



HAL
open science

Contributions à la modélisation mathématiques en écologie microbienne, des données aux modèles

Simon Labarthe

► **To cite this version:**

Simon Labarthe. Contributions à la modélisation mathématiques en écologie microbienne, des données aux modèles. Mathématiques [math]. Université de bordeaux, 2024. tel-04654712

HAL Id: tel-04654712

<https://hal.inrae.fr/tel-04654712>

Submitted on 19 Jul 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 - NonCommercial - ShareAlike 4.0 International License

**Contributions à la modélisation
mathématique en écologie microbienne, des
données aux modèles**

Par *Simon Labarthe*

Habilitation à diriger les recherches
Mathématiques appliquées et calcul scientifique
Présentée et soutenue publiquement le *(10 juillet 2024)*

Jury :

Delphine Ropers, *DR, INRIA, rapportrice*
Marcela Szopos, *PU, Université Paris Cité, rapportrice*
Rafael Muñoz-Tamayo, *DR, INRAE, rapporteur*
Magali Ribot-Barré, *PU, université d'Orléans, examinatrice*
Michel Bergmann, *DR, Inria, examinateur*
Béatrice Laroche, *DR, INRAE, examinatrice*

Ce document est écrit en \LaTeX en utilisant un template écrit par Chloé Goupy. Original disponible à l'adresse <https://fr.overleaf.com/latex/templates/thesis-layout-upariscite/qvmbkpnxvtdn>. J'en ai modifié la couleur et les logos.

Il est à disposition selon les termes de la licence Creative Commons :

© 2024 CC BY-NC 4.0



<https://creativecommons.org/licenses/by-nc/4.0/deed.fr>

Sommaire

		3
1	Introduction	5
1	Microbiote et écologie microbienne	5
2	Approches holistiques ou réductionnistes en écologie microbienne	6
2.1	Données méta-omiques pour les approches holistiques	6
2.2	Réductionnisme pour l'étude contrôlée des communautés	8
3	Problématiques mathématiques pour la modélisation en écologie microbienne	8
3.1	Approche holistique : dérivation de systèmes dynamiques et intégration de données multi-omiques	9
3.2	Approche réductionniste : vers l'utilisation de jumeaux numériques	9
3.3	Interaction du microbiote avec son environnement	11
4	Plan du manuscrit	12
2	Approches holistiques : analyse des liens structures-fonctions	13
1	Données métagénomiques pour la caractérisation fonctionnelle de microbiomes	13
2	NMF pour la simplification de microbiote	14
3	Résultats	15
3.1	Des profils fonctionnels et taxonomiques reconstituant les métagénomomes	15
3.2	Stratification de la matrice de poids	18
4	Perspective : utilisation des résultats de la NMF pour construire un modèle de dynamique de population	18
3	Approches réductionnistes pour l'étude de communautés microbiennes	21
1	Modèles de communautés microbiennes réduites.	21
2	Modèle d'une communauté fermentaire pour la production de fromage	22
3	Approximation de modèle par <i>machine learning</i>	25
4	Interaction du microbiote avec son environnement	29
1	Interactions de micro-nageurs avec leur environnement dans un biofilm	30
1.1	Biofilms et micro-nageurs comme facilitateurs de contrôle	30
1.2	Caractérisation de bactéries nageuses dans un biofilm.	30
2	Modélisation des régulations de la crypte intestinale par le microbiote	32
2.1	Renouvellement de la crypte intestinale modulé par le microbiote.	32
2.2	Un modèle PDMP de la dynamique de la crypte intestinale incluant le microbiote.	34
2.2.1	Description du modèle cellule-centré	34
2.2.2	Évolution des concentrations de dioxygène et de butyrate	35
2.3	Modèles de densités de populations à l'échelle de la crypte	36

2.3.1	Modèle limite en grande population	36
2.3.2	Modèle limite de l'équation de <i>self-aggregation</i> quand le support d'interaction tend vers 0.	38
3	Modèle du microbiote intestinal	39
3.1	Biogéographie du microbiote intestinal.	39
3.2	Modèle du microbiote intestinal en interaction avec son environnement multi-phasique.	41
3.3	Simplification du modèle par analyse asymptotique	42
3.4	Exploration numérique	43
5	Flux microbiens entre holobiontes et épidémiologie	47
1	Des flux microbiens entre holobiontes.	47
2	Modèle populationnel incluant la dynamique d'infection intra-hôte.	48
2.1	Modèle simplifié intra-hôte	48
2.2	Modèle à l'échelle de la population	49
3	Étude analytique et numérique du modèle	50
6	Conclusion	53
1	Des axes de recherche à l'interface des mathématiques et de l'écologie microbienne	53
2	Un panel étendu d'outils mathématiques pour la modélisation	54
3	Un investissement sur les diverses facettes du travail de chercheur	55
3.1	Formation par la recherche	56
3.2	Montage de projets	56
3.3	Animation de la science et expertise	57
4	perspectives	57
	Bibliographie personnelle	61
	Bibliographie générale	65
A	Annexes	69
1	Synthèse des travaux de thèse	69
B	Curriculum Vitae	71
C	Résumé étendu	77
1	Introduction : modélisation de communautés microbiennes	77
2	Approches holistiques : analyse des liens structures-fonctions	78
3	Approches réductionnistes pour l'analyse des communautés	79
4	Interaction du microbiote avec son environnement	81
4.1	Modèle du microbiote intestinal	81
4.2	Modélisation des régulations de la crypte intestinale par le microbiote . .	82
4.3	Interactions de micro-nageurs avec leur environnement dans un biofilm . .	82
5	Flux microbiens entre holobiontes et épidémiologie	83
6	Conclusion	84

C H A P I T R E



Introduction

Ce manuscrit synthétise les recherches que j'ai menées depuis mon recrutement comme chargé de recherche à l'INRAE, sur la modélisation mathématique de communautés microbiennes. Après une thèse à l'Université de Bordeaux 2 et à l'Inria de Bordeaux (équipe Carmen) sur la modélisation de l'électrophysiologie cardiaque et d'arythmies auriculaires, soutenue en décembre 2013, j'ai été recruté dans l'unité MaIAGE (INRAE) en septembre 2014 comme chargé de recherche. J'ai ensuite effectué en 2018 et 2019 un séjour post-doctoral d'environ une année à UC Davis dans le laboratoire d'Andreas Bäumlér, puis j'ai bénéficié d'une mobilité vers l'unité Biogeco (Cestas) en 2020 où je fais depuis partie de l'équipe INRAE/Inria Pléiade. J'ai fait le choix de ne pas présenter dans cette synthèse mes travaux de thèse en modélisation appliquée à l'électrophysiologie cardiaque, dont une brève description peut être trouvée en annexe, mais de centrer ma présentation sur la modélisation de communautés microbiennes et sur certaines problématiques mathématiques adaptées à l'écologie microbienne. Après une brève introduction sur les communautés microbiennes, les expérimentations et les données disponibles, je présenterai quelques grands enjeux de modélisation mathématique en écologie microbienne.

1 Microbiote et écologie microbienne

Les communautés microbiennes, aussi appelées microbiotes, sont des écosystèmes extrêmement complexes, tant par le nombre d'individus qu'ils contiennent, que par leurs diversités taxonomiques ou fonctionnelles. Pour fixer un ordre d'idée au travers de l'exemple du microbiote intestinal, il peut y avoir jusqu'à 10^{11} bactéries par gramme de fèces issues de plusieurs centaines d'espèces [44], pouvant contenir plusieurs millions de gènes [60], soit environ deux ordres de grandeur de plus que le génome humain [41]. Ces communautés forment de véritables écosystèmes en interaction avec un environnement physico-chimique qui peut être régulé par un macro-organisme hôte vivant en symbiose avec la communauté. Ces liens intriqués entre hôte et microbes ont amené le concept d'holobionte [40], [46], l'ensemble symbiotique formé par l'hôte et son microbiote commensal. Sur un même hôte, des microbiotes très différents peuvent ré-

sider dans différents compartiments interconnectés par des flux microbiens. Par exemple, chez l'Homme, les microbiotes de la peau, de la bouche, du colon ou des parties uro-génitales sont connectés à des degrés divers [49]. De même, chez la plante, des communautés très différentes se développent dans la rhizosphère (dans les zones racinaires), dans la phyllosphère (les parties aériennes extra-tissulaires comme la surface des feuilles, des fruits ou des graines) ou à l'intérieur des tissus (endosphère) [36].

Certains pathogènes microbiens ont développé la capacité de perturber cet équilibre, et de détourner les mécanismes de régulation de la symbiose à leur avantage [45], [48]. Ces découvertes ont amené le concept de pathobionte, pour décrire le système composé par l'hôte, son microbiote commensal et le pathogène microbien [53]. Les pathobiontes peuvent être utilisés comme perturbation pour étudier les régulations généralement à l'œuvre pour assurer l'homéostasie du microbiote sain.

Enfin, différents hôtes peuvent être en contact, engendrant des flux microbiens entre leurs communautés symbiotiques ou des flux de pathogènes (voir Fig. 1.1 pour un résumé de ces différents niveaux d'étude des communautés microbiennes). L'étude de ces écosystèmes microbiens dépasse donc l'étude des individus qui les composent, mais s'inscrit dans des approches systémiques étudiant les interactions, les mécanismes de régulations permettant des équilibres homéostatiques, ou au contraire des perturbations pouvant amener à des états de dysbiose. On parle alors d'écologie microbienne, particulièrement propice à des approches de type biologie des systèmes [19].

2 Approches holistiques ou réductionnistes en écologie microbienne

Ces communautés peuvent être étudiées de manière holistique, en analysant des données métagénomiques à l'échelle de la communauté, ou réductionnistes en analysant des communautés microbiennes réduites représentatives d'une communauté complexe (voir Fig. 1.1 pour un résumé graphique).

2.1 Données méta-omiques pour les approches holistiques

Les différentes espèces microbiennes composant un microbiote sont porteuses d'un matériel génétique qui peut être étudié à l'échelle de la communauté. Plutôt qu'isoler les microbes dont on veut séquencer l'ADN, des enzymes sont utilisés pour dégrader les parois cellulaires des microbes de la communauté afin d'accéder à leur matériel génétique intracellulaire, qui se trouve alors mélangé. Ce matériel génétique communautaire est appelé *métagénome*. Par séquençage haut débit, il est possible de séquencer une portion du métagénome, et ainsi reconstituer après traitement bioinformatique tout ou partie des génomes des microbes de la communauté. Il est possible de se focaliser sur la composition taxonomique de la communauté en ciblant spécifiquement des gènes marqueurs de la taxonomie après une phase d'amplification [42], [43], [58], [61] : on parle alors de metabarcoding. Ces gènes marqueurs permettent de regrouper des micro-organismes par leur proximité génétique, en des groupements appelés OTU (Operational Taxonomic Unit) assimilés à une même entité taxonomique [61]. Les OTUs permettent une définition moléculaire de notions taxonomiques telles que l'espèce ou le genre. Il est également possible d'avoir une approche non ciblée et de tenter de cataloguer l'intégralité des gènes microbiens présents dans le métagénome [60], les catalogues les plus récents comportant plusieurs millions de gènes uniques : on parle alors de métagénomique plein génome ou shotgun. Le métagénome intestinal humain a été étudié via de nombreuses cohortes [38], [39], [50], [57], ce qui a permis de mettre en évidence

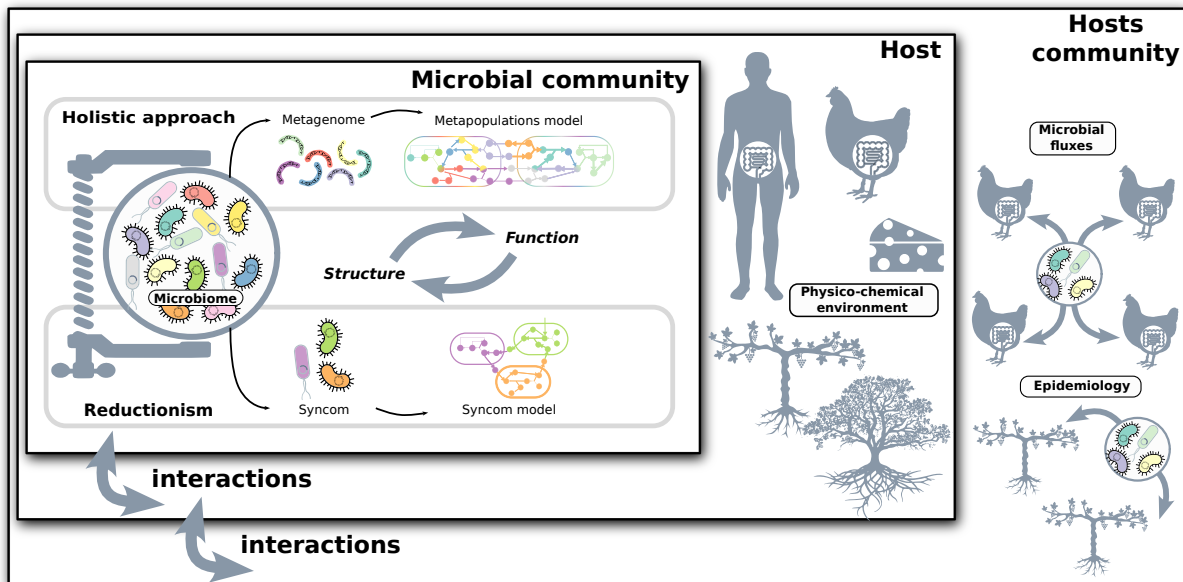


Fig. 1.1 Echelles d'étude d'un microbiote. Un microbiote peut d'abord être étudié par une approche holistique : la communauté est étudiée dans son entier via des données métagénomiques. L'étude de ces données permet d'identifier des invariants taxonomiques ou fonctionnels qui peuvent être utilisés pour bâtir des méta-communautés, c'est à dire des regroupements abstraits d'espèces microbiennes rassemblées pour leurs caractéristiques fonctionnelles ou taxonomiques communes. Un microbiote peut être également étudié par une approche réductionniste : un petit nombre d'espèces microbiennes représentatives sont isolées et assemblées dans des communautés synthétiques, ou SynCom, construites pour représenter le fonctionnement de la communauté. Ces deux paradigmes peuvent être utilisés conjointement pour étudier les liens entre structure taxonomique de la communauté et fonctions régissant sa dynamique. Le microbiote évolue en interaction avec un hôte régulant un environnement physico-chimique caractéristique, permettant de sélectionner et contrôler des populations microbiennes symbiotiques. Enfin, différents hôtes peuvent vivre en interaction : les flux microbiens entre hôtes influent sur la dynamique de leurs microbiotes et sur la propagation de pathogènes.

des associations entre la composition du microbiote, les fonctions microbiennes contenues dans le métagénome et la santé de l'hôte [35], [51].

Les gènes ainsi identifiés représentent des fonctions activables dans la communauté. On parle de potentiel fonctionnel. D'autres données omiques permettent de rendre compte des fonctions effectivement activées dans la communauté dans un certain contexte, comme les données métatranscriptomiques, qui indiquent les gènes du métagénome qui sont transcrits, les données métaprotéomiques, qui déterminent les protéines produites suite à la transcription des gènes, ou les données métabolomiques, qui indiquent les métabolites produits ou transformés par ces protéines.

2.2 Réductionnisme pour l'étude contrôlée des communautés

Une approche expérimentale différente a été proposée pour l'étude des microbiotes : le réductionnisme, qui consiste à assembler des communautés réduites dont on contrôle exactement la composition. On parle alors de communautés synthétiques, ou SynComs. Les approches réductionnistes visent à sélectionner des communautés synthétiques dont le fonctionnement est représentatif des communautés complexes [54]. Leur aspect réduit facilite l'interprétation des résultats et la répétabilité des expérimentations [33]. La sélection de SynComs peut être associée à des approches de culturomique, c'est à dire l'isolation et la culture haut-débit de micro-organismes issus des communautés étudiées [30].

Pour étudier le microbiote intestinal, différentes SynComs ont été proposées [37], [47], [54]. Le critère de sélection des micro-organismes le composant est le plus souvent la couverture des données métagénomiques par les génomes des microbes sélectionnés. Chez la plante, des approches utilisant des communautés synthétiques ont été également proposées [31]. Une fois une SynCom assemblée, elle peut être mise en association avec un hôte stérile (appelé hôte axénique) afin d'étudier les interactions entre la communauté réduite et le macro-organisme hôte [54].

3 Problématiques mathématiques pour la modélisation en écologie microbienne

J'étudie les communautés microbiennes à ces différentes échelles de complexité avec des systèmes d'équations de dynamique de population. Pour une population microbienne $i \in [0, N_b]$ et un métabolite $j \in [0, N_m]$, où N_b et N_m sont respectivement le nombre de micro-organismes (le plus souvent des bactéries, mais pouvant être d'autres micro-organismes comme des champignons, des levures ou des archées) et de métabolites considérés dans le modèle, on peut écrire le système d'équation

$$\partial_t b_i + \operatorname{div}(F_{b,i}(b, m)) = G_{b,i}(b, m) \quad (1.1)$$

$$\partial_t m_j + \operatorname{div}(F_{m,j}(b, m)) = G_{m,j}(b, m) \quad (1.2)$$

Dans ce système, b et m sont les vecteurs de dimension N_b et N_m regroupant les densités de microbes et métabolites considérés. Les termes $F_{b,i}$ et $F_{m,j}$ désignent des opérateurs différentiels décrivant les flux locaux de microbes ou métabolites décrivant leur propagation spatiale. Les fonctions $G_{b,i}$ et $G_{m,j}$ décrivent les taux de croissance microbienne et de consommation ou de production de métabolites. Les équations (1.1)-(1.2) sont couplées par les taux de croissance ou les termes spatiaux.

Ce système doit être considéré comme un système générique pouvant être adapté aux différentes approches d'écologie microbienne présentées dans ce manuscrit, qu'elles soient holistiques ou réductionnistes, centrées sur les interactions avec l'environnement ou sur les flux microbiens entre holobiontes, avec des problématiques mathématiques spécifiques. Tâchons de formaliser ces problématiques mathématiques pour chacune de ces approches.

3.1 Approche holistique : dérivation de systèmes dynamiques et intégration de données multi-omiques

Lorsqu'on étudie une communauté microbienne par une approche holistique, l'enjeu principal est de gérer la très grande dimension des matrices de données omiques, reflétant la grande diversité taxonomique et fonctionnelle des communautés microbiennes. Une matrice de comptage de metabarcoding ou de métagénomique comporte typiquement plusieurs centaines de lignes, chacune représentant un échantillon, et plusieurs milliers (ou millions pour la métagénomique plein génome) de colonnes, représentant un gène ou un taxon. La problématique devient alors

- développer des stratégies de réduction de dimension permettant d'identifier des structures pertinentes pour dériver le système (1.1)-(1.2).

En d'autre terme, il s'agira de définir des regroupements de populations microbiennes ayant des fonctions métaboliques communes, que l'on peut appeler des méta-communautés fonctionnelles.

Cette problématique est *a priori* une problématique de statistique ou de science des données. Toutefois, je présenterai dans ce manuscrit une méthode couplant des techniques de réduction de dimension avec des modèles métaboliques. Cette méthode de réduction de dimension permet alors de déterminer N_b , le nombre de méta-populations fonctionnelles, et N_m , le nombre de métabolites d'intérêt. Mais en déterminant les regroupements fonctionnels pertinents, elle permet également de caractériser les taux de croissance G_b et les rendements de production et de consommation de métabolites G_m , ce qui permet de dériver la structure du système (1.1)-(1.2). Cette manière de construire un modèle, en regroupant des gènes fonctionnels en réseaux métaboliques sans tenir compte des individus qui les portent, s'appelle des modèles "sac-de-gènes" (*bag-of-genes models*) [1].

