédente **3.2.3**, O_n dépend à la fois de I_{n-1} et de I_n , et conditionnellement à ces deux états I_{n-1} et I_n , les observations O_n sont mutuellement indépendantes. Ainsi, le schéma d'émission n'est pas standard (**Fig.3.2**).

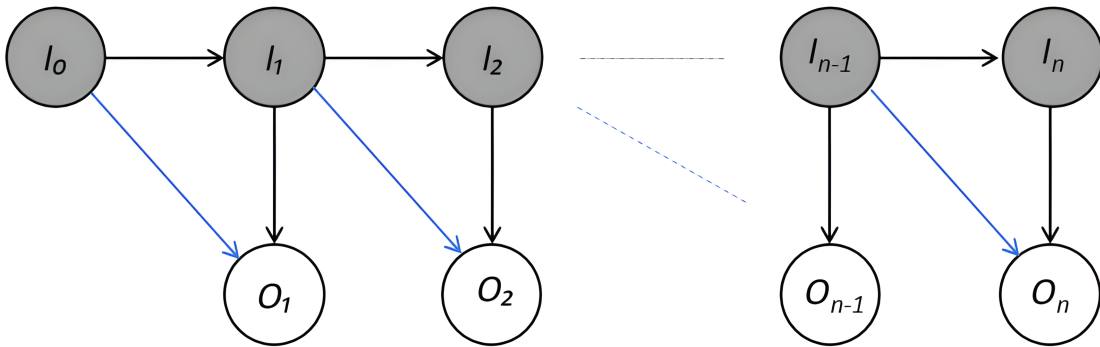


FIGURE 3.2 – Schéma d'émission du HMM.

Pour contourner la difficulté découlant du schéma d'émission non standard, nous utilisons la chaîne de Markov bidimensionnelle correspondant à des paires d'états cachés.

Notons par $(Z_n)_{n \in \mathbb{N}^*} = (I_{n-1}, I_n)_{n \in \mathbb{N}^*}$ la chaîne de Markov non observée à deux dimensions consécutifs à valeurs dans l'espace d'état caché \mathbb{N}^2 . Il est alors facile de prouver que (Z_n, O_n) est un HMM standard avec les caractéristiques suivantes.

Lemme 3.4. *Le processus $(Z_n, O_n)_{n \in \mathbb{N}^*}$ est un modèle de Markov caché dont les caractéristiques sont données par le triple $M = (Q, \psi, \rho)$, où*

i) La matrice de probabilité de transition Q de la chaîne cachée (Z_n) est

$$Q_{(i,j),(i',j')} = \mathbb{P}(Z_{n+1} = (i', j') | Z_n = (i, j)) = p_{i',j'} \delta_{i'=j},$$

pour $(i, j), (i', j') \in \mathbb{N}^2$.

ii) La probabilité d'émission ψ du processus O étant donné le processus Z est

$$\psi_{(i,j)}(o) = \mathbb{P}(O_n = o | Z_n = (i, j)) = \frac{p_{i,(j,o)}}{p_{i,j}},$$

pour $i, j, o \in \mathbb{N}$, où $p_{i,(j,o)} = p_{i,(j,o)}(\Delta t)$ est caractérisée par l'équation 3.2.

iii) La loi initiale ρ du processus d'état Z est

$$\rho_{i,j} = \mathbb{P}(Z_1 = (i, j)) = p_{i,j} \pi_i,$$

pour $i, j \in \mathbb{N}$, où π est la loi invariante de I donnée par l'équation 1.3.

Preuve. *i)* La probabilité de transition Q : soient $(i, j), (i', j') \in \mathbb{N}^2$, nous avons

$$\begin{aligned} Q_{(i,j),(i',j')} &= \mathbb{P}(Z_{n+1} = (i', j') | Z_n = (i, j)) \\ &= \frac{\mathbb{P}(I_n = i', I_{n+1} = j', I_{n-1} = i, I_n = j)}{\mathbb{P}(I_{n-1} = i, I_n = j)}. \end{aligned}$$

En utilisant la propriété de Markov de la chaîne (I_n) , nous obtenons

$$\begin{aligned} Q_{(i,j),(i',j')} &= \frac{\mathbb{P}(I_{n+1} = j' | I_n = i') \mathbb{P}(I_n = i', I_{n-1} = i)}{\mathbb{P}(I_n = j, I_{n-1} = i)} \delta_{i'=j} \\ &= p_{i',j'} \delta_{i'=j}. \end{aligned}$$

ii) La probabilité d'émission ψ : par la définition de la probabilité conditionnelle.

iii) La loi initiale ρ : soit $i, j \in \mathbb{N}$

$$\begin{aligned} \rho_{i,j} &= \mathbb{P}(Z_1 = (i, j)) \\ &= \mathbb{P}(I_1 = j | I_0 = i) \mathbb{P}(I_0 = i) \\ &= p_{i,j} \pi_i. \end{aligned}$$

□

L'algorithme standard pour estimer les paramètres les plus probables $M = (Q, \psi, \rho)$ du HMM compte tenu des observations est l'algorithme forward-backward ou algorithme de Baum-Welch, un cas particulier de l'algorithme d'espérance-maximisation (EM)[25]. Il s'agit d'un algorithme itératif. À l'étape n , étant donné les paramètres actuels $M^n = (Q^n, \psi^n, \rho^n)$, la vraisemblance des observations $\mathbb{P}(O_1 = o_1, \dots, O_T = o_T | M^n)$ est maximisée et le modèle M^n est mis à jour en M^{n+1} en conséquence. Le principal avantage de l'algorithme de Baum-Welch est que la vraisemblance peut être maximisée explicitement.

Malheureusement, l'exécution de la procédure HMM standard du package R HMM sur (Z_n, O_n) produit une instabilité numérique et des erreurs dues aux nombreux zéros dans la matrice de transition Q de Z . Nous avons donc réécrit les fonctions principales en adaptant les formules récursives à la forme particulière de Q afin d'obtenir des estimateurs explicites des $p_{i,j}$.

Nous commençons par définir la probabilité forward $\alpha_{(i,j)}(t)$ comme la probabilité conjointe d'observer les t premiers vecteurs et d'être dans l'état (i, j) à l'instant t . Cette probabilité peut être calculée par la récurrence suivante : pour tout $(i, j) \in \mathbb{N}^2$ et $2 \leq t \leq T$

$$\begin{cases} \alpha_{(i,j)}(1) = p_{i,j} \pi_i \psi_{(i,j)}(o_1), \\ \alpha_{(i,j)}(t) = \psi_{(i,j)}(o_t) p_{i,j} \sum_{i' \in \mathbb{N}} \alpha_{(i',i)}(t-1). \end{cases} \quad (3.19)$$

Ensuite, nous définissons la probabilité backward $\beta_{(i,j)}(t)$ comme la probabilité conditionnelle d'observer les vecteurs après le temps t , étant donné que l'état au temps t est (i, j) . La probabilité backward peut être évaluée à l'aide de la formule du récurrence : pour tout $(i, j) \in \mathbb{N}^2$ et $2 \leq t \leq T$

$$\begin{cases} \beta_{(i,j)}(T) = 1, \\ \beta_{(i,j)}(t-1) = \sum_{j' \in \mathbb{N}} p_{j,j'} \psi_{(j,j')}(o_t) \beta_{(j,j')}(t). \end{cases} \quad (3.20)$$

En utilisant ces deux probabilités, la vraisemblance est simplement calculée en additionnant tous les produits forward et backward, comme indiqué ci-dessous.

$$\mathbb{P}_M(O_{1:T} = o_{1:T}) := \mathbb{P}(O_{1:T} = o_{1:T} | M) = \sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} \alpha_{(i,j)}(1) \beta_{(i,j)}(T).$$

Considérons maintenant la probabilité conditionnelle d'une transition de l'état (i, j) au temps t à l'état (i', j') au temps $t+1$, étant donné les observations $O_{1:T} = o_{1:T}$ et le modèle M

$$\begin{aligned} \xi_{(i,j),(i',j')}(t) &= \mathbb{P}(Z_t = (i, j), Z_{t+1} = (i', j') | O_{1:T} = o_{1:T}, M) \\ &= \frac{\alpha_{(i,j)}(t) p_{(i',j')} \psi_{(i',j')}(o_{t+1}) \beta_{(i',j')}(t+1)}{\sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} \alpha_{(i,j)}(t) \beta_{(i,j)}(t)} \delta_{i'=j}. \end{aligned} \quad (3.21)$$

Soit $\gamma_{(i,j)}(t) = \sum_{j' \in \mathbb{N}} \xi_{(i,j),(j,j')}(t)$, la probabilité conditionnelle d'être dans l'état (i, j) à l'instant t , étant donné les observations $O_{1:T} = o_{1:T}$ et le modèle M :

$$\gamma_{(i,j)}(t) = \frac{\alpha_{(i,j)}(t) \beta_{(i,j)}(t)}{\sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} \alpha_{(i,j)}(t) \beta_{(i,j)}(t)}. \quad (3.22)$$

En termes d'interprétation, $\sum_{t=1}^{T-1} \xi_{(i,j),(i',j')}(t)$ et $\sum_{t=1}^{T-1} \gamma_{(i,j)}(t)$ représentent respectivement des estimateurs du nombre de transitions de (i, j) vers (i', j') et du nombre de transitions issues de (i, j) dans la suite $Z_{1:T}$, fondés sur l'observation de $O_{1:T} = o_{1:T}$.

Le théorème suivant présente les mises à jour itératives de l'algorithme de Baum-Welch pour estimer les caractéristiques du HMM, étant donné le modèle à l'étape n et la suite des observations $O_{1:T} = o_{1:T}$.

Théorème 3.3. *Étant donné les paramètres $M^n = (Q^n, \psi^n, \rho^n)$, les estimations du maximum de vraisemblance de $M^{n+1} = (Q^{n+1}, \psi^{n+1}, \rho^{n+1})$ étant donné les observations (o_1, \dots, o_T)*

sont données, pour $i, j, o \in \mathbb{N}$, par

$$\begin{aligned} Q_{(i,j),(i',j')}^{n+1} &= \frac{\sum_{t=1}^{T-1} \xi_{(i,j),(i',j')}^n(t)}{\sum_{t=1}^{T-1} \gamma_{(i,j)}^n(t)} \delta_{i'=j}, \\ \psi_{(i,j)}^{n+1}(y) &= \frac{\sum_{t=1}^T \mathbb{1}_{o_t=o} \gamma_{(i,j)}^n(t)}{\sum_{t=1}^T \gamma_{(i,j)}^n(t)}, \\ \rho_{i,j}^{n+1} &= \gamma_{(i,j)}^n(1), \\ p_{i,j}^{n+1} &= \frac{\sum_{i' \in \mathbb{N}} Q_{(i',i),(i,j)}^{n+1}}{\sum_{i' \in \mathbb{N}} \mathbb{1}_{Q_{(i',i),(i,j)}^{n+1} \neq 0}}, \end{aligned}$$

où $\alpha_{(i,j)}^n(t)$, $\beta_{(i,j)}^n(t)$, $\xi_{(i,j),(i',j')}^n(t)$ et $\gamma_{(i,j)}^n(t)$ sont définies dans les équations **3.19**, **3.20**, **3.21** et **3.22**, respectivement.

Preuve. Notons par $\mathcal{Z} = \mathbb{N}^2$ et soit $z = (i, j) \in \mathcal{Z}$, étant donné les paramètres du modèle M^n , on cherche à maximiser la probabilité d'observations $\mathbb{P}(O_{1:T} = o_{1:T} | M^{n+1})$ en utilisant l'algorithme EM (Bilmes, 1998). La fonction auxiliaire est donnée par :

$$\Gamma_{O_{1:T}}(M^n, M^{n+1}) = \sum_{z_{1:T} \in \mathcal{Z}^T} \log \mathbb{P}(O_{1:T} = o_{1:T}, Z_{1:T} = z_{1:T} | M^{n+1}) \mathbb{P}(Z_{1:T} = z_{1:T} | O_{1:T}, M^n), \quad (3.23)$$

où

$$\mathbb{P}(O_{1:T} = o_{1:T}, Z_{1:T} = z_{1:T} | M^{n+1}) = \rho_{z_1}^{n+1} \left(\prod_{t=1}^{T-1} Q_{z_t, z_{t+1}}^{n+1} \right) \left(\prod_{t=1}^T \psi_{z_t}^{n+1}(o_t) \right).$$

En remplaçant la probabilité conjointe de $(O_{1:T}, Z_{1:T})$ par son expression dans **3.23**, nous pouvons réécrire la fonction $\Gamma_{O_{1:T}}$ comme suit :

$$\begin{aligned} \Gamma_{O_{1:T}}(M^n, M^{n+1}) &= \sum_{z_{1:T} \in \mathcal{Z}^T} \log \rho_{z_1}^{n+1} \mathbb{P}(Z_{1:T} = z_{1:T} | O_{1:T}, M^n) \\ &\quad + \sum_{z_{1:T} \in \mathcal{Z}^T} \left(\sum_{t=1}^{T-1} \log Q_{z_t, z_{t+1}}^{n+1} \right) \mathbb{P}(Z_{1:T} = z_{1:T} | O_{1:T}, M^n) \\ &\quad + \sum_{z_{1:T} \in \mathcal{Z}^T} \left(\sum_{t=1}^T \log \psi_{z_t}^{n+1}(o_t) \right) \mathbb{P}(Z_{1:T} = z_{1:T} | O_{1:T}, M^n) \\ &= \Gamma_{O_{1:T}}^\rho(M^n, M^{n+1}) + \Gamma_{O_{1:T}}^Q(M^n, M^{n+1}) + \Gamma_{O_{1:T}}^\psi(M^n, M^{n+1}). \end{aligned}$$

i) Maximisation de $\Gamma_{O_{1:T}}^\rho(M^n, M^{n+1})$ par la méthode de multiplicateur de Lagrange pour contrainte $\sum_{(i',j')} \rho_{i',j'}^{n+1} = 1$,

$$\mathcal{L}(\rho, \theta) = \sum_{(i,j)} \log \rho_{i,j}^{n+1} \mathbb{P}(Z_1 = (i, j) | O_{1:T}, M^n) + \theta \left(\sum_{(i',j')} \rho_{i',j'}^{n+1} - 1 \right).$$

En dérivant, on obtient,

$$\begin{aligned} \frac{\partial \mathcal{L}(\rho, \theta)}{\partial \rho_{i,j}^{n+1}} &= \frac{\mathbb{P}(Z_1 = (i, j) | O_{1:T}, M^n)}{\rho_{i,j}^{n+1}} + \theta \\ &= 0, \end{aligned}$$

et

$$\begin{aligned}\frac{\partial \mathcal{L}(\rho, \theta)}{\partial \theta} &= \sum_{(i', j')} \rho_{i', j'}^{n+1} - 1 \\ &= 0.\end{aligned}$$

Les deux équations donne,

$$\theta = -1 \text{ et } \rho_{i, j}^{n+1} = \mathbb{P}(Z_1 = (i, j) | O_{1:T}, M^n) = \gamma_{(i, j)}^n(1).$$

ii) La fonction de Lagrange associée à la maximisation de $\Gamma_{O_{1:T}}^Q(M^n, M^{n+1})$ avec la contrainte $\sum_{(i', j')} Q_{(i, j), (i', j')}^{n+1} = 1$ est définie par :

$$\begin{aligned}\mathcal{L}(Q, \theta) &= \sum_{(i, j)} \sum_{(i', j')} \left(\sum_{t=1}^{T-1} \log Q_{(i, j), (i', j')}^{n+1} \right) \mathbb{P}(Z_t = (i, j), Z_{t+1} = (i', j') | O_{1:T}, M^n) \\ &\quad + \sum_{(i, j)} \theta_{(i, j)} \left(\sum_{(i', j')} Q_{(i, j), (i', j')}^{n+1} - 1 \right).\end{aligned}$$

En dérivant et en mettant les dérivées premières des Lagrangiens à zéro, on obtient,

$$\sum_{t=1}^{T-1} \frac{\mathbb{P}(Z_t = (i, j), Z_{t+1} = (i', j') | O_{1:T}, M^n)}{Q_{(i, j), (i', j')}^{n+1}} + \theta_{(i, j)} = 0,$$

et

$$\sum_{(i', j')} Q_{(i, j), (i', j')}^{n+1} - 1 = 0.$$

La multiplication par $Q_{(i, j), (i', j')}^{n+1}$ et la sommation sur (i', j') de la première équation permettent d'obtenir,

$$\theta_{(i, j)} = - \sum_{t=1}^{T-1} \mathbb{P}(Z_t = (i, j) | O_{1:T}, M^n).$$

Ainsi, $Q_{(i, j), (i', j')}^{n+1}$ est donnée par :

$$Q_{(i, j), (i', j')}^{n+1} = \frac{\sum_{t=1}^{T-1} \mathbb{P}(Z_t = (i, j), Z_{t+1} = (i', j') | O_{1:T}, M^n)}{\sum_{t=1}^{T-1} \mathbb{P}(Z_t = (i, j) | O_{1:T}, M^n)} = \frac{\sum_{t=1}^{T-1} \xi_{(i, j), (i', j')}^n(t)}{\sum_{t=1}^{T-1} \gamma_{(i, j)}^n(t)} \delta_{i'=j}.$$

iii) Le Lagrangien associé à la maximisation de $\Gamma_{O_{1:T}}^\psi(M^n, M^{n+1})$ avec la contrainte $\sum_o \psi_{(i, j)}^{n+1}(o) = 1$ est donné comme suit :

$$\begin{aligned}\mathcal{L}(\psi, \theta) &= \sum_{(i, j)} \left(\sum_{t=1}^T \log \psi_{(i, j)}^{n+1}(o) \right) \mathbb{P}(Z_t = (i, j) | O_{1:T}, M^n) \mathbb{1}_{o_t=o} \\ &\quad + \sum_{(i, j)} \theta_{(i, j)} \left(\sum_o \psi_{(i, j)}^{n+1}(o) - 1 \right).\end{aligned}$$

Comme précédemment, on dérive et on trouve,

$$\sum_{t=1}^T \frac{\mathbb{P}(Z_t = (i, j) | O_{1:T}, M^n) \mathbb{1}_{o_t=o}}{\psi_{(i, j)}^{n+1}(o)} + \theta_{(i, j)} = 0,$$

et

$$\sum_o \psi_{(i,j)}^{n+1}(o) - 1 = 0.$$

Par conséquent,

$$\theta_{(i,j)} = - \sum_{t=1}^T \mathbb{P}(Z_t = (i,j) | O_{1:T}, M^n),$$

et

$$\psi_{(i,j)}^{n+1}(o) = \frac{\sum_{t=1}^T \mathbb{P}(Z_t = (i,j) | O_{1:T}, M^n) \mathbb{1}_{o_t=o}}{\sum_{t=1}^T \mathbb{P}(Z_t = (i,j) | O_{1:T}, M^n)} = \frac{\sum_{t=1}^T \mathbb{1}_{o_t=o} \gamma_{(i,j)}^n(t)}{\sum_{t=1}^{T-1} \gamma_{(i,j)}^n(t)}.$$

iv) Enfin, pour établir un estimateur de $p_{i,j}$, il suffit de remarquer, à partir de la définition de la matrice de transition Q de la chaîne Z , que :

$$\sum_{i'} p_{i,j} = \sum_{i'} Q_{(i',i),(i,j)}.$$

Ainsi,

$$p_{i,j}^{n+1} = \frac{\sum_{i' \in \mathbb{N}} Q_{(i',i),(i,j)}^{n+1}}{\sum_{i' \in \mathbb{N}} \mathbb{1}_{Q_{(i',i),(i,j)}^{n+1} \neq 0}}.$$

□

Après m itérations, nous obtenons les paramètres optimaux du modèle et les estimations $\hat{\lambda}$, $\hat{\mu}$ et $\hat{\nu}$ sont données par

$$(\hat{\lambda}, \hat{\mu}, \hat{\nu}) = g(p_{0,0}^m, p_{0,1}^m, p_{1,0}^m).$$

Comme d'habitude pour l'algorithme EM, le résultat dépend fortement de la condition initiale, et il faut également régler le nombre d'itérations et trouver un moyen approprié de tronquer les sommes infinies. Tous ces points sont abordés dans la section suivante.

3.4 Étude numérique

Nous étudions maintenant les performances numériques de notre procédure d'estimation sur des données synthétiques et réelles. Tous les codes ont été développés en R. Pour mettre en œuvre la fonction Lambert-W, nous utilisons le package `lambertW R` [64]. Pour l'algorithme de Baum-Welch, nous avons adapté les fonctions du package `HMM` [65].

3.4.1 Données synthétiques

Nous commençons par étudier les performances des estimateurs $\hat{\lambda}^n$, $\hat{\mu}^n$ et $\hat{\nu}^n$ sur des données simulées. Afin de quantifier la perte d'information due aux états cachés, nous allons utiliser deux cadres différents. Dans un premier temps, nous allons considérer le cadre où le processus (I_n) est entièrement observé. Nous allons alors calculer les estimateurs en utilisant toutes les données disponibles. Dans un second temps, nous allons utiliser le cadre du HMM. En utilisant les données simulées, nous allons également étudier la convergence des estimateurs en fonction de la taille de l'échantillon n .

Observations complètes

Lorsque la chaîne de Markov à temps discret (I_n) est observée, on peut utiliser les estimations du maximum de vraisemblance pour les probabilités de transition $p_{i,j}$, [66]

$$\hat{p}_{i,j}^n = \frac{\sum_{k=0}^{n-1} \mathbb{1}_{\{I_k=i, I_{k+1}=j\}}}{\sum_{k=0}^{n-1} \mathbb{1}_{\{I_k=i\}}}.$$

Nous avons simulé un LBDPI partant de $I_0 = 0$ avec $\lambda = 0.03$, $\mu = 0.1$ et $\nu = 0.01$, jusqu'à un horizon temporel (continu) H pour plusieurs valeurs de Δt . Pour estimer les probabilités de transition entre les états discrets par la méthode du maximum de vraisemblance (EMV), nous utilisons la fonction `markovchainFit` du package `markovchain` pour les modèles de chaînes de Markov à temps discret [67].

Le tableau (3.1) présente les valeurs estimées des paramètres avec l'erreur standard asymptotique calculée à partir du corollaire 3.2. Nous avons effectué plusieurs simulations

TABLE 3.1 – Estimation des paramètres et de leur erreur standard lorsque le nombre de personnes infectées est observé sur une période Δt et un horizon temporel H .

	$H = 1000$		
Pas du temps Δt	1	7	30
nombre d'observations n	1001	143	34
$\hat{\lambda}^n$	0.027 (0.100)	0.081 (0.053)	0.049 (0.065)
$\hat{\mu}^n$	0.084 (0.035)	0.105 (0.046)	0.070 (0.128)
$\hat{\nu}^n$	0.012 (0.484)	0.007 (0.106)	0.007 (0.076)
	$H = 5000$		
Pas du temps Δt	1	7	30
nombre d'observations n	5001	715	167
$\hat{\lambda}^n$	0.041 (0.045)	0.052 (0.024)	0.064 (0.029)
$\hat{\mu}^n$	0.088 (0.016)	0.132 (0.020)	0.117 (0.058)
$\hat{\nu}^n$	0.009 (0.216)	0.012 (0.047)	0.009 (0.034)
	$H = 50000$		
Pas du temps Δt	1	7	30
nombre d'observations n	50001	7143	1667
$\hat{\lambda}^n$	0.030 (0.014)	0.023 (0.007)	0.020 (0.009)
$\hat{\mu}^n$	0.098 (0.004)	0.097 (0.006)	0.102(0.018)
$\hat{\nu}^n$	0.010 (0.068)	0.010 (0.014)	0.010 (0.011)

du processus de naissance et de mort avec immigration pour différentes tailles d'échantillon n , correspondant à des tailles petites, moyennes et très grandes. Les estimations $\hat{\lambda}^n$, $\hat{\mu}^n$ et $\hat{\nu}^n$ sont calculées avec ces échantillons et comparées aux valeurs réelles du paramètre.

La table (3.1) montre que les estimations convergent vers les vraies valeurs des paramètres et que les erreurs standards asymptotiques sont faibles. Nous constatons également que plus la taille de l'échantillon n est grande (pour un Δt petit et un H fixe, ou pour un H grand et un Δt fixe), plus les estimations sont précises, ce qui est cohérent avec la théorie. Cependant, pour de petits échantillons (moins de 500), les estimations oscillent beaucoup, ce qui suggère que les estimateurs ne sont pas très précis. Cela peut être dû à une forte

variabilité des données ou à un échantillonnage insuffisant.

En conclusion, ces résultats confirment l'efficacité de nos estimateurs et leur capacité à estimer les paramètres du processus de naissance et de mort avec immigration, en particulier pour des échantillons de taille suffisamment grande.

Observations du cumul des nouveaux cas déclarés

Nous étudions maintenant le cas où seuls les cumuls de nouveaux cas déclarés (O_n) sont observés. Pour cela, nous avons simulé un processus de LBDPI partant de $I_0 \sim \pi$ avec des paramètres $\lambda = 0.03$, $\mu = 0.1$ et $\nu = 0.01$ jusqu'à un horizon temporel continu H (**Fig.3.3**). Ensuite, nous avons extrait les cumuls de nouveaux cas déclarés, ce qui correspond au nombre de sauts d'amplitude -1 de I sur l'intervalle $]n\Delta t, (n+1)\Delta t]$, afin d'estimer les paramètres λ , μ et ν à l'aide de notre procédure d'estimation (**Fig.3.4**).

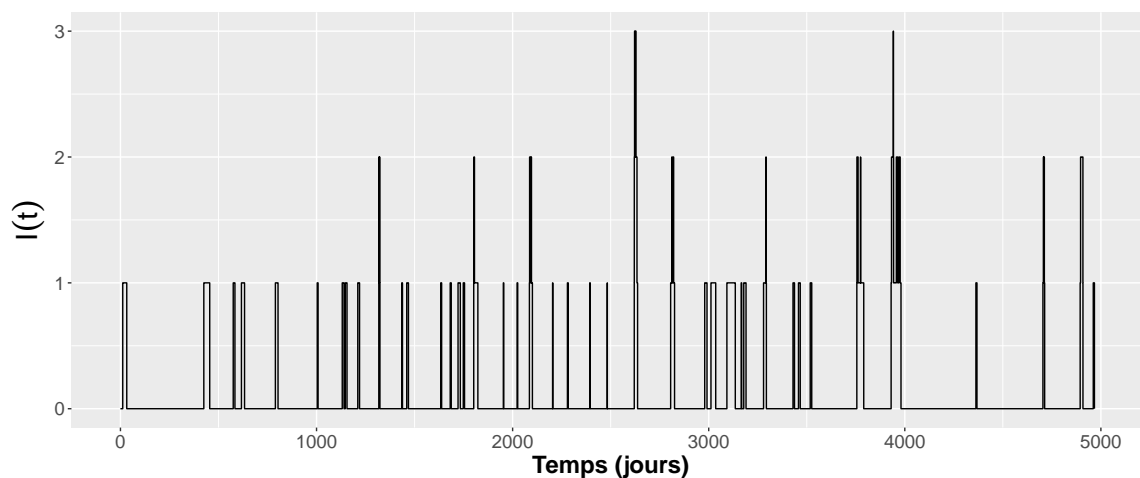


FIGURE 3.3 – Trajectoire du processus naissance-mort avec immigration (LBDPI) I_t sur un horizon $H = 5000$ jours.

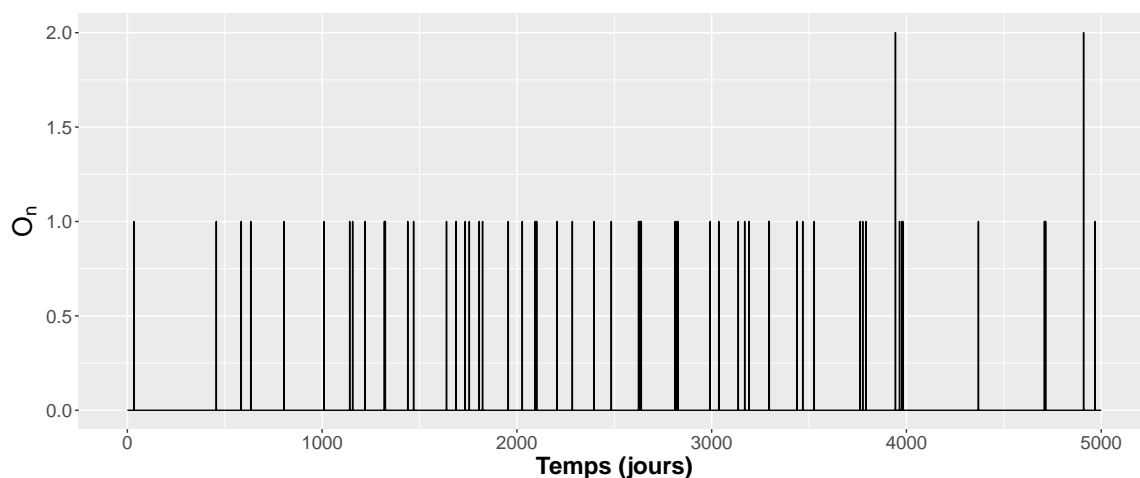


FIGURE 3.4 – Le cumul des nouveaux déclarés O_n correspondant au processus I pour une valeur de $\Delta t = 1$ jour.

Pour le HMM, nous choisissons les paramètres initiaux dans les plages suivantes : $\lambda^{(0)} \in [0.028, 0.034]$, $\mu^{(0)} \in [0.08, 0.15]$ et $\nu^{(0)} \in [0.008, 0.015]$ avec $length.out = 15$ (en générant 15 valeurs initiales pour chaque paramètre), ce qui nous permet d'initialiser les caractéristiques du HMM : $M^{(0)} = (Q^{(0)}, \psi^{(0)}, \rho^{(0)})$, et nous fixons les paramètres $\Delta t = 1$, $maxIterations = 500$ (nombre maximum d'itérations) et $delta = 10^{-9}$ (critère d'arrêt). Nous exécutons l'algorithme HMM pour chaque valeur de $(\lambda_i^{(0)}, \mu_i^{(0)}, \nu_i^{(0)})$ pour $i = 1, \dots, 15$, et nous choisissons le triplet avec la plus grande vraisemblance.

$$(\hat{\lambda}, \hat{\mu}, \hat{\nu}) = \max_{(\lambda^n, \hat{\mu}^n, \hat{\nu}^n)} \mathbb{P}(O|M^n).$$

Sur le plan numérique, on ne peut pas considérer des espaces d'états infinis. Par conséquent, nous avons tronqué les matrices de transition sur l'espace d'état fini $\{0, 1, \dots, N\}$ comme suit : nous calculons les coefficients $(p_{i,j}, 0 \leq i \leq N, 0 \leq j \leq N-1)$ de la matrice de transition par la formule (3.10), et la dernière colonne est définie par :

$$p_{i,N} = 1 - \sum_{j=0}^{N-1} p_{i,j},$$

pour $i = 0, \dots, N$.