Une fois dérivé, le système (1.1)-(1.2) peut être utilisé pour intégrer et co-analyser des données multi-omiques. Ce modèle mécanistique permet de simuler explicitement les processus de production des quantités mesurées : par exemple, des données métabolomiques peuvent être comparées aux densités m . Des données métatranscriptomiques peuvent permettre d'affiner ou de valider les fonctions G_b et G_m . La problématique consiste ici à

- développer des modèles permettant de simuler des observables comparables aux différents types de données omiques assemblées dans des situations réelles.

3.2 Approche réductionniste : vers l'utilisation de jumeaux numériques

Les approches réductionnistes permettent d'obtenir des communautés assez réduites pour mettre en œuvre des modèles très utilisés en biologie des systèmes, les modèles métaboliques basés sur le génome (*genome-scale metabolic models*, GSMM) [10]. A partir du génome complet d'un micro-organisme isolé, les différents gènes sont annotés, c'est à dire qu'ils sont reliés à une enzyme catalysant une réaction biochimique. Les différentes réactions sont ensuite rassemblées en un réseau métabolique sur lequel un problème d'optimisation est défini afin de prédire des flux métaboliques reliant les différents nœuds du réseau (voir Fig. 1.2). Ces flux métaboliques permettent de définir G_b et G_m , second membre des équations (1.1)-(1.2). Cette méthodologie

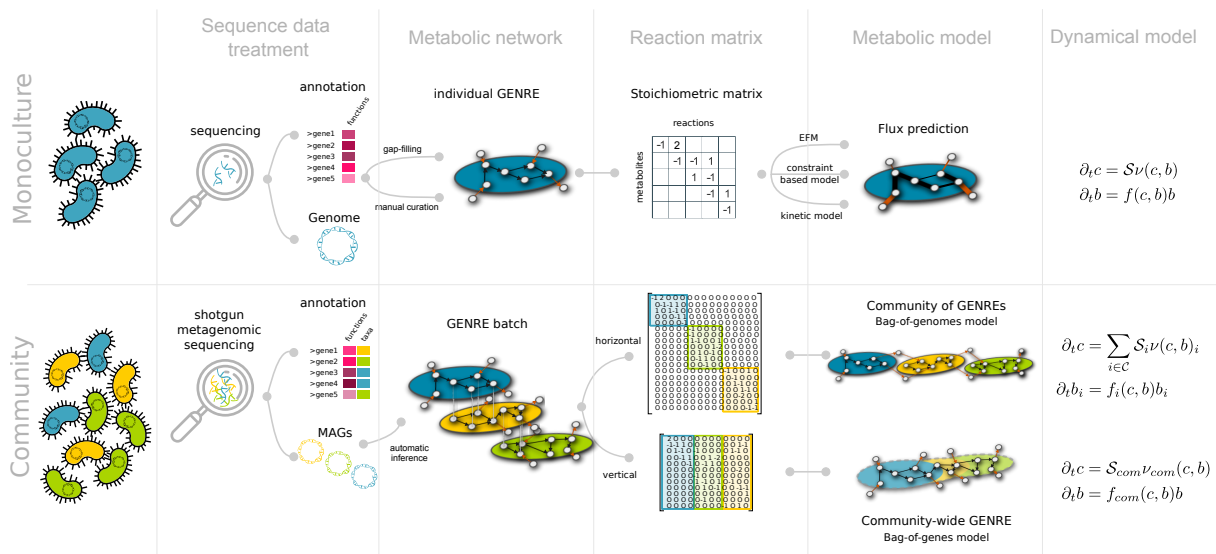


Fig. 1.2

Modélisation du métabolisme microbien, de la séquence aux flux métaboliques.

Lorsqu'il s'agit de modéliser une population d'un seul micro-organisme (panel supérieur), le pipe-line est bien balisé : après isolement et séquençage du micro-organisme, son génome est assemblé et annoté. C'est à dire qu'on met en relation les gènes du génome avec la ou les réactions biochimiques permises par l'enzyme codée par le gène. Ces réactions chimiques sont ensuite combinées pour former un réseau métabolique, dont les nœuds sont les métabolites et les arêtes les réactions. Plusieurs paradigmes de modélisation permettent alors de prédire des flux de métabolites au travers du réseau (FBA, EFM, modèle cinétique...). Ces flux peuvent alors être utilisés dans le second membre de systèmes dynamiques. En communauté (panel inférieur), les premières étapes sont identiques (quoique complexifiées par la diversité des micro-organismes) jusqu'à la constitution de réseaux métaboliques individuels par organismes. Il y a alors deux paradigmes différents pour constituer un modèle de communauté : les réseaux métaboliques individuels peuvent juste être couplés par des métabolites extérieurs (bag-of-genomes models, ou "sac de génomes"), ou bien un réseau métabolique commun à une meta-communauté de micro-organismes partageant les mêmes voies métaboliques peut être assemblé (bag-of-genes models ou "sac de gènes"). Les mêmes méthodes peuvent alors être déployées pour modéliser les flux métaboliques et construire des seconds membres de systèmes dynamiques. Les modèles "sac de gènes" sont à rapprocher des approches holistiques (sec. 2.1 and 3.1), c'est à dire de l'objectif de tirer de la complexité des communautés microbiennes naturelles des résumés (ici des "sac de gènes") qui peuvent servir de points d'entrée à des systèmes dynamiques. Les méthodes "sac de génomes" quant à elles sont à rapprocher des approches réductionnistes (sec. 2.2 et 3.2), c'est à dire de l'objectif de modéliser finement des communautés simplifiées (SynComs). Cette figure est issue de [1]

peut être répliquée sur les différents micro-organismes de la SynCom, ce qui permet d'assembler le système (1.1)-(1.2). Contrairement aux approches holistiques où le vecteur b décrivait la densité de méta-communautés fonctionnelles (les "sacs-de-gènes"), c'est à dire un regroupement abstrait de différents micro-organismes, le vecteur b décrit cette fois la densité de population microbiennes correspondant à des espèces taxonomiques. On parle alors de modèles "sacs-de-génomés" (*bag-of-genomes models*) [1] (voir Fig. 1.2).

Le caractère réduit et contrôlé des SynComs permet de bâtir des dispositifs expérimentaux répliquables à même d'assembler des séries temporelles de différentes données omiques. Ces séries temporelles peuvent être confrontées aux sorties des modèles numériques basés sur les GSMMs qui peuvent les reproduire avec précision. Toutefois, si cette méthodologie est éprouvée pour des modèles mono-espèces, la généralisation à des communautés microbiennes reste un défi. La problématique est alors :

- adapter les méthodes de modélisation dynamique basées sur des GSMM à des communautés microbiennes de façon à reproduire des données omiques, en monoculture et en communauté.

A cette problématique de modélisation s'ajoute une problématique numérique. En effet, les modèles d'optimisation sur lesquels sont basés ces GSMMs doivent être résolus à chaque pas de temps et d'espace des équations (1.1)-(1.2), ce qui alourdit considérablement les temps de calcul. La problématique devient alors

- construire des approximations de G_b et G_m , précis et rapide à résoudre numériquement, permettant d'approcher les solutions de (1.1)-(1.2) en temps raisonnable.

3.3 Interaction du microbiote avec son environnement

Bien que le modèle générique (1.1)-(1.2) considère principalement les interactions microbes-microbes, médiées par des métabolites, ces équations permettent de prendre en compte indirectement les interactions avec l'hôte, particulièrement en définissant un environnement nutritionnel spécifique du macro-organisme symbiotique. Ces équations permettent également de prendre en compte au travers des opérateurs $F_{b,i}$ des interactions physico-chimiques complexes des microbes avec leur environnement, par exemple l'interaction avec un fluide visco-élastique comme du mucus ou bien du chimiotactisme.

Toutefois, l'hôte peut être pris en compte de manière plus explicite moyennant de légères adaptations, soit par des conditions de bord définissant les flux de métabolites échangés entre la partie luminale du milieu et les tissus épithéliaux de l'hôte, soit en rajoutant explicitement de nouvelles équations représentant un compartiment 'hôte'. Enfin, le système (1.1)-(1.2) peut être dupliqué et couplé via des termes de flux pour modéliser l'interaction de différents holobiontes.

Devant la complexité des mécanismes et les multiples échelles d'interactions, la problématique principale associée à la modélisation des holobiontes devient

- construire des modèles multi-échelles et dériver des modèles réduits capturant les phénomènes d'intérêt.
- analyser la sensibilité des modèles aux paramètres clefs.
- dériver des modèles adaptés à l'analyse de données omiques et d'imagerie caractérisant les interactions de micro-organismes avec leur environnement.

4 Plan du manuscrit

Dans le chapitre 2, je présenterai une méthode de réduction de dimension adaptée aux approches holistiques permettant d'analyser des données métagénomiques afin de déterminer la structure du système (1.1)-(1.2) par un modèle de type "sac-de-gènes". Dans le chapitre 3, j'aborderai ensuite des modèles adaptés aux approches réductionnistes au travers d'un exemple de modélisation d'une communauté fermentaire pour la production de fromage. Je présenterai également une méthode de méta-modélisation permettant d'accélérer l'évaluation des fonctions sources des modèles de communautés réduites (1.1)-(1.2). Cette accélération est clef pour la spatialisation de ces modèles basés sur des modèles métaboliques complexes. Je présenterai ensuite dans le chapitre 4 différents modèles étudiant les interactions de populations bactériennes avec leur environnement ou avec l'hôte. Dans la section 1, un modèle individu-centré modélisant l'interaction de bactérie nageuse avec un biofilm exogène permettra de caractériser le comportement de différentes espèces bactériennes en analysant des données de microscopie. Dans la section 2, j'introduirai un modèle stochastique individu-centré de crypte intestinale et son approximation déterministe permettant une accélération substantielle des simulations. Dans la section 3, je présenterai un modèle réduit de mécanique des fluides modélisant l'impact du transit intestinal sur le microbiote intestinal. Enfin, dans le chapitre 5, je présenterai un modèle simplifié d'infection à un pathogène permettant de simuler la distribution de la virulence de l'infection dans une population d'hôtes ainsi que les flux microbiens de pathogènes.

C H A P I T R E
2

Approches holistiques : analyse des liens structures-fonctions

1 Données métagénomiques pour la caractérisation fonctionnelle de microbiomes

Comme vu précédemment, les données métagénomiques sont produites à partir du séquençage haut-débit de l'ensemble du métagénome microbien : des fragments d'ADN sont tirés au hasard et séquencés. Des pipelines bioinformatiques visent alors à regrouper ces fragments en gènes (contigage) puis en génome : on parle alors de Metagenome-Assembled Genomes (MAGs) par opposition avec des génomes qui seraient issus du séquençage de micro-organisme isolés. Cet exercice est parfois assimilé à l'assemblage de milliers de puzzles dont toutes les pièces, comptées en millions, seraient mélangées et dont on n'aurait pas les modèles. Les gènes peuvent être ensuite rassemblés dans des catalogues de gènes uniques. Un catalogue très utilisé des gènes issus du microbiote humain (Integrated Genes Catalog, IGC) comporte environ 10 millions de gènes uniques [50]. Plus récemment, un catalogue de 250 000 génomes (Unified Human Gastrointestinal Genome collection, UHGG) a été assemblé [32] et fait plus que doubler le nombre de protéines précédemment identifiées dans l'IGC. Enfin, ces gènes peuvent être annotés fonctionnellement, c'est à dire reliés à une protéine participant à une réaction biochimique d'un réseau métabolique. Toutefois, l'annotation fonctionnelle des catalogues les plus récents est encore très parcellaire, puisque 40% environ des protéines contenues dans l'UHGG n'ont pas été annotées fonctionnellement.

Enfin, les fragments initiaux peuvent être associés aux gènes du catalogue pour évaluer leur fréquence dans le métagénome : on compte alors pour chaque gène le nombre de fragments lui correspondant, et on renormalise ce comptage avec la taille du gène (en nombre de nucléotides) puisque les gènes les plus longs sont plus susceptibles de générer des fragments plus nombreux. Le produit final de ce pipeline est donc une matrice $X^{(g)}$, de N_s (nombre d'échantillons) lignes et N_g (nombre de gènes du catalogue) colonnes, où $X_{i,j}^{(g)}$ est la fréquence du gène j dans l'échantillon

i. Chaque ligne de cette matrice peut donc être vue comme une caractérisation fonctionnelle de l'échantillon, représenté par la distribution des fréquences des gènes. On note qu'on parle ici d'un *potentiel* fonctionnel : seule une fraction de ces gènes sera activée dans un environnement donné.

Le même procédé peut être fait pour assembler des matrices de fréquences de taxons, c'est à dire d'espèces, de souches, de genres microbiens, selon le niveau taxonomique considéré. Pour cela, les méthodes consistent soit à compter des fragments d'ADN associés à des MAGs, soit à s'intéresser aux fréquences de gènes marqueurs de la taxonomie, soit à amplifier artificiellement un gène cible particulièrement discriminant pour la taxonomie. Cette dernière méthode, connue sous le nom de metabarcoding, ajoute avant le séquençage une étape de PCR qui réalise de nombreuses copies de ce gène cible (par ex. le gène 16S pour les bactéries, ou le gène ITS pour les champignons) : le séquenceur séquence alors quasiment exclusivement ces gènes marqueurs, qui peuvent être rassemblés selon leur distance génétique, indiquant alors des taxons différents (on parle d'OTUs, pour Operational Taxonomic Unit). La fréquence de ces taxons peut alors être assemblée dans des matrices $X^{(otu)}$, de N_s lignes et N_{otu} (nombre d'OTUs uniques identifiés) colonnes, où $X_{i,j}^{(otu)}$ est la fréquence de l'OTU j dans l'échantillon i .

Ces matrices sont grandes, N_s étant de l'ordre du millier, N_g de l'ordre de la dizaine de millions et N_{otu} étant de l'ordre du millier. Il y a alors un fort enjeu de réduction de dimensions.

2 NMF pour la simplification de microbiote

Dans un travail d'analyse de métagénomés intestinaux humains [7], deux étapes de simplification ont été menées. D'abord une phase de simplification basée sur une connaissance biologique *a priori* des fonctions principales impliquées dans la dégradation des fibres alimentaires, principale source nutritionnelle pour les micro-organismes intestinaux, a été mise en œuvre. Des gènes spécifiques du métabolisme des fibres en anaérobiose ont été sélectionnés et agrégés par fonctions à différents niveaux du réseau métabolique de dégradation des fibres. Ces regroupements sont dénommés par la suite des traits fonctionnels agrégés (AFT en anglais). Après sélection d'AFT spécifiques, c'est à dire d'AFT uniquement impliqués dans une branche spécifique du réseau de dégradation des fibres, un réseau métabolique simplifié d'hydrolyse des fibres en anaérobiose a pu être dérivé, sur la base des $N_{AFT} = 101$ AFTs sélectionnés. Cette étape peut être vue comme une étape de modélisation du métabolisme microbien dans l'intestin, permettant d'isoler N_{AFT} groupes de gènes d'intérêt. La matrice $X^{(g)}$ peut être transformée, après addition des colonnes des gènes d'intérêt composant chaque AFT, en une matrice $X^{(AFT)}$ de dimension $N_s \times N_{AFT}$ de fréquence d'AFT (voir Fig. 2.1).

La seconde étape de simplification est une méthode de réduction de dimension (Non-negative Matrix Factorization, ou NMF). La matrice $X^{(AFT)}$ est interprétée comme le mélange pondéré de k profils fonctionnels "universels" caractérisés par un profil de distribution d'AFTs : pour un échantillon i et un AFT j , la fréquence $X_{i,j}^{(AFT)}$ est vue comme la somme des fréquences de l'AFT j dans chacun des k profils fonctionnels, pondérée par un poids W_k du profil dans l'échantillon.

Plus précisément, nous avons résolu le problème suivant : trouver la matrice de poids $W^{(AFT)}$

et la matrice de profils universels $H^{(AFT)}$ tels que

$$W^{(AFT)}, H^{(AFT)} = \underset{\substack{W \in \mathbb{R}^{N_s, k}, H \in \mathbb{R}^{k, N_{AFT}} \\ W \geq 0, H \geq 0 \\ F \cdot H \leq 0}}{\operatorname{arg\,min}} \left(\| (X^{(AFT)} - W \cdot H) \cdot D^{-1} \|_2^2 + \mu \left(\|W\|_2^2 + \|H \cdot D^{-1}\|_{1,2}^2 \right) \right). \quad (2.1)$$

où D est une matrice diagonale de renormalisation et F est une matrice définissant des contraintes linéaires entre gènes telles qu'observées dans des génomes représentatifs du microbiote intestinal. La norme $L_{1,2}$ sur le terme de pénalisation sur H permet d'obtenir une matrice H la plus creuse possible : cette norme assure qu'un AFT donné soit porté par au moins un profil fonctionnel, mais qu'il soit présent dans le plus petit nombre possible de profils. Les hyper-paramètres k et μ sont sélectionnés selon différents critères (bi-cross validation, erreur de reconstruction et stabilité d'identification de H).

Le problème de NMF (2.1) est d'abord résolu sur une base de données d'apprentissage composée de 1153 échantillons comportant des métagénomés de sujets sains ou souffrant de maladies métaboliques ou inflammatoires, ce qui permet d'obtenir la matrice de profil $H^{(AFT)}$. Ensuite, l'"universalité" de ces profils est validée sur une large base de données de 2571 échantillons externes provenant de sujets au statut pathologique varié par régression positive (NNLS, comme non-negative least square). Plus précisément, en notant $X_{test}^{(AFT)}$ la matrice de fréquence des AFTs de la base de donnée de test, on cherche W_{test} avec la régression suivante :

$$W_{test}^{(AFT)} = \underset{\substack{W \in \mathbb{R}^{N_s^{test}, k} \\ W \geq 0}}{\operatorname{arg\,min}} \left(\| (X_{test}^{(AFT)} - W \cdot H^{(AFT)}) \cdot D^{-1} \|_2^2 + \mu \left(\|W\|_2^2 \right) \right). \quad (2.2)$$

Enfin, afin de caractériser des profils taxonomiques co-variant avec les profils fonctionnels $H^{(AFT)}$, nous résolvons le problème de régression positive suivant, sur une matrice $X^{(otu)}$ de fréquences de taxon,

$$H^{(otu)} = \underset{\substack{H \in \mathbb{R}^{k, N_{AFT}} \\ H \geq 0}}{\operatorname{arg\,min}} \left(\| (X^{(otu)} - W^{(AFT)} \cdot H) \cdot D^{-1} \|_2^2 + \mu \left(\|H \cdot D^{-1}\|_{1,2}^2 \right) \right). \quad (2.3)$$

Ces différentes étapes sont résumées dans la Figure 2.1.

3 Résultats

Les différentes matrices obtenues dans les différents problèmes d'optimisation ont leur propre interprétation biologique.

3.1 Des profils fonctionnels et taxonomiques reconstituant les métagénomés

Tout d'abord, la matrice H^{AFT} peut être interprétée comme quatre profils fonctionnels (les lignes de la matrice) dont la fréquence en AFT renseigne sur les caractéristiques fonctionnelles intrinsèques du profil. Le profil 1 est représenté dans la figure 2.2-a : ce profil est tout d'abord caractérisé par une forte prévalence de GH (70%), c'est à dire d'enzymes dégradant les fibres en sucres simples, comparée aux KO (30%) permettant de décomposer les sucres en métabolites

secondaires. Au sein des GH, le profil est particulièrement marqué par une prépondérance des enzymes dégradant les membranes cellulaires des plantes, et des enzymes libérant du galactose. Au sein des KO, certaines voies métaboliques sont prépondérantes, comme l'EMP (Embden-Meyerhoff-Parnas pathway).

La matrice $H^{(otu)}$ quant à elle renseigne sur la composition taxonomique des différents profils fonctionnels. La fréquence en OTU de chaque profil peut être représentée dans des graphiques en camembert (Fig. 2.2-b). Nous voyons ici que le profil 1 est principalement composé de bactérie du phylum bacteroidetes, notamment *Bacteroides vulgatus*, alors que le profil 2 est caractérisé par une forte prévalence des firmicutes, notamment *Eubacterium rectale*.

Enfin, lorsqu'on moyenne $W^{(AFT)}$ sur les lignes, on obtient le vecteur de poids moyen $\bar{W}^{(AFT)} = \frac{1}{N} \sum_{i=1}^N W_i^{(AFT)}$. Pour un profil $k = 1, \dots, 4$ et un AFT $j = 1, \dots, 101$, la quantité $\bar{W}_k^{(AFT)} H_{kj}^{(AFT)}$ représente donc la contribution du profil k à la fréquence de l'AFT j au métagénome. Cette quantité permet de mesurer la contribution écologique des différents profils aux différents AFT considérés dans le modèle (Fig. 2.2-c). Le même procédé peut être effectué sur les micro-organismes. Ainsi, on peut déterminer que le profil 1 est le principal pourvoyeur de gènes codant pour les GH dans les métagénomes (Fig. 2.2-c, panel du haut, barres de droite à partir de "starch and fructans") et de Bacteroidetes (panel du bas, barres de gauche).

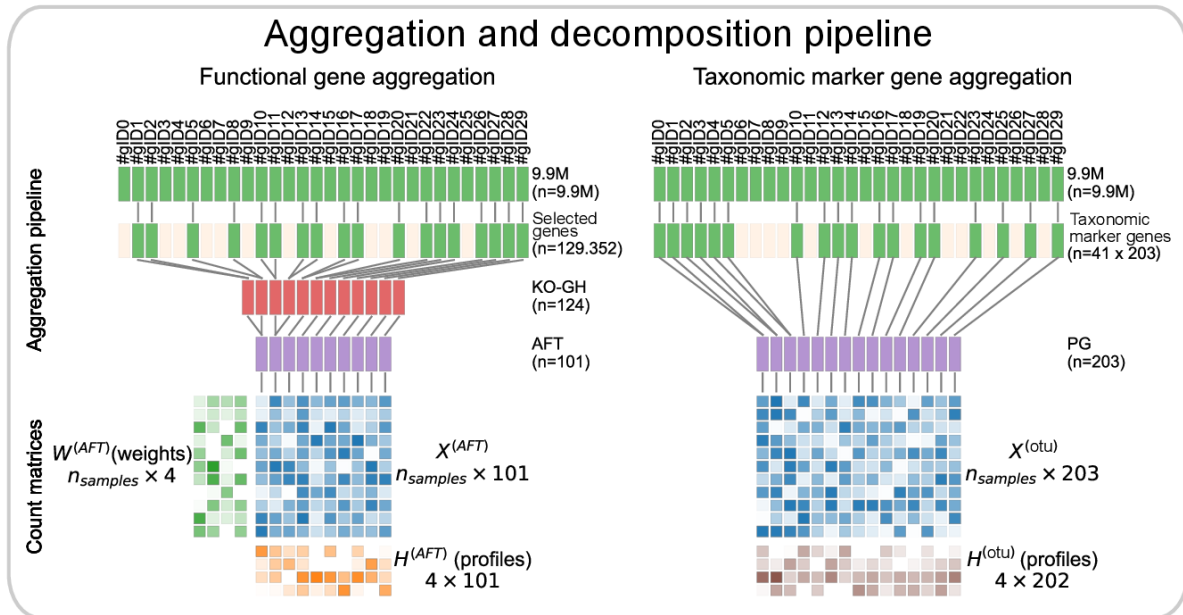
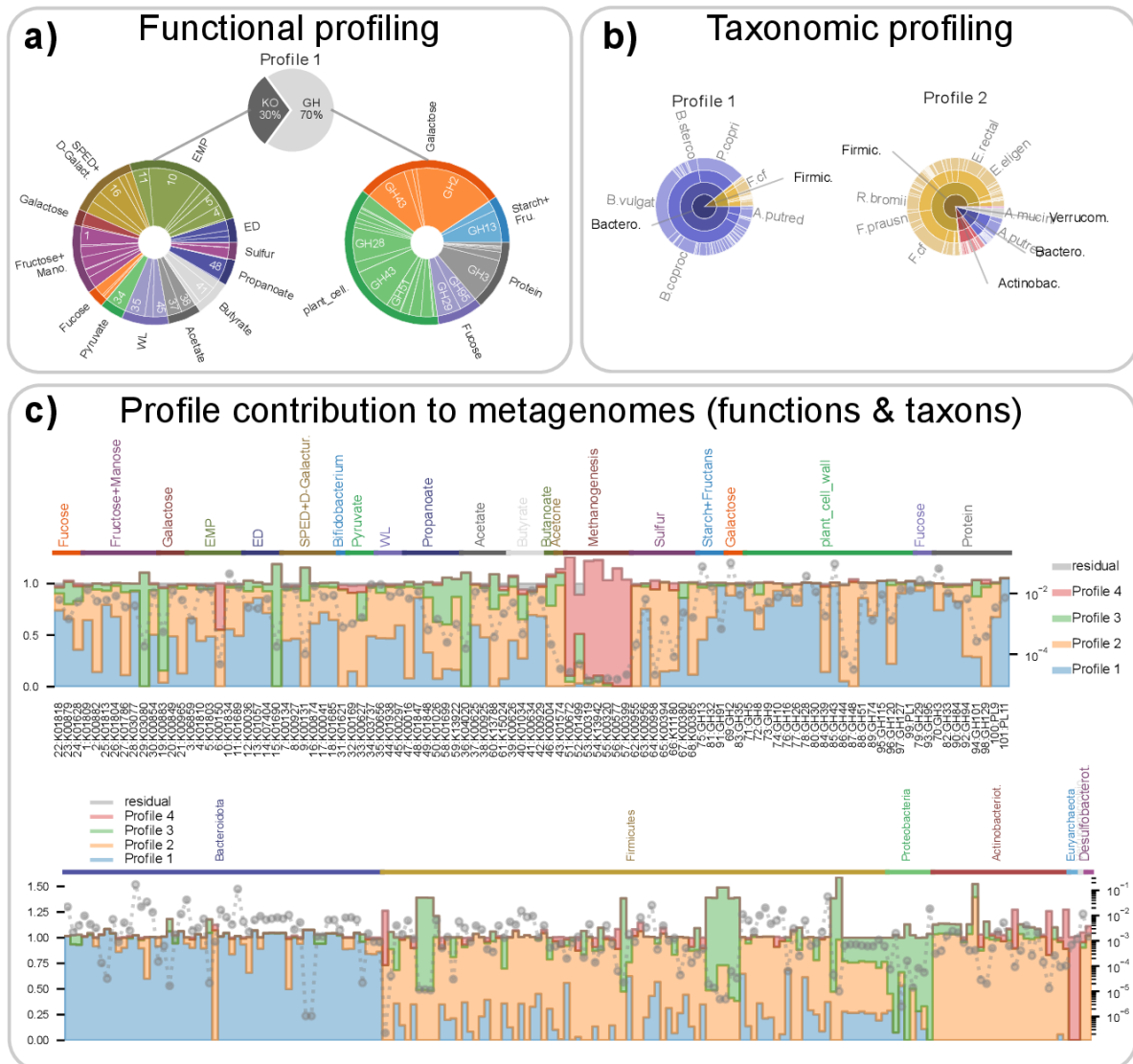


Fig. 2.1

Pipeline de la NMF. Les différents pipelines de construction des matrices de comptage et de décomposition par NMF ou NNLS sont d'abord présentés. Parmi les 9.9M de gènes du catalogue, seuls les gènes sélectionnés sont conservés ($n = 129352$) et leurs comptages sont agrégés par KEGG orthology (KO) ou famille de glycoside hydrolase (GH), qui sont encore agrégés en trait fonctionnel agrégé (AFT). Une matrice de comptage d'AFT $X^{(AFT)}$ est ainsi obtenue et décomposée par NMF ($X^{(AFT)} \simeq W^{(AFT)} H^{(AFT)}$). En repérant dans le catalogue des gènes marqueurs taxonomiques, des matrices de comptage de micro-organismes peuvent être assemblées. On peut alors trouver les micro-organismes co-variant avec les profils fonctionnels (matrice $H^{(AFT)}$) en déterminant $H^{(otu)}$ par régression positive $X^{(otu)} \simeq W^{(AFT)} H^{(otu)}$.


Fig. 2.2

Profils fonctionnels obtenus et contribution des profils au métagénome. a) et b) Les matrices $H^{(AFT)}$ (profils fonctionnels, a) et $H^{(otu)}$ (profils taxonomiques, b) peuvent être caractérisées. Ces graphiques en camembert représentent la proportion des différents AFT ou micro-organismes dans le profil. Pour le profil fonctionnel, on représente d'abord la proportion de KO et de GH, puis les proportions des AFT dans chaque classe (KO ou GH) sont représentées. Nous présentons ici les profils fonctionnels et taxonomiques pour certains profils seulement, arbitrairement choisis pour illustrer leur interprétation biologique. Ces profils représentent un potentiel fonctionnel ou taxonomique. c) En calculant pour chaque profil k et chaque AFT j la proportion $\frac{\bar{W}_k^{(AFT)} H_{k,j}^{(AFT)}}{\bar{X}_j^{(AFT)}}$, où $\bar{W}^{(AFT)} = \frac{1}{N_s} \sum_{i=1}^{N_s} W_i^{(AFT)}$ et $\bar{X}^{(AFT)} = \frac{1}{N_s} \sum_{i=1}^{N_s} X_i^{(AFT)}$ sont les poids et comptages moyens sur l'ensemble de la cohorte, on peut déterminer la contribution de chacun des profils au métagénomes, et ainsi déterminer quel profil est le principal contributeur pour chaque AFT considéré. La même chose peut être réalisée pour les comptages taxonomiques. Les pointillés gris représentent les fréquences des AFT ou des génomes dans les métagénomes : nous voyons ainsi que les AFT les moins représentés sont les gènes correspondant à la méthanogénèse, et qu'ils sont principalement pourvus par le Profil 4.

3.2 Stratification de la matrice de poids

Après avoir considéré les matrices $H^{(AFT)}$ et $H^{(otu)}$, on peut également étudier la matrice de poids $W^{(AFT)}$ qui décrit le poids de chaque profil dans les échantillons. En particulier, nous pouvons voir si les poids des différents profils se stratifient en fonction des groupes d'individus auxquels appartiennent les échantillons. Afficher les poids W_1 , W_2 et W_3 dans un ternary plot (Fig. 2.3-a), c'est à dire un graphe dans lequel les poids normalisés $(W_{i,j}/(W_{i,1} + W_{i,2} + W_{i,3}))_{1 \leq i \leq N_s, 1 \leq j \leq 3}$ représentent les coordonnées barycentriques des points, permet de visualiser les poids prépondérants dans chaque groupe. Nous pouvons voir que le poids W_3 est assez peu représenté chez les sujets sains, qu'ils soient dysbiotiques (N&dys) ou non (N&Not_dys). A l'inverse, le poids relatif du profil W_3 est plus important chez les malades de Crohn (CD&Not_dys), notamment dysbiotique (CD&dys). En regardant les distributions de ces poids dans des *bar plots*, on constate à nouveau cette augmentation significative du profil W_3 chez les malades de Crohn. On constate également une baisse concomitante du profil W_2 dans cette population.

On peut également calculer les poids moyens du groupe $\tilde{W}_k^g = \frac{1}{N_g} \sum_{i \in g} W_{i,k}$, pour $k = 1, \dots, 4$, où g est le groupe et N_g l'effectif du groupe. Ces poids moyens permettent de calculer la contribution de chaque profil aux différents AFT dans l'individu moyen du groupe (Fig. 2.3-b). Par contribution moyenne, j'entends la valeur $\tilde{W}_k^g H_{k,j}^{(AFT)} / \sum_{k=1}^4 \tilde{W}_k^g H_{k,j}^{(AFT)}$. Pour chaque AFT, cette contribution moyenne est re-normalisée pour les profils W_1 , W_2 et W_3 de façon à représenter ce triplet par une couleur dans un triangle Rouge-Vert-Bleu. Par exemple, l'AFT 15, représenté en jaune, est ainsi principalement pourvu par le profil 3 dans le groupe sain non dysbiotique (N&Not_dys, panel de gauche, Fig. 2.3-b). Pour le groupe de malades de Crohn dysbiotique, nous regardons les AFT pour lesquels le vecteur de contribution des profils W_1 , W_2 et W_3 a significativement changé (après correction pour test multiple), et nous conservons les 20 modifications les plus importantes pour faciliter la représentation. Nous pouvons voir que les changements les plus importants sont pour certains AFT dont le principal contributeur passe de W_1 à W_3 (AFT 12, 67, 23 et 28), alors que d'autres AFT (plus bas dans le réseau métabolique de dégradation des fibres) passent de W_2 à W_3 comme contributeur principal.

Ces représentations permettent de facilement identifier quelles fonctions sont principalement affectées par les variations de composition du microbiote observées dans les différents groupes.

4 Perspective : utilisation des résultats de la NMF pour construire un modèle de dynamique de population

Comme les profils contenus dans la matrice H sont directement reliés au réseau métabolique de dégradation des fibres et donc interprétables biologiquement, ils peuvent être utilisés pour bâtir les fonctions sources G_b et G_m d'un système (1.1)-(1.2) contenant $N_b = k$ populations fonctionnelles. Plus précisément, il s'agit de construire un modèle de prédiction des flux métaboliques traversant le réseau métabolique. Nous pouvons pour cela opter pour une modélisation de type cinétique : une première fonction détermine le taux de consommation des différents métabolites d'entrée dans le réseau, et des règles conservatives de partage de flux permettent ensuite de les partager lors des différentes réactions situées en aval des chemins métaboliques. Cette modélisation permettrait de modéliser l'évolution des poids W dans le métagénome. Multiplié à $H^{(AFT)}$, ces poids W permettent de reconstruire des métagénomés simulés pouvant être comparés à des métagénomés expérimentaux. Ce travail sera finalisé et valorisé dans les prochains mois.

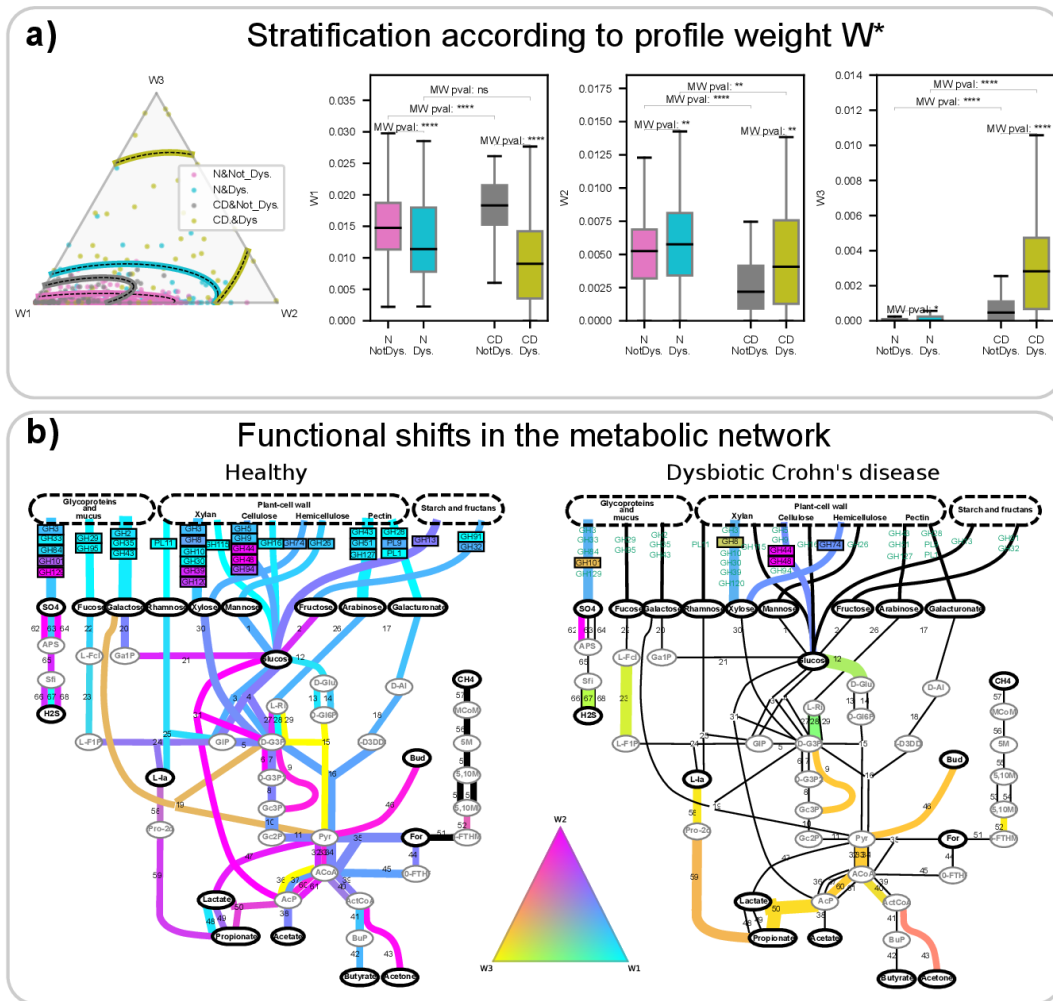


Fig. 2.3

Stratification de la population selon la matrice de poids W , et modifications fonctionnelles correspondantes. a) Nous pouvons représenter les variations des poids W dans des groupes d'individus ayant des status physiopathologiques différents. Les différents échantillons sont d'abord représentés dans un ternary plot : les fréquences $W_{ij}/(\sum_{k=1}^3 W_{ik})$ des profils $j = 1, 2, 3$ peuvent être interprétées comme des coordonnées barycentriques, ce qui permet de représenter les échantillons $i = 1, \dots, N_s$ dans un triangle. Les points sont colorés par groupes : sains et non dysbiotiques (N&Not_dys), sains et dysbiotiques (N&dys), malades de Crohn et non dysbiotiques (CD¬_dys) et malades de Crohn et dysbiotiques (CD&Dys). Une ellipse de confiance à 95% est représentée pour chaque groupe. Nous voyons que W_3 est très peu représenté chez les sujets sains, alors que les sujets malades de Crohn et dysbiotiques ont une forte augmentation de W_3 , et une diminution de W_2 comparé aux sains. Ces observations sont corroborées par les barplots (panel de droite). b) On peut ensuite calculer le poids moyen par groupe, et calculer ensuite la contribution pour chaque groupe des différents AFT, et représenter la part de cette contribution dans un triangle de couleur : plus la couleur de l'AFT tire vers le bleu, plus W_1 pourvoit au comptage de l'AFT. Ces contributions sont représentées pour le groupe sain non dysbiotique (panel de gauche). Comme ces contributions dépendent de W , on peut observer les changements métaboliques caractérisant chaque groupe physiopathologique. Afin de mieux visualiser les changements, on représente les 20 AFT dont la contribution a le plus significativement bougé (après correction pour test multiple). On peut constater que les AFT 12, 23 et 28, majoritairement portés par W_1 chez les sujets sains, sont maintenant substantiellement apportés par W_3 chez les malades de Crohn dysbiotiques (panel de droite).