Comme nous ne nous intéressons qu'aux probabilités de transition de 0 à 0, de 0 à 1 et de 1 à 0, on peut s'attendre à ce que cette procédure de troncature n'ait pas un grand impact sur l'estimation. (**Table.3.2**).

TABLE 3.2 – L'impact du choix de N sur l'estimation.

Le framework HMM est implémenté dans R , en utilisant une machine de 32 GO RAM.

Troncature de l'espace d'état caché N	$H = 5000$		
	3	4	5
$\hat{\lambda}^n$	0.0293	0.0306	0.0301
$\hat{\mu}^n$	0.095	0.11	0.105
$\hat{\nu}^n$	0.009	0.011	0.0105
Temps d'exécution (h)	0.98	1.765	3.595

Nous pouvons remarquer que l'augmentation de la taille de l'espace tronqué N a un impact sur la complexité de l'algorithme EM, avec un temps d'exécution accru. En effet, à chaque incrément de troncature, le schéma itératif de Baum-Welch prend presque le double de temps. Cependant, en termes de précision, nous avons constaté que pour une troncature de l'espace N égale à 5, les estimations sont relativement bonnes par rapport à une troncature égale à 3 ou 4. En général, nous pouvons conclure que la troncature de l'espace n'a pas un grand impact sur l'estimation, mais plutôt sur le temps d'exécution. Enfin, nous fixons $N = 5$ et nous illustrons l'impact de l'horizon temporel H sur notre schéma d'estimation. (**Table.3.3**).

TABLE 3.3 – L’impact de l’horizon temporel H sur l’estimation.

	$N = 5$			
Horizon temporel H	1000	2000	3000	5000
nombre d’observations n	1001	2001	3001	5001
$\hat{\lambda}^n$	0.028	0.0284	0.031	0.0301
$\hat{\mu}^n$	0.085	0.085	0.12	0.105
$\hat{\nu}^n$	0.008	0.008	0.012	0.0105

Il convient de souligner que le choix de l’horizon temporel H a une influence substantielle sur les estimations obtenues. Les résultats de nos simulations ont montré que l’utilisation d’un horizon temporel plus long permettait d’obtenir des estimations plus fiables des paramètres. En effet, en augmentant l’horizon temporel, nous obtenons une plus grande quantité d’informations qui peut être utilisée pour estimer les paramètres. Ainsi, l’augmentation du nombre d’observations n permet d’améliorer la convergence locale de l’algorithme EM vers les vraies valeurs des paramètres.

Cependant, il est important de noter que l’augmentation du nombre d’observations entraîne également une augmentation du temps de calcul nécessaire pour effectuer les estimations. En résumé, le choix de l’horizon temporel et du nombre d’observations est un compromis entre la qualité des estimations et le temps de calcul nécessaire pour les obtenir.

Discussion sur l’identifiabilité des paramètres :

Le problème d’identifiabilité dans un HMM se réfère à la capacité de distinguer différentes configurations du modèle qui produisent des distributions d’observation similaires. En d’autres termes, il peut y avoir plusieurs jeux de paramètres différents qui donnent lieu à des modèles HMM avec les mêmes performances d’observation. Cela peut se produire lorsque certaines combinaisons de paramètres du modèle HMM sont équivalentes en termes de générer les mêmes probabilités d’observation. Par conséquent, lors de l’estimation des paramètres du modèle HMM à partir des données observées, il peut y avoir une incertitude quant à la véritable configuration des paramètres, ce qui rend le modèle non identifiable.

Il est important de noter que l’identifiabilité du modèle HMM dépend de l’ensemble de données utilisées pour estimer les paramètres [68, 69]. Dans certains cas, une plus grande quantité de données peut aider à résoudre le problème d’identifiabilité en fournissant des informations supplémentaires permettant de distinguer entre différentes configurations de paramètres. De plus, dans notre cas, nous avons effectué des estimations sur des données simulées qui sont rassurantes quant à ce point d’identifiabilité, car les paramètres désirés ont été retrouvés.

En outre, en prenant en compte la structure spécifique de notre chaîne multidimensionnelle Z et en réécrivant les fonctions du HMM en conséquence, nous avons pu restreindre l’espace des solutions possibles et obtenir une estimation plus fiable des paramètres. Cette adaptation de la modélisation nous permet de mieux tenir compte des caractéristiques particulières de nos données et d’améliorer l’identifiabilité du modèle HMM.

3.4.2 Données de la typhoïde à Mayotte

Nous étudions maintenant un jeu de données réelles de l'Agence régionale de santé (ARS) française de Mayotte qui représente les cumuls quotidiens de nouveaux cas déclarés de fièvre typhoïde sur l'île entre 2018 et 2022.

Nous disposons ainsi de 1816 observations, avec des valeurs comprises entre 0 et 4 (**Table.3.4**).

TABLE 3.4 – Le nombre de nouveaux cas déclarés de fièvre typhoïde à Mayotte entre 2018 et 2022.

valeur d'observation	0	1	2	3	4
nombre des observations	1596	167	34	12	7

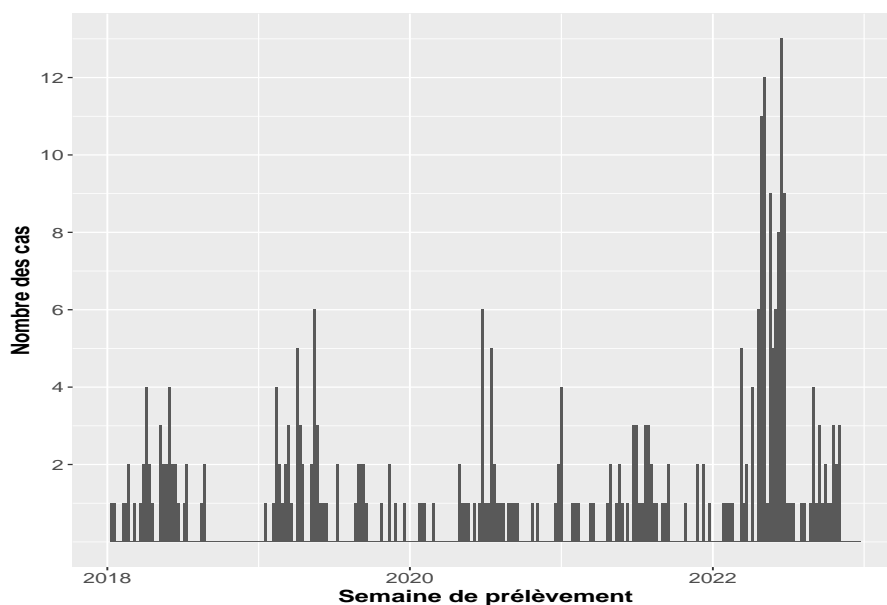


FIGURE 3.5 – Cumul hebdomadaire de nouveaux cas déclarés de fièvre typhoïde entre 2018 et 2022 à Mayotte.

Le modèle stochastique proposé est particulièrement bien adapté aux situations de faible prévalence de la maladie que l'on constate à Mayotte. En outre, des périodes de plusieurs semaines, peuvent s'écouler sans qu'aucun nouveau cas ne soit déclaré, suivies d'une augmentation soudaine du nombre de nouveaux cas. Cette observation suggère que le modèle LBDPI en régime récurrent positif, est approprié pour ce contexte. Il convient de noter que la fièvre typhoïde est une maladie à déclaration obligatoire à Mayotte, ce qui signifie en principe que tous les nouveaux cas déclarés sont observés.

Nous avons opté pour une approche de paramétrage initial basée sur la littérature existante, les experts de la fièvre typhoïde à Mayotte et nos données synthétiques. Plus précisément, nous nous sommes appuyés sur des études telles que [28, 30, 40, 70], qui ont abordé la modélisation de la fièvre typhoïde dans divers contextes géographiques. Nous avons également sollicité l'avis d'experts de la typhoïde à Mayotte pour affiner nos choix. Enfin, nous avons exploré nos données synthétiques pour évaluer la performance du modèle pour différentes combinaisons de paramètres initiaux. Nous avons finalement choisi

des paramètres initiaux dans les intervalles suivants $\lambda^{(0)} = [0.05, 0.08]$, $\mu^{(0)} = [0.11, 0.25]$ et $\nu^{(0)} = [0.015, 0.03]$, en générant une séquence de paramètres de longueur 20 à l'aide de la fonction **length.out**. Ceci nous permet d'initialiser les caractéristiques de notre modèle, à savoir $P^{(0)}$, $Q^{(0)}$, $\psi^{(0)}$ et $\rho^{(0)}$, de la même manière que celle utilisée précédemment pour les données synthétiques.

Pour chaque valeur initiale de $\lambda^{(0)}$, $\mu^{(0)}$ et $\nu^{(0)}$ dans notre grille, nous avons appliqué un critère d'arrêt basé sur le calcul du résidu, fixé à $\delta = 10^{-9}$, et un nombre maximum d'itérations, défini à $maxIterations = 500$, pour l'algorithme de Baum-Welch adapté du théorème **3.3**. Nous avons ensuite extrait les valeurs estimées pour les paramètres λ , μ et ν . Enfin, le triplet de valeurs qui maximise la vraisemblance des observations a été sélectionné comme l'estimation finale (voir le tableau des différentes valeurs de log-vraisemblance obtenues en Annexe (C)).

Les estimations obtenues sont présentées ci-dessous.

$$\hat{\lambda} = 0.054, \quad \hat{\mu} = 0.132 \quad \text{et} \quad \hat{\nu} = 0.017.$$

Notons que la condition de récurrence positive $\lambda < \mu$ est satisfaite. L'estimation de λ indiquant que la transmission de personne à personne joue un rôle significatif dans la propagation de la maladie à Mayotte met en évidence l'importance des interactions directes entre individus infectés et susceptibles dans la transmission de la fièvre typhoïde. Cela souligne l'importance de mesures telles que l'hygiène personnelle, le lavage des mains et la sensibilisation à la maladie pour réduire la transmission de personne à personne.

La durée moyenne entre le moment où une personne devient infectée et le moment où elle déclare sa maladie est de 7.5 jours, ce qui peut être influencé par plusieurs facteurs, tels que la gravité des symptômes, la connaissance et la sensibilisation des individus vis-à-vis de la maladie, l'accès aux soins de santé, et les pratiques de déclaration des cas. Un délai plus court indique une détection et une déclaration plus rapides des cas de typhoïde, ce qui peut contribuer à une réponse plus efficace en termes de prise en charge des individus infectés et de prévention de la propagation de la maladie.

Enfin, en ce qui concerne la transmission par l'eau, l'estimation du taux d'infection dû à la transmission de la bactérie *Salmonella Typhi* suggère que la qualité de l'eau joue un rôle dans la propagation de la fièvre typhoïde à Mayotte. Cela souligne l'importance de garantir l'accès à une eau potable sûre et de mettre en place des mesures de contrôle de la qualité de l'eau pour réduire les risques de transmission par cette voie. L'amélioration du système d'assainissement, y compris la gestion adéquate des eaux usées et des déchets, ainsi que la promotion de bonnes pratiques d'hygiène, sont des mesures essentielles pour limiter la propagation de la fièvre typhoïde et d'autres maladies hydriques à Mayotte.

Il convient de souligner que ces résultats sont basés sur des estimations spécifiques à Mayotte et doivent être interprétés dans le contexte de cette région. Des études supplémentaires et des mesures de surveillance continues sont nécessaires pour mieux comprendre la dynamique de la propagation de la fièvre typhoïde à Mayotte et pour orienter les interventions de santé publique visant à réduire la prévalence de la maladie.

3.4.3 Application à l'estimation de la probabilité d'atteindre un seuil épidémique dans une période de temps

Dans cette section, notre objectif est d'étudier la probabilité qu'un cumul d'individus infectés atteigne ou dépasse un seuil épidémique (représenté par l'état j) au cours d'une période de temps t , en fonction du nombre initial d'infectés (I_0). Cette probabilité

nous permet d'évaluer la progression de l'infection à partir de l'état initial i . En d'autres termes, en utilisant notre modèle LBDPI calibré sur les données de Mayotte, nous cherchons à estimer la probabilité d'observer un nombre d'infectés égal ou supérieur à j dans l'intervalle de temps $[0, t]$, sachant que le processus débute avec $I_0 = i$ individus infectés. Cette mesure est essentielle pour évaluer la dynamique de propagation de l'infection et pour étudier les différents scénarios possibles en fonction des valeurs de i, j et t .

On désigne par $j > 0$ le seuil épidémique de la fièvre typhoïde à Mayotte, qui correspond en moyenne à un nombre d'infections hebdomadaires de l'ordre de la dizaine. Nous définissons la probabilité d'atteindre l'état j dans un intervalle de temps $[0, t]$, sachant que le processus débute à l'état i par :

$$h_{i,j}(t) = \mathbb{P}_i(T_j \leq t) = \mathbb{P}(T_j \leq t \mid I_0 = i),$$

pour $0 \leq i \leq j$, avec T_j est le temps d'atteinte l'état j par le processus $(I_s)_{s \geq 0}$.

Théorème 3.4. *La probabilité $h_{i,j}(t)$ satisfait le système d'équations suivant :*

$$\begin{cases} \frac{dh_{0,j}(t)}{dt} &= -\nu h_{0,j}(t) + \nu h_{1,j}(t), \\ \frac{dh_{i,j}(t)}{dt} &= -((\lambda + \mu)i + \nu)h_{i,j}(t) + (\lambda i + \nu)h_{i+1,j}(t) + \mu i h_{i-1,j}(t) \quad 1 \leq i \leq j-1, \\ h_{j,j}(t) &= 1, \end{cases} \quad (3.24)$$

avec la condition initiale : $h_{i,j}(0) = \delta_{i=j}$.

Preuve. Soit $t > 0$ et $0 \leq i \leq j-1$, on a :

$$h_{i,j}(t) = \mathbb{P}_i(T_j \leq t) = \mathbb{P}_i(T_j \leq t, S_1 \geq t) + \mathbb{P}_i(T_j \leq t, S_1 < t),$$

où S_1 est le premier temps de saut du processus $(I_s)_{s \geq 0}$.

D'une part, nous avons :

$$\begin{aligned} \mathbb{P}_i(T_j \leq t, S_1 \geq t) &= \mathbb{P}_i(T_j \leq t \mid S_1 \geq t) \mathbb{P}_i(S_1 \geq t) \\ &= \mathbb{1}_{i=j} e^{-((\lambda+\mu)i+\nu)t}. \end{aligned} \quad (3.25)$$

D'autre part :

$$\begin{aligned} \mathbb{P}_i(T_j \leq t, S_1 < t) &= \mathbb{P}_i(T_j \leq t \mid S_1 < t) \mathbb{P}_i(S_1 < t) \\ &= \int_0^t \mathbb{P}_i(T_j \leq t \mid S_1 = s) ((\lambda + \mu)i + \nu) e^{-((\lambda+\mu)i+\nu)s} ds. \end{aligned}$$

Par la propriété de Markov, on obtient :

$$\mathbb{P}_i(T_j \leq t \mid S_1 = s) = \frac{\lambda i + \nu}{(\lambda + \mu)i + \nu} \mathbb{P}_{i+1}(T_j \leq t-s) + \frac{\mu i}{(\lambda + \mu)i + \nu} \mathbb{P}_{i-1}(T_j \leq t-s). \quad (3.26)$$

En combinant (3.25) et (3.26), on obtient :

$$h_{i,j}(t) = \mathbb{1}_{i=j} e^{-((\lambda+\mu)i+\nu)t} + \int_0^t \left((\lambda i + \nu) h_{i+1,j}(t-s) + \mu i h_{i-1,j}(t-s) \right) e^{-((\lambda+\mu)i+\nu)s} ds.$$

Ainsi,

$$\frac{dh_{i,j}(t)}{dt} = -((\lambda + \mu)i + \nu)h_{i,j}(t) + (\lambda i + \nu)h_{i+1,j}(t) + \mu i h_{i-1,j}(t).$$

Pour $i = 0$, nous avons :

$$\frac{dh_{0,j}(t)}{dt} = -\nu h_{0,j}(t) + \nu h_{1,j}(t).$$

Pour $i = j$, on a :

$$h_{j,j}(t) = \mathbb{P}_j(T_j \leq t) = 1.$$

A l'instant $t = 0$, on a :

$$h_{i,j}(0) = \mathbb{P}_i(T_j \leq 0) = \delta_{i=j}.$$

□

Implémentation : Schéma de Runge Kutta d'ordre 4

Dans ce qui suit, nous décrivons l'algorithme Runge-Kutta d'ordre 4 (RK4) pour calculer numériquement la probabilité de saturation $h_{i,j}(t)$.

Dans un premier temps, nous remarquons que nous pouvons mettre le système (3.24) sous la forme suivante :

$$\begin{cases} \frac{dH_j(t)}{dt} = F(H_j(t)), \\ H_j(0) = e_j, \end{cases}$$

où :

$$H_j(t) = \begin{pmatrix} h_{0,j}(t) \\ h_{1,j}(t) \\ \vdots \\ h_{i,j}(t) \\ \vdots \\ h_{j-1,j}(t) \\ h_{j,j}(t) \end{pmatrix}, \quad e_j = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix},$$

et

$$F(H_j(t)) = \begin{pmatrix} -\nu h_{0,j}(t) + \nu h_{1,j}(t) \\ -(\lambda + \mu + \nu)h_{1,j}(t) + (\lambda + \nu)h_{2,j}(t) + \mu h_{0,j}(t) \\ \vdots \\ -((\lambda + \mu)i + \nu)h_{i,j}(t) + (\lambda i + \nu)h_{i+1,j}(t) + \mu i h_{i-1,j}(t) \\ \vdots \\ -((\lambda + \mu)(j-1) + \nu)h_{j-1,j}(t) + \mu(j-1)h_{j-2,j}(t) + (\lambda(j-1) + \nu)h_{j,j}(t) \\ 0 \end{pmatrix}.$$

Nous présentons ci-après l'algorithme RK4 .

Algorithme 1 : Algorithme RK4

entrée : paramètres du modèles : λ, μ, ν .

Seuil de saturation j , période du temps t ,

pas de discrétisation en temps Δt .

le vecteur initial : $H_j(0) = e_j$.

sortie : La probabilité de saturation $H_j(t)$.

1 début

2 Initialisation

3 $H_j(0) \leftarrow e_j$

4 pour k de 0 à t par pas de Δt faire

5 $K_1 = F(H_j(k))$

6 $K_2 = F(H_j(k) + \frac{\Delta t}{2} K_1)$

7 $K_3 = F(H_j(k) + \frac{\Delta t}{2} K_2)$

8 $K_4 = F(H_j(k) + \Delta t K_3)$

9 $H_j(k+1) \leftarrow H_j(k) + \frac{\Delta t}{6}(K_1 + 2K_2 + 2K_3 + K_4)$

10 fin

11 renvoyer : $H_j(t)$.

12 fin

Jeu de paramètres : Nous fixons la fenêtre de temps t à 7 jours (une semaine), le seuil épidémique à $j = 15$ cas et prenons un pas de discrétisation d'ordre $h = 10^{-3}$.

Les paramètres λ, μ et ν sont estimés dans la section précédent à partir du modèle de Markov caché **3.2.3** :

$$\hat{\lambda} = 0.054, \quad \hat{\mu} = 0.132, \quad \hat{\nu} = 0.017.$$

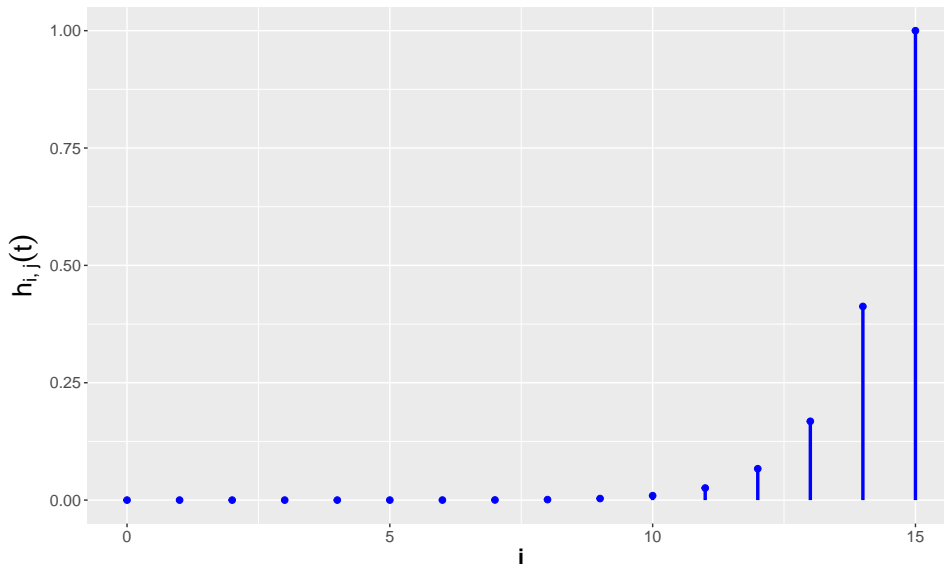


FIGURE 3.6 – Évolution de la probabilité $h_{i,j}(t)$ en fonction d'infectés initiaux i .

Ces résultats **Fig.3.6** montrent que la probabilité estimée d'observer un cumul d'infectés égal ou supérieur à 15 cas à Mayotte au cours d'une semaine augmente à mesure que le nombre initial d'infectés augmente. Cela reflète la tendance naturelle de la maladie à

se propager davantage lorsque le nombre initial d'infectés est plus élevé.

Pour $i = 0$, la probabilité est très faible mais n'est pas nulle en raison de la possibilité de contamination par l'eau, ce qui peut entraîner de nouveaux cas même en l'absence initiale d'infectés.

Ces résultats peuvent être utiles pour évaluer les risques associés à différents niveaux d'infection initiale et pour guider les décisions en matière de prévention et de surveillance de la fièvre typhoïde à Mayotte.

3.5 Conclusion

Nous avons proposé une méthodologie pour estimer les paramètres inconnus d'un LBDPI dans le cas mathématiquement inhabituel mais épidémiologiquement courant où seuls les cumuls de nouveaux cas déclarés sur des périodes données sont disponibles. Notre approche repose à la fois sur des formules analytiques et des itérations numériques. Nous avons étudié les propriétés des estimateurs sur des données synthétiques et appliqué notre procédure à un ensemble de données réelles sur la fièvre typhoïde à Mayotte qui a motivé ce travail. De plus, notre analyse a permis d'estimer la probabilité d'atteindre un seuil épidémique des infectés dans une période de temps, ce qui est essentiel pour prendre des décisions éclairées et mettre en place des mesures de surveillance et de prévention appropriées pour contenir la propagation de la maladie.

Cette approche est un premier pas vers la compréhension de la propagation de la typhoïde à Mayotte. Nous avons proposé un modèle simplifié avec quelques paramètres interprétables.

Dans les chapitres suivants, nous allons poursuivre notre étude en élargissant notre modèle pour prendre en compte des facteurs supplémentaires qui peuvent affecter la propagation de la fièvre typhoïde à Mayotte. Tout d'abord, nous allons inclure la durée d'incubation de la maladie et les individus exposés, ce qui nous permettra de quantifier plus précisément la transmission de la maladie dans la population. Ensuite, nous allons examiner l'effet de la pluviométrie sur l'évolution de la maladie, car il a été observé que les épidémies de fièvre typhoïde à Mayotte surviennent de préférence en saison humide. Ces analyses supplémentaires nous permettront d'obtenir une vision plus complète de la dynamique de la maladie en vue de développer des stratégies de prévention et de contrôle plus efficaces pour réduire l'impact de la fièvre typhoïde sur la santé de la population.

Modélisation et estimation des paramètres d'un modèle multi-chaîne cachée

Sommaire

4.1	Introduction	86
4.2	Modèle et propriété	87
4.2.1	Définition du modèle	87
4.2.2	Propriétés mathématiques du modèle	88
4.3	Estimation des paramètres	92
4.3.1	Observations	93
4.3.2	Stratégie d'estimation	94
4.3.3	Expression analytique des paramètres	94
4.3.4	Modèle de Markov Multi-Chaîne Cachée	97
4.4	Simulation numérique	102
4.4.1	Observations complètes	102
4.4.2	Observations du cumul des nouveaux cas déclarés	103
4.5	Conclusion	107

4.1 Introduction

Dans ce chapitre, nous modifions le modèle présenté dans le chapitre précédent pour prendre en compte la durée d'incubation de la maladie, qui est la période entre la contamination par la bactérie *Salmonella Typhi* et l'apparition des premiers symptômes chez un individu. Cette durée peut varier de quelques jours à plusieurs semaines, et elle peut être influencée par divers facteurs individuels et environnementaux, tels que la charge bactérienne dans l'organisme, l'âge, l'état de santé et d'autres facteurs.

En intégrant cette durée d'incubation dans notre modèle, nous pourrions mieux comprendre la dynamique de transmission de la fièvre typhoïde à Mayotte. Nous serons en mesure de déterminer le délai entre la contamination par la bactérie et le début d'apparition des premiers signes cliniques, tels que la fièvre, les maux de tête, la fatigue, les douleurs abdominales, la constipation ou la diarrhée. Nous nous appuyons sur les caractéristiques médicales de la maladie [71] et sur les données épidémiologiques 1.2.4 fournies par l'ARS de Mayotte pour mener notre étude.

Dans ce nouveau modèle, nous proposons une extension du modèle 3.2 afin de l'affiner et de le rendre plus réaliste. Nous introduisons un nouveau compartiment dans le modèle, correspondant aux individus exposés à la bactérie *Salmonella Typhi*, noté E . Les individus exposés sont définis comme ceux qui ont été en contact avec la bactérie, mais qui ne présentent pas encore de symptômes de la maladie et ils sont non contagieux [71].

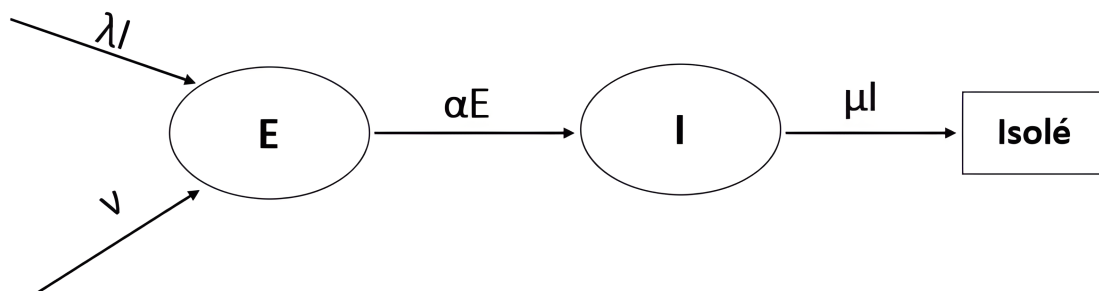


FIGURE 4.1 – **Modèle à deux compartiments : Exposé-Infecté.**

Le nouveau paramètre α du modèle exposé-infecté représente le taux *d'incubation*. Le chapitre présent a pour objectif de modéliser la dynamique des individus exposés et infectés, d'étudier les propriétés théoriques du modèle et de mettre en œuvre une approche paramétrique reposant sur un processus Markovien à temps continu pour estimer les paramètres de ce nouveau modèle. Ces paramètres sont les taux de contamination de personne à personne λ , de contamination par l'environnement ν , d'incubation α et d'isolement μ . Nous utilisons le même schéma d'observation que précédemment pour estimer ces paramètres 3.2.2.

Le reste du chapitre est organisé comme suit, dans la section 4.2, nous décrivons la formulation du modèle stochastique exposé-infecté, ainsi que ses propriétés mathématiques. Dans la section 4.3, nous présentons la procédure d'estimation des paramètres de notre modèle exposé-infecté en suivant une stratégie d'estimation en deux étapes. Dans la section 4.4, nous mettons en place des simulations numériques pour illustrer les résultats obtenus.

4.2 Modèle et propriété

4.2.1 Définition du modèle

Pour décrire notre nouveau modèle exposé-infecté, nous avons repris les hypothèses du modèle précédent (3.2.1), tout en introduisant un nouveau compartiment d'individus exposés **E**. Ces individus ne sont pas contagieux, mais peuvent devenir infectés à un taux individuel α (taux d'incubation). Les individus infectés et contagieux **I** sont considérés comme isolés à partir de la date d'hospitalisation (Fig.4.1).

La dynamique de la maladie est caractérisée par une phase d'incubation, de contagiosité et d'hospitalisation. Pendant la phase d'incubation, l'individu exposé (il a été en contact avec la maladie mais n'est pas encore infectieux) ne présente pas encore de symptômes de la maladie, et il devient infectieux après une durée moyenne $1/\alpha$, où $\alpha > 0$ est le taux d'incubation individuel. Ensuite, il y a une phase contagieuse où l'individu infecté devient contagieux et peut transmettre la maladie à un taux $\lambda > 0$ à une autre personne. La phase d'hospitalisation est caractérisée par l'isolement de l'individu à un taux $\mu > 0$, où il ne peut plus contaminer d'autres personnes. Pendant cette phase, l'individu reçoit un traitement à base d'antibiotiques et peut se rétablir en l'espace de 1 à 5 jours, en fonction de la gravité de l'infection et de la réponse du patient au traitement [71]. La contamination via l'environnement génère également de nouveaux infectés via un processus de Poisson de taux $\nu > 0$ (Fig.4.2).

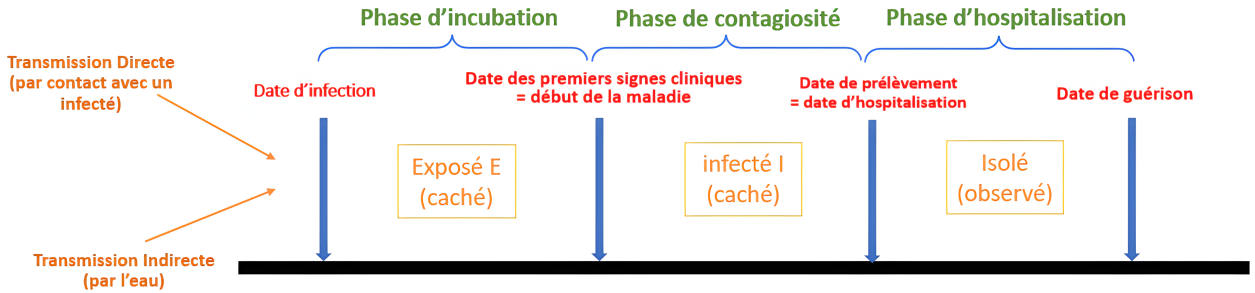


FIGURE 4.2 – Dynamique de la fièvre typhoïde à Mayotte dans le modèle à deux compartiments.

Le modèle peut être caractérisé par un processus de Markov bivarié en temps continu $(E_t, I_t)_{t \geq 0}$, à valeurs dans $\mathbb{S} = \mathbb{N} \times \mathbb{N}$, où E_t et I_t représentent respectivement le nombre des individus exposés et infectieux à l'instant t au sein de la population.

Étant donné l'état courant (e, i) , les transitions possibles du notre processus sont :

- $(e, i) \rightarrow (e + 1, i)$ avec un taux de $\lambda i + \nu$.
- $(e, i) \rightarrow (e - 1, i + 1)$ avec un taux de αe .
- $(e, i) \rightarrow (e, i - 1)$ avec un taux de μi .

Ainsi, le générateur infinitésimal Q du processus $(E_t, I_t)_{t \geq 0}$ est donné par :

$$\begin{aligned}
 Q_{(e,i),(e+1,i)} &= \lambda i + \nu, & \text{pour } e, i \geq 0, \\
 Q_{(e,i),(e-1,i+1)} &= \alpha e, & \text{pour } e \geq 1, i \geq 0, \\
 Q_{(e,i),(e,i-1)} &= \mu i, & \text{pour } e \geq 0, i \geq 1, \\
 Q_{(e,i),(e,i)} &= -((\lambda + \mu)i + \alpha e + \nu), & \text{pour } e, i \geq 0,
 \end{aligned} \tag{4.1}$$

où les paramètres λ , ν , α et μ sont positifs.

Nous supposons que $(E_0, I_0) = (e, i) \in \mathbb{S}$ et nous désignons par

$$p_{(e,i),(e',i')}(t) = \mathbb{P}\left((E_t, I_t) = (e', i') \mid (E_0, I_0) = (e, i)\right) \quad (e', i') \in \mathbb{S}, t \geq 0,$$

les probabilités de transition de $(E_t, I_t)_{t \geq 0}$. Elles satisfont l'équation de Kolmogorov progressive suivante :

$$\begin{aligned} \frac{d}{dt} p_{(e,i),(e',i')}(t) &= (\lambda i' + \nu) p_{(e,i),(e'-1,i')}(t) + \alpha(e' + 1) p_{(e,i),(e'+1,i'-1)}(t) + \mu(i' + 1) p_{(e,i),(e',i'+1)}(t) \\ &\quad - ((\lambda + \mu)i' + \alpha e' + \nu) p_{(e,i),(e',i')}(t), \end{aligned} \quad (4.2)$$

avec la condition initiale $p_{(e,i),(e,i)}(0) = 1$ et $p_{(e,i),(e',i')}(0) = 0$ pour $(e', i') \neq (e, i)$.

Dans la sous-section suivante, notre intérêt se porte sur l'étude théorique du comportement moyen et asymptotique du processus $(E_t, I_t)_{t \geq 0}$. Nous allons explorer les propriétés mathématiques de ce modèle et comprendre comment il se comporte à long terme. Nous allons également illustrer l'évolution de ce modèle au cours du temps grâce à des simulations numériques. Cette analyse approfondie nous permettra de prédire l'évolution de l'épidémie dans le temps en fonction des paramètres.

4.2.2 Propriétés mathématiques du modèle

Nous étudions maintenant quelques propriétés mathématiques du processus $(E_t, I_t)_{t \geq 0}$.

Théorème 4.1. (*Comportement moyen de $(E_t, I_t)_{t \geq 0}$*). *Pour tout $e_0, i_0 \geq 0$, si $\lambda = \mu$, alors*

$$\begin{cases} \mathbb{E}_{(e_0, i_0)}[E_t] = \frac{\lambda \nu}{\lambda + \alpha} t - c \left(e^{-(\lambda + \alpha)t} + \frac{\lambda}{\alpha} \right) + i_0 \frac{\lambda}{\alpha} + \frac{\nu}{\lambda + \alpha}, \\ \mathbb{E}_{(e_0, i_0)}[I_t] = \frac{\alpha \nu}{\lambda + \alpha} t + c \left(e^{-(\lambda + \alpha)t} - 1 \right) + i_0, \end{cases}$$

et si $\lambda \neq \mu$, alors

$$\begin{cases} \mathbb{E}_{(e_0, i_0)}[E_t] = c_1 e^{-\frac{-\mu - \alpha + \sqrt{(\mu - \alpha)^2 + 4\alpha\lambda}}{2} t} + c_2 e^{-\frac{\mu + \alpha + \sqrt{(\mu - \alpha)^2 + 4\alpha\lambda}}{2} t} + \frac{\mu \nu}{\alpha(\mu - \lambda)}, \\ \mathbb{E}_{(e_0, i_0)}[I_t] = c_3 e^{-\frac{-\mu - \alpha + \sqrt{(\mu - \alpha)^2 + 4\alpha\lambda}}{2} t} + c_4 e^{-\frac{\mu + \alpha + \sqrt{(\mu - \alpha)^2 + 4\alpha\lambda}}{2} t} + \frac{\nu}{\mu - \lambda}, \end{cases}$$

où

$$\begin{aligned} c &= \left(e_0 + i_0 - \frac{\nu}{\lambda + \alpha} \right) \frac{\alpha}{\lambda + \alpha}, \\ c_1 &= \frac{1}{a} \left(i_0 + \frac{\mu - \alpha + a}{2} e_0 - \frac{\nu}{\mu - \lambda} \left(1 + \frac{\mu(\mu - \alpha + a)}{2\alpha} \right) \right), \\ c_2 &= e_0 - c_1 - \frac{\mu \nu}{\alpha(\mu - \lambda)}, \\ c_3 &= \frac{-\mu + \alpha + a}{2a} \left(i_0 + \frac{\mu - \alpha + a}{2} e_0 - \frac{\nu}{\mu - \lambda} \left(1 + \frac{\mu(\mu - \alpha + a)}{2\alpha} \right) \right), \\ c_4 &= -\frac{\mu - \alpha + a}{2} c_2, \\ a &= \sqrt{(\mu - \alpha)^2 + 4\alpha\lambda}. \end{aligned}$$

Preuve. Soit $(e_0, i_0) \in \mathbb{S}$, pour toute fonction f positive ou bornée sur \mathbb{S} et pour tout $e, i, \geq 0$, nous avons :

$$\mathbb{E}_{(e_0, i_0)}[f(E_t, I_t)] = \sum_{(e, i) \in \mathbb{S}} p_{(e_0, i_0)(e, i)}(t) f(e, i).$$

En utilisant l'équation de Kolmogorov progressive, on obtient :

$$\begin{aligned} \frac{d}{dt} \mathbb{E}_{(e_0, i_0)}[f(E_t, I_t)] &= \sum_{(e, i) \in \mathbb{S}} p'_{(e_0, i_0)(e, i)}(t) f(e, i) \\ &= \sum_{(e, i) \in \mathbb{S}} \left(\sum_{(e', i') \in \mathbb{S}} p_{(e_0, i_0)(e', i')} Q_{(e', i')(e, i)} \right) f(e, i) \\ &= \sum_{(e', i') \in \mathbb{S}} p_{(e_0, i_0)(e', i')} \left(\sum_{(e, i) \in \mathbb{S}} Q_{(e', i')(e, i)} f(e, i) \right) \\ &= \sum_{(e', i') \in \mathbb{S}} p_{(e_0, i_0)(e', i')} \left((\lambda i' + \nu) (f(e' + 1, i') - f(e', i')) \right. \\ &\quad \left. + \alpha e' (f(e' - 1, i' + 1) - f(e', i')) + \mu i' (f(e', i' - 1) - f(e', i')) \right) \\ &= \mathbb{E}_{(e_0, i_0)} \left[(\lambda I_t + \nu) (f(E_t + 1, I_t) - f(E_t, I_t)) \right. \\ &\quad \left. + \alpha E_t (f(E_t - 1, I_t + 1) - f(E_t, I_t)) + \mu I_t (f(E_t, I_t - 1) - f(E_t, I_t)) \right]. \end{aligned}$$

Pour $f(E_t, I_t) = E_t$, nous trouvons :

$$\begin{aligned} \frac{d}{dt} \mathbb{E}_{(e_0, i_0)}[E_t] &= \mathbb{E}_{(e_0, i_0)}[\lambda I_t + \nu - \alpha E_t] \\ &= -\alpha \mathbb{E}_{(e_0, i_0)}[E_t] + \lambda \mathbb{E}_{(e_0, i_0)}[I_t] + \nu. \end{aligned} \quad (4.3)$$

Pour $f(E_t, I_t) = I_t$, nous avons :

$$\begin{aligned} \frac{d}{dt} \mathbb{E}_{(e_0, i_0)}[I_t] &= \mathbb{E}_{(e_0, i_0)}[\alpha E_t - \mu I_t] \\ &= \alpha \mathbb{E}_{(e_0, i_0)}[E_t] - \mu \mathbb{E}_{(e_0, i_0)}[I_t]. \end{aligned} \quad (4.4)$$

En combinant les deux équations **4.3** et **4.4**, nous obtenons le système d'équations différentielles suivant :

$$\begin{cases} \frac{d}{dt} \mathbb{E}_{(e_0, i_0)}[E_t] &= -\alpha \mathbb{E}_{(e_0, i_0)}[E_t] + \lambda \mathbb{E}_{(e_0, i_0)}[I_t] + \nu, \\ \frac{d}{dt} \mathbb{E}_{(e_0, i_0)}[I_t] &= \alpha \mathbb{E}_{(e_0, i_0)}[E_t] - \mu \mathbb{E}_{(e_0, i_0)}[I_t]. \end{cases} \quad (4.5)$$

En résolvant le système **(4.5)**, nous pouvons obtenir les expressions des espérances suivantes :

Si $\lambda = \mu$, alors

$$\begin{cases} \mathbb{E}_{(e_0, i_0)}[E_t] = \frac{\lambda \nu}{\lambda + \alpha} t - c \left(e^{-(\lambda + \alpha)t} + \frac{\lambda}{\alpha} \right) + i_0 \frac{\lambda}{\alpha} + \frac{\nu}{\lambda + \alpha}, \\ \mathbb{E}_{(e_0, i_0)}[I_t] = \frac{\alpha \nu}{\lambda + \alpha} t + c \left(e^{-(\lambda + \alpha)t} - 1 \right) + i_0. \end{cases}$$

Et si $\lambda \neq \mu$, alors

$$\begin{cases} \mathbb{E}_{(e_0, i_0)}[E_t] = c_1 e^{\frac{-\mu - \alpha + \sqrt{(\mu - \alpha)^2 + 4\alpha\lambda}}{2}t} + c_2 e^{-\frac{\mu + \alpha + \sqrt{(\mu - \alpha)^2 + 4\alpha\lambda}}{2}t} + \frac{\mu\nu}{\alpha(\mu - \lambda)}, \\ \mathbb{E}_{(e_0, i_0)}[I_t] = c_3 e^{\frac{-\mu - \alpha + \sqrt{(\mu - \alpha)^2 + 4\alpha\lambda}}{2}t} + c_4 e^{-\frac{\mu + \alpha + \sqrt{(\mu - \alpha)^2 + 4\alpha\lambda}}{2}t} + \frac{\nu}{\mu - \lambda}. \end{cases}$$

Les constantes réelles c et c_i pour $i = 1, \dots, 4$ sont déterminées à partir de la condition initiale $(E_0, I_0) = (e_0, i_0)$. □

Corollaire 4.1. *Asymptotiquement et conditionnellement à $(E_0, I_0) = (e_0, i_0)$, dans le cas où $\lambda \geq \mu$, le comportement moyen du processus $(E_t, I_t)_{t \geq 0}$ est explosif. Lorsque $\lambda > \mu$, les nombres moyens des individus exposés et infectés augmentent de manière exponentielle (**Fig.4.3**). Cela indique une croissance non contrôlée de l'épidémie, où le nombre d'infections continue d'augmenter rapidement au fil du temps. Si $\lambda = \mu$, les deux espérances augmentent de manière linéaire (**Fig.4.4**). c'est-à-dire que le nombre moyen d'individus exposés et infectés augmente proportionnellement au temps. Cependant, le processus moyen se stabilise dans le cas où $\lambda < \mu$ (**Fig.4.5**), ce qui conduit à une stabilisation du nombre moyen d'individus exposés et infectés à un niveau constant donné par :*

$$\begin{cases} \lim_{t \rightarrow +\infty} \mathbb{E}_{(e_0, i_0)}[E_t] = \frac{\mu\nu}{\alpha(\mu - \lambda)}, \\ \lim_{t \rightarrow +\infty} \mathbb{E}_{(e_0, i_0)}[I_t] = \frac{\nu}{\mu - \lambda}. \end{cases}$$

Cette situation correspond à un équilibre endémique des moyennes.

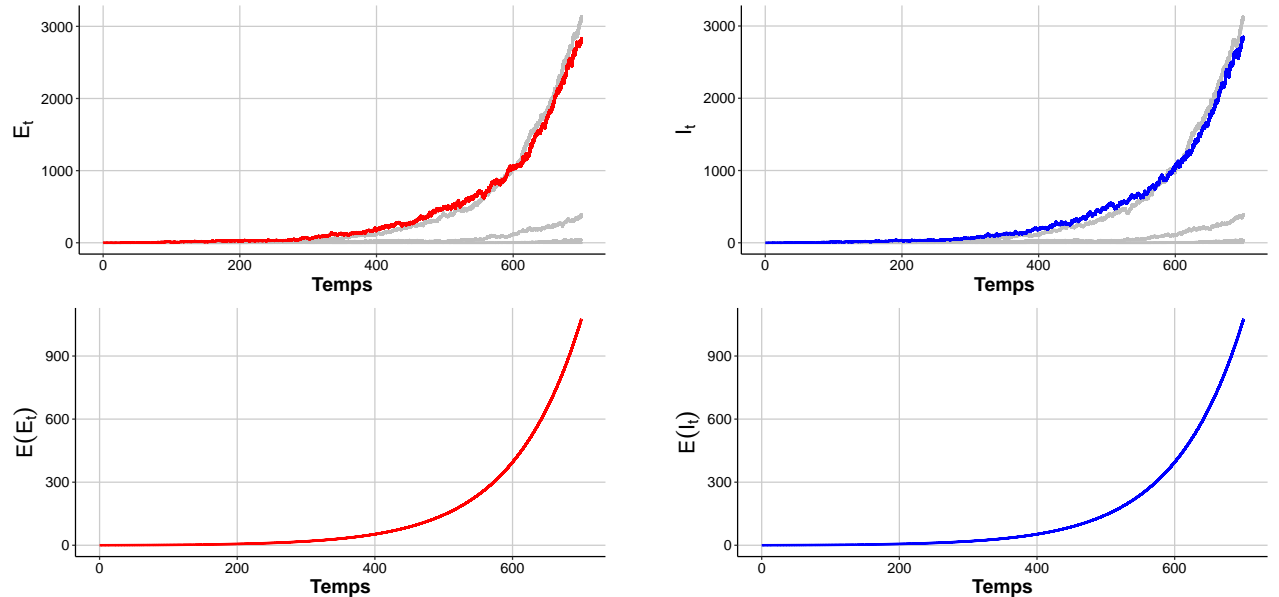


FIGURE 4.3 – Évolution du processus $(E_t, I_t)_{t \geq 0}$ et sa moyenne $(\mathbb{E}[E_t], \mathbb{E}[I_t])_{t \geq 0}$ pour les paramètres $\lambda = 0.27$, $\mu = 0.25$, $\alpha = 0.26$ et $\nu = 0.02$.

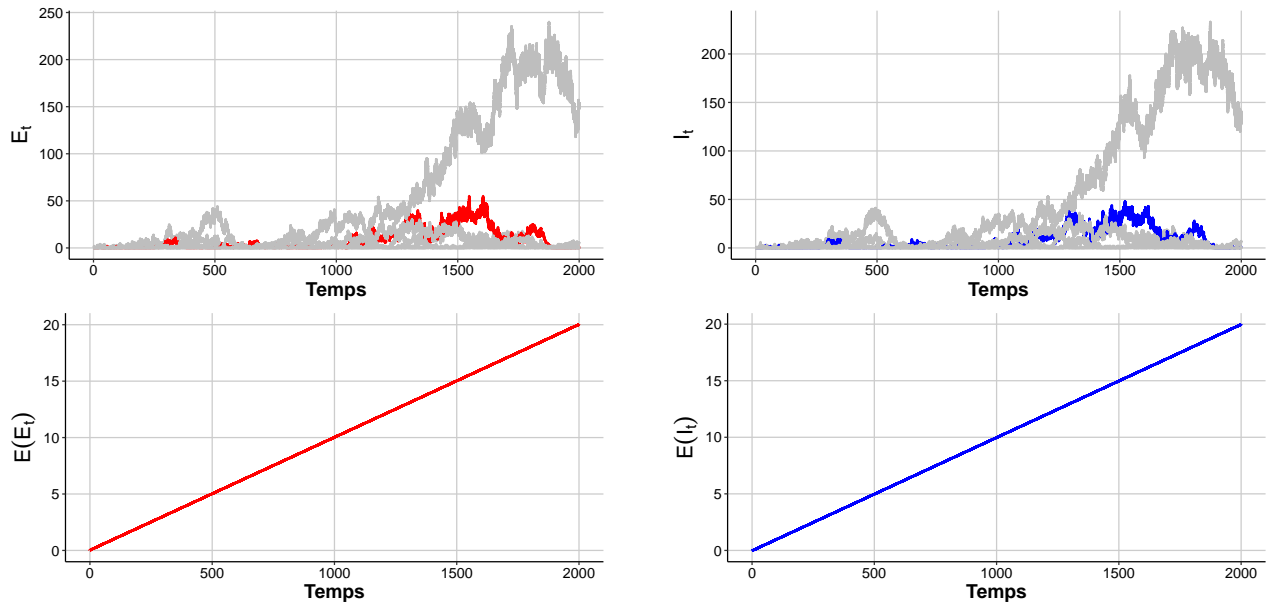


FIGURE 4.4 – Évolution du processus $(E_t, I_t)_{t \geq 0}$ et sa moyenne $(\mathbb{E}[E_t], \mathbb{E}[I_t])_{t \geq 0}$ pour les paramètres $\lambda = \mu = \alpha = 0.2$ et $\nu = 0.02$.

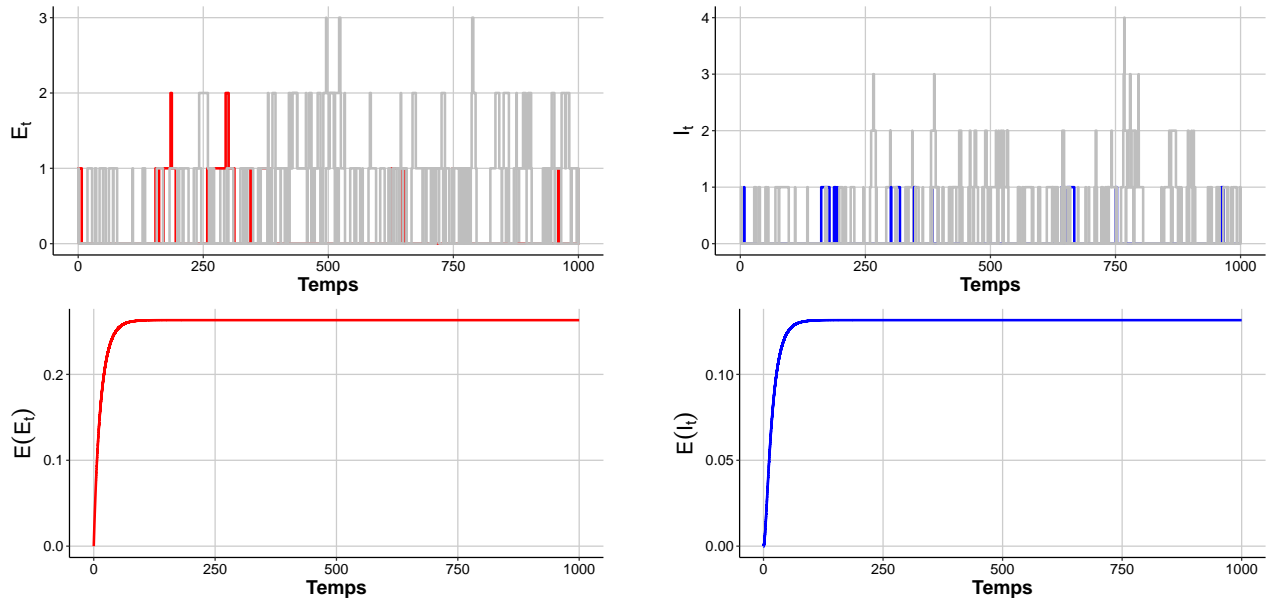


FIGURE 4.5 – Évolution du processus $(E_t, I_t)_{t \geq 0}$ et sa moyenne $(\mathbb{E}[E_t], \mathbb{E}[I_t])_{t \geq 0}$ pour les paramètres $\lambda = 0.048$, $\mu = 0.2$, $\alpha = 0.1$ et $\nu = 0.02$.

Dans la suite, nous allons identifier le régime récurrent positif du modèle et étudier l'existence d'une loi invariante pour le processus exposé-infecté. Pour étudier cette propriété, nous allons utiliser l'approche de Lyapunov, qui a été développée pour étudier la stabilité des systèmes dynamiques [72].

Le théorème suivant présente une condition suffisante pour garantir l'existence d'une loi invariante pour le processus $(E_t, I_t)_{t \geq 0}$.

Théorème 4.2. (Comportement asymptotique de $(E_t, I_t)_{t \geq 0}$). Si $\lambda < \alpha$ et $\frac{\lambda + \alpha}{2} < \mu$, alors $(E_t, I_t)_{t \geq 0}$ est récurrent positif et admet une unique loi invariante.

Preuve. Le processus $(E_t, I_t)_{t \geq 0}$ est clairement irréductible, son générateur \mathcal{L} est donné par : pour toute fonction $f : \mathbb{N}^2 \rightarrow \mathbb{R}$

$$\begin{aligned} \mathcal{L}f(e, i) &= (\lambda i + \nu) [f(e + 1, i) - f(e, i)] + \alpha e [f(e - 1, i + 1) - f(e, i)] \\ &\quad + \mu i [f(e, i - 1) - f(e, i)]. \end{aligned}$$

On applique ce générateur \mathcal{L} à la fonction $f(e, i) = e^2 + i^2$, on obtient :

$$\begin{aligned} \mathcal{L}f(e, i) &= (\lambda i + \nu) [(e + 1)^2 - e^2] + \alpha e [(e - 1)^2 - e^2] + \alpha e [(i + 1)^2 - i^2] \\ &\quad + \mu i [(i - 1)^2 - i^2]. \end{aligned}$$

Par conséquent, on peut facilement voir qu'il existe un (e_0, i_0) , tel que lorsque $(e, i) \geq (e_0, i_0)$,

$$\begin{aligned} \mathcal{L}f(e, i) &\leq 2\lambda e i - 2\alpha e^2 + 2\alpha e i - 2\mu i^2 \\ &\leq (\lambda - \alpha)e^2 + (\lambda + \alpha - 2\mu)i^2 \\ &\leq \max\{\lambda - \alpha, \lambda + \alpha - 2\mu\} (e^2 + i^2), \end{aligned}$$

où $\max\{\lambda - \alpha, \lambda + \alpha - 2\mu\} < 0$, puisque $\lambda < \alpha$ et $\frac{\lambda + \alpha}{2} < \mu$.

Ainsi,

$$\mathcal{L}f(e, i) \leq \max\{\lambda - \alpha, \lambda + \alpha - 2\mu\} (e^2 + i^2) + C 1_{\{(e, i) \geq (e_0, i_0)\}},$$

où $C > 0$.

Par conséquent, le processus $(E_t, I_t)_{t \geq 0}$ satisfait la condition de Foster-Lyapunov [72, Théorème 7.1], donc, la chaîne de Markov (E_t, I_t) admet une unique loi invariante et converge en \mathbb{L}^2 lorsque t tend vers l'infini. \square

Après avoir présenté notre modèle exposé-infecté ainsi que ses propriétés mathématiques et identifié et illustré les différents régimes possibles du processus à l'aide de simulations, notre prochaine étape consiste à intégrer le schéma d'observation dans ce modèle exposé-infecté afin de construire une procédure d'estimation.

4.3 Estimation des paramètres

Dans cette section, nous présentons notre méthode d'estimation des paramètres du modèle exposé-infecté en adaptant la démarche réalisée tout au long du chapitre précédent 3, qui ne s'applique pas directement cette fois-ci. Tout d'abord, nous décrivons les caractéristiques de nos observations liées au processus des exposés et infectés 4.3.1. Nous expliquons ensuite notre stratégie d'estimation ainsi que les différentes étapes impliquées dans cette méthode 4.3.2. Nous présentons également les expressions analytiques des estimateurs obtenus à partir des moments du processus $(E_t, I_t)_{t \geq 0}$ 4.3.3. Nous adaptons l'algorithme de Baum-Welch à notre schéma d'observations spécifique, ce qui nous permet d'estimer les probabilités de transition de la chaîne cachée couplée à partir des données cumulées des cas isolés et de déduire les différentes estimations des moments de la chaîne couplée 4.3.4. Enfin, nous obtenons des estimations des paramètres originaux par "plug-in".

4.3.1 Observations

Les nombres d'individus exposés E_t et infectés I_t ne sont pas directement observables. Nous avons à notre disposition uniquement les cumuls de nouveaux cas déclarés par jour, correspondant aux sauts d'amplitude -1 du processus I_t . Nous utilisons les mêmes notations que dans le chapitre précédent pour les observations (voir Partie.3.2.2). $O_{]s,t]}$ représente le nombre de sauts d'amplitude -1 du processus $(I_t)_{t \geq 0}$ sur l'intervalle $]s, t]$. Le processus joint $(E_t, I_t, O_{]0,t]})_{t \geq 0}$ est caractérisé par les probabilités de transition suivantes : pour $t > 0$

$$p_{(e,i),(e',i',o)}(t) := \mathbb{P} \left((E_t, I_t) = (e', i'), O_{]0,t]} = o \mid (E_0, I_0) = (e, i) \right).$$

Ces probabilités satisfont une équation de Kolmogorov présentée dans le lemme suivant.

Lemme 4.1. *Pour $t > 0$ et $e, i, e', i', o \in \mathbb{N}$, on a*

$$\begin{aligned} \frac{d}{dt} p_{(e,i),(e',i',o)}(t) &= (\lambda i + \nu) p_{(e+1,i),(e',i',o)}(t) + \alpha e p_{(e-1,i+1),(e',i',o)}(t) + \mu i p_{(e,i-1),(e',i',o-1)}(t) \\ &\quad - ((\lambda + \mu)i + \alpha e + \nu) p_{(e,i),(e',i',o)}(t), \end{aligned} \quad (4.6)$$

pour $i \leq i' + o$, avec $p_{(e,i),(e',i',o)}(t) = 0$ quand $i > i' + o$ et avec la condition initiale $p_{(e,i),(e',i',o)}(0) = \delta_{i=i'} \delta_{o=0}$.

Preuve. Nous utilisons la caractérisation infinitésimale de la distribution du processus joint.

Pour $h > 0$ et $i \leq i' + o$, on a

$$\begin{aligned} &p_{(e,i),(e',i',o)}(t+h) \\ &= \mathbb{P} \left((E_{t+h}, I_{t+h}) = (e', i'), O_{]0,t+h]} = o \mid (E_0, I_0) = (e, i) \right) \\ &= \mathbb{P} \left((E_{t+h}, I_{t+h}) = (e', i'), O_{]h,t+h]} = o, (E_h, I_h) = (e+1, i), O_{]0,h]} = 0 \mid (E_0, I_0) = (e, i) \right) \\ &\quad + \mathbb{P} \left((E_{t+h}, I_{t+h}) = (e', i'), O_{]h,t+h]} = o, (E_h, I_h) = (e-1, i+1), O_{]0,h]} = 0 \mid (E_0, I_0) = (e, i) \right) \\ &\quad + \mathbb{P} \left((E_{t+h}, I_{t+h}) = (e', i'), O_{]h,t+h]} = o-1, (E_h, I_h) = (e, i-1), O_{]0,h]} = 1 \mid (E_0, I_0) = (e, i) \right) \\ &\quad + \mathbb{P} \left((E_{t+h}, I_{t+h}) = (e', i'), O_{]h,t+h]} = o, (E_h, I_h) = (e, i), O_{]0,h]} = 0 \mid (E_0, I_0) = (e, i) \right) + 0(h) \\ &= \mathbb{P} \left((E_h, I_h) = (e+1, i), O_{]0,h]} = 0 \mid (E_0, I_0) = (e, i) \right) \mathbb{P} \left((E_{t+h}, I_{t+h}) = (e', i'), O_{]h,t+h]} = o \mid \right. \\ &\quad \left. (E_h, I_h) = (e+1, i) \right) + \mathbb{P} \left((E_h, I_h) = (e-1, i+1), O_{]0,h]} = 0 \mid (E_0, I_0) = (e, i) \right) \mathbb{P} \left((E_{t+h}, \right. \\ &\quad \left. I_{t+h}) = (e', i'), O_{]h,t+h]} = o \mid (E_h, I_h) = (e-1, i+1) \right) + \mathbb{P} \left((E_h, I_h) = (e, i-1), O_{]0,h]} = 1 \mid \right. \\ &\quad \left. (E_0, I_0) = (e, i) \right) \mathbb{P} \left((E_{t+h}, I_{t+h}) = (e', i'), O_{]h,t+h]} = o-1 \mid (E_h, I_h) = (e, i-1) \right) + \mathbb{P} \left((E_h, \right. \\ &\quad \left. I_h) = (e, i), O_{]0,h]} = 0 \mid (E_0, I_0) = (e, i) \right) \mathbb{P} \left((E_{t+h}, I_{t+h}) = (e', i'), O_{]h,t+h]} = o \mid (E_h, I_h) = (e, i) \right) \\ &\quad + 0(h) \end{aligned}$$

$(E_t, I_t)_{t \geq 0}$ est un processus de Markov homogène, ce qui donne :

$$\begin{aligned} p_{(e,i),(e',i',o)}(t+h) &= (\lambda i + \nu) h p_{(e+1,i),(e',i',o)}(t) + \alpha e h p_{(e-1,i+1),(e',i',o)}(t) + \mu i h p_{(e,i-1),(e',i',o-1)}(t) \\ &\quad + (1 - ((\lambda + \mu)i + \alpha e + \nu)h) p_{(e,i),(e',i',o)}(t) + 0(h). \end{aligned}$$

Le résultat souhaité est donc obtenu. \square

Les équations **4.2** et **4.6** ne possèdent pas de solution analytique, ce qui nous empêche, d'une part, d'exprimer la distribution marginale des observations en fonction des paramètres λ, μ, α et ν , et d'autre part, d'obtenir une expression explicite des probabilités de transition par la méthode des fonctions génératrices des moments comme dans le chapitre précédent. En conséquence, une méthode d'estimation directe par maximum de vraisemblance n'est pas réalisable avec ce schéma d'observations. Bien que la fonction génératrice des moments du processus $(E_t, I_t)_{t \geq 0}$ puisse être calculée (voir Annexe **D**), il n'est pas possible d'identifier la loi de $(E_t, I_t)_{t \geq 0}$. Pour contourner ces difficultés, nous allons adopter une stratégie d'estimation présentée dans la sous-section suivante.