C H A P I T R E



Approches réductionnistes pour l'étude de communautés microbiennes

1 Modèles de communautés microbiennes réduites.

Une approche réductionniste pour analyser des interactions microbiennes consiste à sélectionner parmi la communauté un petit nombre d'individus dont le potentiel fonctionnel est représentatif de la communauté. Une fois isolés, ces N_b organismes peuvent être cultivés en communautés synthétiques (SynCom) de manière contrôlée et répétable. Du point de vue de la modélisation, l'avantage est de pouvoir séquencer complètement les individus de la SynCom et dériver après annotation fonctionnelle du génome un réseau métabolique représentatif de chaque espèce. Ce réseau métabolique peut être représenté par une matrice de réaction R contenant autant de lignes n_m que de métabolites et autant de colonnes n_r que de réactions dans le réseau, chaque élément de la matrice étant le coefficient stœchiométrique pour le métabolite m dans la réaction r (voir Fig. 1.2). Il s'agit ensuite de prédire les flux métaboliques de métabolites transformés par le réseau pour dériver les fonctions $G_{b,i}$ et $G_{m,j}$, c'est à dire les taux de croissance des équations de dynamiques de population (1.1)-(1.2), avec les relations

$$G_{b,i} = \nu_b^{(i)} b_i \quad G_{m,j} = \sum_{i=1}^{N_b} \nu_{m,j}^{(i)} b_i \quad (3.1)$$

où $\nu_b^{(i)}$ est le flux de production de biomasse pour la bactérie i et $\nu_{m,j}^{(i)}$ est le flux de production (ou de consommation) du métabolite j par la bactérie i . Une méthode classique pour prédire les flux $\nu_b^{(i)}$ et $\nu_{m,j}^{(i)}$ est la méthode de Flux Balance Analysis (FBA) résolvant le problème de programmation linéaire

$$\text{Trouver } \nu_b^*, \nu^* \text{ tels que } \nu_b^*, \nu^* = \begin{array}{l} \operatorname{argmax} \quad \nu_b. \\ R \cdot \nu = 0 \\ \nu_l(b, m) \leq \nu \leq \nu_u(b, m) \end{array} \quad (3.2)$$

Les contraintes $\nu_l(b, m)$ et $\nu_u(b, m)$ sont des fonctions dépendant des variables d'état du système (1.1)-(1.2) permettant de coupler fortement les équations (1.1)-(1.2) et 3.2 via la fonction

$$\mathcal{F} : (\nu_l(b, m), \nu_u(b, m)) \mapsto \nu_b^*, \nu^*. \quad (3.3)$$

L'avantage de cet environnement de modélisation, connu sous la dénomination de *dynamic FBA*, est qu'il prend en compte l'ensemble des réactions métaboliques contenues dans le génome de chaque espèce. Il permet, en résolvant les équations (1.1)-(1.2) couplées pour le calcul de la fonction source au problème (3.2), de reproduire des données omiques obtenues sur des SynCom cultivées expérimentalement afin de comprendre finement le fonctionnement de la communauté.

2 Modèle d'une communauté fermentaire pour la production de fromage

Afin de fixer les idées, nous allons étudier la modélisation de SynCom réduite au travers d'un exemple : le modèle d'une communauté fermentaire pour la production de fromage. Ce modèle est une représentation mathématique d'une expérimentation de microbiologie (une ombre numérique, pour reprendre la terminologie issue de la littérature sur les jumeaux numériques). Ce modèle n'est pas spatialisé. C'est un système d'équations différentielles ordinaires : les opérateurs différentiels spatiaux des équations (1.1)-(1.2) ne sont pas considérés. Une SynCom de trois bactéries, deux bactéries lactiques (*L.lactis* et *L.plantarum*) et une bactérie d'affinage (*P.freudenreichii*), est définie. Le génome de chaque bactérie est séquencé, à partir duquel un réseau métabolique est assemblé (Fig. 3.1.a). Ces réseaux sont utilisés, via des fonctions $\nu_l(b, m|\theta)$ et $\nu_u(b, m|\theta)$ à définir et paramétrer et via l'équation (3.2), pour assembler les fonctions sources des systèmes dynamiques (voir eq. (3.1)).

Pour caractériser le fonctionnement de chaque bactérie, des expérimentations de croissance sur du lait sont d'abord menées. Du lait estensemencé par une seule souche bactérienne. Les populations bactériennes, ainsi que le pH (pour les bactéries lactiques) sont suivis en dynamique. La production de certains métabolites est également mesurée en fin de croissance pour *P.freudenreichii*, pour lequel le lait est également supplémenté en lactate. Ces expériences en mono-culture sont utilisées comme données d'apprentissage pour définir et inférer les fonctions $\nu_l(b, m|\theta)$ et $\nu_u(b, m|\theta)$ (Fig. 3.1 b). Après inférence, les modèles de mono-culture sont précis ($R^2 = 0.99$, Fig. 3.2 a).

Un deuxième jeu de données est ensuite assemblé, avec un suivi dynamique des populations microbiennes, de métabolites et du pH lors de la fabrication d'un fromage. Ces données de co-culture, indépendantes des données de mono-culture, servent pour la validation du modèle communautaire (Fig. 3.2 c), avec également une bonne précision ($R^2 = 0.98$, Fig. 3.2 a). Ce modèle permet de déterminer la contribution de chaque micro-organisme à la production des métabolites, notamment ceux impliqués dans les qualités gustatives du fromage (Fig. 3.2 b).

Ce travail a été effectué dans le cadre du projet TANGO, porté par H.Falentin, et de la thèse de M.Lecomte. Il a fait l'objet d'une publication [2] et de plusieurs communications dans des conférences.

Un travail similaire dans l'esprit a été réalisé dans le cadre de la thèse d'A.Paulay, avec l'inclusion d'un module modélisant la dégradation des protéines dans un modèle métabolique d'une bactérie commensale du microbiote intestinal [3]. Ce module, validé sur des données de croissance microbienne, a permis de capturer l'activité protéolytique des bactéries, et de mieux comprendre leur utilisation des acides aminés dérivés de la protéolyse.

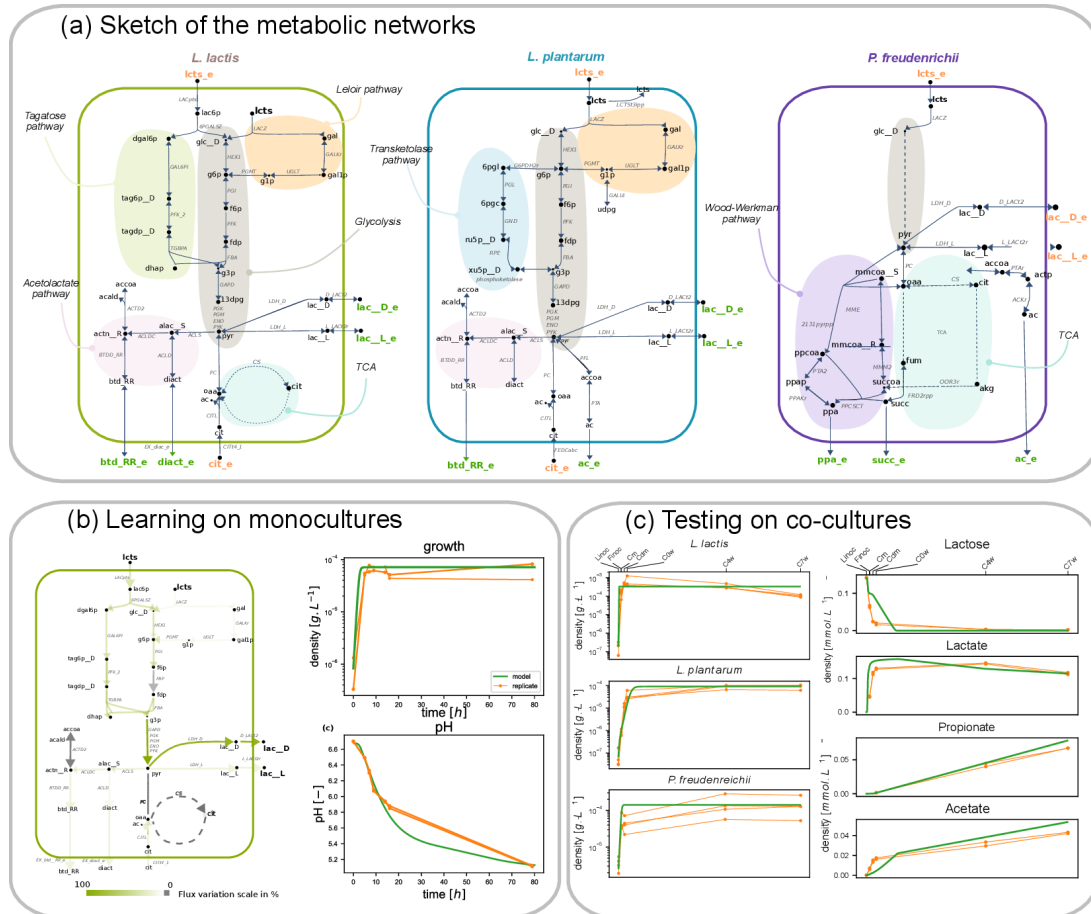


Fig. 3.1

Modélisation de la communauté fromagère simplifiée. (a) Réseaux métaboliques. Une vue simplifiée des trois réseaux métaboliques est présentée. Pour chaque réseau, les voies métaboliques principales sont représentées, ainsi que les métabolites utilisés comme substrats (vert) et les métabolites excrétés (orange). (b) Apprentissage. les paramètres d'un modèle dFBA sont inférés pour chaque micro-organismes à partir de données obtenues en mono-culture. Des souches pures sont cultivées sur du lait, et leurs courbes de croissance sont mesurées sur 2 réplicats expérimentaux, ainsi que l'acidification du milieu (pH). (c) Validation. lorsque les modèles de chaque souche sont combinés en communauté, et que des paramètres de saturation sont adaptés aux données, le modèle obtenu reproduit avec précision la dynamique de chaque souche, ainsi que les concentrations des métabolites principaux obtenues dans les expérimentations en co-culture. Figures issues de [2].

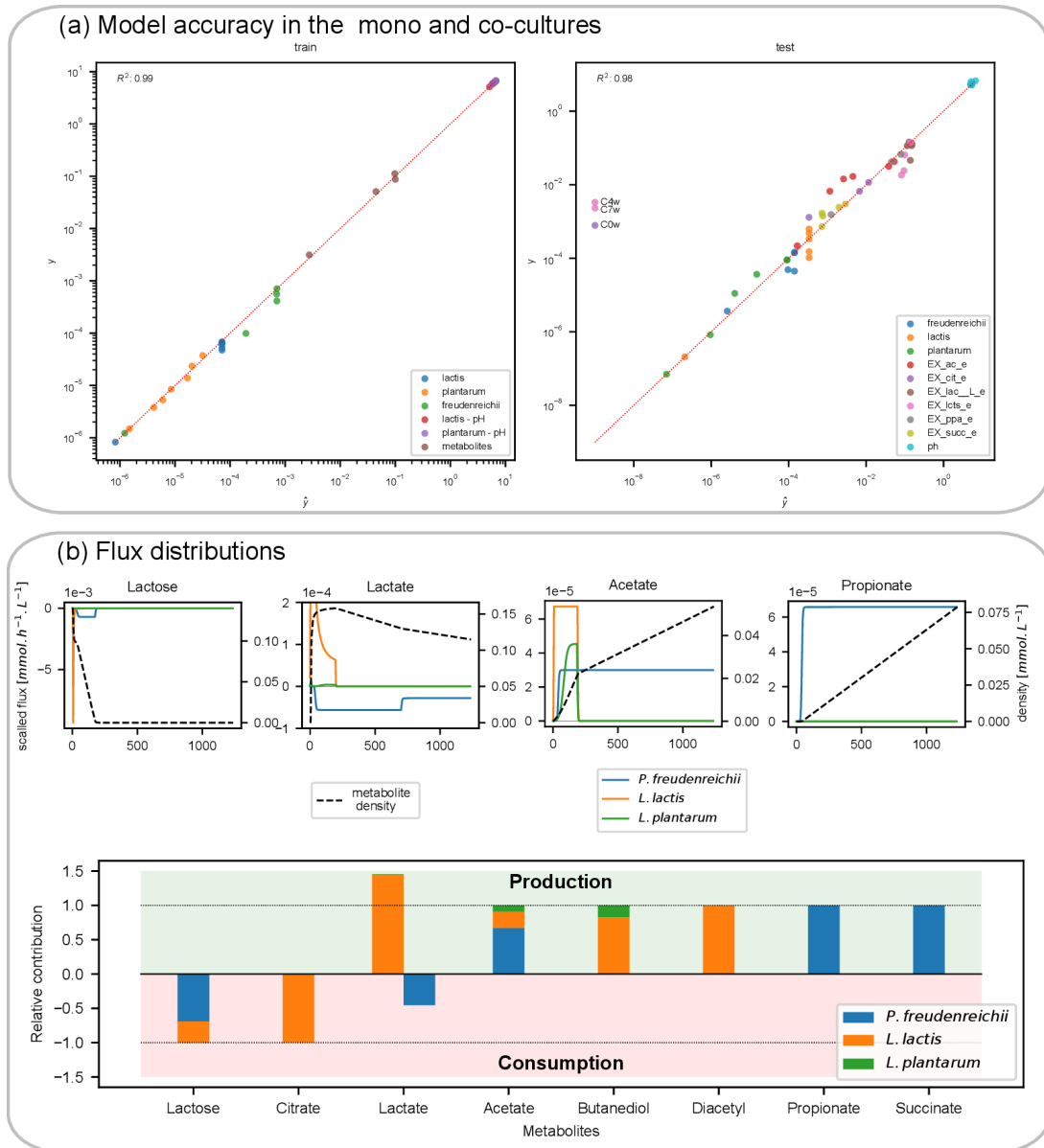


Fig. 3.2

Vérification de la précision des modèles, et contribution de chaque souche à la production de composés organoleptiques. (a) *Précision*. Les données simulées (abscisses) sont affichées en regard des données expérimentales (ordonnées) pour le jeu d'entraînement en mono-culture (gauche) et pour le jeu de validation en co-culture (droite). Le coefficient de détermination est indiqué pour chaque jeu de donnée. Les différents composés sont codés par couleur. (b) *Répartition des flux et contribution des différentes souches*. Pour chaque souche, l'évolution des flux de production et de consommation est représentée pour différents composés (panel du haut). La contribution de chaque souche à la production (ou consommation) totale de chaque métabolite est ensuite représentée (panel du bas). Figures issues de [2].

3 Approximation de modèle par *machine learning*.

La résolution du problème (1.1)-(1.2) couplé à (3.2) nécessite la résolution du problème d'optimisation (3.2) à chaque pas d'espace et de temps du système d'EDP (1.1)-(1.2). Bien que des solveurs de problèmes de programmation linéaire particulièrement efficaces soient disponibles, ce couplage représente une charge computationnelle prohibitive lorsque la communauté est grande, lorsque les réseaux métaboliques sont très grands, comme les réseaux de cellules humaines, ou pour la résolution d'EDP, où un modèle FBA doit être résolu à chaque pas de temps et d'espace.

Nous avons donc développé une approche de métamodélisation permettant de calculer une solution approchée du problème (3.2), et donc de la fonction \mathcal{F} définie en (3.3), pour un coût numérique bien moindre (voir Fig. 3.3 a). Concrètement, nous avons assemblé une base d'apprentissage composée d'une matrice $\mathbf{X} = (\nu_l(b_s, m_s), \nu_u(b_s, m_s))_{1 \leq s \leq N_s}$ de contraintes du problème (3.2) calculées pour N_s points de l'espace d'état (b, m) du système (1.1)-(1.2), ainsi que de la matrice $\mathbf{Y} = \mathcal{F}(\mathbf{X})$. Ensuite, nous avons approché \mathcal{F} dans un espace de RKHS particulier, appelé ANOVA-RKHS. L'ANOVA-RKHS, comme tous les espaces RKHS, permet de discrétiser le problème d'approximation de la fonction \mathcal{F} , via le *Representer theorem*. L'ANOVA-RKHS permet également de construire une approximation de la décomposition de Hoefding de la fonction \mathcal{F} , ce qui permet de sélectionner les groupes de co-variables participant de façon prépondérante à la reconstruction de \mathcal{F} . Plus précisément, il s'agit de résoudre le problème d'inférence, pour une colonne j de \mathbf{Y} ,

$$\hat{\theta}_{0,j}, (\hat{\theta}_{p,j})_{p \in \mathcal{P}} := \underset{\substack{\theta_{0,j} \in \mathbb{R} \\ \theta_{p,j} \in \mathbb{R}^{N_s}, \forall p \in \mathcal{P}}}{\operatorname{argmin}} \quad \|\mathbf{Y}_j - (\theta_{0,j} \mathbb{1} + \sum_{p \in \mathcal{P}} K_p \theta_{p,j})\|_2^2 + \lambda \sum_{p \in \mathcal{P}} \|\theta_{p,j}\|_2. \quad (3.4)$$

Dans cette équation, \mathcal{P} est l'ensemble des multi-indices sur les colonnes de \mathbf{X} , K_p est une matrice de Gram d'un noyau reproduisant k_p défini sur les colonnes de \mathbf{X} correspondant au multi-indice p et λ est un hyperparamètre de régularisation pour le terme group-lasso $\sum_{p \in \mathcal{P}} \|\theta_{p,j}\|_2$. Ce terme group-lasso permet la sélection de variable (voir Fig. 3.3 b). Avant inférence, une normalisation min-max est appliquée sur les co-variables \mathbf{X} .

Après inférence, une approximation de \mathcal{F} en un nouveau point (ν_l, ν_u) externe à la base d'apprentissage peut être obtenu avec la formule explicite

$$\hat{\mathcal{F}}_j(\nu_l, \nu_u) := \hat{\theta}_{0,j} + \sum_{p \in \mathcal{P}} F_p(\nu_l, \nu_u) \cdot \hat{\theta}_{p,j} \quad (3.5)$$

où $F_p(\nu_l, \nu_u)$ est le vecteur de dimension N_s obtenu par l'évaluation du noyau $k_p((\nu_l, \nu_u), \cdot)$ en les N_s points de \mathbf{X} . Cette approximation est beaucoup plus rapide à calculer que la résolution du problème FBA correspondant.

Dans l'article [5], nous avons mis en œuvre cette méthode de métamodélisation pour une SynCom de deux bactéries dans un système décrivant une infection à la *Salmonelle*. Après sélection du paramètre group-lasso (voir Fig. 3.4 a) et du nombre d'observations dans la base d'apprentissage (voir Fig. 3.4 b), nous avons obtenu une accélération par un facteur 50 du temps de calcul pour une précision acceptable dans l'approximation de la dynamique du système. Le métamodèle obtenu donne un coefficient de détermination de 0.99 sur une base de données de validation indépendante (voir Fig. 3.4 c). Lorsque le FBA est remplacé par son métamodèle dans le système dynamique, les sorties du modèle obtenu sont également très fidèles : les courbes d'évolution sont pratiquement confondues (voir Fig. 3.4 d), alors que le temps de calcul est drastiquement réduit. Cet axe de recherche a bénéficié d'une action Exploratoire de l'Inria (Aex SLIMMEST) qui a financé le post-doc de Pablo Ugalde Salas que j'ai encadré.