4.3.2 Stratégie d'estimation

Comme mentionné dans le chapitre précédent, notre problème d'estimation relève toujours de la classe des problèmes d'information cachée. La suite d'observations $(O_n := O_{\lfloor n\Delta t, (n+1)\Delta t \rfloor})_{n \in \mathbb{N}}$ est la même, mais la suite d'états cachés est cette fois-ci bivariée et couplée $(E_{n\Delta t}, I_{n\Delta t})_{n \in \mathbb{N}}$. Nous avons donc affaire à un modèle de type multi-chaîne cachée. Nous définissons pour la suite le processus $N_t := \sum_{k=0}^{\lfloor t \rfloor} O_k$ décrivant le nombre cumulé de nouveaux cas isolés de 0 à l'instant t . Notre objectif est d'estimer le quadruplet $(\lambda, \mu, \alpha, \nu)$. La première étape de notre méthode d'estimation consiste à aborder le cas où les variables E_t et I_t sont observées à chaque instant t . Nous étudions le comportement moyen du processus joint (E_t, I_t, N_t) , nous sommes en mesure d'exprimer explicitement leurs moments en fonction des paramètres λ, μ, α et ν . Sous la condition de stabilité mentionnée dans le corollaire **4.1**, le système d'espérances obtenu est inversible, ce qui nous permet de déduire une expression explicite des paramètres en termes des moments du processus (E_t, I_t, N_t) (**4.3.3**).

Nous pouvons alors estimer les paramètres par plug-in à partir de tout estimateur des moments.

Comme précédemment **3.3.2**, la deuxième phase prend en compte le cadre des données manquantes en considérant la chaîne de Markov $(E_{n\Delta t}, I_{n\Delta t}, O_n)_{n \in \mathbb{N}}$ comme un modèle de Markov multi-chaîne cachée (en anglais : Hidden Multi-Chain Markov Model (HMCMM)). Nous avons réécrit les caractéristiques du modèle en tenant compte notre contexte particulier et adapté l'algorithme de Baum Welsh permettant d'estimer la probabilité de transition de la chaîne de Markov cachée couplée à partir de la suite d'observations (**4.3.4**). Nous estimons ensuite les moments du processus à l'aide de simulations de Monte Carlo et utilisons les formules analytiques obtenues lors de la première étape pour obtenir des estimations des paramètres du modèle exposé-infecté.

4.3.3 Expression analytique des paramètres

Le théorème suivant présente les expressions analytiques des paramètres $(\lambda, \mu, \alpha, \nu)$ en termes des moments du processus $(E_t, I_t, N_t)_{t \geq 0}$.

Théorème 4.3. *Si $\lambda < \mu$, alors les paramètres λ, μ, α et ν sont donnés par les formules*

suivantes

$$\begin{aligned}
\lambda &= \frac{N^* \left(\frac{R^*}{E^* I^*} - 1 \right) (E^* + I^*)}{I^* \left(1 + \left(\frac{R^*}{E^* I^*} - 1 \right) (E^* + I^*) \right)}, \\
\mu &= \frac{N^*}{I^*}, \\
\alpha &= \frac{N^*}{E^*}, \\
\nu &= I^* (\mu - \lambda),
\end{aligned} \tag{4.7}$$

où

$$N^* = \lim_{t \rightarrow +\infty} \frac{\mathbb{E}_{(e_0, i_0, n_0)}[N_t]}{t}, \quad E^* = \lim_{t \rightarrow +\infty} \mathbb{E}_{(e_0, i_0, n_0)}[E_t], \quad I^* = \lim_{t \rightarrow +\infty} \mathbb{E}_{(e_0, i_0, n_0)}[I_t],$$

et

$$R^* = \lim_{t \rightarrow +\infty} \mathbb{E}_{(e_0, i_0, n_0)}[E_t I_t],$$

pour $e_0, i_0, n_0 \in \mathbb{N}$.

Preuve. Soit $(E_t, I_t, N_t)_{t \geq 0}$ un processus markovien à valeurs dans $\tilde{S} = \mathbb{N} \times \mathbb{N} \times \mathbb{N}$, de générateur \tilde{Q} donné par : pour tout $e, i, n \geq 0$:

$$\begin{aligned}
\tilde{Q}_{(e, i, n), (e+1, i, n)} &= \lambda i + \nu, \\
\tilde{Q}_{(e, i, n), (e-1, i+1, n)} &= \alpha e, \\
\tilde{Q}_{(e, i, n), (e, i-1, n+1)} &= \mu i, \\
\tilde{Q}_{(e, i, n), (e, i, n)} &= -((\lambda + \mu)i + \alpha e + \nu).
\end{aligned}$$

Soit $(e_0, i_0, n_0) \in \tilde{S}$, pour toute fonction f positive ou bornée sur \tilde{S} et pour tout $e, i, n \geq 0$, nous avons :

$$\mathbb{E}_{(e_0, i_0, n_0)}[f(E_t, I_t, N_t)] = \sum_{(e, i, n) \in \tilde{S}} p_{(e_0, i_0, n_0)(e, i, n)}(t) f(e, i, n).$$

En utilisant l'équation de Kolmogorov progressive, on obtient :

$$\begin{aligned}
\frac{d}{dt} \mathbb{E}_{(e_0, i_0, n_0)}[f(E_t, I_t, N_t)] &= \sum_{(e, i, n) \in \tilde{S}} p'_{(e_0, i_0, n_0)(e, i, n)}(t) f(e, i, n) \\
&= \sum_{(e, i, n) \in \tilde{S}} \left(\sum_{(e', i', n') \in \tilde{S}} p_{(e_0, i_0, n_0)(e', i', n')} \tilde{Q}_{(e', i', n')(e, i, n)} \right) f(e, i, n) \\
&= \sum_{(e', i', n') \in \tilde{S}} p_{(e_0, i_0, n_0)(e', i', n')} \left(\sum_{(e, i, n) \in \tilde{S}} \tilde{Q}_{(e', i', n')(e, i, n)} f(e, i, n) \right) \\
&= \sum_{(e', i', n') \in \tilde{S}} p_{(e_0, i_0, n_0)(e', i', n')} \left((\lambda i' + \nu) (f(e' + 1, i', n') - f(e', i', n')) \right. \\
&\quad \left. + \alpha e' (f(e' - 1, i' + 1, n') - f(e', i', n')) \right. \\
&\quad \left. + \mu i' (f(e', i' - 1, n' + 1) - f(e', i', n')) \right) \\
&= \mathbb{E}_{(e_0, i_0, n_0)} \left[(\lambda I_t + \nu) (f(E_t + 1, I_t, N_t) - f(E_t, I_t, N_t)) \right. \\
&\quad \left. + \alpha E_t (f(E_t - 1, I_t + 1, N_t) - f(E_t, I_t, N_t)) + \mu I_t (f(E_t, I_t - 1, N_t + 1) \right. \\
&\quad \left. - f(E_t, I_t, N_t)) \right].
\end{aligned}$$

En considérant des fonctions $f(E_t, I_t, N_t)$, telles que $E_t, I_t, N_t, E_t^2, E_t I_t$ et I_t^2 , on peut établir le système d'équations différentielles suivant :

$$\left\{ \begin{array}{l} \frac{d}{dt} \mathbb{E}_{(e_0, i_0, n_0)}[E_t] = -\alpha \mathbb{E}_{(e_0, i_0, n_0)}[E_t] + \lambda \mathbb{E}_{(e_0, i_0, n_0)}[I_t] + \nu, \\ \frac{d}{dt} \mathbb{E}_{(e_0, i_0, n_0)}[I_t] = \alpha \mathbb{E}_{(e_0, i_0, n_0)}[E_t] - \mu \mathbb{E}_{(e_0, i_0, n_0)}[I_t], \\ \frac{d}{dt} \mathbb{E}_{(e_0, i_0, n_0)}[N_t] = \mu \mathbb{E}_{(e_0, i_0, n_0)}[I_t], \\ \frac{d}{dt} \mathbb{E}_{(e_0, i_0, n_0)}[E_t^2] = (2\nu + \alpha) \mathbb{E}_{(e_0, i_0, n_0)}[E_t] + \lambda \mathbb{E}_{(e_0, i_0, n_0)}[I_t] - 2\alpha \mathbb{E}_{(e_0, i_0, n_0)}[E_t^2] \\ \quad + 2\lambda \mathbb{E}_{(e_0, i_0, n_0)}[E_t I_t] + \nu, \\ \frac{d}{dt} \mathbb{E}_{(e_0, i_0, n_0)}[E_t I_t] = -\alpha \mathbb{E}_{(e_0, i_0, n_0)}[E_t] + \nu \mathbb{E}_{(e_0, i_0, n_0)}[I_t] + \alpha \mathbb{E}_{(e_0, i_0, n_0)}[E_t^2] \\ \quad - (\mu + \alpha) \mathbb{E}_{(e_0, i_0, n_0)}[E_t I_t] + \lambda \mathbb{E}_{(e_0, i_0, n_0)}[I_t^2], \\ \frac{d}{dt} \mathbb{E}_{(e_0, i_0, n_0)}[I_t^2] = \alpha \mathbb{E}_{(e_0, i_0, n_0)}[E_t] + \mu \mathbb{E}_{(e_0, i_0, n_0)}[I_t] + 2\alpha \mathbb{E}_{(e_0, i_0, n_0)}[E_t I_t] - 2\mu \mathbb{E}_{(e_0, i_0, n_0)}[I_t^2]. \end{array} \right.$$

En résolvant le système (4.8), nous pouvons obtenir les expressions des espérances suivantes :

$$\begin{aligned} \mathbb{E}_{(e_0, i_0, n_0)}[E_t] &= c_1 e^{-\frac{\mu+\alpha+\sqrt{(\mu-\alpha)^2+4\alpha\lambda}}{2}t} + c_2 e^{-\frac{-\mu-\alpha+\sqrt{(\mu-\alpha)^2+4\alpha\lambda}}{2}t} + \frac{\mu\nu}{\alpha(\mu-\lambda)}, \\ \mathbb{E}_{(e_0, i_0, n_0)}[I_t] &= c_3 e^{-\frac{\mu+\alpha+\sqrt{(\mu-\alpha)^2+4\alpha\lambda}}{2}t} + c_4 e^{-\frac{-\mu-\alpha+\sqrt{(\mu-\alpha)^2+4\alpha\lambda}}{2}t} + \frac{\nu}{\mu-\lambda}, \\ \mathbb{E}_{(e_0, i_0, n_0)}[N_t] &= c_5 + c_6 e^{-\frac{\mu+\alpha+\sqrt{(\mu-\alpha)^2+4\alpha\lambda}}{2}t} + c_7 e^{-\frac{-\mu-\alpha+\sqrt{(\mu-\alpha)^2+4\alpha\lambda}}{2}t} + \frac{\mu\nu}{\mu-\lambda}t, \\ \mathbb{E}_{(e_0, i_0, n_0)}[E_t I_t] &= c_8 e^{-(\mu+\alpha)t} + c_9 e^{-\frac{\mu+\alpha+\sqrt{(\mu-\alpha)^2+4\alpha\lambda}}{2}t} + c_{10} e^{-\frac{-\mu-\alpha+\sqrt{(\mu-\alpha)^2+4\alpha\lambda}}{2}t} \\ &\quad + c_{11} e^{-(\mu+\alpha+\sqrt{(\mu-\alpha)^2+4\alpha\lambda})t} + c_{12} e^{(-\mu-\alpha+\sqrt{(\mu-\alpha)^2+4\alpha\lambda})t} + \frac{\mu\nu((\mu+\alpha)\nu + \alpha\lambda)}{\alpha(\mu-\lambda)^2(\mu+\alpha)}, \end{aligned}$$

où c_i pour $i = 1, \dots, 12$ sont des constantes réelles.

Ainsi, si $\lambda < \mu$, alors

$$\begin{aligned} E^* &:= \lim_{t \rightarrow +\infty} \mathbb{E}_{(e_0, i_0, n_0)}[E_t] = \frac{\mu\nu}{\alpha(\mu-\lambda)}, \\ I^* &:= \lim_{t \rightarrow +\infty} \mathbb{E}_{(e_0, i_0, n_0)}[I_t] = \frac{\nu}{\mu-\lambda}, \\ N^* &:= \lim_{t \rightarrow +\infty} \frac{\mathbb{E}_{(e_0, i_0, n_0)}[N_t]}{t} = \frac{\mu\nu}{\mu-\lambda}, \\ R^* &:= \lim_{t \rightarrow +\infty} \mathbb{E}_{(e_0, i_0, n_0)}[E_t I_t] = \frac{\mu\nu((\mu+\alpha)\nu + \alpha\lambda)}{\alpha(\mu-\lambda)^2(\mu+\alpha)}. \end{aligned} \tag{4.8}$$

Ces quatre équations (4.8) forment un système inversible en fonction des paramètres $(\lambda, \mu, \alpha, \nu)$, car $\lambda \neq \mu$.

$$\begin{aligned}
\lambda &= \frac{N^* \left(\frac{R^*}{E^* I^*} - 1 \right) (E^* + I^*)}{I^* \left(1 + \left(\frac{R^*}{E^* I^*} - 1 \right) (E^* + I^*) \right)}, \\
\mu &= \frac{N^*}{I^*}, \\
\alpha &= \frac{N^*}{E^*}, \\
\nu &= I^* (\mu - \lambda).
\end{aligned}$$

Ainsi, nous pouvons déduire des estimateurs (de type moment) des paramètres λ, μ, ν et α . \square

Le processus N_t est observé mais les valeurs $(E_n, I_n)_{n \in \mathbb{N}} := (E_{n\Delta t}, I_{n\Delta t})_{n \in \mathbb{N}}$ sont cachées dans notre contexte, nous devons donc recourir à une méthode adéquate afin d'estimer les valeurs E^*, I^* et R^* . C'est la deuxième étape de notre méthodologie, pour laquelle nous utiliserons un modèle de Markov caché approprié.

4.3.4 Modèle de Markov Multi-Chaîne Cachée

Le modèle de Markov Multi-Chaîne Cachée (en anglais : Hidden Multi-Chain Markov Model (**HMCMM**)) est un outil statistique très puissant pour le traitement de données séquentielles multivariées. Il est considéré comme une extension du modèle de Markov caché classique en introduisant des probabilités conditionnelles couplées entre les suites d'états de plusieurs HMM. Cette extension permet de capturer les dépendances complexes entre les différentes suites d'états, ce qui en fait un choix très utile pour de nombreuses applications [73, 74, 75, 76].

Le but de cette section est d'adapter les algorithmes du HMM afin d'estimer les probabilités de transition de la chaîne cachée $(E_n, I_n)_{n \in \mathbb{N}}$ à partir des observations $(O_n)_{n \in \mathbb{N}}$. Ensuite, nous utiliserons ces estimations pour obtenir les valeurs E^*, I^* et R^* à l'aide d'une méthode de Monte Carlo. Dans notre modèle de Markov caché exposé-infecté (**Fig.4.6**), chaque chaîne représente l'état d'une population (exposée, infectée ou isolée) et où les chaînes sont couplées entre elles. À chaque instant n , l'état caché des infectés dépend de l'état de sa propre chaîne et de la chaîne de couplage des exposés à l'instant $n - 1$. De même, l'état caché des exposés dépend de l'état de sa propre chaîne et de la chaîne de couplage des infectés à l'instant $n - 1$. Ce couplage entre les chaînes reflète l'interaction entre les populations exposée et infectée dans la propagation de la maladie. Cependant, la valeur observée O_n à chaque instant dépend non seulement de l'état actuel I_n , mais aussi de l'état précédent I_{n-1} , ce qui représente un schéma d'émission non standard.

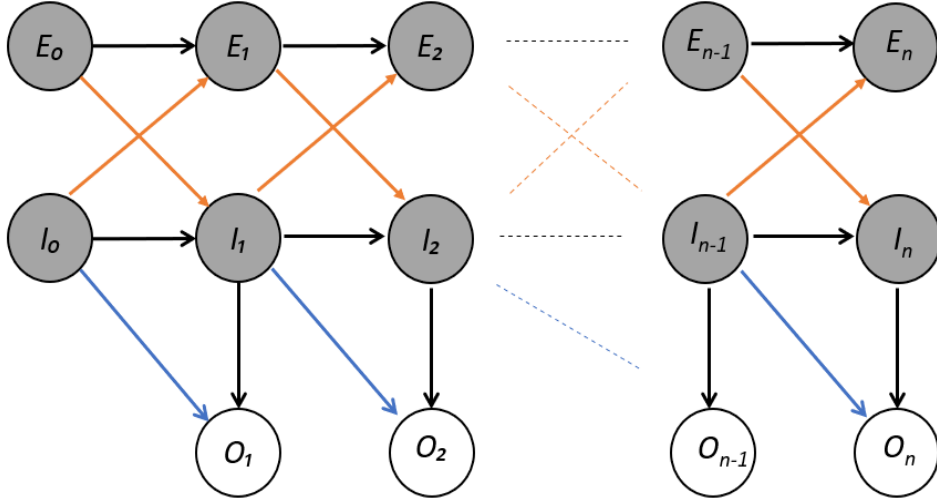


FIGURE 4.6 – Modèle de Markov multi chaîne cachée (exposé-infecté).

Nous pouvons ramener notre HMCMM à un HMM avec un schéma d'observations standard en introduisant la chaîne tridimensionnelle $(Z_n)_{n \in \mathbb{N}^*} = (E_{n-1}, I_{n-1}, I_n)_{n \in \mathbb{N}^*}$, qui représente une chaîne de Markov cachée à valeurs dans l'espace d'état \mathbb{N}^3 . Cette transformation permet de démontrer que la suite (Z_n, O_n) est un modèle de Markov caché standard avec les caractéristiques suivantes.

Lemme 4.2. *Le processus $(Z_n, O_n)_{n \in \mathbb{N}^*}$ est un modèle de Markov caché dont les caractéristiques sont données par le triple $M = (Q, \psi, \rho)$, où*

i) La matrice de probabilité de transition Q de la chaîne cachée (Z_n) est

$$Q_{(e,i,j),(e',i',j')} = \frac{P_{(e',i'),(\cdot,j')} P_{(e,i),(e',i')}}{P_{(e,i),(\cdot,j)}} \delta_{i'=j} \quad (e, i, j), (e', i', j') \in \mathbb{N}^3,$$

où $p_{(e,i),(e',i')} = p_{(e,i),(e',i')}(\Delta t)$ est la matrice de transition de la chaîne $(E_n, I_n)_{n \in \mathbb{N}}$ et $P_{(e,i),(\cdot,j)} = \sum_k P_{(e,i),(k,j)}$.

ii) La probabilité d'émission ψ du processus O sachant le processus Z est

$$\psi_{(e,i,j)}(o) = \mathbb{P}(O_n = o | Z_n = (e, i, j)) = \frac{P_{(e,i),(\cdot,j,o)}}{P_{(e,i),(\cdot,j)}} \quad e, i, j, o \in \mathbb{N},$$

où $p_{(e,i),(\cdot,j,o)} = \sum_k P_{(e,i),(k,j,o)}(\Delta t)$ est caractérisée par l'équation 4.6.

iii) La loi initiale ρ du processus d'état Z

$$\rho_{e,i,j} = \mathbb{P}(Z_1 = (e, i, j)) = p_{(e,i),(\cdot,j)} \pi_{(e,i)} \quad e, i, j \in \mathbb{N},$$

où π est la loi initiale de la chaîne $(E_n, I_n)_{n \in \mathbb{N}}$.

Preuve. *i)* La probabilité de transition Q : soient $(e, i, j), (e', i', j') \in \mathbb{N}^3$, nous avons

$$\begin{aligned} Q_{(e,i,j),(e',i',j')} &= \mathbb{P}(Z_{n+1} = (e', i', j') | Z_n = (e, i, j)) \\ &= \frac{\mathbb{P}(E_n = e', I_n = i', I_{n+1} = j', E_{n-1} = e, I_{n-1} = i, I_n = j)}{\mathbb{P}(E_{n-1} = e, I_{n-1} = i, I_n = j)}. \end{aligned}$$

En utilisant la propriété de Markov de la chaîne (E_n, I_n) , nous obtenons

$$\begin{aligned} Q_{(e,i,j),(e',i',j')} &= \frac{\mathbb{P}(I_{n+1} = j' | E_n = e', I_n = i') \mathbb{P}(E_n = e', I_n = i' | E_{n-1} = e, I_{n-1} = i)}{\mathbb{P}(I_n = j | E_{n-1} = e, I_{n-1} = i)} \delta_{i'=j} \\ &= \frac{p_{(e',i'),(\cdot,j')} p_{(e,i),(e',i')}}{p_{(e,i),(\cdot,j)}} \delta_{i'=j}. \end{aligned}$$

ii) La probabilité d'émission ψ : c'est évident par la formule de la probabilité conditionnelle.

iii) La loi initiale ρ : soit $e, i, j \in \mathbb{N}$

$$\begin{aligned} \rho_{e,i,j} &= \mathbb{P}(Z_1 = (e, i, j)) \\ &= \mathbb{P}(I_1 = j | (E_0, I_0) = (e, i)) \mathbb{P}((E_0, I_0) = (e, i)) \\ &= p_{(e,i),(\cdot,j)} \pi_{(e,i)}. \end{aligned}$$

□

Dans la suite, nous visons à estimer les paramètres du modèle de Markov caché (Z_n, O_n) , à savoir la matrice de transition p de la chaîne (E_n, I_n) , en maximisant la fonction de vraisemblance $\mathbb{P}(O_1 = o_1, \dots, O_T = o_T | M)$ à l'aide de l'algorithme de Baum-Welch ou l'algorithme forward-backward. Comme dans le chapitre précédent, nous avons réécrit les fonctions principales de l'algorithme EM, en adaptant les formules récursives à la structure particulière de notre modèle pour obtenir des estimateurs explicites des probabilités de transition $p_{(e,i),(e',i')}$.

Pour implémenter cet algorithme de maximisation de la vraisemblance des observations, nous devons d'abord définir les probabilités Forward et Backward.

La probabilité Forward peut être calculée par la récursion suivante : pour tout $e, i, j \in \mathbb{N}$ et $2 \leq t \leq T$

$$\begin{cases} \alpha_{(e,i,j)}(1) = p_{(e,i),(\cdot,j)} \pi_{(e,i)} \psi_{(e,i,j)}(o_1), \\ \alpha_{(e,i,j)}(t) = \psi_{(e,i,j)}(o_t) \sum_{(e',i') \in \mathbb{N}^2} \frac{p_{(e,i),(\cdot,j)} p_{(e',i'),(e,i)}}{p_{(e',i'),(\cdot,i)}} \alpha_{(e',i',i)}(t-1). \end{cases}$$

La probabilité Backward peut être évaluée par la formule récursive : $e, i, j \in \mathbb{N}$ et $1 \leq t \leq T-1$

$$\begin{cases} \beta_{(e,i,j)}(T) = 1, \\ \beta_{(e,i,j)}(t) = \sum_{(e',j') \in \mathbb{N}^2} \frac{p_{(e',j'),(\cdot,j')} p_{(e,i),(e',j)}}{p_{(e,i),(\cdot,j)}} \psi_{(e',j,j')} (o_{t+1}) \beta_{(e',j,j')}(t+1). \end{cases}$$

En utilisant les probabilités Forward et Backward, la vraisemblance peut être calculée en additionnant le produit de ces deux probabilités pour tous les états de la chaîne cachée, comme indiqué par la formule suivante :

$$\mathbb{P}(O_1 = o_1, \dots, O_T = o_T | M) = \sum_{(e,i,j) \in \mathbb{N}^3} \alpha_{(e,i,j)}(1) \beta_{(e,i,j)}(T).$$

Le théorème suivant présente le schéma itératif de l'algorithme de Baum-Welch adapté.

Théorème 4.4. *Étant donné les paramètres $M^n = (Q^n, \psi^n, \rho^n)$, les estimations du maximum de vraisemblance de $M^{n+1} = (Q^{n+1}, \psi^{n+1}, \rho^{n+1})$ étant donné les observations (o_1, \dots, o_T) sont données, pour $e, i, j, o \in \mathbb{N}$, par*

$$\begin{aligned} Q_{(e,i,j),(e',i',j')}^{n+1} &= \frac{\sum_{t=1}^{T-1} \xi_{(e,i,j),(e',i',j')}^n(t)}{\sum_{t=1}^T \gamma_{(e,i,j)}^n(t)} \delta_{i'=j}, \\ \psi_{(e,i,j)}^{n+1}(o) &= \frac{\sum_{t=1}^T \mathbb{1}_{o_t=o} \gamma_{(e,i,j)}^n(t)}{\sum_{t=1}^T \gamma_{(e,i,j)}^n(t)}, \\ \rho_{e,i,j}^{n+1} &= \gamma_{(e,i,j)}^n(1), \\ p_{(e,i),(e',i')}^{n+1} &= \frac{\sum_{j' \in \mathbb{N}} Q_{(e,i,j),(e',j',j')}^{n+1} \rho_{e,i,j}^{n+1}}{\sum_{j \in \mathbb{N}} \rho_{e,i,j}^{n+1}}, \end{aligned}$$

où

$$\begin{aligned} \xi_{(e,i,j),(e',i',j')}^n(t) &= \mathbb{P}(Z_t = (e, i, j), Z_{t+1} = (e', i', j') | O_{1:T} = o_{1:T}, M^n) \\ &= \frac{\alpha_{(e,i,j)}^n(t) p_{(e',i'),(.,j')}^n p_{(e,i),(e',i')}^n \psi_{(e',i',j')}^n(o_{t+1}) \beta_{(e',i',j')}^n(t+1)}{p_{(e,i),(.,j)}^n \sum_{(e,i,j) \in \mathbb{N}^3} \alpha_{(e,i,j)}^n(t) \beta_{(e,i,j)}^n(t)} \delta_{i'=j}, \end{aligned}$$

et

$$\begin{aligned} \gamma_{(e,i,j)}^n(t) &= \mathbb{P}(Z_t = (e, i, j) | O_{1:T} = o_{1:T}, M^n) \\ &= \frac{\alpha_{(e,i,j)}^n(t) \beta_{(e,i,j)}^n(t)}{\sum_{(e,i,j) \in \mathbb{N}^3} \alpha_{(e,i,j)}^n(t) \beta_{(e,i,j)}^n(t)}, \end{aligned}$$

avec $\alpha_{(e,i,j)}^n(t), \beta_{(e,i,j)}^n(t)$ sont les probabilités Forward et Backward à l'itération n .

Preuve. Notons par $\mathcal{Z} = \mathbb{N}^3$ l'espace d'état de la chaîne $(Z_n)_{n \in \mathbb{N}^*}$ et $z = (e, i, j) \in \mathcal{Z}$. On cherche à maximiser la vraisemblance d'observations $\mathbb{P}(O_{1:T} = o_{1:T} | M^{n+1})$. En appliquant l'algorithme EM [77] à la maximisation de cette probabilité, on est amené à maximiser la fonction auxiliaire suivante :

$$\Gamma_{O_{1:T}}(M^n, M^{n+1}) = \sum_{z_{1:T} \in \mathcal{Z}^T} \log \mathbb{P}(O_{1:T} = o_{1:T}, Z_{1:T} = z_{1:T} | M^{n+1}) \mathbb{P}(Z_{1:T} = z_{1:T} | O_{1:T}, M^n).$$

Sachant que

$$\mathbb{P}(O_{1:T} = o_{1:T}, Z_{1:T} = z_{1:T} | M^{n+1}) = \rho_{z_1}^{n+1} \left(\prod_{t=1}^{T-1} Q_{z_t, z_{t+1}}^{n+1} \right) \left(\prod_{t=1}^T \psi_{z_t}^{n+1}(o_t) \right),$$

la fonction $\Gamma_{O_{1:T}}$ se ré-écrit

$$\begin{aligned} \Gamma_{O_{1:T}}(M^n, M^{n+1}) &= \sum_{z_{1:T} \in \mathcal{Z}^T} \log \rho_{z_1}^{n+1} \mathbb{P}(Z_{1:T} = z_{1:T} | O_{1:T}, M^n) \\ &\quad + \sum_{z_{1:T} \in \mathcal{Z}^T} \left(\sum_{t=1}^{T-1} \log Q_{z_t, z_{t+1}}^{n+1} \right) \mathbb{P}(Z_{1:T} = z_{1:T} | O_{1:T}, M^n) \\ &\quad + \sum_{z_{1:T} \in \mathcal{Z}^T} \left(\sum_{t=1}^T \log \psi_{z_t}^{n+1}(o_t) \right) \mathbb{P}(Z_{1:T} = z_{1:T} | O_{1:T}, M^n) \\ &= \Gamma_{O_{1:T}}^\rho(M^n, M^{n+1}) + \Gamma_{O_{1:T}}^Q(M^n, M^{n+1}) + \Gamma_{O_{1:T}}^\psi(M^n, M^{n+1}). \end{aligned}$$

On remarque alors que la fonction $\Gamma_{O_{1:T}}$ peut se décomposer en trois fonctions ayant des paramètres distincts et indépendants. Par conséquent, nous pouvons optimiser analytiquement ces trois fonctions à l'aide de multiplicateurs de Lagrange. Soient $\mathcal{L}(\rho, \theta)$, $\mathcal{L}(Q, \theta)$ et $\mathcal{L}(\psi, \theta)$ les Lagrangiens associés respectivement à $\Gamma_{O_{1:T}}^\rho$, $\Gamma_{O_{1:T}}^Q$ et $\Gamma_{O_{1:T}}^\psi$.

$$\begin{aligned}\mathcal{L}(\rho, \theta) &= \sum_{(e,i,j)} \log \rho_{(e,i,j)}^{n+1} \mathbb{P}(Z_1 = (e, i, j) | O_{1:T}, M^n) + \theta \left(\sum_{(e',i',j')} \rho_{(e',i',j')}^{n+1} - 1 \right). \\ \mathcal{L}(Q, \theta) &= \sum_{(e,i,j)} \sum_{(e',i',j')} \left(\sum_{t=1}^{T-1} \log Q_{(e,i,j),(e',i',j')}^{n+1} \right) \mathbb{P}(Z_t = (e, i, j), Z_{t+1} = (e', i', j') | O_{1:T}, M^n) \\ &\quad + \sum_{(e,i,j)} \theta_{(e,i,j)} \left(\sum_{(e',i',j')} Q_{(e,i,j),(e',i',j')}^{n+1} - 1 \right). \\ \mathcal{L}(\psi, \theta) &= \sum_{(e,i,j)} \left(\sum_{t=1}^T \log \psi_{(e,i,j)}^{n+1}(o) \right) \mathbb{P}(Z_t = (e, i, j) | O_{1:T}, M^n) \mathbb{1}_{o_t=o} \\ &\quad + \sum_{(e,i,j)} \theta_{(e,i,j)} \left(\sum_o \psi_{(e,i,j)}^{n+1}(o) - 1 \right).\end{aligned}$$

En mettant les dérivées premières des Lagrangiens à zéro, et en procédant de manière similaire à la preuve du théorème **3.3**, on peut montrer que $\Gamma_{O_{1:T}}$ est maximisé lorsque

$$\begin{aligned}Q_{(e,i,j),(e',i',j')}^{n+1} &= \frac{\sum_{t=1}^{T-1} \xi_{(e,i,j),(e',i',j')}^n(t)}{\sum_{t=1}^T \gamma_{(e,i,j)}^n(t)} \delta_{i'=j}, \\ \psi_{(e,i,j)}^{n+1}(o) &= \frac{\sum_{t=1}^T \mathbb{1}_{o_t=o} \gamma_{(e,i,j)}^n(t)}{\sum_{t=1}^T \gamma_{(e,i,j)}^n(t)}, \\ \rho_{e,i,j}^{n+1} &= \gamma_{(e,i,j)}^n(1).\end{aligned}$$

D'autre part, en utilisant le lemme **4.2**, on peut facilement déduire que

$$\sum_{j' \in \mathbb{N}} Q_{(e,i,j),(e',j',j')}^{n+1} \rho_{e,i,j}^{n+1} = p_{(e,i),(e',i')}^{n+1} \pi_{(e,i)}^{n+1}.$$

Or

$$\sum_{j \in \mathbb{N}} \rho_{e,i,j}^{n+1} = \pi_{(e,i)}^{n+1}.$$

Par conséquent :

$$p_{(e,i),(e',i')}^{n+1} = \frac{\sum_{j' \in \mathbb{N}} Q_{(e,i,j),(e',j',j')}^{n+1} \rho_{e,i,j}^{n+1}}{\sum_{j \in \mathbb{N}} \rho_{e,i,j}^{n+1}}.$$

□

À chaque itération, l'algorithme de Baum-Welch met à jour les paramètres du modèle en utilisant les formules données dans le théorème **(4.4)** et réévalue la log-vraisemblance des observations, jusqu'à ce qu'une convergence vers les valeurs optimales soit atteinte. L'obtention de la matrice de transition \hat{p} optimale permet de simuler la chaîne de Markov (E_n, I_n) avec une taille suffisamment grande de n pour récupérer les moments E^* , I^* et R^* . Finalement, le théorème **(4.3)** peut être utilisé pour estimer les paramètres du modèle exposé-infecté λ, μ, α et ν .

4.4 Simulation numérique

Nous allons étudier les performances numériques des estimateurs $\hat{\lambda}^n, \hat{\mu}^n, \hat{\alpha}^n$ et $\hat{\nu}^n$ sur des données synthétiques. Nous commencerons par examiner le cadre où la chaîne bivariée (E_n, I_n) est observé, puis nous nous tournons vers le cadre HMCMM.

4.4.1 Observations complètes

Pour estimer les moments E^*, I^* et R^* de la chaîne de Markov observée (E_n, I_n) , nous utilisons la méthode de Monte Carlo qui consiste à simuler un grand nombre de trajectoires de la chaîne à partir desquelles nous pouvons estimer les moments empiriques. Nous simulons plusieurs trajectoires du processus (E_t, I_t) partant d'un état initial $(E_0, i_0) = (0, 0)$ avec des valeurs de paramètres fixées à $\lambda = 0.05, \mu = 0.2, \alpha = 0.1$ et $\nu = 0.015$, sur différents horizons du temps H . Pour chaque trajectoire, avec un pas de temps fixé à $\Delta t = 1j$, nous obtenons la chaîne discrète (E_n, I_n) et les observations O_n qui décrivent le nombre de sauts négatifs du processus marginal I_t sur l'intervalle $]n\Delta t, (n+1)\Delta t]$. Nous utilisons ensuite les moyennes empiriques de la chaîne pour estimer les moments, puis nous obtenons des estimations des paramètres en les injectant dans la formule (4.4). Cette méthode nous permet d'obtenir une distribution des estimateurs (Fig.4.7), à partir de laquelle nous pouvons calculer des intervalles de confiance.

Afin de réduire le temps de calcul des simulations de Monte Carlo, nous utilisons la fonction "dopar" du package "doparallel" [78] pour paralléliser nos calculs. Cette méthode permet de diviser le travail en plusieurs tâches indépendantes qui peuvent être exécutées simultanément, ce qui a permis de réduire considérablement le temps nécessaire pour effectuer les simulations.

Le tableau (4.1) présente les valeurs estimées des paramètres ainsi que leurs intervalles de confiance pour différents horizons H (nombre des observations), pour 10000 trajectoires simulées.

TABLE 4.1 – Estimations des paramètres et leurs intervalles de confiance sur 10000 trajectoires du processus exposé-infecté.

Nombre des observations $n = 1000$		
Les paramètres	Les estimations	Intervalle de confiance (95%)
$\lambda = 0.05$	0.049	[0.047, 0.052]
$\mu = 0.2$	0.1999	[0.1996, 0.2019]
$\alpha = 0.1$	0.10	[0.0996, 0.1005]
$\nu = 0.015$	0.0147	[0.0145, 0.0149]
Nombre des observations $n = 10000$		
$\lambda = 0.05$	0.050	[0.049, 0.0506]
$\mu = 0.2$	0.20	[0.1998, 0.2001]
$\alpha = 0.1$	0.100	[0.0998, 0.1002]
$\nu = 0.015$	0.015	[0.0149, 0.01506]

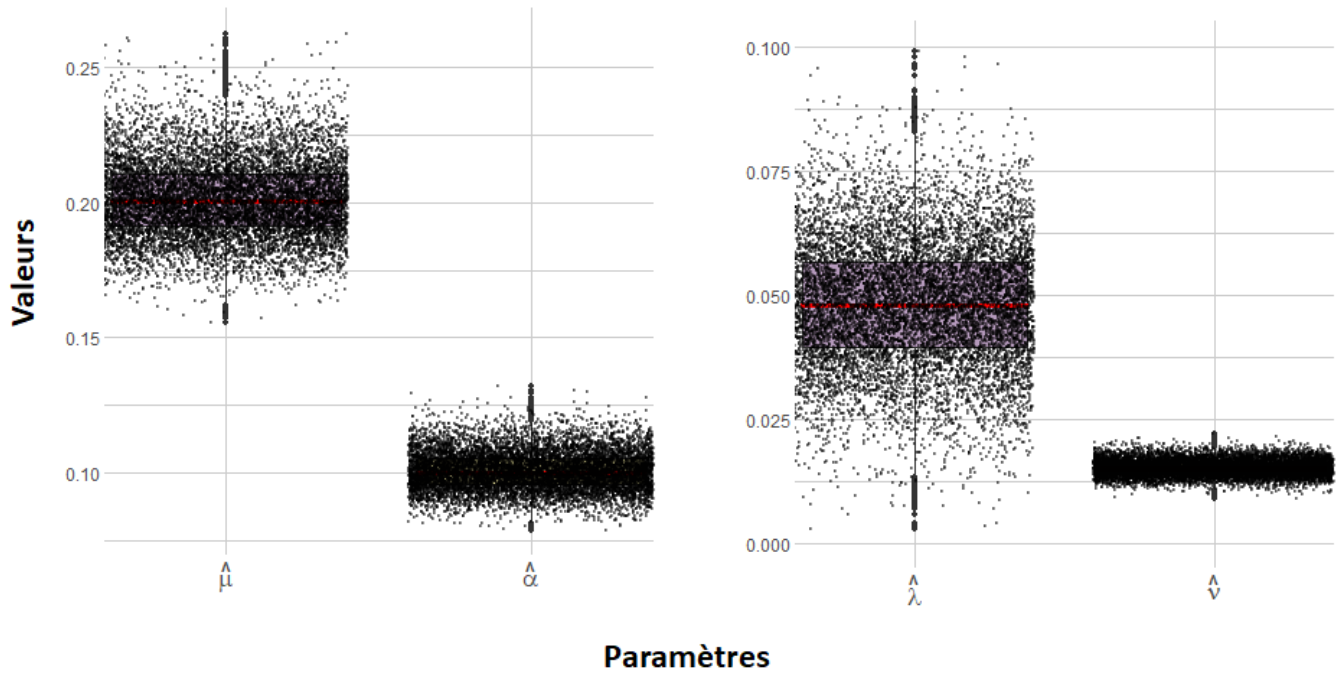


FIGURE 4.7 – Distributions des 10000 estimateurs simulés sur une longueur $H = 10000$.

Il convient de noter que la précision des estimations dépend non seulement du nombre de trajectoires simulées, mais aussi de la longueur de chaque trajectoire. En effet, plus le nombre de trajectoires simulées est élevé, plus la précision de l'estimation sera élevée et l'intervalle de confiance sera réduit, ce qui signifie que l'estimation des paramètres sera plus précise et plus fiable. Cependant, il est important de noter que la longueur de chaque trajectoire peut également affecter la précision de l'estimation, en particulier si la longueur est trop courte, ce qui peut entraîner une variance élevée des estimations. Il est donc important de trouver un compromis entre le nombre de trajectoires et la longueur de chaque trajectoire pour obtenir des estimations précises des paramètres avec des intervalles de confiance resserrés.

4.4.2 Observations du cumul des nouveaux cas déclarés

Nous allons maintenant nous concentrer sur le cadre HMCMM, plus adapté à notre contexte d'étude, où seuls les nouveaux cas déclarés cumulés (O_n) sont observés. Pour cela, nous avons simulé un processus exposé-infecté défini par son générateur infinitésimal (4.1), à partir du point initial $(0, 0)$, avec des paramètres $\lambda = 0.05$, $\mu = 0.2$, $\alpha = 0.1$ et $\nu = 0.015$, sur une période de temps H (Fig.4.8). Nous avons ensuite extrait les observations correspondant aux cumuls des nouveaux sauts -1 du processus du processus I_t (les cumuls de nouveaux cas déclarés Fig.4.9), afin d'exécuter notre schéma adaptatif de Baum-Welch et estimer les paramètres recherchés.

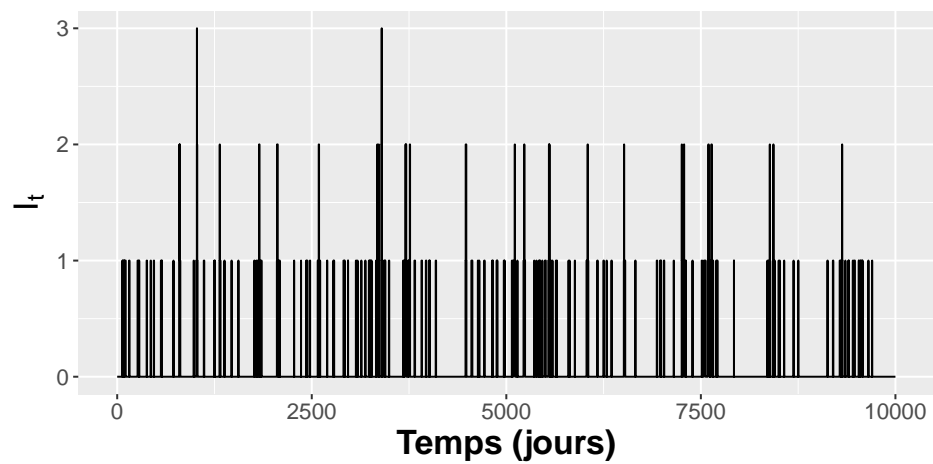
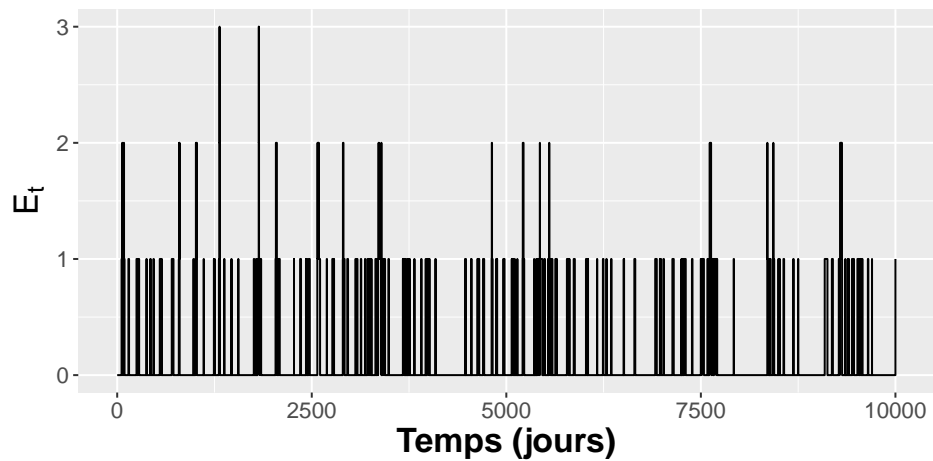


FIGURE 4.8 – Trajectoire du processus exposé-infecté (E_t, I_t) sur un horizon $H = 10000$ jours.

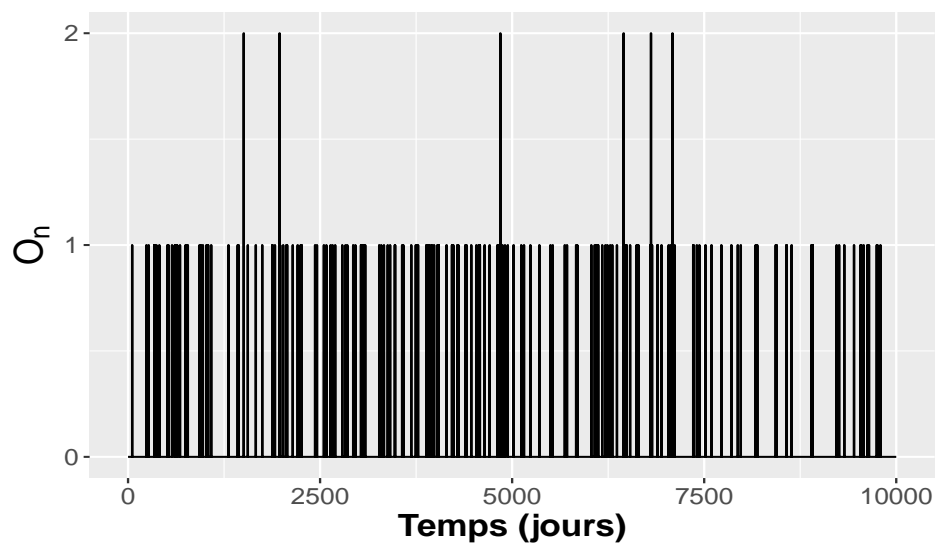


FIGURE 4.9 – Les cumuls des nouveaux déclarés O_n sur une période de $\Delta t = 1$ jour.

Pour le reste des simulations, Nous supposons que le nombre d'états cachés des processus E_t et I_t est égal à $N + 1$, et nous tronquons les matrices de transition et d'émission respectivement sur les espace d'états finis $\{0, 1, \dots, N\} \times \{0, 1, \dots, N\}$ et $\{0, 1, \dots, N\} \times \{0, 1, \dots, N\} \times \{0, 1, \dots, M\}$, où M représente l'état maximal observé. Il est important de noter que l'algorithme Baum-Welch est sensible aux choix des paramètres initiaux, car il peut converger vers des optima locaux plutôt que globaux. Si les paramètres initiaux sont trop éloignés des vraies valeurs des paramètres, cela peut entraîner une non-convergence de l'algorithme. Par conséquent, il est important de choisir des paramètres initiaux qui sont proches et dans le même ordre de grandeur que les vraies valeurs des paramètres autant que possible. Cela peut être fait en utilisant des connaissances préalables sur le système étudié ou en effectuant des simulations initiales pour obtenir des valeurs raisonnables pour les paramètres. Dans notre étude, nous avons décidé de sélectionner les paramètres initiaux dans les plages suivantes : $\lambda^{(0)} \in [0.04, 0.07]$, $\mu^{(0)} \in [0.185, 0.25]$, $\alpha^{(0)} \in [0.09, 0.130]$ et $\nu^{(0)} \in [0.013, 0.02]$ avec *length.out* = 15 (en générant 15 valeurs initiales pour chaque paramètre), cela nous permet d'initialiser les caractéristiques de notre modèle HMCMM de la façon suivante : Pour chaque valeur de quadruplet $(\lambda_i^{(0)}, \mu_i^{(0)}, \alpha_i^{(0)}, \nu_i^{(0)})$ pour $i = 1, \dots, 15$, nous effectuons des simulations de Monte Carlo parallélisées sur 10000 trajectoires différentes pour estimer les coefficients des matrices $P^{(0)}$ et $\psi^{(0)}$. Ensuite, nous redéfinissons la dernière colonne de chaque matrice en utilisant les formules suivantes :

$$p_{(e,i)(N,N)} = 1 - \sum_{e'=0}^{N-1} \sum_{i'=0}^{N-1} p_{(e,i)(e',i')},$$

$$\psi_{(e,i,j)}(M) = 1 - \sum_{o=0}^{M-1} \psi_{(e,i,j)}(o),$$

pour $e, i, j = 0, \dots, N$.

Pour les autres paramètres du notre schéma itératif, nous fixons $\Delta t = 1$, *maxIterations* = 500 (nombre maximum d'itérations) et *delta* = 10^{-9} (critère d'arrêt). Nous exécutons l'algorithme de Baum-Welch adapté pour chaque valeur de $(\lambda_i^{(0)}, \mu_i^{(0)}, \alpha_i^{(0)}, \nu_i^{(0)})$ pour $i = 1, \dots, 15$, et nous choisissons finalement la matrice de transition qui maximise la vraisemblance des observations.

$$\hat{p}^m = \max_{\hat{p}^n} \mathbb{P}(O|M^n).$$

Nous utilisons ensuite cette matrice optimale \hat{p}^m pour simuler la chaîne de Markov (E_n, I_n) avec une taille suffisamment grande de n pour récupérer les moments E^* , I^* et R^* . Enfin, nous avons appliqué le théorème (4.3) pour estimer les paramètres du modèle exposé-infecté λ, μ, α et ν .

L'impact du choix des paramètres de troncature N et de l'horizon temporel H est examiné dans les tableaux 4.2 et 4.3. Ces paramètres sont importants pour la précision des estimations.

TABLE 4.2 – Estimation des paramètres pour différente valeur de N sur une trajectoire d’horizon $H = 1000$.

Nombre des observations	$n = 1000$		
Troncature de l’espace d’état caché	$N = 3$	$N = 4$	$N = 5$
$\hat{\lambda}^n$	0.059	0.057	0.058
$\hat{\mu}^n$	0.245	0.248	0.197
$\hat{\alpha}^n$	0.120	0.120	0.092
$\hat{\nu}^n$	0.0182	0.0185	0.014

TABLE 4.3 – Estimation des paramètres pour différente valeur de N sur une trajectoire d’horizon $H = 10000$.

Nombre des observations	$n = 10000$		
Troncature de l’espace d’état caché	$N = 3$	$N = 4$	$N = 5$
$\hat{\lambda}^n$	0.050	0.049	0.05
$\hat{\mu}^n$	0.189	0.196	0.209
$\hat{\alpha}^n$	0.098	0.101	0.102
$\hat{\nu}^n$	0.0146	0.015	0.015

Ces tableaux montrent les résultats d’estimation des paramètres du modèle pour différentes valeurs de N (le nombre d’états cachés) et de H (l’horizon temporel). Les estimations sont basées sur une trajectoire de simulation de longueur H et pour chaque valeur de N , l’algorithme de Baum-Welch est exécuté pour estimer les paramètres du modèle. On observe que les estimations des paramètres ne varient pas de manière significative avec le choix de N , pour $H = 10000$, les résultats montrent que pour les trois valeurs de N testées, les estimations des paramètres sont assez stables et proches les unes des autres, excepté pour μ^n qui a une légère variation. Donc, la troncature d’espace d’état n’a pas un impact significatif sur l’estimation, ceci est dû au fait que la majorité des états observés sont des 0 et 1. Cependant, une grande valeur de N augmente également le temps de calcul nécessaire pour obtenir les résultats.

En revanche, le choix du nombre d’observations n (ou de l’horizon temporel H) a un impact significatif sur les estimations obtenues. Les résultats de nos simulations montrent qu’un horizon temporel plus long conduit à des estimations plus précises des paramètres. En effet, un horizon temporel plus long fournit une plus grande quantité d’informations qui peuvent être utilisées pour estimer les paramètres. Par conséquent, l’augmentation du nombre d’observations n améliore d’une part la convergence locale de l’algorithme EM vers la vraie valeur de la matrice de transition et d’autre part l’estimation de N^* qui dépend essentiellement par l’horizon H et du nombre d’observations. Toutefois, il est important de noter que l’augmentation du nombre d’observations entraîne également une augmentation significative du temps d’exécution.

Comme précédemment mentionné dans le chapitre 3 (**3.4.1**), nos expériences numériques ont démontré que l’identifiabilité des moments est garantie avec notre stratégie d’estimation et notre adaptation du modèle. Cela confirme que notre approche permet

d'estimer les paramètres de manière fiable et de retrouver les caractéristiques désirées. Cependant, il convient de noter que nous n'avons pas pu vérifier explicitement si la matrice optimale obtenue par l'algorithme de Baum-Welch adapté est une meilleure approximation de la vraie matrice. Cela est dû au fait que nous n'avons pas pu trouver une expression explicite de la matrice p dans ce contexte spécifique.

Les résultats de notre étude numérique révèlent que la qualité de l'estimation des paramètres du modèle exposé-infecté dépend essentiellement de la quantité de données disponibles. Malheureusement, le jeu de données fourni par l'ARS de Mayotte s'est avéré insuffisant pour obtenir des estimations fiables et précises des paramètres clés du modèle, tels que λ, μ, α et ν , et pour saisir pleinement leur influence sur la dynamique de transmission de la fièvre typhoïde à Mayotte.

4.5 Conclusion

Dans ce chapitre, nous avons proposé une modélisation de la transmission de la fièvre typhoïde basée sur un modèle à deux compartiments (exposé-infecté), permettant de prendre en compte le facteur de la durée d'incubation de la maladie. Nous avons commencé par décrire notre modèle puis étudié ses propriétés mathématiques telles que son comportement moyen et asymptotique, nous avons également illustré ses propriétés par des simulations numériques permettant de quantifier le comportement qualitatif du modèle. Ensuite, nous avons présenté notre méthode d'estimation qui consiste tout d'abord à exprimer les paramètres du modèle en termes de moments puis nous nous sommes placés dans le cadre de multi-chaîne de Markov cachée afin d'adapter l'algorithme de Baum-Welch à notre schéma d'observations et d'estimer nos paramètres du modèle exposé-infecté. Enfin, des simulations numériques ont été réalisées pour évaluer la performance de notre méthodologie d'estimation.

Cependant, l'application de notre approche au jeu de données de Mayotte fourni par l'ARS n'a pas été possible en raison il y a trop peu de données pour obtenir des résultats fiables. Ces résultats soulignent l'importance d'une surveillance épidémiologique continue et de la collecte de données pour mieux comprendre et surveiller la propagation de la maladie.

Chapitre 5

Effet de la pluviométrie sur l'évolution de la fièvre typhoïde à Mayotte

Sommaire

5.1	Introduction	109
5.2	Description des données de précipitations	109
5.3	Détection de rupture	110
5.3.1	Méthode CUSUM	111
5.3.2	Application aux données de précipitations à Mayotte	111
5.4	Estimation des paramètres d'un modèle de Markov caché à deux régimes	114
5.5	Conclusion	118

5.1 Introduction

Dans ce chapitre, nous allons proposer une extension de notre modèle présenté dans le chapitre 3, appliqué au jeu de données de la fièvre typhoïde à Mayotte fourni par ARS. Cette fois, nous allons inclure l'effet de la pluviométrie sur l'évolution de la maladie. En effet, nous avons utilisé des données publiques sur les précipitations journalières à Mayotte [79] et nous allons étudier l'impact de ces conditions météorologiques sur notre modèle d'estimation. Nous avons constaté que la pluviométrie peut avoir un impact important sur la contamination des sources d'eau potable par les eaux de ruissellement, ce qui peut augmenter la transmission de la fièvre typhoïde.

La pluviométrie à Mayotte est relativement importante, avec une moyenne annuelle de précipitations d'environ 2000 mm. La saison des pluies s'étend généralement de novembre à avril, avec des précipitations abondantes et fréquentes. La saison sèche s'étend de mai à octobre, et pendant cette période, les précipitations sont rares. Ces conditions météorologiques peuvent avoir un impact sur la transmission des maladies infectieuses, notamment la fièvre typhoïde. En effet, les précipitations peuvent favoriser la contamination de l'environnement par la bactérie *Salmonella Typhi*, en particulier dans les zones où les conditions sanitaires sont précaires. C'est pourquoi il est important de considérer cette variable dans notre étude pour identifier son influence sur les paramètres de notre modèle, notamment le taux de contamination par l'environnement ν .

Dans la section 5.2, nous commençons par décrire les données de précipitations utilisées dans cette étude. Ensuite, nous tentons de modéliser ces données de précipitations à Mayotte (Fig.5.1) en utilisant des modèles classiques de séries chronologiques telles que ARMA (en anglais : AutoRegressive Moving Average) et SARIMA (en anglais : Seasonal AutoRegressive Integrated Moving Average) ou des techniques de lissage par des splines (voir Annexe E). Malheureusement, ces modèles n'ont pas réussi à capturer la variabilité des données. Nous avons donc décidé d'utiliser ces données uniquement pour détecter les deux régimes saisonniers de Mayotte, à savoir la saison sèche et la saison humide, à l'aide d'une méthode de détection de rupture classique présentée dans la section 5.3. Ensuite, dans la section 5.4, nous décrivons notre problème de Markov caché sous les deux régimes afin d'estimer les paramètres du modèle à partir du même jeu de données présenté dans 3.4.2.

5.2 Description des données de précipitations

Les données de précipitations utilisées dans cette étude proviennent de la source *historique-meteo.net* [79] et elles sont basées sur les enregistrements quotidiens de plusieurs régions de Mayotte, notamment Mamoudzou, Dapani, Chiconi, Mtsamboro et Bouéni. Ces données fournissent des informations sur les quantités de pluie enregistrées chaque jour dans ces différentes régions.

Afin d'obtenir une mesure agrégée représentative de l'ensemble de l'île, les données de précipitations de chaque région ont été combinées en calculant la moyenne des précipitations quotidiennes. Cette approche permet de déterminer une estimation globale des précipitations sur l'ensemble de Mayotte pour chaque jour.

Une première observation de la Fig.5.1 révèle une variabilité importante dans la quan-

tivité de précipitations hebdomadaires à Mayotte, avec des fluctuations régulières présentant des pics et des creux. Cette observation suggère l'existence de deux saisons de précipitations distinctes. De plus, nous remarquons que les épidémies de fièvre typhoïde à Mayotte ont tendance à se produire préférentiellement pendant la saison humide de chaque année, indiquant que la pluviométrie peut jouer un rôle dans la propagation de la maladie.

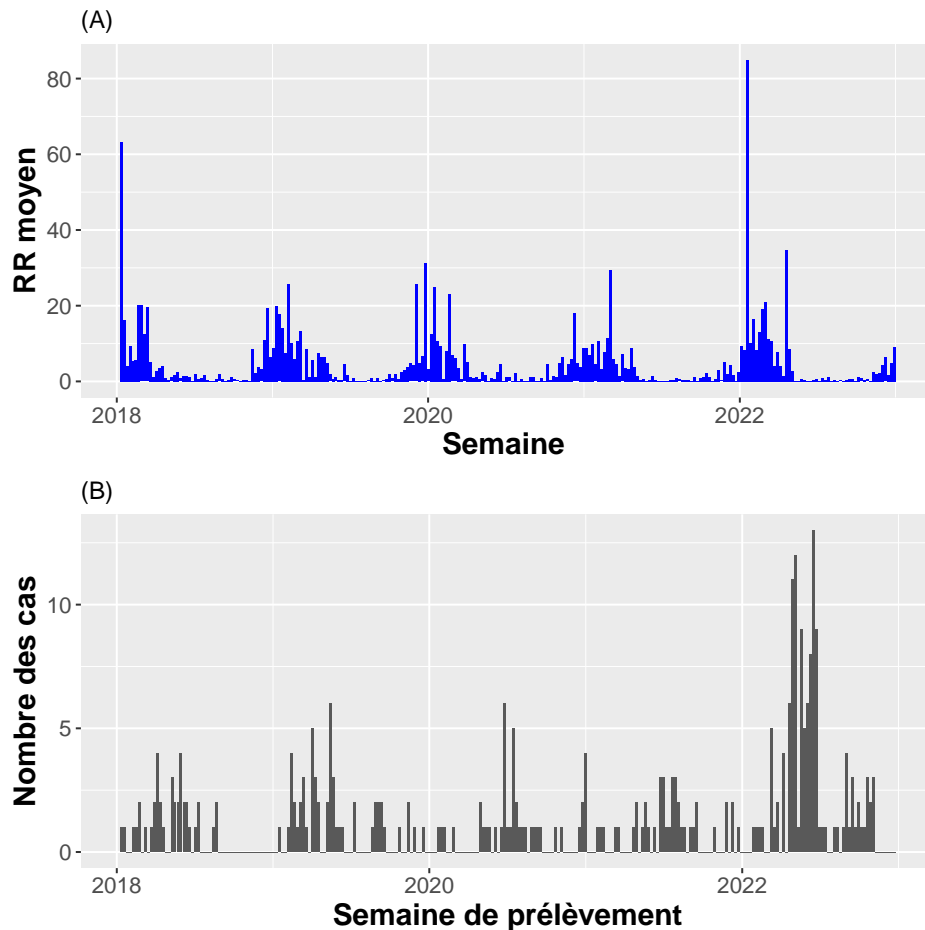


FIGURE 5.1 – - RR : La hauteur de précipitation en mm.
 - (A) : Moyenne des précipitations hebdomadaires à Mayotte entre 2018 et 2022.
 - (B) : Nombre hebdomadaire des nouveaux cas déclarés de la fièvre typhoïde à Mayotte entre 2018 et 2022.

5.3 Détection de rupture

L'objectif de cette section est d'identifier les régimes de deux saisons climatiques à Mayotte (sèche et humide), plus précisément de détecter les instants où la série de précipitations présente des changements brusques sur une période donnée. Pour ce faire, nous avons utilisé la statistique CUSUM (en anglais : Cumulative Sum) qui représente une méthode non paramétrique (la distribution de la série est supposée inconnue et n'est pas nécessairement normale) de détection des ruptures.

5.3.1 Méthode CUSUM

La méthode CUSUM est une méthode statistique de détection de rupture utilisée pour détecter des changements dans une série chronologique. L'approche consiste à tester de manière séquentielle la présence de rupture en utilisant l'écriture récursive de la statistique de détection de *Page* [80]. Elle consiste à calculer la somme cumulative des différences entre les moyennes avant et après chaque instant dans la série chronologique. Plus précisément, soit X_t la valeur de la série chronologique à l'instant t , et soit \bar{X}_1 et \bar{X}_2 les moyennes des valeurs de la série avant et après un instant de rupture supposé. La méthode CUSUM calcule la somme cumulative S_t comme suit :

$$S_t = \max(0, S_{t-1} + (X_t - \bar{X}_1) - (X_t - \bar{X}_2)), \quad (5.1)$$

où $S_0 = 0$ est la condition initiale. Si la série ne présente pas de changement, la somme cumulative restera proche de zéro. En revanche, si un changement survient, la somme cumulative augmentera et dépassera un seuil h prédéfini, ce qui indiquera la détection d'une rupture dans la série chronologique. La méthode conventionnelle pour choisir un seuil de détection constant h est basée sur l'inégalité de Wald [81]. Effectivement, selon Egea-Roca et al [82], le seuil de détection h doit être choisi de manière à respecter l'inégalité : $h \leq -\log(\alpha)$, où α est le niveau de risque souhaité pour la détection de rupture. En fixant un α donné (par exemple 0.05), on peut donc calculer la valeur maximale possible de h . Si le seuil choisi est supérieur à cette valeur maximale, alors le risque de fausse détection augmente. En pratique, il est souvent recommandé de choisir un seuil de détection légèrement inférieur à cette valeur maximale pour réduire ce risque.