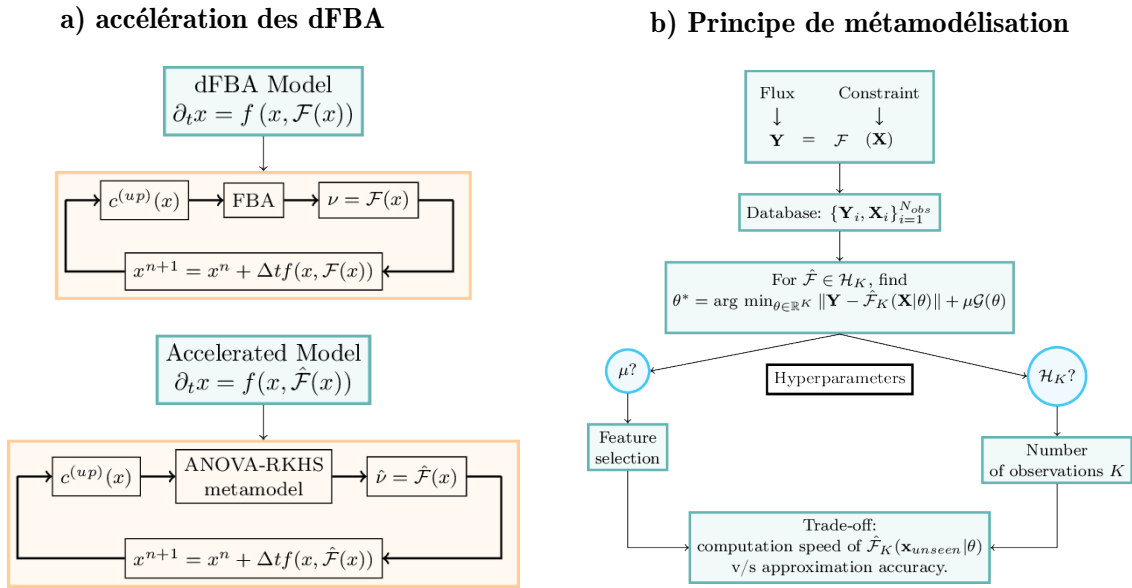


Fig. 3.3

Description de la stratégie de métamodélisation. *a) accélération du dFBA.* Le dFBA (panel supérieur) se caractérise par le couplage d'un modèle métabolique de type FBA avec un système dynamique. Numériquement, cela revient à itérer un schéma d'intégration temporel dans lequel un FBA est résolu à chaque pas de temps. L'alternative proposée (panel inférieur) consiste à remplacer le modèle FBA par un métamodèle plus sobre numériquement, ce qui accélère l'intégration temporelle. *b) cadre de métamodélisation.* Le cadre général de métamodélisation est défini, où \mathbf{Y} est la sortie du modèle FBA \mathcal{F} étant données les entrées \mathbf{X} . Une base de données d'apprentissage est alors assemblée en échantillonnant l'espace des variables d'entrée (\mathbf{X}_i) et en calculant les sorties du FBA \mathbf{Y}_i correspondantes, avec N_{obs} observations. Le métamodèle est alors défini comme la solution d'un problème de régression non linéaire non paramétrique dans un espace fonctionnel de dimension fini \mathcal{H}_K (ANOVA-RKHS) de dimension K , avec un terme de régularisation \mathcal{G} . En pratique, une régularisation de type group-lasso est retenue pour sélectionner des variables d'entrées. Ce problème de régression a deux hyperparamètres à déterminer : le coefficient de régularisation μ , qui détermine le nombre de variables d'entrée sélectionnées, et la dimension K des espaces fonctionnels, qui est directement reliée au nombre d'observations N_{obs} dans les méthodes RKHS. Sélectionner un plus petit nombre de co-variables, ou diminuer la dimension K , réduit la charge computationnelle nécessaire pour évaluer le métamodèle en un nouveau point, accélérant ainsi les temps de calculs. Mais cela grève également la précision de l'approximation fournie par le métamodèle. Un compromis doit donc être recherché entre accélération et erreur d'approximation.

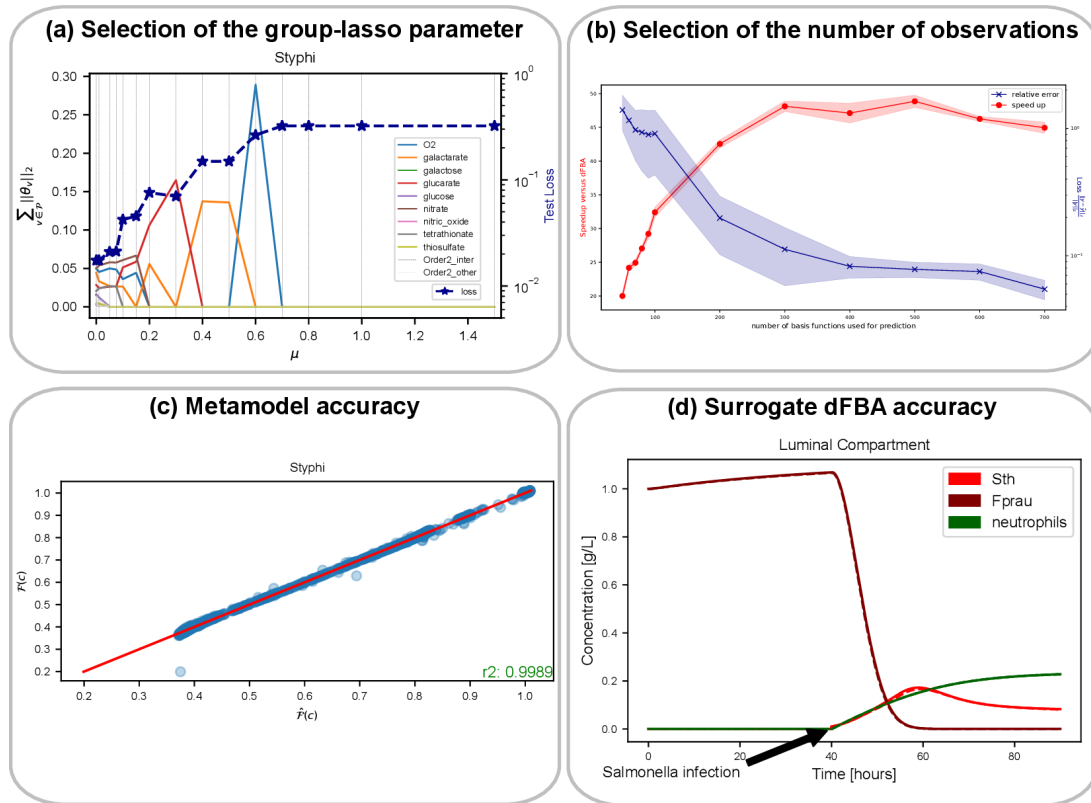


Fig. 3.4 Sélection d'hyperparamètres et précision du métamodèle. *a)* Sélection du paramètre group-lasso. Nous présentons ici un "lasso path" : la contribution de chaque groupe de paramètres est affichée en abscisse, alors que le paramètre de régularisation du terme group-lasso est représenté en ordonnée. Les différentes courbes représentent la contribution de chaque groupe de variables d'entrée : soit des variables individuelles (indiquées par un code couleur), soit des groupes de variables (interactions d'ordre 2) représentés par des lignes en pointillés gris. Lorsque le paramètre de régularisation augmente, la majorité des variables voit sa contribution s'annuler, alors que ne persistent que les plus significatives. En pointillés bleus (échelle en ordonnées à droite) est représentée l'évolution de la fonction coût. Lorsque des groupes de variables s'annulent, la fonction coût augmente. *b)* Sélection du nombre d'observations. À valeur de paramètre de group-lasso fixée, nous faisons varier le nombre d'observations dans le jeu d'apprentissage. Nous représentons en rouge l'accélération du temps de calcul et en bleu l'erreur d'approximation du métamodèle. Comme attendu, l'erreur d'approximation diminue avec le nombre d'observations dans le jeu d'apprentissage. L'évolution de l'accélération est plus inattendue : pour de faibles valeurs du nombre d'observations, l'accélération est faible. Ceci est dû au fait que de nombreux groupes de variables sont conservés pour cette valeur de régularisation group-lasso. L'accélération atteint un maximum pour $N_{obs} = 500$, puis décroît à nouveau ensuite. *c)* Précision du métamodèle. Pour une colonne de Y correspondant à la croissance d'une bactérie, nous représentons les valeurs observées dans le jeu de données de validation (ordonnée) et prédites par le métamodèle (abscisse). Le coefficient de détermination est indiqué. Le métamodèle est très précis. *d)* Précision du système dynamique quand le modèle est remplacé par le métamodèle. Nous représentons la dynamique de trois variables d'état (la concentration de 2 bactéries et de neutrophiles) dans le système lorsque la dynamique est calculée par un dFBA (traits pleins) ou lorsqu'elle est calculée par un métamodèle (traits pointillés). Les courbes sont quasiment confondues.

Interaction du microbiote avec son environnement

Jusqu'à présent, nous nous sommes principalement intéressés aux interactions entre micro-organismes au sein d'un écosystème microbien, que ce soit par des approches holistiques (Chap. 2) ou réductionnistes (Chap. 3). Dans un cas comme dans l'autre, nous nous sommes notamment focalisés sur les interactions nutritionnelles, induites par le métabolisme des micro-organismes. Ces interactions sont médiées par les métabolites présents dans l'environnement, produits ou consommés par les différents microbes, structurant ainsi des réseaux trophiques au sein de la communauté. Cette grille d'analyse des interactions microbiennes au sein de la communauté est très avantageuse, puisqu'elle permet d'intégrer au travers des réseaux métaboliques des différents individus les différentes données omiques qui peuvent être assemblées. Toutefois, cette vision est partielle. D'une part, des interactions autres que métaboliques peuvent être déterminantes. Nous verrons par exemple dans la section 1 un contexte microbien où les interactions d'intérêt pour expliquer les phénotypes observés sont mécaniques. D'autre part, les micro-organismes ne sont pas les seuls acteurs de l'holobionte : l'hôte met en place des régulations dynamiques pour maîtriser son microbiote commensal. Dans la section 2, un modèle dynamique des cryptes intestinales de l'hôte sera présenté. Enfin, dans la section 3, un modèle du microbiote intestinal dans son environnement sera introduit, faisant la synthèse de ces différentes interactions, qu'elles soient métaboliques ou mécaniques, entre micro-organismes ou avec l'hôte.

Une autre grille de lecture, plus mathématique, peut être appliquée aux sections de ce chapitre. Dans la section 1, nous nous intéresserons à l'inférence de modèles individu-centrés afin de déterminer des caractéristiques physiologiques des microbes à l'échelle d'une population. Dans la section 2, nous nous focaliserons sur une question de simplification de modèle, en présentant le passage entre un modèle stochastique individu-centré de cellules épithéliales, et son approximation déterministe continue sous la forme d'une EDP sur la densité de population. Enfin, dans la section 3, nous introduirons un modèle couplant des équations de dynamique de population avec des équations issues de la mécanique des fluides, dont nous introduirons ensuite une limite asymptotique sur une géométrie simplifiée.

1 Interactions de micro-nageurs avec leur environnement dans un biofilm

1.1 Biofilms et micro-nageurs comme facilitateurs de contrôle

Certaines bactéries ont la capacité de former des biofilms : les bactéries sécrètent de larges protéines qui s'entremêlent pour former un mélange visqueux, la matrice extracellulaire, dans lequel elles s'agglomèrent. Sous forme de biofilm, la communauté bactérienne est beaucoup plus résistante aux stress mécaniques ou chimiques (comme un biocide ou un désinfectant). Lorsque les bactéries sont pathogènes, la formation de biofilm pose d'importants problèmes sanitaires, comme par exemple dans des applications médicales ou agroalimentaires. Il y a quelques années, des collaborateurs spécialisés dans l'étude des biofilms par microscopie ont fait l'observation que des bactéries nageuses avaient la possibilité de nager dans des biofilms de *Staphylocoques dorés*, laissant derrière eux des pores ouvertes dans le biofilm et la matrice extracellulaire. Ils ont alors démontré par une preuve de concept que ces pores amélioreraient la diffusivité et la pénétrabilité, et donc l'efficacité de biocides permettant d'éradiquer le biofilm. Ils ont alors criblé une centaine de bactéries nageuses pour déterminer leur impact comme facilitateur de traitement de biofilm.

1.2 Caractérisation de bactéries nageuses dans un biofilm.

Afin d'identifier des cocktails de bactéries nageuses optimisant le traitement des biofilms par biocide, il est devenu primordial de caractériser de manière quantifiable leur nage, notamment leurs interactions avec le biofilm. Dans le cadre du post-doc de Guillaume Ravel, financé par le département Mathnum de l'INRAe et que j'ai co-encadré avec Angelo Iollo et Michel Bergmann, nous avons développé un modèle de marche aléatoire de bactéries nageuses, en interaction avec le biofilm sous-jacent. L'inférence de ce modèle à partir de données de microscopie confocale a permis de caractériser l'interaction de ces nageuses avec le biofilm : des bactéries nageuses de trois espèces de *Bacillus* (*B.pumilus*, *B.sphaericus* et *B.cereus*) sont inoculées dans des biofilm de *Staphylococcus aureus* (Fig. 4.1 a). Les trajectoires de ces bactéries sont ensuite reconstruites, et différentes données sont calculées sur ces trajectoires, comme la vitesse, l'accélération ou l'angle entre deux directions de nage. Ces descripteurs des trajectoires sont co-localisés avec des descripteurs de l'environnement biophysique, comme la densité de biofilm hôte ou des gradients de densité de *Staphylococcus aureus*.

Ces trajectoires sont également modélisées, en associant descripteurs de trajectoires et co-variables environnementales, avec le modèle suivant :

$$d\mathbf{v} = \underbrace{\gamma(\alpha(b(\mathbf{X})) - \|\mathbf{v}\|) \frac{\mathbf{v}}{\|\mathbf{v}\|}}_{\text{sélection de vitesse}} dt + \underbrace{\beta \frac{\nabla b(\mathbf{X})}{\|\nabla b(\mathbf{X})\|}}_{\text{sélection de direction}} dt + \underbrace{\eta dt}_{\text{terme aléatoire}}. \quad (4.1)$$

C'est un modèle de conservation du mouvement supposant que les variations de vitesses \mathbf{v} des micro-nageurs sont issues de deux forces déterministes perturbées par un terme aléatoire gaussien d'intensité η . La première force sélectionne le module de la vitesse de nage, l'amenant vers la valeur $\alpha(b)$ dépendant de la densité locale de biofilm b , à la position \mathbf{X} du nageur, avec un temps de relaxation $1/\gamma$. Ce terme suppose donc que la densité locale de biofilm détermine une vitesse "atteignable" que le nageur est capable de détecter et d'atteindre (Fig. 4.1 b). La fonction α est une fonction affine $\alpha(b) = v_0 + b(v_1 - v_0)$. La seconde force influe sur la direction : elle suppose que le micro-nageur est capable de détecter les gradients de densité de biofilm et d'orienter sa nage vers des gradients favorables avec intensité β . Dans cette équation, les vecteurs vitesse,

accélération, et la densité locale de biofilm b sont donnés par les données d'imagerie. Il s'agit alors d'inférer γ , β et la variance de ϵ , sachant les variables d'état \mathbf{X} , \mathbf{v} et $d\mathbf{v}$, ainsi que les co-variables b et ∇b . Des exemples de sortie de modèles sont calculées sur de fausses images de biofilm, en éteignant un à un les paramètres pour observer leur impact sur les trajectoires (Fig 4.1 c).

On ré-écrit l'équation (4.1) sous la forme suivante, avec s l'espèce de nageur, et i l'identifiant du nageur

$$A_i^s(t) = \gamma(v_0^s + b(t, X_i^s(t))(v_1^s - v_0^s) - \|V_i^s(t)\|) \frac{V_i^s(t)}{\|V_i^s(t)\|} + \beta^s \frac{\nabla b(t, X_i^s(t))}{\|\nabla b(t, X_i^s(t))\|} + \eta^s \quad (4.2)$$

$$:= f_A(\theta^s, b(t, X_i^s(t)), V_i^s(t), X_i^s(t)) + \eta^s \quad (4.3)$$

où

$$\theta^s := (\gamma^s, v_0^s, v_1^s, \beta^s)$$

sont des paramètres inférés pour chaque espèce, afin de caractériser la nage de la population dans son ensemble. Dans cette équation, nous introduisons La fonction f_A peut être vue comme une déviation déterministe de la marche aléatoire, rassemblant tous les mécanismes du modèle. L'inférence des paramètres θ^s peut être définie dans un cadre Bayésien par la résolution du problème de régression non linéaire

$$A_i^s(t) \sim \mathcal{N}(f_A(\theta^s | b(t, X_i^s(t)), V_i^s(t), X_i^s(t)), \epsilon^s) \quad (\text{vraisemblance}) \quad (4.4)$$

à partir des données observées $b(t, X)$, $X_i^s(t)$, $V_i^s(t)$ et $A_i^s(t)$, et les distributions normales tronquées *a priori*

$$\theta^s \sim \mathcal{N}(0, 1) \quad (4.5)$$

$$\epsilon^s \sim \mathcal{N}(0, 1). \quad (4.6)$$

et les contraintes

$$\gamma^s \geq 0, \quad v_0^s \geq 0, \quad v_1^s \geq 0, \quad \epsilon^s \geq 0.$$

La visualisation directe des trajectoires de nage, ou des couples "angle/vitesse", donne déjà des descriptions qualitatives des caractéristiques de nage (Fig. 4.2 a). L'inférence de ces paramètres sur une population de nageurs permet de caractériser la nage de l'espèce de manière plus fine, notamment en quantifiant l'importance relative des différents mécanismes de nage, en s'affranchissant des caractéristiques individuelles de chaque trajectoire de nage. Après inférence, des trajectoires de nage de mêmes caractéristiques peuvent être simulées (Fig. 4.2 b). La valeur des paramètres inférés permet en particulier de savoir si la bactérie nageuse est réactive aux changements de densité de biofilm (γ), si elles changent de direction en fonction des densités observées (β), ou si la marche est très aléatoire (ϵ). L'inférence a été effectuée par inférence Bayésienne sur les 3 populations de *Bacillus* avec le logiciel Stan (Fig. 4.2 c). Elle a permis de relier les caractéristiques d'interaction avec le biofilm (c'est à dire les différents paramètres de l'équation) avec des caractéristiques physiologiques des espèces de micro-nageurs, notamment la forme du complexe de flagelles.

Le modèle et la méthode d'inférence ont été valorisés par un article [11] et plusieurs présentations en conférence (IBSB2022, CEMRACS 2021, ECMTB2018). J'ai également participé à un article de revue sur les mécanismes spatiaux dans les biofilms [15] et dans les biofilms dans un environnement de micro-gravité [8].

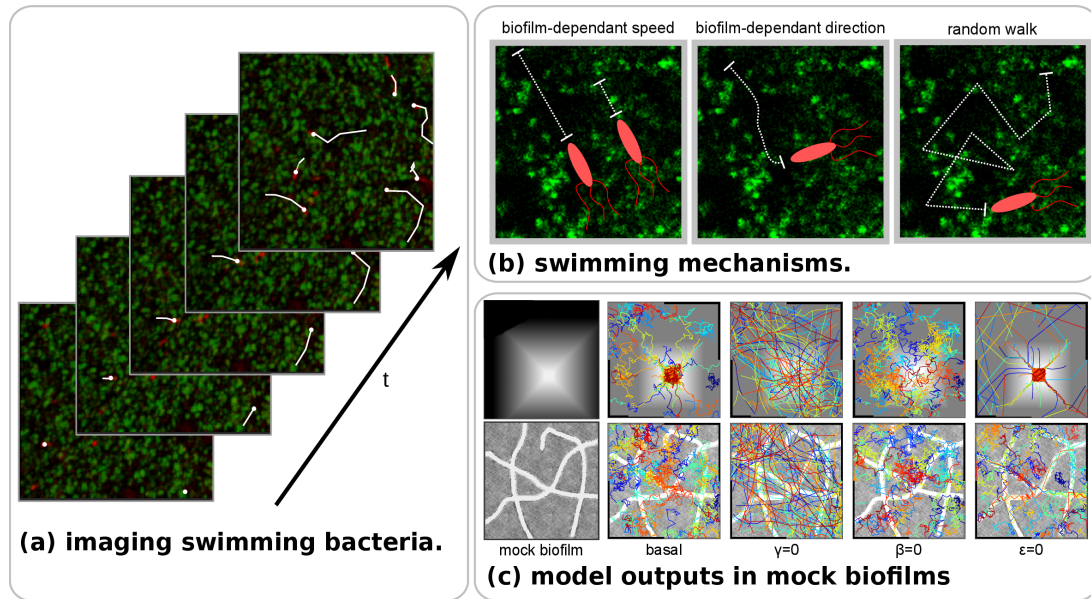


Fig. 4.1 Données et caractéristiques du modèle. *a)* Série temporelle d'images de microscopie confocale. Des bactéries nageuses de trois espèces différentes (en rouge) sont observées dans un biofilm exogène de *S. aureus* (vert). La trajectoire de chaque bactérie est reconstruite avec un logiciel de traitement d'images. À partir de ces trajectoires, les vitesses, accélérations et angles de changement de direction sont calculés. *b)* Mécanismes inclus dans le modèle. Le modèle inclut un mécanisme de sélection de vitesse (en fonction de la densité locale de biofilm, le nageur nage plus ou moins vite), de changement de direction (il descend les gradients de densité), ou de marche aléatoire. *c)* Sorties de modèle sur pseudo-images de biofilms, en situation basale ou en éteignant un des paramètres. Deux domaines modèles sont présentés : l'un avec des gradients de densité convergeant vers le centre du domaine, l'autre avec des trouées rectilignes dans un biofilm. Des sorties de modèles sont présentées en situation basale (avec des valeurs de paramètres homogènes), ou en annulant le paramètre γ , β ou ϵ . Ces figures sont issues de [11].

2 Modélisation des régulations de la crypte intestinale par le microbiote

2.1 Renouvellement de la crypte intestinale modulé par le microbiote.

Les cryptes intestinales sont des villosités composées de quelques centaines de cellules épithéliales composant la surface du colon. C'est également le lieu de l'essentiel des mécanismes déterminant l'impact du microbiote sur l'hôte. En effet, les cellules épithéliales métabolisent principalement des composés produits par le microbiote, les acides gras à chaîne courte (AGCC). De plus, ces AGCC régulent également la différenciation des cellules souches en cellules différenciées. À l'inverse, les cellules de l'hôte régulent les niveaux d'oxygène dans la crypte, ce qui impacte les populations bactériennes.

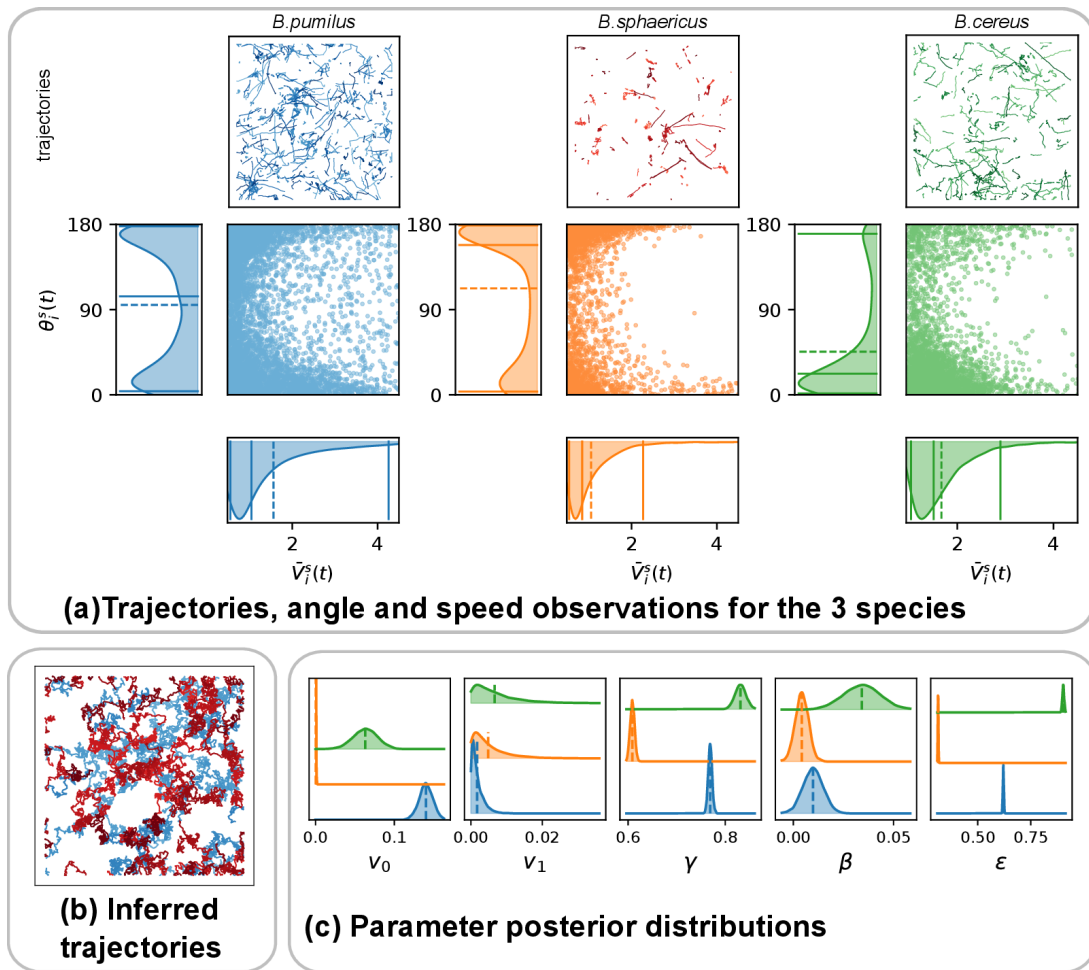


Fig. 4.2 Caractérisation de la nage des espèces de micro-nageurs. *a)* Visualisation de trajectoires et de descripteurs. Les trajectoires de tous les nageurs observés durant une même fenêtre temporelle sont juxtaposées pour chaque espèce. Les couples 'angles/vitesses' sont également affichés pour tous les points de temps observés. A gauche, nous présentons les distributions de direction de nage durant les trajectoires, et en bas les distributions de vitesse. Cet affichage permet de discriminer les espèces : certaines favorisent la ligne droite, d'autres réalisent beaucoup d'allers-retours. Les vitesses de nage sont également discriminantes. *b)* simulations de trajectoires après inférence. Après inférence des paramètres, le modèle permet de simuler de nouvelles trajectoires. Notons ici que l'inférence ne permet pas de reconstituer les trajectoires, mais de simuler des trajectoires de même caractéristiques. *c)* Distribution a posteriori des paramètres après inférences. Les distributions a posteriori (après échantillonnage par STAN) sont présentées pour chaque espèce. Ces valeurs de paramètres permettent également de bien différencier les espèces sur leur capacité à adapter leur nage aux caractéristiques du biofilm qu'elles rencontrent pendant leur nage. Ces figures sont issues de [11].

2.2 Un modèle PDMP de la dynamique de la crypte intestinale incluant le microbiote.

Afin d'intégrer les observations biologiques réalisées à l'échelle cellulaire, un modèle individu-centré de la crypte intestinale est formalisé : chaque cellule est modélisée individuellement par des variables d'état décrivant sa position dans l'espace et son type cellulaire. Afin de coupler des mécanismes stochastiques tels que les événements de différenciation, de division ou de mort, avec des mécanismes déterministes comme les efforts mécaniques des cellules les unes sur les autres, nous avons opté pour un modèle de type *Piecewise-Deterministic Markov Process* (PDMP) : entre les événements stochastiques décrits par des sauts dans l'espace d'état, l'évolution du système est formalisée par des systèmes dynamiques déterministes. Le couplage avec les populations microbiennes est effectué dans la paramétrisation des taux et des intensités de saut qui dépendent des densités locales en AGCC. Nous avons ensuite dérivé rigoureusement, sous hypothèse de grande population, une approximation déterministe du système de PDMP sous la forme d'une EDP non locale. Enfin, sous hypothèse de petit rayon d'interaction entre cellules, nous avons formellement dérivé une deuxième approximation sous forme d'une équation de réaction-diffusion non linéaire.

2.2.1 Description du modèle cellule-centré

Chaque cellule du modèle est caractérisée par la position $z \in [0, z_{max}]$ de son centre dans la crypte, et son type $l \in \mathcal{T}$, où \mathcal{T} est l'ensemble fini des types cellulaires. Une cellule k à la position z^k et de type l^k est représentée par une mesure de Dirac

$$\delta_{x^k}(dx), \text{ où } x^k = (z^k, l^k)$$

et la population des n cellules au temps t , notée ν_t , est une mesure ponctuelle finie correspondant à la somme de ces Diracs :

$$\nu_t = \sum_{k=1}^n \delta_{x^k(t)}(dx). \quad (4.7)$$

Le couplage avec le microbiote passe par l'action de molécules diffusant dans la crypte, dont la répartition est déterminée par l'activité microbienne : la concentration de butyrate $c_b(z, t)$ et de dioxygène $c_o(z, t)$.

Modélisation stochastique du destin cellulaire Les cellules peuvent se diviser, se différencier et mourir, la combinaison de cette succession d'événements constituant le destin cellulaire. Ces événements sont modélisés par des processus de saut, qui sont modélisés par la description de leur amplitude et de leur taux de saut, qui prennent en compte des régulations biologiques.

On note $\mathcal{J} = \{\text{'division'}, \text{'différenciation'}, \text{'extrusion'}\}$ l'ensemble des types de sauts, et $(\nu_t)_{t \geq 0}$ la mesure ponctuelle qui représente la population de cellules. Quand au temps T , un saut de type k a lieu pour la cellule i située en $x^i = (z^i, l^i)$, on modifie la population par un terme $\mu_k(x^i)$ qui est une mesure ponctuelle finie de \mathcal{X} :

$$\nu_T = \nu_{T^-} + \mu_k(x^i).$$

La mesure $\mu_k(x^i)$ représente l'amplitude du saut, et détermine l'état final (position, type cellulaire, ou disparition) de la cellule i .

Les taux de sauts sont représentés par des lois exponentielles $Exp(q_k(z, l, \nu_t * D_k(z)))$, où la fonction $q_k(z, l, \nu_t * D_k(z))$ décrit le taux de sauts de type k , qui dépend de la position dans la

crypte, du type cellulaire, mais également du reste de la population convoluée avec un noyau de convolution D_k comptant les cellules du voisinage (par exemple pour décrire des régulations densité-dépendantes). Pour rappel, pour $t \geq 0$, en notant T_k le temps auquel l'évènement de type k a lieu, on a

$$P(T_k \leq t) = 1 - \exp(-\lambda t) \text{ et } \mathbb{E}[T_k] = \frac{1}{q_k(z, l, \nu_t * D_k(z))}.$$

Modèle déterministe inter-saut : interactions mécaniques entre cellules Les cellules se déplacent sur l'axe z sous l'effet d'interactions mécaniques entre cellules qui induisent une force de répulsion lorsqu'elles sont trop proches.

En conséquence, l'expression de la force \tilde{F} exercée par la cellule j sur la cellule i est la suivante :

$$\tilde{F}(z^i - z^j) \begin{cases} = 0 & \text{si } |z^i - z^j| > a, \\ > 0 & \text{si } z^i - z^j > 0, \\ = 0 & \text{si } z^i - z^j = 0, \\ < 0 & \text{si } z^i - z^j < 0. \end{cases} \quad (4.8)$$

Pour des raisons physiologiques et géométriques, les forces d'interactions s'annulent au bord, ce qui est pris en compte par une fonction ϕ

$$\phi(z) = \begin{cases} \left(\frac{z}{r_0}\right)^2 & \text{si } 0 \leq z \leq r_0, \\ 1 & \text{si } r_0 \leq z \leq z_{max} - r_0, \\ \left(\frac{z_{max} - z}{r_0}\right)^2 & \text{si } z_{max} - r_0 \leq z \leq z_{max}, \end{cases} \quad (4.9)$$

où r_0 représente la taille caractéristique de la zone de modulation des forces d'interactions mécaniques. La fonction ϕ est une fonction de modulation multiplicative des forces qui agissent sur la cellule i , si bien que la force totale agissant sur la cellule i dans une population de n cellules est

$$\phi(z^i) \sum_{j=1}^n \tilde{F}(z^i - z^j).$$

Sous l'effet de ces interactions, les cellules vont se déplacer continûment sur l'axe z , dans un régime non-inertiel.

Alors, pour $1 \leq i \leq n$,

$$\frac{dz_i}{dt}(t) = \phi(z^i(t)) \sum_{j=1}^n F(z^i(t) - z^j(t)) \quad (4.10)$$

où on a posé $F(\cdot) := \frac{1}{\eta_{bm}} \tilde{F}(\cdot)$ et η_{bm} est un coefficient de friction avec la membrane basale strictement positif.

2.2.2 Évolution des concentrations de dioxygène et de butyrate

On modélise la dynamique du dioxygène et du butyrate par un système d'équations de réaction-diffusion sur $[-a/2, z_{max} + a/2]$

$$\partial_t c_b - \sigma \partial_{zz} c_b = -4\gamma_\beta^\infty \frac{c_b^4 c_o}{c_b^4 c_o + K_\beta^5} \times (\nu_t * \psi_a(z)) \quad (4.11)$$

et

$$\partial_t c_o - \sigma \partial_{zz} c_o = -\gamma_\beta^\infty \frac{c_b^4 c_o}{c_b^4 c_o + K_\beta^5} \times (\nu_t * \psi_a(z)) \quad (4.12)$$

où γ_β^∞ est la vitesse de consommation maximale de ces molécules par les cellules épithéliales, K_β est la constante de Michaelis de la réaction, ν_t est la population de cellules, et ψ_a un noyau de convolution permettant de mesurer la quantité de cellules à une distance inférieure à $\frac{a}{2}$ de la position z .

Le système est complété par les conditions de bord

$$c_o(0, t) = c_o^{bas}(t) \quad \text{en } z = -a/2, \quad c_o(z_{max}, t) = c_o^{lum}(t) \quad \text{en } z = z_{max} + a/2. \quad (4.13)$$

$$-\sigma \partial_z c_b = 0 \quad \text{en } z = -a/2, \quad c_b(z_{max}, t) = c_b^{lum}(t) \quad \text{en } z = z_{max} + a/2. \quad (4.14)$$

où $c_o^{bas}(t)$ et $c_o^{lum}(t)$ sont respectivement les concentrations de dioxygène à la base de la crypte déterminée par l'état de l'hôte, et à la jonction avec la lumière intestinale déterminée par l'activité du microbiote.

Le modèle individu-centré se résout par un algorithme de Gillespie. Le modèle individu centré permet de calculer des trajectoires individuelles (cf fig. 4.3-a), mais également de dériver des métriques quantitatives à l'échelle de la population qui peuvent être comparées à des mesures expérimentales, comme des taux de renouvellement des cellules (cf fig. 4.3-b). De multiples réalisations du PDMP permettent de calculer des statistiques sur des observables, en particulier des distributions autour de comportement moyens (cf fig. 4.3-b).

2.3 Modèles de densités de populations à l'échelle de la crypte

Nous introduisons maintenant deux modèles de densité de population. A l'inverse du modèle précédent qui était décrit sous forme de PDMP, c'est à dire sous la forme d'un processus de saut couplé à des EDP, ces modèles continus déterministes décrivent directement l'évolution de la densité de la population.

2.3.1 Modèle limite en grande population

Un premier modèle déterministe est obtenu comme limite en grande population du précédent, sous la forme d'un système couplé entre des EDP à valeur mesure pour les populations des différents types cellulaires et les équations de réaction diffusion pour les substrats. Il s'agit d'une équation de type *self-aggregation* avec des termes sources non-locaux. En notant ξ la limite de la population totale dans l'espace des fonctions $\mathcal{D}([0, T], \mathcal{M}_{F^+}(\mathcal{X}))$, avec $\mathcal{M}_{F^+}(\mathcal{X})$ l'espace des mesures positives finies sur \mathcal{X} , elle vérifie la relation suivante

$$\begin{aligned} \langle \xi_t, f_t \rangle &= \langle \xi_0, f_0 \rangle + \int_0^t \langle \xi_s, \phi(\xi_s * F) \partial_z f_s + \partial_t f_s \rangle ds \\ &\quad + \sum_{k \in \mathcal{J}} \int_0^t \sum_{l \in \mathcal{T}} \int_{\mathcal{Z}} \langle \mu_k(z, l), f_s \rangle q_k(z, l, \xi_s * D_k(z)) \xi_s(dz, l) ds \end{aligned} \quad (4.15)$$

pour tout $f = (f^{(1)}, \dots, f^{(|\mathcal{T}|)}) \in (\mathcal{C}^1(\mathbb{R}^+ \times [0, z_{max}], \mathbb{R}))^{|\mathcal{T}|}$, et où

$$\langle \xi_t, f_t \rangle = \sum_{l \in \mathcal{T}} \int_0^{z_{max}} f^{(l)}(t, z) \xi_t^{(l)}(dz)$$

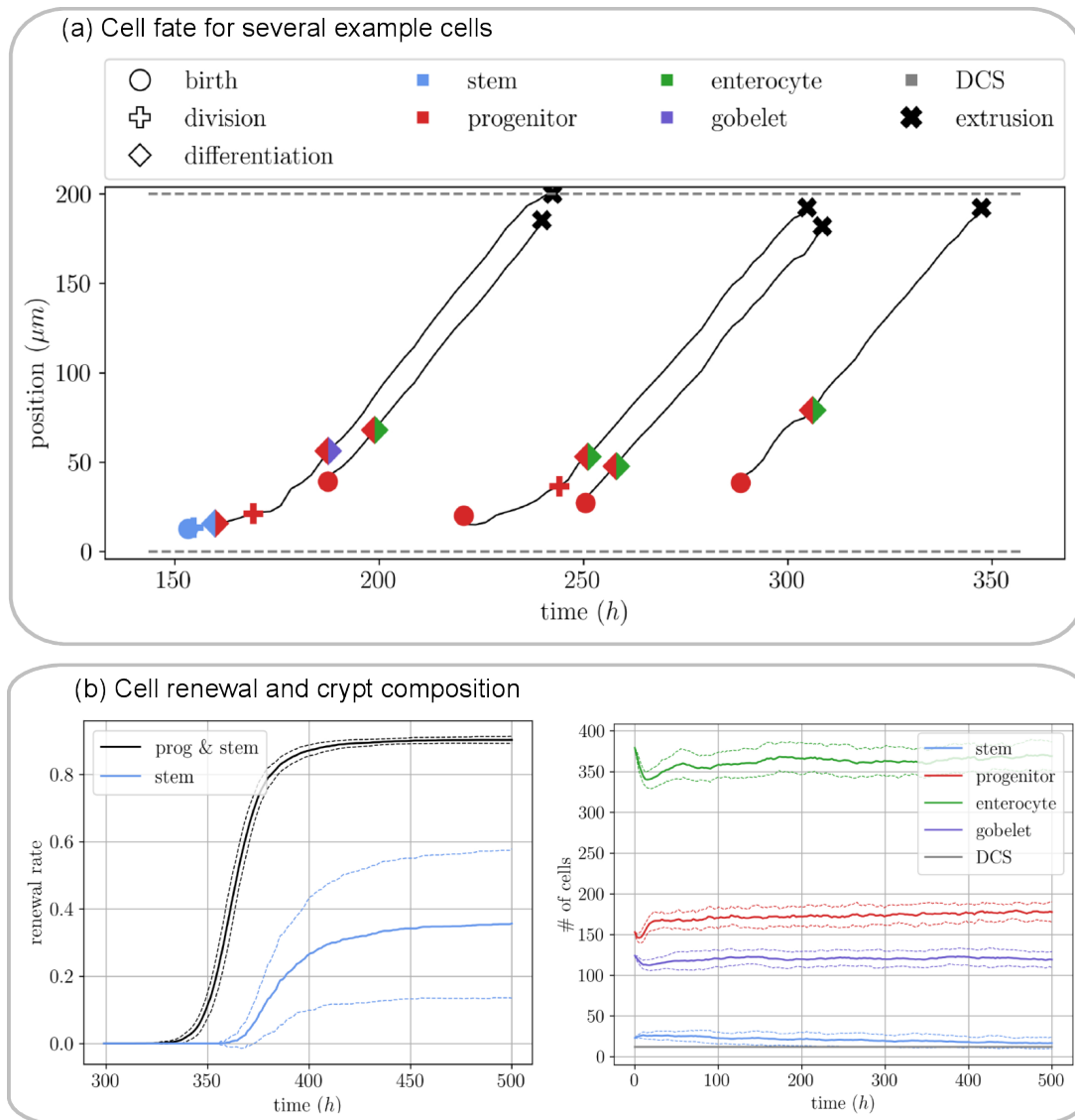


Fig. 4.3 **Exploration numérique du PDMP.** *a) Exemple de réalisation de trajectoires cellulaires.* Nous montrons pour quelques cellules leur trajectoire, depuis la création de la cellule par division, jusqu'à leur extrusion. La figure représente la migration cellulaire le long de la crypte (position, en ordonnée) au cours du temps (abscisse), ainsi que les différents évènements stochastiques qui ont émaillé leur histoire de vie. *b) Statistiques descriptives à l'échelle de la population.* Le taux de renouvellement est d'abord représenté (panel de gauche), défini comme la proportion de cellules ayant été extrudées, parmi une large population de cellules identifiées au temps initial, au cours du temps. L'évolution du nombre total des différentes populations de cellules est également représentée (panel de droite). Pour ces deux figures, la moyenne est indiquée en trait plein, et l'écart type en pointillés. Ces figures sont issues de [9].

avec $\xi_t^{(l)}$ la mesure ξ_t restreinte aux cellules de type l .