Il existe plusieurs variantes de la méthode CUSUM en fonction des statistiques utilisées pour détecter les ruptures. Les deux versions les plus couramment utilisées sont la méthode CUSUM par moyenne et la méthode CUSUM par variance.

La méthode CUSUM par moyenne est basée sur la différence entre la moyenne actuelle et la moyenne précédente de la série de données. Elle est donnée par la formule 5.1.

La méthode CUSUM par variance, quant à elle, est basée sur la différence entre la variance actuelle et la variance précédente de la série de données. Elle est donnée par la formule suivante :

$$S_t = \max(0, S_{t-1} + (X_t - \sigma_1^2) - (X_t - \sigma_2^2)),$$

où σ_1^2 et σ_2^2 sont les variances de la série avant et après l'instant t . De même, si la statistique CUSUM dépasse le seuil h fixé, cela signifie qu'une rupture a été détectée.

Il existe également une méthode CUSUM combinée qui utilise à la fois la différence entre la moyenne et la différence entre la variance de la série de données pour détecter les ruptures. Cette méthode peut être plus robuste dans certaines situations où les ruptures peuvent affecter à la fois la moyenne et la variance de la série de données.

5.3.2 Application aux données de précipitations à Mayotte

Maintenant que nous avons présenté la méthode de détection CUSUM, nous pouvons l'appliquer pour sélectionner les instants des ruptures et ainsi identifier les régimes saisonniers de Mayotte.

En pratique, on utilise le package '**changepoint**' du logiciel R [83], il offre principalement trois fonctions pour la détection de rupture, notamment :

- La fonction **cpt.mean** pour la détection de rupture dans des séries chronologiques univariées basées sur le changement de la moyenne.

- La fonction **cpt.var** pour la détection de rupture dans des séries chronologiques univariées basées sur le changement de la variance.
- La fonction **cpt.meanvar** pour la détection de rupture dans des séries chronologiques univariées basées sur le changement de la moyenne et de la variance simultanément.

Pour chaque fonction, Les changements sont déterminés à l'aide de la méthode fournie, qui peut être un point de changement unique (AMOC) ou des points de changement multiples utilisant des méthodes exactes (PELT ou SegNeigh) ou approximatives (BinSeg) :

- La méthode **AMOC** (en anglais : Analysis of MOment Change) est une technique utilisée pour localiser un unique point de rupture dans une série d'observations. Elle est aussi connue sous le nom de CUSUM de base. Le point de rupture est alors déterminé comme le point où la somme cumulée atteint son maximum.
- **PELT** (en anglais : Pruned Exact Linear Time) : il s'agit d'une méthode exacte qui cherche à minimiser une fonction de coût basée sur la somme des carrés des écarts entre les données et les prévisions à chaque segment.
- **BinSeg** (en anglais : Binary Segmentation) : il s'agit d'une méthode qui divise la série en sous-segments de taille égale jusqu'à atteindre la partition optimale qui minimise une fonction de coût basée sur les carrés des écarts entre les données et les prévisions à chaque segment.

Le choix de la méthode dépend donc des caractéristiques de la série à analyser et de l'objectif de l'analyse. Il est souvent recommandé de tester plusieurs méthodes (voir Annexe **F**) pour comparer les résultats et choisir celle qui convient le mieux.

Dans notre étude sur la série de précipitations à Mayotte $R := (R_t)_{0 \leq t \leq T}$ **5.1**, la méthode la plus performante pour détecter les changements de régime est la fonction **cpt.meanvar** couplée avec la méthode de **PELT**. Les résultats obtenus sont présentés dans la Table **5.1**, qui donne les dates des ruptures T_i (en semaines) sur $J = [0, T]$, ainsi que les moyennes de précipitations associées (**Fig.5.2**).

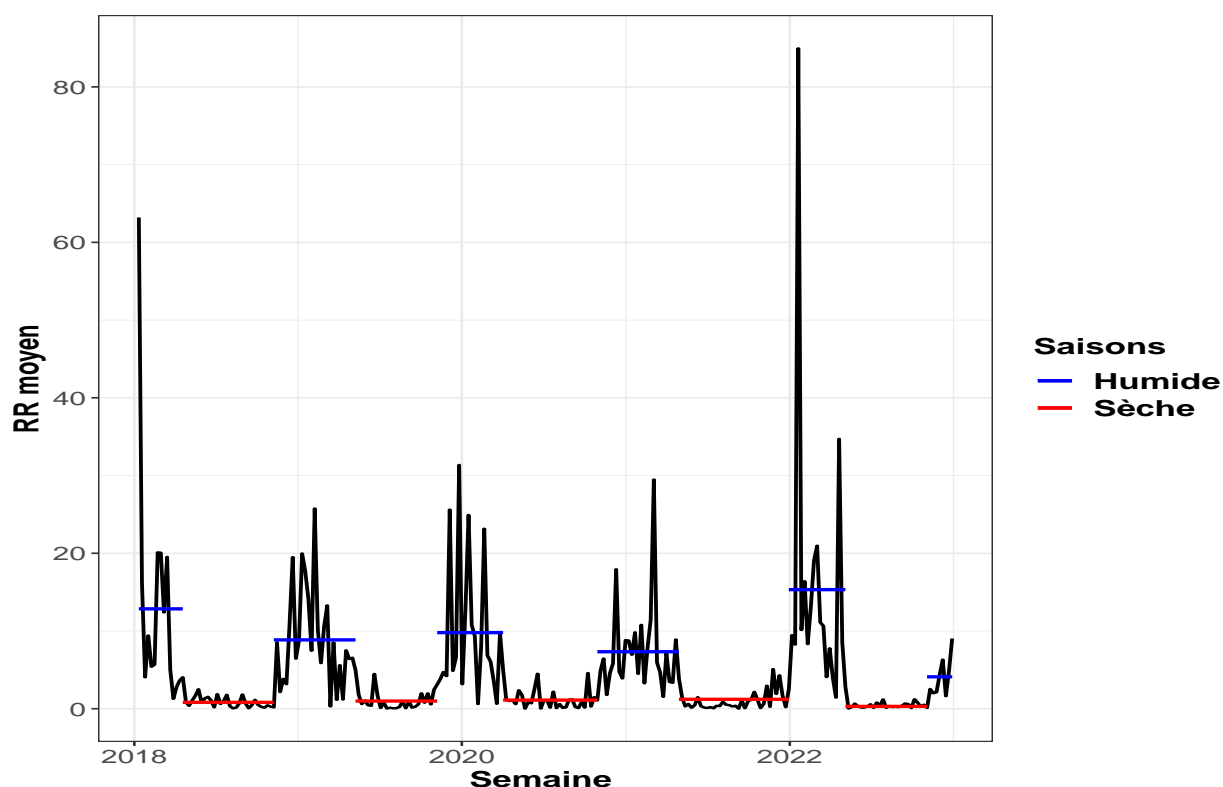


FIGURE 5.2 – Détection des ruptures de données pluviométriques à Mayotte entre 2018 et 2022.

TABLE 5.1 – Les régimes saisonniers à Mayotte et leurs RR moyen associé.

Période par semaines $J_i =]T_i, T_{i+1}]$	RR moyen	Saison
$[0, 15]$	12.85	Humide
$]15, 44]$	0.82	Sèche
$]44, 70]$	8.85	Humide
$]70, 96]$	0.99	Sèche
$]96, 117]$	9.80	Humide
$]117, 147]$	1.11	Sèche
$]147, 173]$	7.34	Humide
$]173, 208]$	1.41	Sèche
$]208, 226]$	15.32	Humide
$]226, 252]$	0.30	Sèche
$]252, 260]$	8.19	Humide

Pour évaluer la significativité des différences entre les moyennes des taux de RR obtenus, nous avons recouru à une analyse de variance (ANOVA) à un facteur. Les résultats de cette ANOVA ont révélé une très faible valeur de p-value ($p < 2.10^{-16}$), indiquant que les moyennes sont statistiquement distinctes. Par la suite, nous avons procédé à des tests post-hoc, afin de cerner précisément quelles paires de groupes présentent des différences significatives entre elles. Parmi les tests post-hoc couramment utilisés, nous avons

employé le test de Tukey. Les résultats de ce test confirment l'existence de différences significatives entre les moyennes des groupes "Humide" et "Sec", comme illustré dans la Figure 5.4.

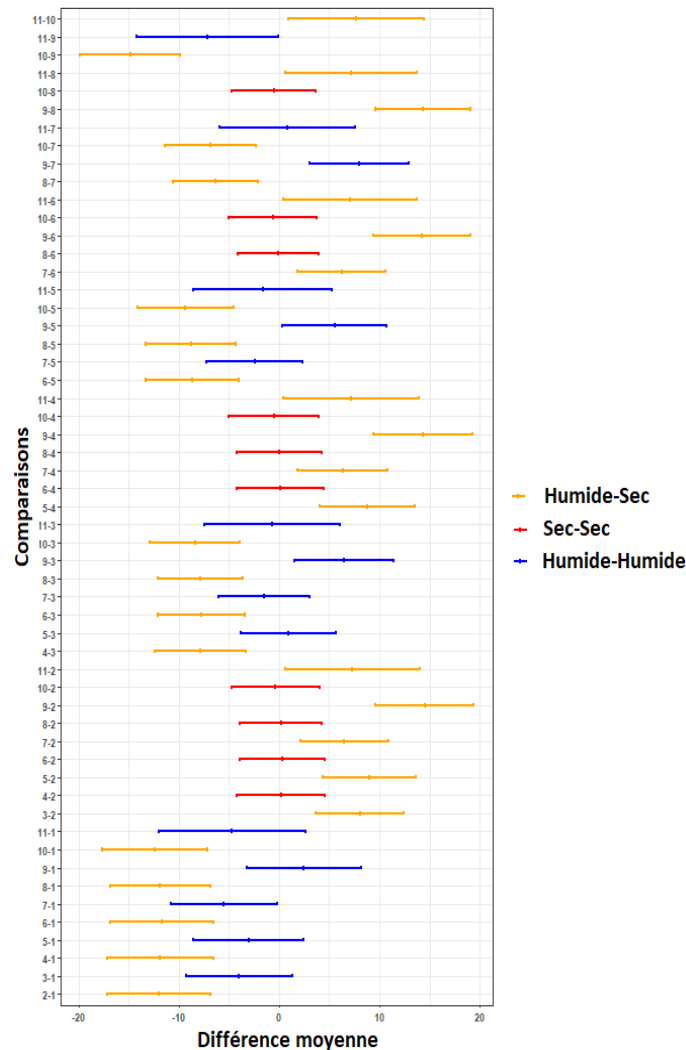


FIGURE 5.3 – Résultats du test Tukey.

5.4 Estimation des paramètres d'un modèle de Markov caché à deux régimes

Après avoir identifié les instants des changements dans les données de pluviométrie à Mayotte et défini les différentes périodes correspondant à chaque régime, nous procédons maintenant à l'extension de notre modèle de Markov caché, tel que présenté dans la section 3.3.2. Cette fois-ci, nous incorporons les données de pluviométrie représentant les deux régimes saisonniers, dans le but d'estimer les trois paramètres du modèle spécifiques à chaque régime (λ , μ et ν). Notre objectif est de quantifier l'impact de cette variable météorologique sur la propagation de la fièvre typhoïde à Mayotte.

Nous distinguons deux régimes saisonniers, le régime humide et le régime sec, que

nous notons respectivement J^H et $J^S = J \setminus J^H$. Nous utilisons la fonction $m(t)$ pour représenter le régime en vigueur à l'instant $t \in J$:

$$m(t) = \begin{cases} 1 & \text{si } t \in J^H, \\ 2 & \text{si } t \in J^S. \end{cases}$$

Afin de réutiliser la même méthodologie d'estimation présentée en détail dans la section **3.3**, nous réécrivons les caractéristiques de notre modèle de Markov caché sous les deux régimes (**Fig.5.4**). Dans ce cas, on note par

$$p_{i,j}^{m(t)} = \mathbb{P}(I_{t+1}^{m(t+1)} = j | I_t^{m(t)} = i), \quad i, j \in \mathbb{N}, t \in J,$$

la probabilité de transition de la chaîne I_t . Plus précisément,

$$p_{i,j}^{m(t)} = \begin{cases} p_{i,j}^{(1)}(\lambda_1, \mu_1, \nu_1) & \text{si } t \in J^H, \\ p_{i,j}^{(2)}(\lambda_2, \mu_2, \nu_2) & \text{si } t \in J^S, \end{cases}$$

où $p_{i,j}^{(1)}(\lambda_1, \mu_1, \nu_1)$ et $p_{i,j}^{(2)}(\lambda_2, \mu_2, \nu_2)$ sont les probabilités de transition pour les régimes humide et sec, respectivement, et sont données par le théorème **3.1**.

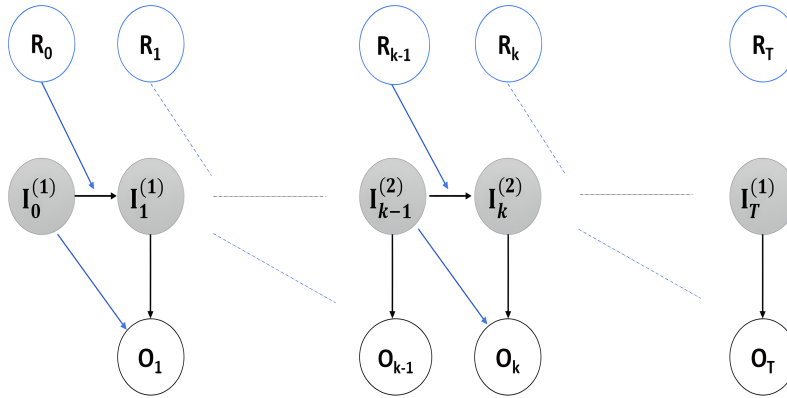


FIGURE 5.4 – **Modèle de Markov caché à deux régimes.**

Les paramètres $M = (Q^{(1)}, \psi^{(1)}, \rho^{(1)}, Q^{(2)}, \psi^{(2)}, \rho^{(2)})$ du HMM (Z_t, O_t, R_t) sont données comme suit :

i) La matrice de transition de la chaîne $(Z_t^{m(t)})_{t \geq 1} = (I_{t-1}^{m(t-1)}, I_t^{m(t)})_{t \geq 1}$:

$$Q_{(i,j),(i',j')}^{m(t)} = p_{i',j'}^{m(t)} \delta_{i'=j},$$

pour $(i, j), (i', j') \in \mathbb{N}^2$.

ii) La probabilité d'émission :

$$\psi_{(i,j)}^{m(t)}(o) = \mathbb{P}(O_t = o | Z_t^{m(t)} = (i, j)) = \frac{p_{i,(j,o)}^{m(t)}}{p_{i,j}^{m(t-1)}},$$

pour $i, j, o \in \mathbb{N}$, où $p_{i,(j,o)} = p_{i,(j,o)}(\Delta t)$ est caractérisée par l'équation **3.2**.

iii) La distribution initiale :

$$\rho_{i,j}^{m(1)} = p_{i,j}^{m(0)} \pi_i^{m(0)},$$

pour $i, j \in \mathbb{N}$, où $\pi_i^{m(0)}$ est la loi initiale de la chaîne I_t sous le régime $m(0)$.

En utilisant ces paramètres, nous pouvons exprimer la loi jointe de (Z, O) par la formule suivante, pour tout $z_{1:T} \in (\mathbb{N}^2)^T$ et $O_{1:T} \in \mathbb{N}^T$:

$$\begin{aligned} \mathbb{P}_M \left(Z_{1:T}^{m(1):m(T)} = z_{1:T}, O_{1:T} = o_{1:T} \right) &= \rho^{m(1)}(z_1) \prod_{t=1}^{T-1} Q^{m(t)}(z_t, z_{t+1}) \prod_{t=1}^T \psi_{z_t}^{m(t)}(o_t) \\ &= \rho^{m(1)}(z_1) \prod_{t \in J^H}^{T-1} Q^{(1)}(z_t, z_{t+1}) \prod_{t \in J^H}^T \psi_{z_t}^{(1)}(o_t) \\ &\quad \prod_{t \in J^S}^{T-1} Q^{(2)}(z_t, z_{t+1}) \prod_{t \in J^S}^T \psi_{z_t}^{(2)}(o_t). \end{aligned}$$

Ainsi, la fonction auxiliaire de l'algorithme EM s'écrit comme suit :

$$\Gamma_{O,R}(M, M') = \Gamma_{O,R}(M^{(1)}, M'^{(1)}) + \Gamma_{O,R}(M^{(2)}, M'^{(2)}),$$

où $M^{(i)}$ pour $i = 1, 2$ représente les paramètres du HMM sous le régime humide et sec, respectivement.

Par conséquent, en utilisant la même technique d'optimisation décrite dans la preuve du théorème **3.3**, nous pouvons déduire que les ré-estimateurs de Baum-Welch adapté, étant donné le modèle $M = (M^{(1)}, M^{(2)})$, les observations $O_{1:T}$ et les régimes J^H et J^S , sont données par :

$$\begin{aligned} Q_{(i,j),(i',j')}^{(1)} &= \frac{\sum_{t \in J^H}^{T-1} \xi_{(i,j),(i',j')}^{m(t)}(t)}{\sum_{t \in J^H}^{T-1} \gamma_{(i,j)}^{m(t)}(t)} \delta_{i'=j}, \\ Q_{(i,j),(i',j')}^{(2)} &= \frac{\sum_{t \in J^S}^{T-1} \xi_{(i,j),(i',j')}^{m(t)}(t)}{\sum_{t \in J^S}^{T-1} \gamma_{(i,j)}^{m(t)}(t)} \delta_{i'=j}, \\ \psi_{(i,j)}^{(1)}(y) &= \frac{\sum_{t \in J^H}^{T-1} \mathbb{1}_{o_t=o} \gamma_{(i,j)}^{m(t)}(t)}{\sum_{t \in J^H}^{T-1} \gamma_{(i,j)}^{m(t)}(t)}, \\ \psi_{(i,j)}^{(2)}(y) &= \frac{\sum_{t \in J^S}^{T-1} \mathbb{1}_{o_t=o} \gamma_{(i,j)}^{m(t)}(t)}{\sum_{t \in J^S}^{T-1} \gamma_{(i,j)}^{m(t)}(t)}, \\ \rho_{i,j}^{m(1)} &= \gamma_{(i,j)}^{m(1)}(1), \\ p_{i,j}^{(1)} &= \frac{\sum_{i' \in \mathbb{N}} Q_{(i',i),(i,j)}^{(1)}}{\sum_{i' \in \mathbb{N}} \mathbb{1}_{Q_{(i',i),(i,j)}^{(1)} \neq 0}}, \\ p_{i,j}^{(2)} &= \frac{\sum_{i' \in \mathbb{N}} Q_{(i',i),(i,j)}^{(2)}}{\sum_{i' \in \mathbb{N}} \mathbb{1}_{Q_{(i',i),(i,j)}^{(2)} \neq 0}}, \end{aligned}$$

où

$$\begin{aligned} \gamma_{(i,j)}^{m(t)} &= \frac{\alpha_{(i,j)}^{m(t)}(t) \beta_{(i,j)}^{m(t)}(t)}{\sum_{i,j \in \mathbb{N}} \alpha_{(i,j)}^{m(t)}(t) \beta_{(i,j)}^{m(t)}(t)}, \\ \xi_{(i,j),(i',j')}^{m(t)}(t) &= \frac{\alpha_{(i,j)}^{m(t)}(t) p_{(i',j')}^{m(t)} \psi_{(i',j')}^{m(t+1)}(o_{t+1}) \beta_{(i',j')}^{m(t+1)}(t+1)}{\sum_{i,j \in \mathbb{N}} \alpha_{(i,j)}^{m(t)}(t) \beta_{(i,j)}^{m(t)}(t)} \delta_{i'=j}, \end{aligned}$$

et les probabilités forward et backward α et β sont définies récursivement pour i, j dans \mathbb{N} et $2 \leq t \leq T$ par

$$\begin{cases} \alpha_{(i,j)}^{m(1)}(1) = p_{i,j}^{m(0)} \pi_i^{m(0)} \psi_{(i,j)}^{m(1)}(o_1), \\ \alpha_{(i,j)}^{m(t)}(t) = \psi_{(i,j)}^{m(t)}(o_t) p_{i,j}^{m(t)} \sum_{i' \in \mathbb{N}} \alpha_{(i',i)}^{m(t-1)}(t-1), \\ \beta_{(i,j)}^{m(T)}(T) = 1, \\ \beta_{(i,j)}^{m(t-1)}(t-1) = \sum_{j' \in \mathbb{N}} p_{j,j'}^{m(t)} \psi_{(j,j')}^{m(t)}(o_t) \beta_{(j,j')}^{m(t)}(t). \end{cases}$$

Enfin, les estimations des paramètres dans chaque régime sont obtenus à partir du théorème **3.2** : dans le régime humide J^H :

$$(\hat{\lambda}_1, \hat{\mu}_1, \hat{\nu}_1) = g(p_{0,0}^{(1)}, p_{0,1}^{(1)}, p_{1,0}^{(1)}),$$

et dans le régime sec J^S :

$$(\hat{\lambda}_2, \hat{\mu}_2, \hat{\nu}_2) = g(p_{0,0}^{(2)}, p_{0,1}^{(2)}, p_{1,0}^{(2)}).$$

Pour la partie numérique, nous avons choisi les mêmes intervalles initiaux pour $\lambda_1^{(0)}$ et $\lambda_2^{(0)}$ dans l'intervalle $[0.05, 0.08]$, et $\mu_1^{(0)}$ et $\mu_2^{(0)}$ dans l'intervalle $[0.11, 0.25]$. Cependant, nous avons spécifié des intervalles différents pour $\nu_1^{(0)}$ dans $[0.015, 0.03]$ et $\nu_2^{(0)}$ dans $[0.0015, 0.003]$, car le paramètre ν (le taux de contamination par l'eau) est directement lié à la pluviométrie et est considéré comme étant le plus représentatif pour quantifier l'impact de cette variable météorologique sur la transmission de la fièvre typhoïde à Mayotte. Les autres paramètres de l'algorithme sont détaillés dans la section **3.4.2**.

Enfin, les estimations des paramètres qui maximisent la vraisemblance des observations pour ce modèle HMM à deux régimes sont les suivantes :

Dans le régime humide :

$$(\hat{\lambda}_1, \hat{\mu}_1, \hat{\nu}_1) = (0.066, 0.21, 0.05),$$

Dans le régime sec :

$$(\hat{\lambda}_2, \hat{\mu}_2, \hat{\nu}_2) = (0.061, 0.1, 0.02).$$

Les résultats numériques montrent que dans le régime humide, le taux de contamination de personne à personne est plus élevé ($\hat{\lambda}_1 = 0.066$) que dans le régime sec ($\hat{\lambda}_2 = 0.061$), ce qui peut être expliqué par le fait que l'humidité favorise la survie des bactéries et donc augmente le risque de contamination. De plus, le taux d'isolement ou de déclaration des cas de typhoïde est plus élevé dans le régime humide ($\hat{\mu}_1 = 0.21$) que dans le régime sec ($\hat{\mu}_2 = 0.1$), ce qui peut être dû à une sensibilisation accrue de la population et/ou une surveillance renforcée pendant la saison des pluies. Enfin, le taux de contamination par l'eau est également plus élevé dans le régime humide ($\hat{\nu}_1 = 0.05$) que dans le régime sec ($\hat{\nu}_2 = 0.02$), ce qui suggère que la pluviométrie a un impact significatif sur la transmission de la typhoïde à Mayotte. Cela est cohérent avec la littérature scientifique [84, 85, 86] et les remarques des médecins de la fièvre typhoïde à Mayotte. En effet, lorsque qu'il pleut, l'eau de pluie peut s'écouler dans les rivières, les lacs, les puits ou les réservoirs d'eau. Si ces sources d'eau ne sont pas correctement traitées ou protégées, elles peuvent être contaminées par des bactéries et autres agents pathogènes provenant des eaux usées, des animaux ou des sols.

De plus, la pluie peut également entraîner des inondations, ce qui peut contaminer les

sources d'eau potable par les eaux usées ou les matières fécales des animaux. Dans les zones où l'assainissement et le traitement de l'eau sont insuffisants, les épisodes de pluie peuvent donc augmenter la contamination de l'eau par les bactéries responsables de la typhoïde, ce qui peut expliquer en partie l'impact de la pluie sur la transmission de cette maladie.

Cependant, il est important de noter que ces résultats ne sont basés que sur un échantillon limité de données et peuvent être améliorés avec des données plus nombreuses.

En conclusion, ces résultats montrent l'importance de prendre en compte les variables météorologiques dans l'étude de la transmission des maladies infectieuses hydriques et soulignent la nécessité de renforcer la surveillance et la prévention de la typhoïde pendant la saison des pluies.

5.5 Conclusion

Au cours de ce chapitre, nous avons étudié l'impact de la pluviométrie et de la saisonnalité sur la transmission de la fièvre typhoïde à Mayotte en se basant sur une méthode de détection de rupture, à savoir CUSUM et une extension de notre HMM **3.3.2**.

Nous avons commencé par une analyse exploratoire des données, en traçant les graphiques de la pluviométrie et des cas de typhoïde au fil du temps. Ensuite, nous avons utilisé la méthode CUSUM pour détecter les changements ou bien les ruptures dans les données de précipitations à Mayotte et identifier les différentes périodes de deux régimes saisonniers, humide et sec.

Nous avons ensuite décrit notre modèle HMM à deux régimes afin d'estimer les trois paramètres λ , μ et ν dans chaque régime et caractériser et comparer la transmission de la maladie par rapport au changement de saison. Les estimations des paramètres pour ces deux régimes ont montré que le taux de contamination par l'eau était plus élevé dans le régime humide que dans le régime sec, suggérant ainsi que la pluviométrie peut augmenter la contamination de l'eau par les bactéries. Cette observation explique également les pics épidémiques enregistrés pendant la période pluvieuse chaque année.

Enfin, nous avons discuté des résultats obtenus et de leur pertinence dans le contexte de la lutte contre la typhoïde. Nous avons souligné l'importance de prendre en compte la pluviométrie et la saisonnalité dans la conception des stratégies de lutte contre la typhoïde, et avons conclu que les deux méthodes de détection de rupture étudiées peuvent être des outils utiles dans ce domaine.

Perspectives

Dans nos perspectives de recherche, nous souhaitons approfondir notre compréhension de l'effet de la variation des niveaux de précipitations sur l'évolution de la fièvre typhoïde. Nous nous pencherons sur l'identification d'un lien potentiel entre les précipitations et l'apparition des pics épidémiques de la maladie. Nos observations (voir **Fig.6.1**) ont mis en évidence que le niveau moyen des précipitations à Mayotte est variable, avec des années présentant des pics épidémiques plus importants, comme observé en 2022 avec des niveaux de précipitations élevés par rapport aux années précédentes.

Nous avons également remarqué qu'après chaque épisode de fortes pluies, il existe un décalage temporel suivi d'une augmentation des cas de fièvre typhoïde. Ce décalage pourrait s'expliquer par plusieurs facteurs potentiels, tels que la période d'incubation et de contagiosité de la maladie, ainsi que le processus de développement et de multiplication des bactéries nécessitant un certain laps de temps. Il est possible que les conditions environnementales créées lors des périodes de fortes pluies favorisent la survie et la propagation de la bactérie *Salmonella typhi*. Ainsi, le délai nécessaire pour que ces conditions environnementales favorables conduisent à une augmentation significative des cas pourrait expliquer le décalage observé. Ces constatations suggèrent donc que d'autres facteurs doivent être pris en compte dans notre analyse statistique.

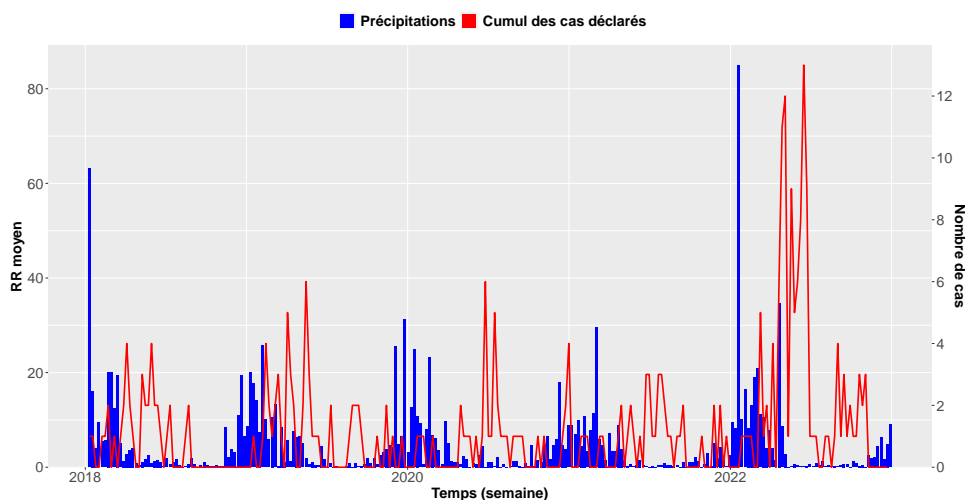


FIGURE 6.1 – Présentation de la moyenne des précipitations et des cumuls de nouveaux cas déclarés à Mayotte de 2018 à 2022.

En outre, il serait intéressant, en fonction des données disponibles dans le futur, d'explorer l'influence d'autres variables liées à l'accès à l'eau sur la concentration des bactéries responsables de la fièvre typhoïde, ainsi que sur le processus de transmission par l'eau contaminée. Parmi ces variables, nous pourrions nous concentrer sur la température, le pH et la conductivité de l'eau, qui sont des indicateurs importants de sa qualité (**Fig.6.2**). La température et le pH de l'eau peuvent jouer un rôle significatif dans la survie et la multiplication des bactéries. Des études antérieures [87, 88] ont montré que des températures plus élevées peuvent favoriser la croissance de *Salmonella typhi*. D'autres études [89, 90] suggèrent que des valeurs de pH spécifiques peuvent favoriser la survie de *Salmonella typhi*.