La dérivation de ce modèle déterministe s'obtient en étudiant le processus normalisé $\bar{\nu}_t^N$ défini par

$$\bar{\nu}_t^N = \frac{1}{N} \nu_t^N$$

où N est un paramètre d'échelle multiplicatif paramétrant le nombre de cellules introduites au temps initial.

Il s'agit d'abord de montrer la stabilité de ce processus, c'est à dire de montrer que la population totale reste bornée par une valeur indépendante de N , mais dépendante du temps. La convergence du processus normalisé vers la solution de l'équation (4.15) s'obtient ensuite avec un raisonnement par "compacité", qui se déroule en trois étapes. Tout d'abord, on montre l'unicité des solutions positives de l'équation (4.15). Ensuite, on montre la tension de la suite des distributions de $(\bar{\nu}^N)_N$ sur $\mathcal{D}([0, T], \mathcal{M}_{F^+}(\mathcal{X}))$. Par le théorème de Prokhorov, la tension de la suite des lois de $(\bar{\nu}^N)_{N \in \mathbb{N}^*}$ sur $\mathcal{D}([0, T], \mathcal{M}_{F^+}(\mathcal{X}))$ implique que l'on peut en extraire une sous-suite qui converge en loi vers une mesure positive $\bar{\nu}$. Il reste alors à montrer que $\bar{\nu}$ est l'unique solution positive de (4.15).

Si on suppose de plus que, $\forall l \in \mathcal{T}$ l'espace des types cellulaires, les mesures $\xi_t^{(l)}(dz)$ sur \mathcal{Z} admettent une densité notée $\rho_l(z, t)$ alors (4.15) est une formulation faible de :

$$\forall l \in \mathcal{T}, \quad \partial_t \rho_l(z, t) + \partial_z (\phi(z)(\rho(t) * F)(z) \rho_l(z, t)) = \sum_{k \in \mathcal{J}} h_k[\rho(t)](z, l, c_b(t)) \quad (4.16)$$

où, pour tout évènement $k \in \mathcal{J}$, $h_k[\rho(t)] = \pm q_k(z, l, \rho_k * D_k(z))$, c'est à dire plus ou moins le taux d'occurrence de l'évènement, selon le signe de l'amplitude du saut pour la population k impliquée dans l'évènement.

Pour les concentrations, on obtient des valeurs limites dans $\mathcal{C}([0, T], H \times H)$ qui sont des solutions faibles de

$$\partial_t c_b - \sigma \partial_{zz} c_b = -3\gamma_\beta^\infty \frac{c_b^4 c_o}{c_b^4 c_o + K_\beta^5} \times (\rho_l * \psi_a)$$

et

$$\partial_t c_o - \sigma \partial_{zz} c_o = -\gamma_\beta^\infty \frac{c_b^4 c_o}{c_b^4 c_o + K_\beta^5} \times (\rho_l * \psi_a)$$

où l est l'indice des cellules consommatrices, avec des conditions de bord inchangées par rapport au modèle précédent.

2.3.2 Modèle limite de l'équation de *self-aggregation* quand le support d'interaction tend vers 0.

Lorsque le support d'interaction est d'ordre $\frac{a}{n}$, et lorsque la force d'interaction est d'ordre n , où n est un entier paramétrant ces grandeurs, la suite de solutions (ρ_l^n, c^n) de l'équation de *self-aggregation*, sous hypothèse additionnelle de régularité des solutions et du noyau d'interaction, tend faiblement vers la solution de l'équation $\forall l \in \mathcal{T}$, pour tout $f \in \mathcal{C}^1(\mathbb{R}^+ \times [0, z_{max}], \mathbb{R})$,

$$\begin{aligned} \langle \rho_l(t, \cdot), f_t \rangle &= \langle \rho_l(0, \cdot), f_0 \rangle + \int_0^t \langle \rho_l(s, \cdot), -\phi(\cdot) \bar{W} \partial_z \rho(s, \cdot) \partial_z f_s(\cdot) + \partial_t f_s(\cdot) \rangle ds \\ &+ \sum_{k \in \mathcal{J}} \int_0^t \sum_{l \in \mathcal{T}} \int_{\mathcal{Z}} \langle \mu_k(z, l), f_s \rangle q_k(z, l, \bar{D}_k \rho(s, z), c(s, z)) \rho_l(s, z) dz ds \quad (4.17) \end{aligned}$$

où $\rho = \sum_{l \in \mathcal{T}} \rho_l$ est la densité de cellules totale, $\bar{W} := \frac{1}{a} \int_{-a}^a \int_{-a}^z F(\frac{y}{a}) dy dz$ et $\bar{D}_k = \int_{-a}^a D_k(\frac{y}{a}) dy$, et, pour $m \in \{b, o\}$,

$$\partial_t c_m - \sigma_m \partial_{zz} c_m = \sum_{l \in \mathcal{T}} M_{ml}(c) \rho_l \quad (4.18)$$

avec

$$\sum_{l \in \mathcal{T}} M_{ml}(c) \rho_l = \begin{cases} -4\gamma_\beta^\infty \frac{c_b^4 c_o}{c_b^4 c_o + K_\beta^5} \times \rho_{(gc,ent)} & \text{si } m = b, \\ -\gamma_\beta^\infty \frac{c_b^4 c_o}{c_b^4 c_o + K_\beta^5} \times \rho_{(gc,ent)} & \text{si } m = o. \end{cases} \quad (4.19)$$

Les conditions aux bords restent inchangées. C'est une équation de type milieu poreux comportant un terme de diffusion croisée non linéaire. A la limite, la fonction source n'est plus non-locale.

Cette dérivation s'obtient de manière formelle à partir de l'équation (4.16), en notant que lorsque le support d'interaction tend vers 0, le terme $\xi_s * F$ tend vers $\bar{W} \partial_z \rho$.

Une version forte de l'équation (4.17) s'écrit

$$\forall l \in \mathcal{T}, \quad \partial_t \rho_l(z, t) - \nabla \cdot (\phi \bar{W} \partial_z \rho(t) \rho_l) = \sum_{k \in \mathcal{J}} \bar{h}_k[\rho(t)](z, l, c_b(t)) \quad (4.20)$$

où, pour tout évènement $k \in \mathcal{J}$, $h_k[\rho(t)] = \pm q_k(z, l, \bar{D}_k \rho_k(s, z))$, c'est à dire plus ou moins le taux d'occurrence de l'évènement, selon le signe de l'amplitude du saut pour la population k impliquée dans l'évènement.

La précision du modèle obtenu avec l'équation (4.20) des milieux poreux (PME) est vérifiée en comparant les sorties du modèle avec le modèle PDMP individu-centré (cf Fig. 4.4 a), pour différentes familles cellulaires et pour l'occurrence de différents évènements. Le modèle PME fournit une approximation acceptable du modèle PDMP, mais à un coût numérique très inférieur. En effet, la résolution numérique du PDMP est particulièrement coûteuse, du fait du nombre important d'évènements et de cellules qui doivent être simulées pour une réalisation du PDMP, mais également du fait des nombreuses réalisations qui doivent être effectuées pour obtenir de bonnes estimations des statistiques sur les sorties de modèle. Comme le modèle PME est rapide à résoudre, on peut en faire une exploration numérique systématique, en réalisant une analyse de sensibilité. Nous utilisons pour cela la méthode de Morris.

Ce travail a été réalisé principalement dans le cadre de la thèse de Léo Darrigade, que j'ai co-encadré avec Béatrice Laroche [34]. Le modèle PDMP a été valorisé dans [9] où le modèle PDMP a été formalisé et exploré numériquement, et où l'approximation déterministe a été utilisée pour réaliser une analyse de sensibilité du modèle. Enfin, la preuve de la convergence du modèle PDMP vers le modèle de *self-aggregation* a été valorisé dans un pre-print actuellement en review [4]. La résolution et l'exploration du modèle PME ont principalement été réalisées dans le cadre de la thèse de Marie Haghebaert, encadrée par B.Laroche.

3 Modèle du microbiote intestinal

3.1 Biogéographie du microbiote intestinal.

Dans la section 2, nous avons longuement détaillé l'environnement biochimique du microbiote intestinal, qui peut être approché en première approximation par les métabolites et les réactions impliqués dans le réseau métabolique de dégradation des fibres, principal substrat pour les populations microbiennes, et par un environnement anaérobie. Cependant, l'environnement physique du microbiote intestinal joue également un rôle important dans sa dynamique. Tout d'abord, le

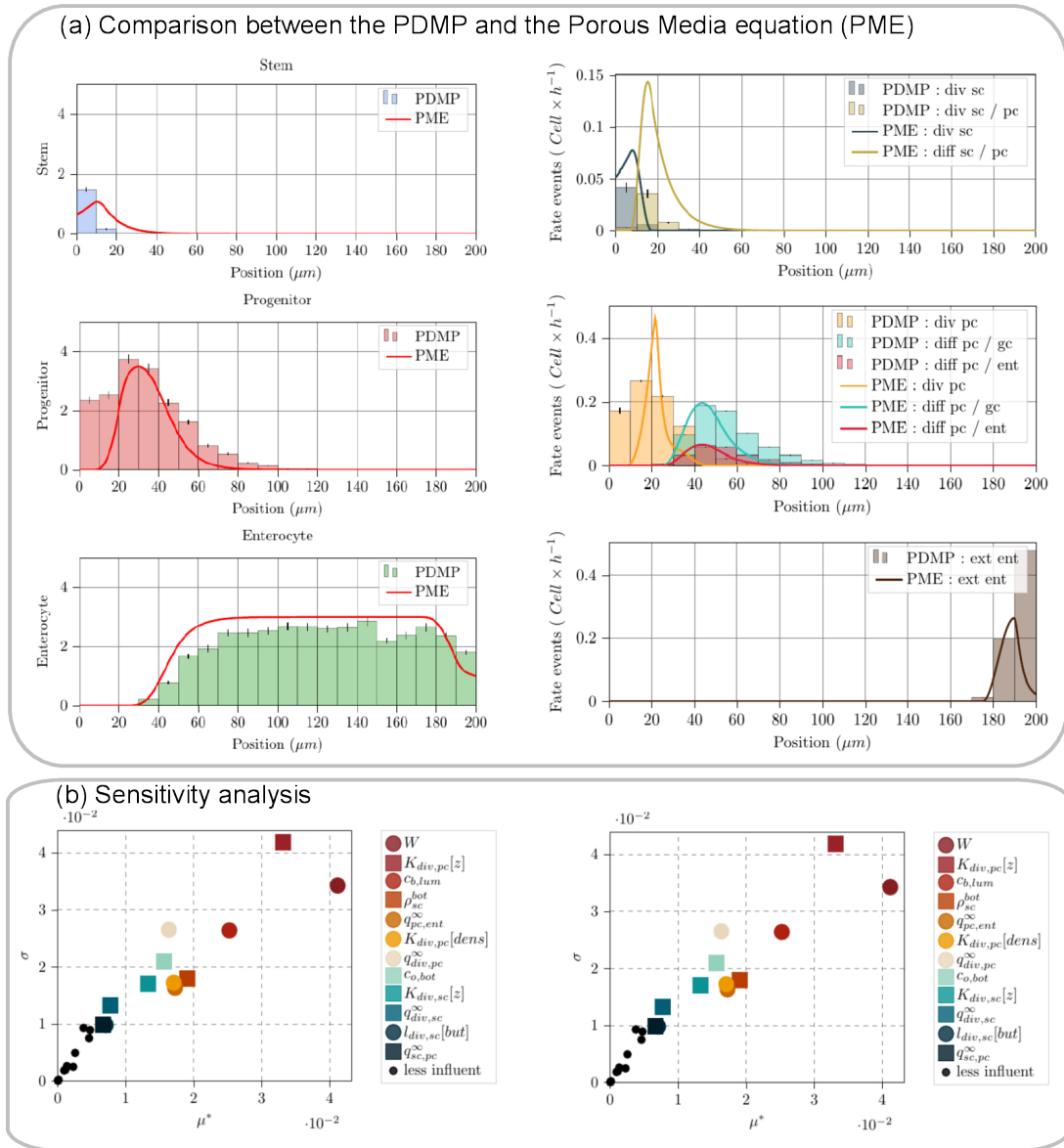


Fig. 4.4

Vérification du modèle PME et exploration numérique. *a)* Vérification du modèle PME. L'équation des milieux poreux (PME) (4.20) est utilisée et comparée au modèle PDMP correspondant. La distribution des populations (colonne de gauche) et celle des événements (colonne de droite) le long de la crypte sont représentées lorsque le modèle est à l'état stationnaire. Les moyennes et écarts types du nombre de cellules sont représentés par un graphique en barre à différentes hauteurs de la crypte, et la densité correspondante calculée avec le PME est représentée en trait rouge. *b)* Analyse de sensibilité. Une analyse de sensibilité est conduite sur le modèle PME, en utilisant la méthode de Morris. Cette méthode consiste à calculer la moyenne μ^* et la variance σ des valeurs absolues des effets élémentaires, c'est à dire des variations de modèles obtenues après des incréments des valeurs de paramètres. Les paramètres en haut à droite du graphe sont donc les paramètres qui impactent le plus les sorties de modèle. Ces figures sont issues de [9].

transit intestinal, provoqué par le flux constant de chyme, c'est à dire les résidus de digestion gagnant le gros intestin, tend à lessiver le microbiote intestinal. Ce flux est également modélisé par le péristaltisme, c'est à dire par les trains de contractions des muscles de la paroi intestinale, qui parcourent le colon des parties distales aux parties proximales, et inversement, contribuant au transit. Par ailleurs, une couche de mucus, un fluide visqueux enveloppant les cellules épithéliales de l'intestin, permet à certaines bactéries de se protéger du transit intestinal, en leur offrant un environnement visqueux, riche en carbohydrates. Enfin, certaines bactéries sont capables de se déplacer activement dans cet environnement vers un métabolite cible par chimiotactisme, ce qui contribue aux bilans de mouvement globaux. Nous avons ainsi un milieu multi-phasique de différentes viscosités dont la dynamique détermine un champ d'advection des populations microbiennes, qui peuvent elles-mêmes se mouvoir activement afin de trouver des zones plus favorables à leur survie. Tous ces phénomènes contribuent activement à la biogéographie du microbiote, c'est à dire à la répartition spatiale du microbiote dans un état stationnaire.

3.2 Modèle du microbiote intestinal en interaction avec son environnement multi-phasique.

Il s'agit dans ce contexte de définir les opérateurs de transport $F_{b,i}$ et $F_{m,j}$ du système général de dynamique de population (1.1)-(1.2) défini en introduction, que nous réécrivons ici pour rappel.

$$\partial_t b_i + \operatorname{div}(F_{b,i}(b, m)) = G_{b,i}(b, m) \quad (4.21)$$

$$\partial_t m_j + \operatorname{div}(F_{m,j}(b, m)) = G_{m,j}(b, m) \quad (4.22)$$

Nous utilisons un modèle de mélange pour modéliser ce système : en notant l'ensemble des différentes phases du fluide multi-phasique $\mathcal{I} := \{l_{muc}, l_{lum}, b_1, \dots, b_{N_b}\}$, i.e. respectivement le mucus, le liquide luminal les N_b populations bactériennes. Nous notons $f_k, k \in \mathcal{I}$ la fraction volumique correspondant à chaque phase, et nous supposons que $\sum_{k \in \mathcal{I}} f_k = 1$, ce qui correspond à une hypothèse de saturation du milieu.

Les équations (4.21)-(4.22) se réécrivent alors

$$\partial_t f_m - \operatorname{div}(\sigma \nabla f_m) + \operatorname{div}(u_m f_m) = G_{f,m}, \quad (4.23)$$

$$\partial_t m_j - \operatorname{div}(\sigma_j \nabla m_j) + \operatorname{div}(m_j \tilde{u}) = G_j \quad (4.24)$$

où la vitesse u_m se décompose avec la formule $u_m = u + \vartheta_{m,chem}$, où u est une vitesse du mélange commune à toutes les phases, et ϑ est une vitesse chimiotactique. La vitesse \tilde{u} est une vitesse moyenne

$$\tilde{u} = \sum_{m \in \mathcal{I}} f_m u_m = u + \sum_{m \in \mathcal{I}} f_m \vartheta_{m,chem}.$$

Les fonctions sources décrivent le métabolisme microbien. Dans ce travail, elles sont obtenues via des termes non-linéaires issus de la modélisation des bioprocédés [59]. Ces équations sont des équations de réaction-diffusion-convection, dont le champs de vitesse de convection reste à déterminer.

On suppose que la dynamique de u est décrite par l'équation de Stokes

$$\nabla p - \operatorname{div}(\mu(f_{l_{muc}}, f_{l_{lum}})D(u)) = 0 \quad (4.25)$$

où p est un champ de pression, $D(u) = \frac{1}{2} (\nabla u + \nabla u^T)$, et μ est la viscosité locale dépendant de la composition locale du fluide en liquide luminal et mucus. En sommant les équations du système (1.1)-(1.2) et en utilisant la condition de saturation, on obtient la condition de compressibilité

$$\operatorname{div}(u) = -\operatorname{div}\left(\sum_{i=1}^{N_b} f_{b_i} \mathcal{V}_{b_i}\right).$$

3.3 Simplification du modèle par analyse asymptotique

En approchant la géométrie de l'intestin par un cylindre dont on suppose que le rayon R est négligeable devant la longueur L , on peut dériver une approximation analytique des champs de vitesse u et \mathcal{V}_{b_i} . Plus précisément, en notant $\epsilon = \frac{R}{L}$ le ratio d'aspect de l'intestin, on cherche un vecteur solution de la forme

$$\begin{aligned} b_i &= b_i^{(0)} + \epsilon b_i^{(1)} + \epsilon^2 b_i^{(2)} + \dots & m_j &= m_j^{(0)} + \epsilon m_j^{(1)} + \epsilon^2 m_j^{(2)} + \dots \\ u_z &= u_z^{(0)} + \epsilon u_z^{(1)} + \epsilon^2 u_z^{(2)} + \dots & u_r &= u_r^{(0)} + \epsilon u_r^{(1)} + \epsilon^2 u_r^{(2)} + \dots \end{aligned}$$

En injectant ces solutions dans les équations (4.23)-(4.25) et en rassemblant les termes associés à un même exposant d'épsilon, on obtient des solutions simplifiées pour les termes dominants de la décomposition asymptotique des solutions. Cette dérivation est une adaptation de la dérivation des écoulements laminaires à partir des équations de Stokes au cas où la viscosité est variable.

La vitesse du mélange est alors donnée par les formules explicites suivantes

$$u_z^{(0)}(r, z) = -\frac{\Lambda(r, z)}{\kappa(z)} \left(R \int_0^z \sum_{i \in \mathcal{I}} \gamma_i(R, y) dy - R^2 U_{z, in} \right) \quad (4.26)$$

$$u_r^{(0)}(r, z) = \left(-\frac{1}{r} \int_0^r s \partial_z u_z^{(0)}(s, z) ds - \sum_{i \in \mathcal{I}} f_i(r, z) \sum_{j \in M \text{ et}} \lambda_{i,j} \Upsilon_j(r, z) \right), \quad (4.27)$$

où

$$\Lambda(r, z) = \int_r^R \frac{s}{\mu(s, z)} ds, \quad \kappa(z) = \int_0^R s \Lambda(s, z) ds, \quad U_{z, in} = \frac{1}{R^2} \int_0^R s \sum_{i \in \mathcal{I}} \gamma_i(s, 0) ds.$$

et γ_i représente le flux de chaque phase à travers l'épithélium, et Υ représente la vitesse chimiotactique de la phase i vers le métabolite j . Nous remarquons que le champs de vitesse asymptotique, tout en ayant une expression analytique explicite, garde trace des principaux paramètres de la mécanique des fluides : l'hétérogénéité de viscosité μ , les conditions de bords γ_i à travers la muqueuse intestinale, l'influx moyen $U_{z, in}$ et la composante radiale du chimiotactisme avec le terme Υ .