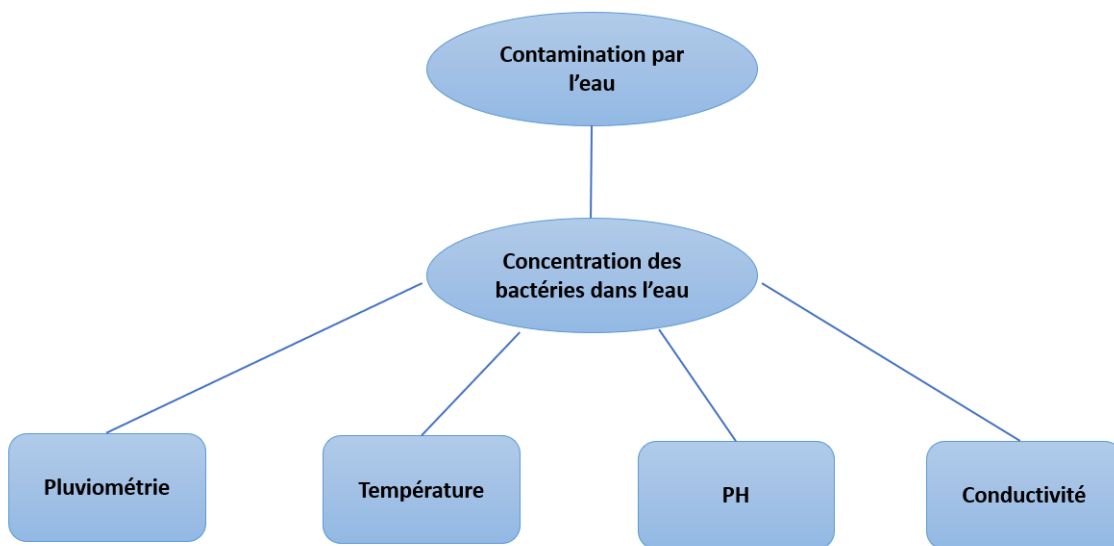


FIGURE 6.2 – Différents facteurs associés à la qualité de l'eau.

Une autre perspective intéressante consisterait à examiner les facteurs socio-économiques et comportementaux qui pourraient influencer la transmission de la fièvre typhoïde. Par exemple, nous pourrions étudier l'impact de la densité de population, des conditions d'assainissement, de l'accès à l'eau potable et des pratiques d'hygiène sur la prévalence de la maladie. Comprendre comment ces facteurs interagissent avec les variables météorologiques pourrait permettre de développer des stratégies de prévention et de contrôle plus efficaces.

Une autre piste de recherche prometteuse consisterait à réaliser une étude théorique sur la stabilité du modèle PDMP présenté dans la section 2.3. Cette étude permettrait de déterminer le comportement stationnaire du modèle, d'identifier les différents régimes possibles et d'étudier les caractéristiques de la population à l'équilibre. Une analyse approfondie de la stabilité nous permettrait de mieux comprendre les dynamiques de la fièvre typhoïde et d'apporter des informations essentielles pour la prévention et le contrôle de la maladie.

En outre, nous pourrions envisager la mise en œuvre de procédures de contrôle stochastique optimal dans différents scénarios épidémiques. Ces approches de contrôle permettraient d'identifier les stratégies les plus efficaces pour réduire la propagation de la

maladie. En prenant en compte les paramètres spécifiques du modèle PDMP et en utilisant des techniques d'optimisation adaptées, nous pourrions concevoir des interventions ciblées visant à minimiser l'impact de la fièvre typhoïde dans la population étudiée.

Enfin, il serait intéressant d'étendre notre étude aux autres maladies infectieuses présentes à Mayotte, en incluant d'autres pathologies présentes sur le territoire étudié, telles que les maladies de la peau, afin d'avoir une vision plus complète de la situation sanitaire. L'intégration de paramètres liés à l'exploitation du Système national des données de santé (SNDS) permettrait une généralisation du modèle à d'autres territoires que Mayotte. Cela permettrait d'élargir l'applicabilité de notre recherche et d'adapter les résultats à des contextes différents, contribuant ainsi à une meilleure compréhension et gestion globale des maladies infectieuses.

B Matrice Jacobienne Dg

Nous calculons la matrice jacobienne de Dg . Tout d'abord, les dérivées partielles des fonctions q , r et u présentées dans la preuve du théorème **3.2** sont données par :

$$\begin{aligned} \frac{\partial q}{\partial p_{0,0}} &= -\frac{p_{0,1} (\ln(p_{0,0}) + 1)}{(p_{0,0} \ln(p_{0,0}))^2} W \left(\frac{\left(\frac{p_{0,0}}{p_{0,1}} + 1\right) \ln(p_{0,0})}{p_{0,1}} \right) \\ &+ \frac{p_{0,1}}{p_{0,0} \ln(p_{0,0})} (\ln(p_{0,0}) + 1) \left(\frac{1}{p_{0,1}} + \frac{1}{p_{0,0} \ln(p_{0,0})} \right) \frac{W \left(\frac{\left(\frac{p_{0,0}}{p_{0,1}} + 1\right) \ln(p_{0,0})}{p_{0,1}} \right)}{W \left(\frac{\left(\frac{p_{0,0}}{p_{0,1}} + 1\right) \ln(p_{0,0})}{p_{0,1}} \right) + 1}. \end{aligned}$$

Puisque

$$W \left(\frac{\left(\frac{p_{0,0}}{p_{0,1}} + 1\right) \ln(p_{0,0})}{p_{0,1}} \right) = q \frac{p_{0,0} \ln(p_{0,0})}{p_{0,1}}, \quad (1)$$

on obtient

$$\begin{aligned} \frac{\partial q}{\partial p_{0,0}} &= \frac{q (\ln(p_{0,0}) + 1)}{qp_{0,0} \ln(p_{0,0}) + p_{0,1}} \left(1 + \frac{p_{0,1}}{p_{0,0} \ln(p_{0,0})} - \frac{qp_{0,0} \ln(p_{0,0}) + p_{0,1}}{p_{0,0} \ln(p_{0,0})} \right) \\ &= \frac{q (\ln(p_{0,0}) + 1)}{qp_{0,0} \ln(p_{0,0}) + p_{0,1}} (1 - q). \end{aligned}$$

Ensuite, on a

$$\begin{aligned} \frac{\partial q}{\partial p_{0,1}} &= \frac{1}{p_{0,0} \ln(p_{0,0})} W \left(\frac{\left(\frac{p_{0,0}}{p_{0,1}} + 1\right) \ln(p_{0,0})}{p_{0,1}} \right) \\ &- \frac{p_{0,1}}{p_{0,0} \ln(p_{0,0})} \left(\frac{p_{0,0} \ln(p_{0,0})}{(p_{0,1})^2} + \frac{1}{p_{0,1}} \right) \frac{W \left(\frac{\left(\frac{p_{0,0}}{p_{0,1}} + 1\right) \ln(p_{0,0})}{p_{0,1}} \right)}{W \left(\frac{\left(\frac{p_{0,0}}{p_{0,1}} + 1\right) \ln(p_{0,0})}{p_{0,1}} \right) + 1}, \end{aligned}$$

et en utilisant l'équation (1) nous obtenons

$$\begin{aligned} \frac{\partial q}{\partial p_{0,1}} &= \frac{q}{p_{0,1}} - \frac{q}{qp_{0,0} \ln(p_{0,0}) + p_{0,1}} \left(\frac{p_{0,0} \ln(p_{0,0})}{p_{0,1}} + 1 \right) \\ &= \frac{q}{p_{0,1} \left(q + \frac{p_{0,1}}{p_{0,0} \ln(p_{0,0})} \right)} (q - 1), \end{aligned}$$

et enfin on a

$$\frac{\partial q}{\partial p_{1,0}} = 0.$$

Les expressions des dérivées partielles de r sont données comme suit :

$$\begin{aligned} \frac{\partial r}{\partial p_{0,0}} &= \frac{1}{p_{0,0} \ln(q)} - \frac{\ln(p_{0,0})}{(\ln(q))^2 q} \frac{\partial q}{\partial p_{0,0}} \\ &= \frac{r}{p_{0,0} \ln(p_{0,0})} - r \frac{(\ln(p_{0,0}) + 1)(1 - q)}{\ln(q)(qp_{0,0} \ln(p_{0,0}) + p_{0,1})}, \\ \frac{\partial r}{\partial p_{0,1}} &= \frac{\ln(p_{0,0})}{(\ln(q))^2 q} \frac{\partial q}{\partial p_{0,1}} \\ &= r \frac{(q - 1)}{p_{0,1} \ln(q) \left(q + \frac{p_{0,1}}{p_{0,0} \ln(p_{0,0})} \right)}, \\ \frac{\partial r}{\partial p_{1,0}} &= 0. \end{aligned}$$

Les dérivées partielles de u s'expriment respectivement comme suit

$$\begin{aligned} \frac{\partial u}{\partial p_{0,0}} &= \frac{p_{1,0}}{p_{0,0}} = \frac{1 - u}{p_{0,0}}, \\ \frac{\partial u}{\partial p_{0,1}} &= 0, \\ \frac{\partial u}{\partial p_{1,0}} &= -\frac{1}{p_{0,0}}. \end{aligned}$$

En définissant les fonctions suivantes

$$\begin{aligned} \alpha &= \frac{q(\ln(p_{0,0}) + 1)(q - 1)}{\Delta t (q p_{0,0} \ln(p_{0,0}) + p_{0,1})(q - u)}, \\ \beta &= \frac{1 - q}{\Delta t p_{0,1} \left(q + \frac{p_{0,1}}{p_{0,0} \ln(p_{0,0})} \right) (q - u)}, \\ \eta &= \frac{1 - u}{\Delta t p_{0,0} u (q - u)}, \end{aligned}$$

on peut réécrire de façon concise les dérivées partielles de λ , μ et ν (de la fonction g) comme suit :

$$\begin{aligned}
\frac{\partial \lambda}{\partial p_{0,0}} &= \eta(q-1) - \frac{\alpha}{q}(1-q) + \lambda \frac{\alpha \Delta t (q-u)}{1-q} + \lambda \alpha \Delta t + \lambda \frac{1-u}{p_{0,0}(q-u)} \\
&= \lambda(1-u) \left(\frac{\alpha \Delta t}{1-q} + \frac{1}{p_{0,0}(q-u)} \right) - (1-q) \left(\frac{\alpha}{q} + \eta \right), \\
\frac{\partial \lambda}{\partial p_{0,1}} &= \lambda \frac{\beta \Delta t q (1-u)}{1-q} - \beta(1-q), \\
\frac{\partial \lambda}{\partial p_{1,0}} &= -\frac{\lambda}{p_{0,0}(q-u)} + \eta \frac{1-q}{1-u}, \\
\frac{\partial \mu}{\partial p_{0,0}} &= -\eta(1-u) - \frac{\alpha}{q}(1-u) - \frac{\mu}{p_{0,0}} + \mu \alpha \Delta t + \frac{\mu(1-u)}{p_{0,0}(q-u)} \\
&= \mu \left(\alpha \Delta t + \frac{1-q}{p_{0,0}(q-u)} \right) - (1-u) \left(\frac{\alpha}{q} + \eta \right), \\
\frac{\partial \mu}{\partial p_{1,0}} &= \mu \beta \Delta t q - \beta(1-u), \\
\frac{\partial \mu}{\partial p_{0,1}} &= \mu \frac{1-q}{p_{0,0}(u-1)(q-u)} + \eta, \\
\frac{\partial \nu}{\partial p_{0,0}} &= \frac{\nu}{p_{0,0} \ln(p_{0,0})} + \nu \frac{\alpha \Delta t (q-u)}{q \ln(q)} + \nu(1-u) \left(\frac{\alpha \Delta t}{1-q} + \frac{1}{p_{0,0}(q-u)} \right) \\
&\quad - r(1-q) \left(\frac{\alpha}{q} + \eta \right) \\
&= \nu \left(\frac{1}{p_{0,0} \ln(p_{0,0})} + \frac{\alpha \Delta t (q-u)}{q \ln(q)} + (1-u) \left(\frac{\alpha \Delta t}{1-q} + \frac{1}{p_{0,0}(q-u)} \right) \right) \\
&\quad - r(1-q) \left(\frac{\alpha}{q} + \eta \right), \\
\frac{\partial \nu}{\partial p_{0,1}} &= \nu \beta \Delta t \left(\frac{q(1-u)}{1-q} - \frac{q-u}{\ln(q)} \right) - r \beta (1-q), \\
\frac{\partial \nu}{\partial p_{1,0}} &= -\frac{\nu}{p_{0,0}(q-u)} + r \eta \frac{1-q}{1-u}.
\end{aligned}$$

C Log-vraisemblance $l(O|M^n)$

TABLE 1 – La log-vraisemblance $l(O|M^n)$ pour différentes valeurs initiales des paramètres.

$(\lambda^{(0)}, \mu^{(0)}, \nu^{(0)})$	$l(Y M^n)$
(0.0500, 0.1100, 0.0150)	-394.1107
(0.0516, 0.1174, 0.0157)	-394.4322
(0.0532, 0.1247, 0.0166)	-394.0771
(0.0546, 0.1321, 0.0174)	-393.7732
(0.0563, 0.1395, 0.0182)	-394.1762
(0.0579, 0.1468, 0.0189)	-394.5896
(0.0595, 0.1542, 0.0197)	-394.9808
(0.0611, 0.1616, 0.0205)	-395.3653
(0.0626, 0.1689, 0.0213)	-395.7492
(0.0642, 0.1763, 0.0221)	-396.1328
(0.0658, 0.1837, 0.0229)	-396.5154
(0.0674, 0.1911, 0.0237)	-397.0610
(0.0689, 0.1984, 0.0245)	-397.4057
(0.0705, 0.2058, 0.0253)	-396.1538
(0.0721, 0.2132, 0.0261)	-395.9935
(0.0737, 0.2205, 0.0268)	-396.1987
(0.0753, 0.2279, 0.0276)	-396.4099
(0.0768, 0.2353, 0.0284)	-396.6318
(0.0784, 0.2426, 0.0292)	-396.8629
(0.0800, 0.2500, 0.0300)	-397.1020

D Fonction génératrice des moments du processus $(E_t, I_t)_{t \geq 0}$

La fonction génératrice des moments du processus $(E_t, I_t)_{t \geq 0}$ est donnée par : pour tout $(r, s, t) \in \mathbb{R}^2 \times [0, +\infty[$

$$G_{(e,i)}(r, s, t) = \sum_{k,l=0}^{+\infty} r^k s^l p_{(e,i),(k,l)}(t).$$

En utilisant l'équation de Kolmogorov progressive (4.2), on trouve :

$$\begin{aligned} \frac{\partial G_{(e,i)}(r, s, t)}{\partial t} &= \sum_{k,l=0}^{\infty} r^k s^l \sum_{(e',i')} p_{(e,i),(e',i')} Q_{(e',i'),(k,l)} \\ &= \sum_{k,l=0} r^{k+1} s^l (\lambda l + \nu) p_{(e,i),(k,l)} + \sum_{k=1,l=0} r^{k-1} s^{l+1} \alpha k p_{(e,i),(k,l)} \\ &\quad + \sum_{k=0,l=1} r^k s^{l-1} \mu l p_{(e,i),(k,l)} - \sum_{k,l=0} r^k s^l ((\lambda + \mu)l + \alpha k + \nu) p_{(e,i),(k,l)} \\ &= \lambda s r \sum_{k=0,l=1} r^k l s^{l-1} p_{(e,i),(k,l)} + \nu r \sum_{k,l=0} r^k s^l p_{(e,i),(k,l)} + \alpha s \sum_{k=1,l=0} k r^{k-1} s^l p_{(e,i),(k,l)} \\ &\quad + \mu \sum_{k=0,l=1} r^k l s^{l-1} p_{(e,i),(k,l)} - (\lambda + \mu) s \sum_{k=0,l=1} r^k l s^{l-1} p_{(e,i),(k,l)} \\ &\quad - \alpha r \sum_{k=1,l=0} k r^{k-1} s^l p_{(e,i),(k,l)} - \nu \sum_{k,l=0} r^k s^l p_{(e,i),(k,l)} \\ &= \alpha (s - r) \sum_{k=1,l=0} k r^{k-1} s^l p_{(e,i),(k,l)} + (\lambda s (r - 1) - \mu (s - 1)) \sum_{k=0,l=1} r^k l s^{l-1} p_{(e,i),(k,l)} \\ &\quad + \nu (r - 1) \sum_{k,l=0} r^k s^l p_{(e,i),(k,l)}. \end{aligned}$$

Ainsi, la fonction génératrice des moments satisfait l'équation différentielle partielle suivante :

$$\begin{aligned} \frac{\partial G_{(e,i)}(r, s, t)}{\partial t} &= \alpha (s - r) \frac{\partial G}{\partial r} + (\lambda s (r - 1) - \mu (s - 1)) \frac{\partial G}{\partial s} \\ &\quad + \nu (r - 1) G_{(e,i)}(r, s, t), \end{aligned} \quad (2)$$

avec la condition initiale $G_{(e,i)}(r, s, 0) = r^e s^i$.

Nous utilisons ensuite la méthode des caractéristiques pour résoudre cette équation.

Étape 1 : Les caractéristiques de l'équation des dérivées partielles **2** sont données par :

$$\begin{aligned} \frac{dr}{dt} &= -(\lambda s (r - 1) - \mu (s - 1)), \\ \frac{ds}{dt} &= -\alpha (s - r). \end{aligned}$$

En intégrant, on obtient,

$$\begin{aligned} k &= (s (\lambda + \mu - \lambda r) - \mu) e^{-(\lambda + \mu - \lambda r)t}, \\ k' &= \alpha (r - s) e^{-\alpha t}. \end{aligned}$$

Étape 2 : Tout au long des caractéristiques, $G_{(e,i)}$ satisfait

$$\frac{dG_{(e,i)}(r, s, t)}{dt} = \nu (1 - r) G_{(e,i)}.$$

L'intégration par rapport à s donne

$$\begin{aligned} G_{(e,i)}(r, s, t) &= C(k, k')e^{(1-r)\nu t} \\ &= C((s(\lambda + \mu - \lambda r) - \mu), \alpha(r - s)e^{-\alpha t})e^{(1-r)\nu t}. \end{aligned} \quad (3)$$

Nous devons maintenant utiliser la condition initiale $G_{(e,i)}(r, s, 0) = r^e s^i$ pour déterminer la fonction C .

$$r^e s^i = C((s(\lambda + \mu - \lambda r) - \mu), \alpha(r - s)).$$

Posons

$$\begin{cases} w &= s(\lambda + \mu - \lambda r) - \mu, \\ v &= \alpha(r - s). \end{cases}$$

Ce système est identifiable si et seulement si

$$r + s \neq \frac{\lambda + \mu}{\lambda} \quad \text{et} \quad s < \frac{\mu^2}{4\lambda(\lambda + \mu)}.$$

Ainsi, pour $b \in \mathbb{R}$, tel que $0 \leq r \leq 1, 0 \leq s \leq b < \frac{\mu^2}{4\lambda(\lambda + \mu)}$ et $r + s \neq \frac{\lambda + \mu}{\lambda}$, on a

$$\lambda s^2 - s(\lambda + \mu - \frac{\lambda v}{\alpha}) + (\mu + w) = 0,$$

ce qui implique,

$$\begin{aligned} s &= \frac{(\lambda + \mu - \frac{\lambda v}{\alpha}) - \sqrt{\Delta}}{2\lambda}, \\ r &= \frac{(\lambda + \mu + \frac{\lambda v}{\alpha}) - \sqrt{\Delta}}{2\lambda}, \end{aligned}$$

avec

$$\Delta = (\lambda + \mu - \frac{\lambda v}{\alpha})^2 - 4\lambda(\mu + w).$$

Par conséquent,

$$C(w, v) = \left(\frac{(\lambda + \mu + \frac{\lambda v}{\alpha}) - \sqrt{\Delta}}{2\lambda} \right)^e \left(\frac{(\lambda + \mu - \frac{\lambda v}{\alpha}) - \sqrt{\Delta}}{2\lambda} \right)^i. \quad (4)$$

En combinant les deux équations **3** et **4**, on trouve l'expression finale de $G_{(e,i)}(r, s, t)$,

$$\begin{aligned} G_{(e,i)}(r, s, t) &= \left(\frac{(\lambda + \mu + \lambda(r - s)e^{-\alpha t}) - \sqrt{\Delta(r, s, t)}}{2\lambda} \right)^e \left(\frac{(\lambda + \mu - \lambda(r - s)e^{-\alpha t}) - \sqrt{\Delta(r, s, t)}}{2\lambda} \right)^i \\ &\quad \times e^{(1-r)\nu t}, \end{aligned}$$

où

$$\Delta(r, s, t) = (\lambda + \mu - \lambda(r - s)e^{-\alpha t})^2 - 4\lambda(\mu + (s(\lambda + \mu - \lambda r) - \mu)e^{-(\lambda + \mu - \lambda r)t}).$$

E Séries chronologiques

Nous présentons ici différents modèles classiques de séries chronologiques utilisés pour modéliser de manière paramétrique les précipitations à Mayotte.

- **Modèle SARIMA :**

Le modèle SARIMA (*Seasonal Autoregressive Integrated Moving Average*) est une méthode statistique utilisée pour analyser et prévoir les séries chronologiques présentant des tendances saisonnières. Il combine les concepts de l'autorégression (AR), de la moyenne mobile (MA) et de l'intégration (I) pour capturer à la fois les variations saisonnières et les relations entre les valeurs observées successives. Le modèle SARIMA prend en compte les effets saisonniers, les tendances à court terme et les variations aléatoires pour fournir des prévisions précises sur les données temporelles.

Pour l'identification du modèle, nous utilisons l'approche de Box-Jenkins en suivant les étapes suivantes :

1. *Identification de la saisonnalité* : L'étape initiale consiste à examiner les données et à identifier la période de saisonnalité. Cela peut être effectué en analysant les graphiques de la série temporelle et en recherchant des motifs et des régularités saisonnières.

2. *Stationnarité* : Les modèles SARIMA nécessitent que la série temporelle soit stationnaire, c'est-à-dire que ses propriétés statistiques ne varient pas avec le temps. Si la série n'est pas stationnaire, des transformations telles que la différenciation peuvent être appliquées pour obtenir une série stationnaire.

3. *Identification des ordres AR, MA et saisonnier* : À cette étape, les ordres autoregressifs (AR) et de moyenne mobile (MA) sont identifiés en utilisant l'autocorrélation (ACF) et l'autocorrélation partielle (PACF). Ces graphiques aident à déterminer les valeurs des retards significatifs qui contribuent aux modèles AR et MA.

4. *Estimation du modèle* : Une fois que les ordres AR, MA et saisonnier ont été identifiés, le modèle SARIMA est estimé à l'aide de méthodes telles que la méthode des moindres carrés ou la méthode de maximisation de la vraisemblance. L'estimation fournit les coefficients du modèle qui captureront les caractéristiques de la série temporelle.

5. *Diagnostic du modèle* : Après l'estimation, il est important de diagnostiquer le modèle pour vérifier s'il satisfait aux hypothèses de la stationnarité résiduelle et de l'absence d'autocorrélation. Les graphiques des résidus, l'autocorrélation des résidus et les tests statistiques aident à évaluer la qualité du modèle.

6. *Réévaluation et itérations* : Si le modèle ne satisfait pas aux critères de diagnostic, il est nécessaire de réévaluer les ordres et d'ajuster le modèle. Ce processus d'itération est effectué jusqu'à ce qu'un modèle satisfaisant soit obtenu.

En suivant ces étapes, l'approche de Box-Jenkins permet d'identifier et d'estimer un modèle SARIMA approprié pour modéliser et prévoir les séries chronologiques avec des composantes saisonnières.

En pratique, nous utilisons le package **fpp3**, développé en tandem avec le manuel de Hyndman et Athanasopoulos [91]. Ce package permet l'installation et le chargement automatique d'autres packages essentiels pour la visualisation et l'analyse des séries temporelles.

Pour visualiser simultanément l'autocorrélation (ACF), l'autocorrélation partielle (PACF) et la série chronologique originale, nous utilisons la fonction `tsdisplay`.

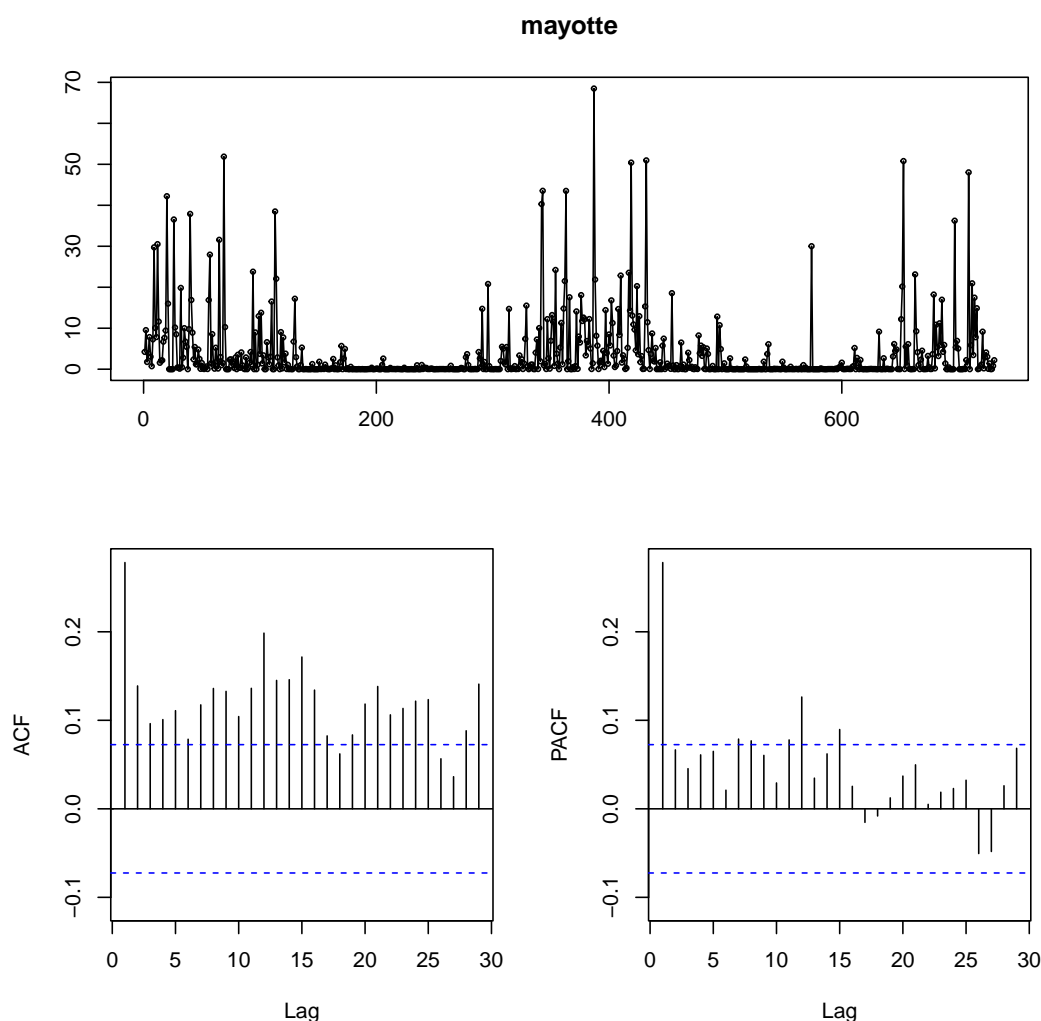


FIGURE 3 – Présentation de la série hebdomadaire des précipitations et de ses graphiques ACF et PACF.

Les lignes bleues tracées représentent les limites de l'intervalle de confiance à 95%. Si une barre de corrélation dépasse ces limites, cela signifie que la corrélation est statistiquement significative, c'est-à-dire qu'elle est significativement différente de zéro.

En observant la présence d'une tendance et d'une composante saisonnière dans la série, on peut conclure qu'elle est non stationnaire. Pour éliminer la tendance, une différenciation d'ordre 1 ($d = 1$) est nécessaire, et pour supprimer la saisonnalité, une différenciation saisonnière d'ordre 1 ($D = 1$) est recommandée. Ainsi, il est approprié de transformer la série initiale en effectuant une double différenciation afin d'obtenir une série stationnaire. En examinant l'autocorrélation (ACF) et l'autocorrélation partielle (PACF), nous pouvons identifier les ordres des termes non saisonniers (p, q) et des termes saisonniers (P, Q) du modèle.

Modèle : $SARIMA(1,1,1)(5,1,0)[3]$

L'utilisation de la fonction "Arima" du package "forecast" permet d'estimer les coefficients du modèle SARIMA ainsi que la variance résiduelle, notée σ_ϵ , en utilisant la méthode du maximum de vraisemblance.

Pour évaluer la significativité des coefficients du modèle, nous vérifions les p-values des tests de Student.

```

z test of coefficients:

      Estimate Std. Error   z value Pr(>|z|)
ar1    0.176758   0.036621    4.8267 1.388e-06 ***
ma1   -1.000000   0.003943  -253.6124 < 2.2e-16 ***
sar1  -0.894460   0.037008  -24.1695 < 2.2e-16 ***
sar2  -0.782948   0.048535  -16.1317 < 2.2e-16 ***
sar3  -0.587172   0.052275  -11.2325 < 2.2e-16 ***
sar4  -0.318645   0.048743   -6.5373 6.264e-11 ***
sar5  -0.097489   0.037293   -2.6141 0.008946 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      2.5 %      97.5 %
ar1    0.1049825  0.24853265
ma1   -1.0077278 -0.99227141
sar1  -0.9669938 -0.82192611
sar2  -0.8780743 -0.68782175
sar3  -0.6896277 -0.48471559
sar4  -0.4141790 -0.22311161
sar5  -0.1705825 -0.02439577

```

FIGURE 4 – Estimation des coefficients du modèle et leurs intervalles de confiance.

On remarque que les coefficients estimés sont significatifs au seuil 5% (p -value < 0.05).

Enfin, nous validons la normalité des résidus en utilisant le test de Blancher des résidus (Test de Portmanteau).

Le graphique des p-values du test de Ljung et Box **Fig.5** (graphique inférieur) montre que l'hypothèse de blancheur des résidus est acceptée, car les p-values associées aux différents décalages ne sont pas statistiquement significatives (p -value > 0.05). Cela confirme que les résidus ne présentent pas de corrélation sérielle significative.

L'autocorrélogramme indique que les résidus présentent une forte corrélation pour un décalage de $h = 0$, mais une faible corrélation avec les décalages ultérieurs ($h > 0$). Cela suggère qu'il ne reste pas de structure de corrélation significative dans les résidus après avoir tenu compte du modèle.

Le graphique **6** montre clairement que l'ajustement du modèle proposé est insuffisant pour capturer la structure temporelle des données de manière satisfaisante.

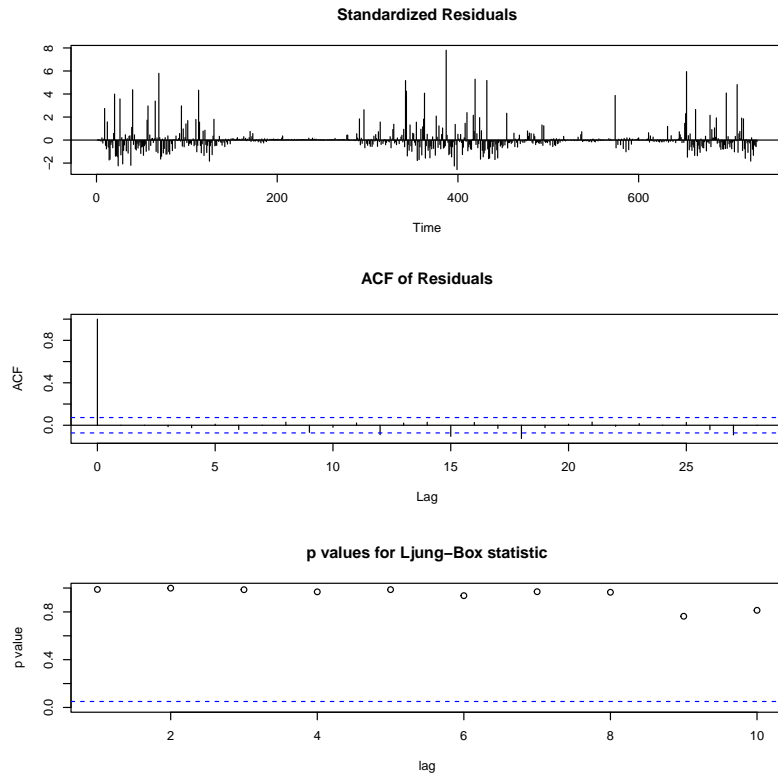


FIGURE 5 – Analyse des résidus standardisés.

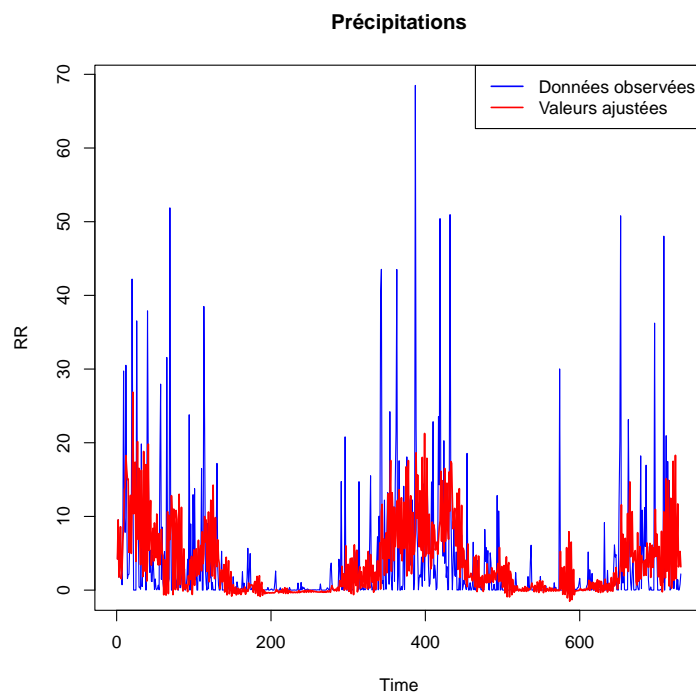


FIGURE 6 – série brute vs série ajustée.

Modèle de Spline cubique de lissage :

C'est une méthode utilisée pour ajuster une courbe ou une fonction à un ensemble de données. Il s'agit d'une approche non paramétrique qui permet de modéliser des relations complexes entre les variables sans imposer de structure spécifique.

Le modèle de spline cubique de lissage divise la courbe en segments plus petits et ajuste un polynôme cubique à chaque segment. Les polynômes cubiques sont choisis de manière à ce que la courbe résultante soit lisse et régulière. Le modèle utilise des noeuds, qui sont des points où les polynômes cubiques se rejoignent, pour contrôler la flexibilité de l'ajustement. Plus il y a de noeuds, plus le modèle est flexible et capable de capturer des variations complexes dans les données. En revanche, trop de noeuds peuvent entraîner un surajustement.

L'estimation des paramètres dans le modèle de spline cubique de lissage est réalisée en minimisant une fonction d'erreur pénalisée, généralement la somme des carrés des résidus. Cette approche permet de trouver les valeurs optimales des coefficients des polynômes cubiques pour obtenir la meilleure correspondance entre le modèle et les données.

$$S(f) = \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int f''(x)^2 dx$$

Le choix de paramètre λ est effectué par la méthode de Validation croisée généralisée.

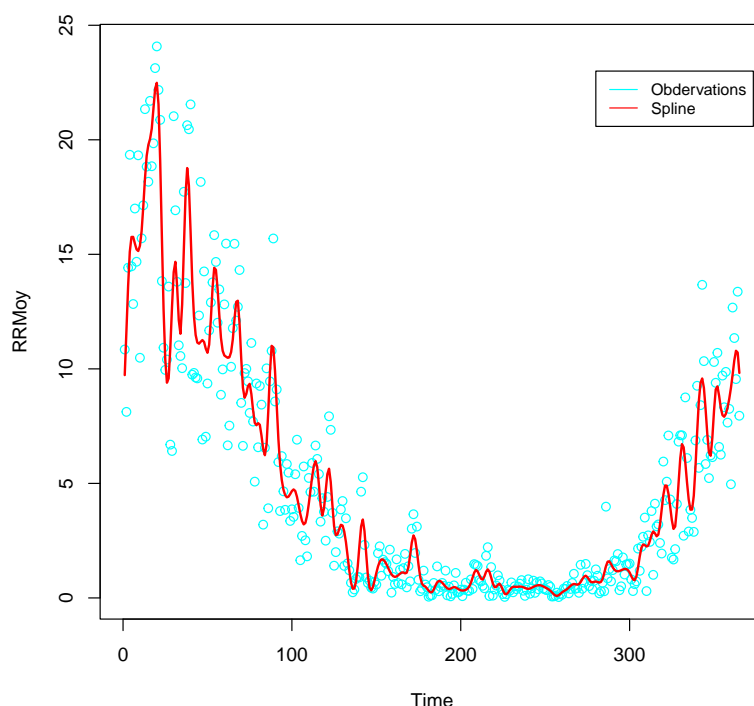


FIGURE 7 – Ajustement du modèle de Spline cubique sur les données de précipitations.

Nous pouvons clairement constater que la courbe lissée ne parvient pas à capturer les variations brusques présentes dans les données, cela indique que le modèle ne correspond pas adéquatement aux données.

F Détection de rupture

Dans cette section de l'annexe, nous présentons les différentes méthodes et fonctions utilisées pour détecter les ruptures dans les données de précipitations à Mayotte.

Détection de ruptures en moyenne :

Nous utilisons la fonction `cpt.mean` du package `changepoint` avec les méthodes suivantes :

1. *PELT* : C'est une méthode de détection de rupture qui offre une approche exacte pour identifier les points de rupture dans une série de données. Elle est robuste aux valeurs aberrantes et peut s'adapter à différentes structures de données.

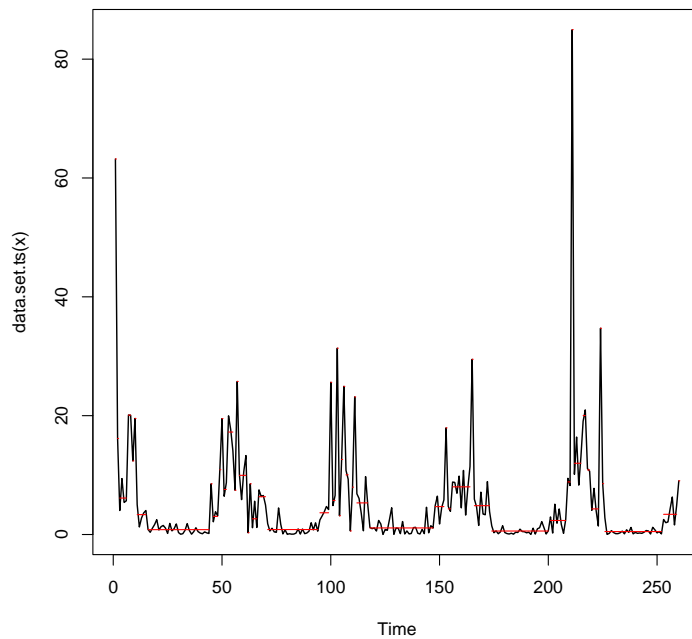


FIGURE 8 – Détection de ruptures en moyenne par la méthode PELT.

2. *BinSeg* : Elle utilise une approche itérative pour trouver la segmentation optimale qui maximise la similarité des valeurs à l'intérieur des segments et minimise la différence entre les segments adjacents. BinSeg est connue pour sa simplicité et sa rapidité de calcul.

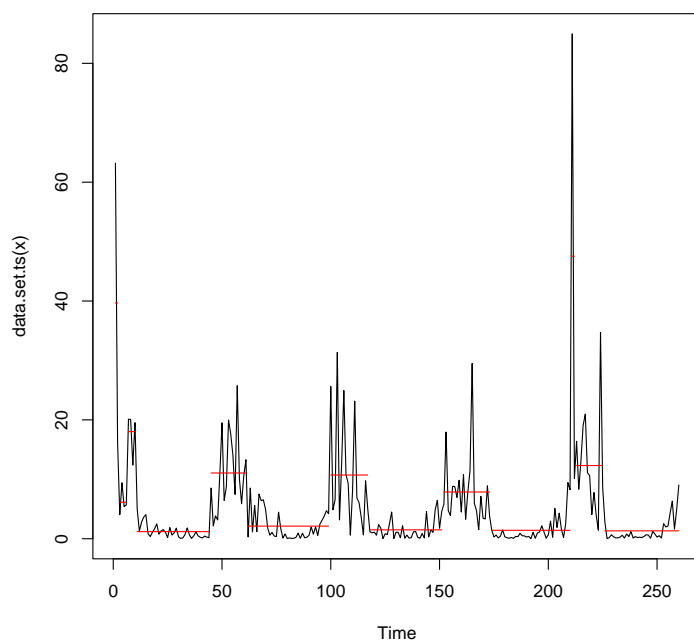


FIGURE 9 – Détection de ruptures en moyenne par la méthode BinSeg.

Détection de ruptures en moyenne-variance :

Dans ce cas, nous utilisons la fonction `cpt.meanvar` du package `changepoint`.

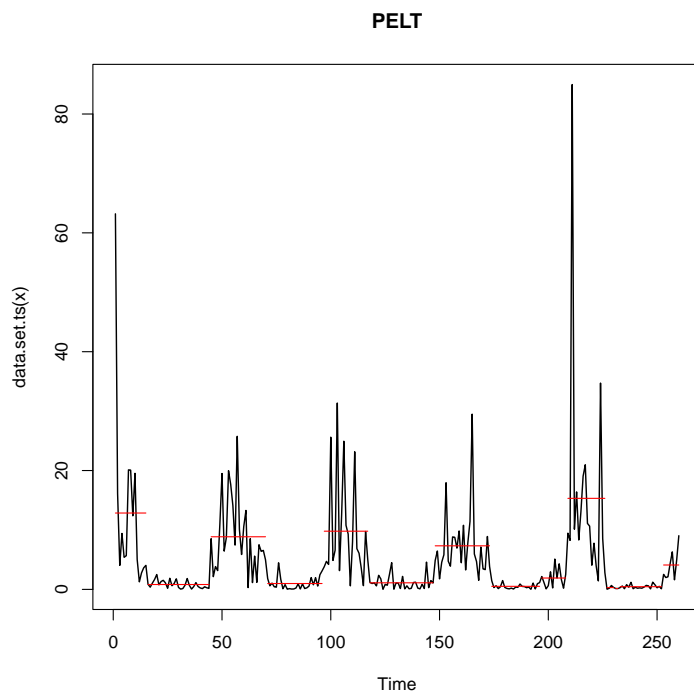


FIGURE 10 – Détection de ruptures en moyenne-variance par la méthode PELT.

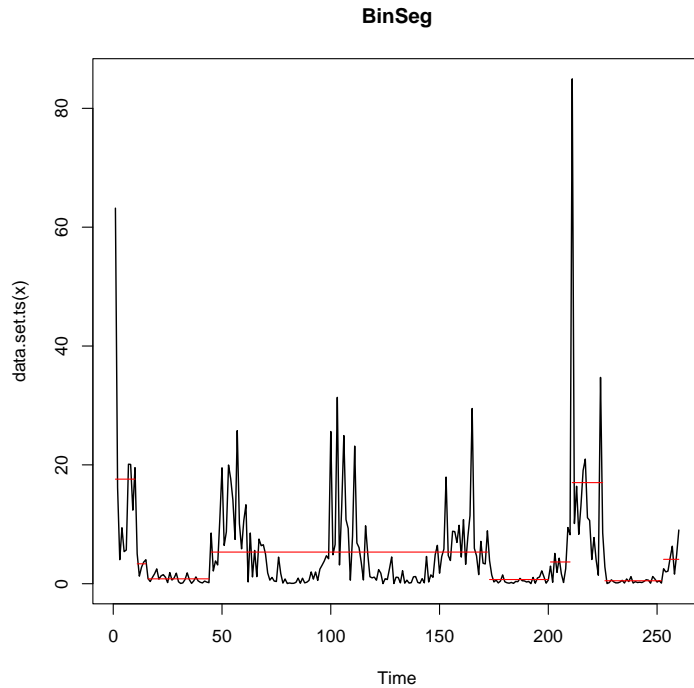


FIGURE 11 – Détection de ruptures en moyenne-variance par la méthode BinSeg.

Comme prévu, les résultats obtenus à partir des fonctions mentionnées ci-dessus varient considérablement en fonction des paramètres choisis.

Pour sélectionner l'algorithme de segmentation et les paramètres associés, nous nous baserons sur deux critères :

1. *Critères théoriques* : Nous évaluerons quel modèle est le plus approprié pour nos données en prenant en compte des critères tels que la constance de la variance, la normalité des résidus et la capacité à identifier des segments de longueurs variables, entre autres.
2. *Critères empiriques* : Nous choisirons la méthode qui produit les résultats les plus exploitables sur des données réelles.

Bibliographie

- [1] D. Bernoulli, “Essai d’une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l’inoculation pour la prévenir,” *Histoire de l’Acad., Roy. Sci.(Paris) avec Mem*, pp. 1–45, 1760.
- [2] W. O. Kermack and A. G. McKendrick, “A contribution to the mathematical theory of epidemics,” *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, vol. 115, no. 772, pp. 700–721, 1927.
- [3] D. G. Kendall, “Deterministic and stochastic epidemics in closed populations,” in *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, vol. 4, pp. 149–165, University of California Press Berkeley, 1956.
- [4] R. Anderson, “May rm infectious diseases of humans,” *Dynamics and Control*, vol. 281992, 1991.
- [5] N. T. Bailey *et al.*, *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.
- [6] F. Brauer, C. Castillo-Chavez, and C. Castillo-Chavez, *Mathematical models in population biology and epidemiology*, vol. 2. Springer, 2012.
- [7] <https://fr.wikipedia.org/wiki/Mayotte>.
- [8] https://fr.wikipedia.org/wiki/D%C3%A9mographie_de_Mayotte#cite_note-diploweb-7.
- [9] <https://www.insee.fr/fr/statistiques/3713016>.
- [10] <https://www.mayotte.ars.sante.fr/les-bornes-fontaines-monetiques>.
- [11] “Fièvre typhoïde à mayotte. point épidémiologique,” *Point de situation au 18/08/2022*.
- [12] R. M. Anderson and R. M. May, *Infectious diseases of humans : dynamics and control*. Oxford university press, 1991.
- [13] M. H. Davis, “Piecewise-deterministic markov processes : A general class of non-diffusion stochastic models,” *Journal of the Royal Statistical Society : Series B (Methodological)*, vol. 46, no. 3, pp. 353–376, 1984.
- [14] K. Pakdaman, M. Thieullen, and G. Wainrib, “Fluid limit theorems for stochastic hybrid systems with application to neuron models,” *Advances in Applied Probability*, vol. 42, no. 3, pp. 761–794, 2010.
- [15] A. Crudu, A. Debussche, A. Muller, and O. Radulescu, “Convergence of stochastic gene networks to hybrid piecewise deterministic processes,” 2012.

- [16] M. Costa, “A piecewise deterministic model for a prey-predator community,” 2016.
- [17] F. Dufour, B. de Saporta, and H. Zhang, *Numerical methods for simulation and optimization of piecewise deterministic Markov processes : application to reliability*. John Wiley & Sons, 2015.
- [18] S. Lessard, *Processus stochastiques*. ellipses, 2014.
- [19] B. D. Saporta, *Modélisation en écologie, poly de cours, Université de Montpellier*,. 2019-2020.
- [20] L. R. Rabiner, “A tutorial on hidden markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [21] K. Hallouli, L. Likforman-Sulem, and M. Sigelle, “Reconnaissance de caractères manuscrits par réseaux bayésiens dynamiques,” in *Conférence Internationale Franco-phone sur l’Ecrit et le Document (CIFED 04)*, 2004.
- [22] R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison, *Biological sequence analysis : probabilistic models of proteins and nucleic acids*. Cambridge university press, 1998.
- [23] R. Bhar and S. Hamori, *Hidden Markov models : Applications to financial economics*, vol. 40. Springer Science & Business Media, 2004.
- [24] P. A. Devijver, “Baum’s forward-backward algorithm revisited,” *Pattern Recognition Letters*, vol. 3, no. 6, pp. 369–373, 1985.
- [25] L. E. Baum and T. Petrie, “Statistical inference for probabilistic functions of finite state markov chains,” *The annals of mathematical statistics*, vol. 37, no. 6, pp. 1554–1563, 1966.
- [26] D. T. Lauria, B. Maskery, C. Poulos, and D. Whittington, “An optimization model for reducing typhoid cases in developing countries without increasing public spending,” *Vaccine*, vol. 27, no. 10, pp. 1609–1621, 2009.
- [27] D. Kalajdziewska and M. Y. Li, “Modeling the effects of carriers on transmission dynamics of infectious diseases,” *Math. Biosci. Eng.*, vol. 8, no. 3, pp. 711–722, 2011.
- [28] S. Mushayabasa, “Impact of vaccines of controlling typhoid fever in kassana-nankana district of upper east region of ghana : insight from a mathematical model,” 2011.
- [29] H. Abboubakar and R. Racke, “Mathematical modelling and optimal control of typhoid fever,” 2019.
- [30] V. E. Pitzer, C. C. Bowles, S. Baker, G. Kang, V. Balaji, J. J. Farrar, and B. T. Grenfell, “Predicting the impact of vaccination on the transmission dynamics of typhoid in south asia : a mathematical modeling study,” *PLoS neglected tropical diseases*, vol. 8, no. 1, p. e2642, 2014.
- [31] <https://www.bioutils.ch/protocoles/13-analyse-de-l?eau>.
- [32] D. Sur, L. von Seidlein, B. Manna, S. Dutta, A. K. Deb, B. L. Sarkar, S. Kanungo, J. L. Deen, M. Ali, D. R. Kim, *et al.*, “The malaria and typhoid fever burden in the slums of kolkata, india : data from a prospective community-based study,” *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 100, no. 8, pp. 725–733, 2006.
- [33] M. I. Khan, A. Pach 3rd, G. M. Khan, D. Bajracharya, S. Sahastrabuddhe, W. Bhutta, R. Tahir, S. Soofi, C. B. Thapa, N. Joshi, *et al.*, “Typhoid vaccine introduction : An evidence-based pilot implementation project in nepal and pakistan,” *Vaccine*, vol. 33, pp. C62–C67, 2015.

- [34] J. A. Crump, F. G. Youssef, S. P. Luby, M. O. Wasfy, J. M. Rangel, M. Taalat, S. A. Oun, and F. J. Mahoney, “Estimating the incidence of typhoid fever and other febrile illnesses in developing countries,” *Emerging infectious diseases*, vol. 9, no. 5, pp. 539–544, 2003.
- [35] K. P. Neil, S. V. Sodha, L. Lukwago, S. O-tipo, M. Mikoleit, S. D. Simington, P. Mukobi, S. Balinandi, S. Majalija, J. Ayers, *et al.*, “A large outbreak of typhoid fever associated with a high rate of intestinal perforation in kasese district, uganda, 2008–2009,” *Clinical infectious diseases*, vol. 54, no. 8, pp. 1091–1099, 2012.
- [36] A. Akullian, E. Ng’eno, A. I. Matheson, L. Cosmas, D. Macharia, B. Fields, G. Bigogo, M. Mugoh, G. John-Stewart, J. L. Walson, *et al.*, “Environmental transmission of typhoid fever in an urban slum,” *PLoS neglected tropical diseases*, vol. 9, no. 12, p. e0004212, 2015.
- [37] A. Gebreyesus, K. Adane, L. Negash, T. Asmelash, S. Belay, M. Alemu, and M. Saravanan, “Prevalence of salmonella typhi and intestinal parasites among food handlers in mekelle university student cafeteria, mekelle, ethiopia,” *Food Control*, vol. 44, pp. 45–48, 2014.
- [38] J. John, C. J. Van Aart, and N. C. Grassly, “The burden of typhoid and paratyphoid in india : systematic review and meta-analysis,” *PLoS neglected tropical diseases*, vol. 10, no. 4, p. e0004616, 2016.
- [39] P. Roumagnac, F.-X. Weill, C. Dolecek, S. Baker, S. Brisse, N. T. Chinh, T. A. H. Le, C. J. Acosta, J. Farrar, G. Dougan, *et al.*, “Evolutionary history of salmonella typhi,” *Science*, vol. 314, no. 5803, pp. 1301–1304, 2006.
- [40] V. E. Pitzer, N. A. Feasey, C. Msefula, J. Mallewa, N. Kennedy, Q. Dube, B. Denis, M. A. Gordon, and R. S. Heyderman, “Mathematical modeling to assess the drivers of the recent emergence of typhoid fever in blantyre, malawi,” *Clinical Infectious Diseases*, vol. 61, no. suppl_4, pp. S251–S258, 2015.
- [41] O. Diekmann, J. A. P. Heesterbeek, and J. A. Metz, “On the definition and the computation of the basic reproduction ratio r_0 in models for infectious diseases in heterogeneous populations,” *Journal of mathematical biology*, vol. 28, pp. 365–382, 1990.
- [42] P. Van den Driessche and J. Watmough, “Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission,” *Mathematical biosciences*, vol. 180, no. 1-2, pp. 29–48, 2002.
- [43] C. T. Codeço, S. Lele, M. Pascual, M. Bouma, and A. I. Ko, “A stochastic model for ecological systems with strong nonlinear response to environmental drivers : application to two water-borne diseases,” *Journal of The Royal Society Interface*, vol. 5, no. 19, pp. 247–252, 2008.
- [44] A. S. Novozhilov, G. P. Karev, and E. V. Koonin, “Biological applications of the theory of birth-and-death processes,” *Briefings in bioinformatics*, vol. 7, no. 1, pp. 70–85, 2006.
- [45] S. Nee, “Birth-death models in macroevolution,” *Annu. Rev. Ecol. Evol. Syst.*, vol. 37, pp. 1–17, 2006.
- [46] J. L. Thorne, H. Kishino, and J. Felsenstein, “An evolutionary model for maximum likelihood alignment of dna sequences,” *Journal of Molecular Evolution*, vol. 33, no. 2, pp. 114–124, 1991.

- [47] F. W. Crawford, V. N. Minin, and M. A. Suchard, “Estimation for general birth-death processes,” *Journal of the American Statistical Association*, vol. 109, no. 506, pp. 730–747, 2014.
- [48] P. Moran, “The estimation of the parameters of a birth and death process,” *Journal of the Royal Statistical Society : Series B (Methodological)*, vol. 15, no. 2, pp. 241–245, 1953.
- [49] J. F. Reynolds, “On estimating the parameters of a birth-death process,” *Australian Journal of Statistics*, vol. 15, no. 1, pp. 35–43, 1973.
- [50] N. Keiding, “Maximum likelihood estimation in the birth-and-death process,” *The Annals of Statistics*, vol. 3, no. 2, pp. 363–372, 1975.
- [51] T. Perkins, “Maximum likelihood trajectories for continuous-time markov chains,” in *Advances in Neural Information Processing Systems* (Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, eds.), vol. 22, Curran Associates, Inc., 2009.
- [52] E. R. Immel, *Problems of estimation and hypothesis testing in connection with birth-and-death stochastic processes*. PhD thesis, Doctoral Thesis, University of California, Los Angeles. Résumé : Ann. Math ?, 1951.
- [53] F. W. Crawford and M. A. Suchard, “Transition probabilities for general birth-death processes with applications in ecology, genetics, and evolution,” *Journal of mathematical biology*, vol. 65, no. 3, pp. 553–580, 2012.
- [54] J. Xu, P. Guttorp, M. Kato-Maeda, and V. N. Minin, “Likelihood-based inference for discretely observed birth-death-shift processes, with applications to evolution of mobile genetic elements,” *Biometrics*, vol. 71, no. 4, pp. 1009–1021, 2015.
- [55] A. C. Davison, S. Hautphenne, and A. Kraus, “Parameter estimation for discretely observed linear birth-and-death processes,” *Biometrics*, vol. 77, no. 1, pp. 186–196, 2021.
- [56] I. Bouzalmat, B. de Saporta, and S. M. Manou-Abi, “Parameter estimation for a hidden linear birth and death process with immigration,” *arXiv preprint arXiv :2303.00531*, 2023.
- [57] J. Fritz, J. LaSalle, and L. Sirovich, *Partial differential equations*. Applied Mathematical Sciences, 1977.
- [58] S. Tavaré, “The linear birth-death process : an inferential retrospective,” *Advances in Applied Probability*, vol. 50, no. A, pp. 253–269, 2018.
- [59] N. T. Bailey, *The elements of stochastic processes with applications to the natural sciences*, vol. 25. John Wiley & Sons, 1991.
- [60] S. M. Ross, *Introduction to probability models*. Academic press, 2014.
- [61] R. M. Corless, G. H. Gonnet, and D. Hare, “On the lambert w function,” *Advances in Computational Mathematics*, vol. 5, pp. 329–359, 1996.
- [62] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. Knuth, “Lambert’s w function in maple,” *Maple Technical Newsletter*, vol. 9, no. 1, pp. 12–22, 1993.
- [63] I. Teodorescu, “Maximum likelihood estimation for markov chains.” arXiv preprint arXiv :0905.4131, 2009.
- [64] G. M. Goerg, *LambertW : Probabilistic Models to Analyze and Gaussianize Heavy-Tailed, Skewed Data*, 2020. R package version 0.6.6.
- [65] M. L. Himmelmann, “Package ?hmm ?,” 2022.

- [66] K. Athreya and C. Fuh, “Bootstrapping markov chains : countable case,” *Journal of Statistical Planning and Inference*, vol. 33, no. 3, pp. 311–331, 1992.
- [67] S. B. Y. M. T. D. Y. I. C. V. J. T. G. R. J. B. D. E. S. M. Giorgio Alfredo Spedicato, Tae Seung Kang, “Markovchain : Easy handling discrete time markov chains. 2022,” 2022.
- [68] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*, vol. 4. Springer, 2006.
- [69] K. P. Murphy, *Machine learning : a probabilistic perspective*. MIT press, 2012.
- [70] H. Abboubakar and R. Racke, “Mathematical modeling, forecasting, and optimal control of typhoid fever transmission dynamics,” *Chaos, Solitons & Fractals*, vol. 149, p. 111074, 2021.
- [71] https://www.inrs.fr/publications/bdd/eficatt/fiche.html?refINRS=EFICATT_Fièvretyphoïde.
- [72] S. P. Meyn and R. L. Tweedie, “Stability of markovian processes iii : Foster–lyapunov criteria for continuous-time processes,” *Advances in Applied Probability*, vol. 25, no. 3, pp. 518–548, 1993.
- [73] P. Kumar, H. Gauba, P. P. Roy, and D. P. Dogra, “Coupled hmm-based multi-sensor data fusion for sign language recognition,” *Pattern Recognition Letters*, vol. 86, pp. 1–8, 2017.
- [74] S. Ghosh, J. Li, L. Cao, and K. Ramamohanarao, “Septic shock prediction for icu patients via coupled hmm walking on sequential contrast patterns,” *Journal of bio-medical informatics*, vol. 66, pp. 19–31, 2017.
- [75] R. Sarno and K. R. Sungkono, “Coupled hidden markov model for process discovery of non-free choice and invisible prime tasks,” *Procedia Computer Science*, vol. 124, pp. 134–141, 2017.
- [76] M. Brand, “Coupled hidden markov models for modeling interacting processes,” tech. rep., Citeseer, 1997.
- [77] J. A. Bilmes *et al.*, “A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models,” *International computer science institute*, vol. 4, no. 510, p. 126, 1998.
- [78] R. Calaway, R. Analytics, S. Weston, D. Tenenbaum, and M. R. Calaway, “Package?doparallel?,” 2015.
- [79] <https://www.historique-meteo.net/afrique/mayotte/>.
- [80] E. S. Page, “Continuous inspection schemes,” *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.
- [81] A. Wald, “Statistical decision functions which minimize the maximum risk,” *Annals of Mathematics*, pp. 265–280, 1945.
- [82] D. Egea-Roca, G. Seco-Granados, and J. A. López-Salcedo, “Comprehensive overview of quickest detection theory and its application to gnss threat detection,” *Gyroscope and Navigation*, vol. 8, pp. 1–14, 2017.
- [83] R. Killick and I. Eckley, “changept : An r package for changepoint analysis,” *Journal of statistical software*, vol. 58, no. 3, pp. 1–19, 2014.

- [84] M. D. Sobsey, C. E. Stauber, L. M. Casanova, J. M. Brown, and M. A. Elliott, "Point of use household drinking water filtration : a practical, effective solution for providing sustained access to safe drinking water in the developing world," *Environmental science & technology*, vol. 42, no. 12, pp. 4261–4267, 2008.
- [85] K. Cann, D. R. Thomas, R. Salmon, A. Wyn-Jones, and D. Kay, "Extreme water-related weather events and waterborne disease," *Epidemiology & Infection*, vol. 141, no. 4, pp. 671–686, 2013.
- [86] O. Oguntoke, O. Aboderin, and A. Bankole, "Association of water-borne diseases morbidity pattern and water quality in parts of ibadan city, nigeria," *Tanzania Journal of health research*, vol. 11, no. 4, 2009.
- [87] M. Hatta and H. L. Smits, "Detection of salmonella typhi by nested polymerase chain reaction in blood, urine, and stool samples," *The American journal of tropical medicine and hygiene*, vol. 76, no. 1, pp. 139–143, 2007.
- [88] G. C. Buckle, C. L. F. Walker, and R. E. Black, "Typhoid fever and paratyphoid fever : Systematic review to estimate global morbidity and mortality for 2010.," *Journal of global health*, vol. 2, no. 1, 2012.
- [89] I. Gantois, R. Ducatelle, F. Pasmans, F. Haesebrouck, R. Gast, T. J. Humphrey, and F. Van Immerseel, "Mechanisms of egg contamination by salmonella enteritidis," *FEMS microbiology reviews*, vol. 33, no. 4, pp. 718–738, 2009.
- [90] T. Akritidou, S. Akkermans, S. Gaspari, N. D. Azraini, C. Smet, T. Van de Wiele, and J. F. Van Impe, "Effect of gastric ph and bile acids on the survival of listeria monocytogenes and salmonella typhimurium during simulated gastrointestinal digestion," *Innovative Food Science & Emerging Technologies*, vol. 82, p. 103161, 2022.
- [91] R. J. Hyndman and G. Athanasopoulos, *Forecasting : principles and practice*. OTexts, 2018.