Les équations (4.23)-(4.24) se simplifient également avec les équations suivantes, en coordonnées polaires

$$\sum_{i \in \mathcal{I}} f_i = 1 \quad (4.28)$$

$$\partial_t f_i - \frac{1}{r} \partial_r (r \sigma \partial_r f_i) + \operatorname{div}_r (u_s f_i) + \frac{1}{r} \partial_r (r \vartheta_{i,r} f_i) = F_i \quad (4.29)$$

$$\partial_t c_j - \frac{1}{r} \partial_r (r \sigma_j \partial_r c_j) + \tilde{u}_s \cdot \nabla_r c_j = G_j, \quad (4.30)$$

$$(4.31)$$

Dans ces équations, la diffusion est dégénérée, dans la mesure où la diffusion selon z a disparu. Cette simplification engendre une accélération des calculs, puisque le maillage de calcul comporte peu de cellules dans la direction radiale.

3.4 Exploration numérique

Ce modèle asymptotique permet une accélération des temps de calcul de plusieurs ordres de grandeur, rendant ainsi possible une exploration numérique systématique. Cela permet en particulier d'explorer la répartition spatiale des différentes grandeurs impliquées dans le modèle et de les comparer aux données de la littérature. La répartition du mucus est par exemple physiologique (cf Fig. 4.5). De plus, nous réalisons une analyse de sensibilité des sorties du modèle, en particulier des densités de population, aux paramètres impliqués dans la biogéographie des populations microbiennes : le péristaltisme, la viscosité du mucus, le niveau d'apport en fibres alimentaires, et les vitesses de chimiotactisme. En échantillonnant l'espace des paramètres avec la méthode FAST99, nous pouvons calculer les indices de Sobol des différents paramètres, c'est à dire la part de variance des variables de sorties expliquée par les variations du paramètre étudié. Nous pouvons observer que les différents paramètres de mécanique des fluides n'influent pas de manière comparable sur les densités de populations le long du colon. Si le péristaltisme influence surtout les densités de population aux extrémités du colon, le chimiotactisme détermine principalement la partie proximale alors que l'apport en fibre impacte fortement la partie transverse (cf Fig. 4.6).

Ces modèles ont été valorisés par la publication [18] pour le modèle d'écoulement, et par [14] pour le modèle du microbiote intestinal. Cette publication contient également une analyse de sensibilité du modèle aux différents paramètres permettant d'identifier les paramètres déterminant la biogéographie des populations bactériennes dans l'intestin. Un autre exemple d'interaction hôte-microbiote est un modèle compartimentaire modélisant l'impact du microbiote intestinal sur le cycle du cholestérol dans les différents organes de l'hôte [12], ou la glycogénèse [16].

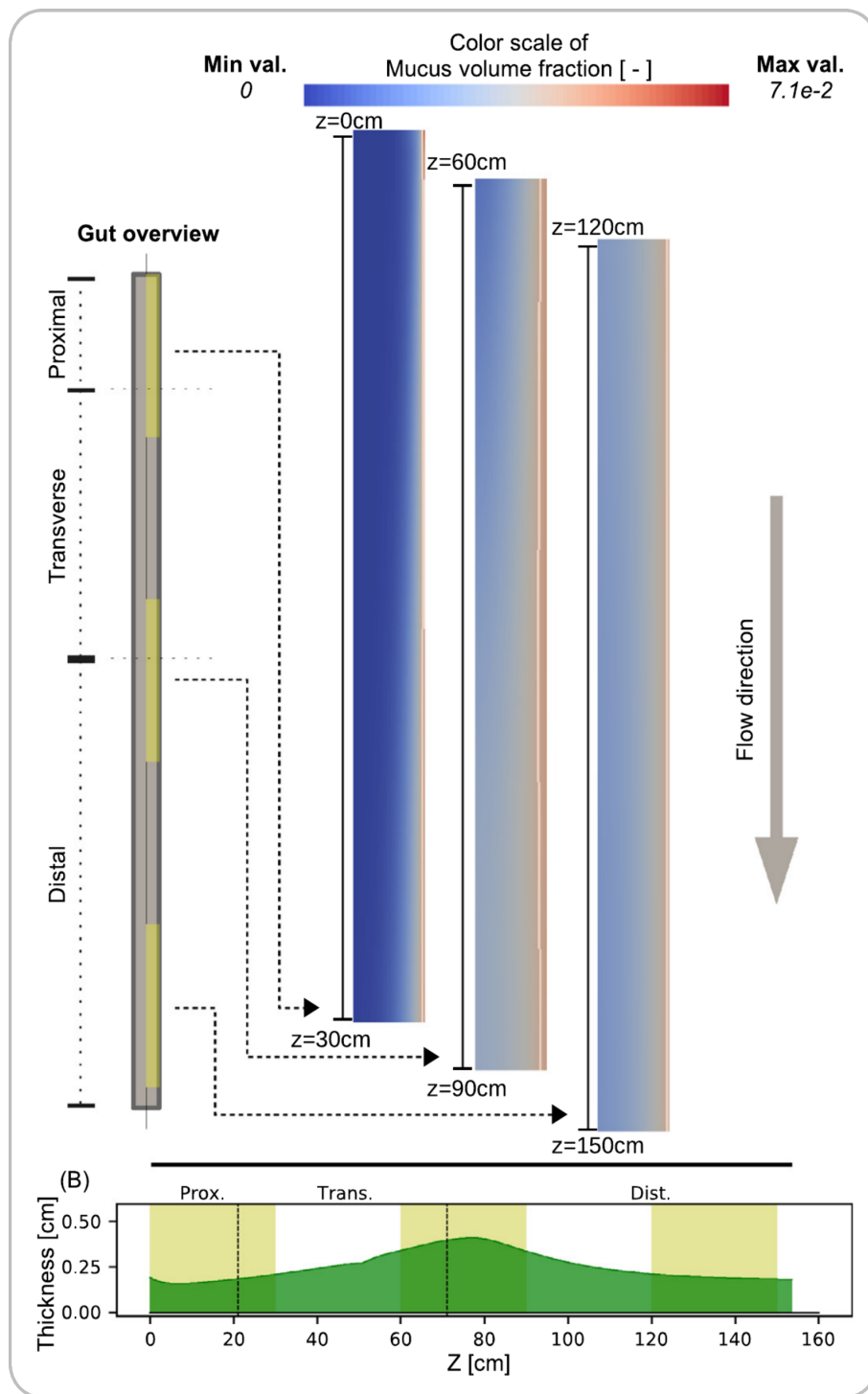


Fig. 4.5

Répartition spatiale de la couche de mucus dans l'intestin. Nous représentons la distribution spatiale de la densité de mucus (échelle de couleur en haut) en trois portions du gros intestin (proximale, transverse et distale). Dans le panel du bas, nous représentons la densité radiale moyenne le long de l'intestin. Nous constatons que la couche de mucus a des épaisseurs différentes, résultat de l'hydrodynamique, mais également de la consommation du mucus par les bactéries intestinales.

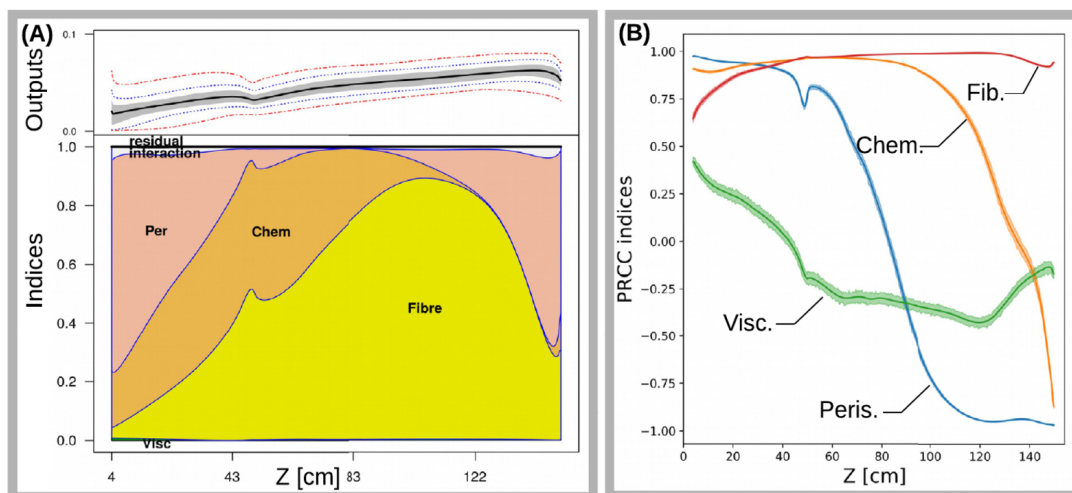


Fig. 4.6 Analyse de sensibilité du modèle EDP du microbiote intestinal. *a)* Indices de Sobol. Nous représentons les indices de Sobol pour différents paramètres définissant le péristaltisme, le chimiotactisme, l'apport en fibres ou la viscosité du mucus. Ces indices de Sobol sont calculés en chaque position z le long du colon. *b)* indices PRCC. Nous représentons les indices PRCC représentant la corrélation partielle des paramètres avec les sorties. Si les indices de Sobol indiquent la contribution du paramètre à la variance de la sortie, le PRCC indique si le paramètre influence positivement ou négativement les sorties de modèle.

CHAPITRE
5

Flux microbiens entre holobiontes et épidémiologie

1 Des flux microbiens entre holobiontes.

Jusqu'ici, nous nous sommes intéressés à la caractérisation de microbiotes, c'est à dire à des communautés microbiennes, ou d'holobiontes, c'est à dire une communauté microbienne en interaction avec un hôte. Nous avons pu étudier la diversité inter-hôtes des microbiotes, notamment des microbiotes intestinaux (cf Sec. 2). Si la structure des communautés microbiennes symbiotiques dépendent pour une large part des caractéristiques génétiques de l'hôte (qui participe à la sélection des communautés microbiennes par son système immunitaire ou par la régulation de niches nutritionnelles favorables aux symbiotes), elle dépend également de flux microbiens provenant de l'environnement extérieur, notamment des interactions avec d'autres hôtes. Ces interactions sont importantes pour les micro-organismes commensaux qui participent à la santé de l'hôte mais également pour les microbes pathogènes pour comprendre leur diffusion et leur épidémiologie.

Cette question de la transmission de micro-organismes entre différents hôtes se décline dans de multiples contextes en écologie microbienne. Par exemple en santé des plantes, où la question de la transmission naturelle de micro-organismes de la plante mère au plantule via la graine se pose, afin d'élucider les mécanismes naturels de transmission, mais également afin d'élaborer des moyens d'agir sur le microbiote de la graine, pour transmettre à la plante fille des phénotypes d'intérêt, médiés par le microbiote (meilleure fixation de l'azote, résistance à la sécheresse, protection contre des pathogènes...).

De même en santé humaine, mieux comprendre les phénomènes de transmission et de colonisation des micro-organismes permet de mieux piloter les techniques de greffe de microbiote (transplantation de microbiote fécal), qui a permis des percées fulgurantes dans le traitement de maladies infectieuses, comme les infections à *Clostridium difficile*. On parle alors "d'impression" de microbiote (microbial engraftment), lorsqu'on cherche à appliquer une communauté exogène à une communauté résidente. Comme pour les greffes d'organe, on cherche alors à mieux piloter

les mécanismes d'implantation, afin d'éviter les phénomènes de rejet de la greffe microbienne.

En santé animale, il est important de comprendre la diffusion de pathogènes entériques comme des *Salmonella* dans des élevages, dans le but de contrôler la présence de ces pathogènes dans la chaîne alimentaire. Par la suite, nous allons décrire un modèle d'épidémiologie, appliqué à la transmission de pathogène microbien, et prenant en compte la diversité inter-individuelle de la réponse du microbiote commensal à l'infection.

2 Modèle populationnel incluant la dynamique d'infection intra-hôte.

Dans le cadre d'un projet européen One Health (Momir-PPC), nous avons proposé un projet CEMRACS lors de la session 2018 s'intéressant à la modélisation de la propagation de *Salmonella* dans les élevages de volaille. Un modèle simplifié de la dynamique intra-hôte du microbiote intestinal lors d'une infection à *Samonella* a d'abord été proposé, étendu ensuite à une population d'hôtes. Ce modèle a ensuite été étudié analytiquement et numériquement.

2.1 Modèle simplifié intra-hôte

Supposons un microbiote simplifié composé d'une population pathogène, de densité p et d'une population commensale, de densité b , en compétition avec le pathogène. Lors de l'infection, *Salmonella* déclenche une inflammation qui lui permet d'acquérir des nutriments qui sont des produits finaux des mécanismes inflammatoires, comme de l'oxygène ou de l'oxyde nitrique [48]. L'oxygène au contraire est délétère pour la plupart des micro-organismes commensaux. Notons d la densité d'oxygène.

La dynamique du pathogène peut alors s'écrire

$$p' = -bp(\mu - ab - C_1d). \quad (5.1)$$

Dans cette équation, le terme $(\mu - ab - C_1d)$ détermine la compétition entre la population commensale et le pathogène. Le paramètre μ détermine l'intensité maximale de cette compétition, qui est modulée par un terme d'inhibition ab représentant la compétition entre microbes commensaux qui réduit l'effet barrière, et par un terme C_1d , représentant l'effet délétère de l'oxygène sur les populations commensales.

La densité d'oxygène suit quant à elle l'équation d'évolution suivante

$$d' = \frac{1}{\tau} \left(-d + C_2 \frac{p^n}{p^n + p_*^n} \right) \quad (5.2)$$

Dans cette équation, $\tau > 0$ est un paramètre déterminant l'échelle caractéristique de la dynamique de l'oxygène, qui est la résultante d'un premier terme, $-d$, de dégradation naturelle de l'oxygène, et d'un second terme de production, décrit par la fonction sigmoïde $C_2 \frac{p^n}{p^n + p_*^n}$ paramétrée par l'intensité C_2 , le "centre" p_* de la sigmoïde, et n , qui détermine sa pente. La production d'oxygène se déclenche donc quand $p \gg p_*$.

En faisant l'hypothèse que la densité totale du microbiote $K = p + b$ reste constante au cours du temps, et que la dynamique de l'oxygène est très rapide, i.e. $0 < \tau \ll 1$, on peut éliminer b et d de l'équation (5.1) pour obtenir l'équation d'évolution sur la population de pathogène

$$p' = F(p) \quad \text{avec} \quad F(p) = -(K - p)p \left(A - \alpha p - C \frac{p^n}{p^n + p_*^n} \right) \quad (5.3)$$

où $A = \alpha K - \mu$ et $C = C_1 C_2$.

2.2 Modèle à l'échelle de la population

Afin d'inclure du bruit biologique dans la dynamique du pathogène, l'équation (5.3) est transformée sous la forme d'une SDE

$$dP = \underbrace{F(p)}_{\text{drift déterministe}} dt + \underbrace{\sigma dB}_{\text{bruit gaussien}}. \quad (5.4)$$

où dB est un bruit blanc gaussien unitaire, et $\sigma > 0$ un paramètre scalaire.

Cette équation considère une perturbation gaussienne autour d'un drift déterministe de la densité intra-hôte en pathogène P . Ce drift prend en compte différents mécanismes de pathogénicité comme l'inflammation (regroupés dans le terme $C \frac{p^n}{p^n + p_*^n}$, où p_* et C sont des paramètres), ou l'interaction avec le microbiote résident (regroupée dans les autres paramètres K , A et α). Selon les valeurs de ces paramètres, ce drift peut avoir 2 à 5 états stationnaires, et la valeur $P = 0$ est instable si et seulement si $A > 0$.

Comme $p \mapsto F(p)$ est uniformément lipschitzienne et à croissance linéaire, et comme σ est constant en temps, les résultats standards de la théorie des SDE s'appliquent [55] pour obtenir l'équation de Fokker-Plank (ou de Kolmogorov forward)

$$\begin{aligned} \partial_t u(t, p) &= -\partial_p(F(p)u(t, p)) + \frac{\sigma^2}{2} \partial_{pp}^2 u(t, p) \\ u(0, p) &= u_{ini}(p) \end{aligned}$$

où $u(t, p)$ est la densité de probabilité de $p(t)$ conditionnellement à la densité de probabilité u_{ini} de la densité initiale en pathogène.

Cette équation est complétée par des conditions de bord de flux nuls (comme $F(0) = F(K) = 0$) :

$$\partial_p u(t, 0) = \partial_p u(t, K) = 0$$

Ces conditions de bord permettent à la population de pathogènes de rester bornée entre 0 et K . Du point de vue de la théorie des SDE, ces conditions correspondent à l'ajout de conditions de bords réflexives à la SDE (5.4) (voir [52]).

On considère maintenant une très grande population d'hôtes, pouvant être décrite par une densité de population $s(t, p)$, ce qui signifie que $\int_a^b s(t, p) dp$ est le nombre d'hôtes ayant une densité de pathogènes comprise entre a et b au temps t . La population s est donc structurée en temps et en densité de pathogènes.

On suppose de plus que la population d'hôtes est constante au court du temps, avec N individus. Ainsi $\int_0^K s(t, p) dp = N$ pour $t > 0$. En supposant que la densité de pathogènes de chaque hôte est décrite par la même densité $u(t, p)$, nous avons $s(t, p) = Nu(t, p)$, où $s \in L^2(0, T; L^2(0, K))$ est décrite par l'EDP de diffusion convection

$$\partial_t s(t, p) = -\partial_p(F(p)s(t, p)) + \frac{\sigma^2}{2} \partial_{pp}^2 s(t, p) \quad (5.5)$$

$$s(0, p) = S_{ini}(p) = Nu_{ini}(p) \quad \text{avec} \quad u_{ini} \in L^2(0, K), \quad \int_0^K u_{ini}(p) dp = 1$$

$$\partial_p s(t, 0) = \partial_p s(t, K) = 0$$

A l'échelle de la population, les variations de la densité de population selon le niveau de pathogènes intra-hôte suivent le drift déterministe de l'infection F ou évoluent aléatoirement via le terme de diffusion.

Cette équation peut être couplée à un "réservoir" décrivant le niveau environnemental de pathogènes dans l'élevage, ce qui permet de prendre explicitement en considération les flux de pathogènes inter-hôtes

$$\partial_t s(t, p) = -\partial_p((F(p) + \beta_{in}(p)r(t) - \beta_{ex}(p))s(t, p)) + \frac{\sigma^2}{2}\partial_{pp}^2 s(t, p) \quad (5.6)$$

$$\partial_t r(t) = -\left(\gamma^* + \int_0^K s(t, p)\beta_{in}(p)dp\right)r(t) + \int_0^K s(t, p)\beta_{ex}(p)dp \quad (5.7)$$

Dans cette équation, r est la densité en pathogènes du réservoir, β_{in} et β_{ex} sont des fonctions décrivant l'ingestion et l'excrétion de pathogènes et γ^* est un terme de mort de pathogène en milieu extérieur.

3 Étude analytique et numérique du modèle

La distribution stationnaire de l'équation (5.5) s'obtient de façon analytique. En effet, si un état stationnaire s_∞ est atteint, alors, il est solution de l'EDO

$$\partial_p(-F(p)s_\infty(p) + \frac{\sigma^2}{2}\partial_p s_\infty(p)) = 0 \quad \text{avec} \quad \int_0^K s_\infty(p)dp = N$$

Cette EDO peut s'intégrer sur $(0, p)$, en utilisant la condition de bord $\partial s_\infty(0) = 0$ et le fait que $F(0) = 0$, ce qui nous permet d'obtenir l'EDO

$$-F(p)s_\infty(p) + \frac{\sigma^2}{2}\partial_p s_\infty(p) = 0 \quad (5.8)$$

dont la solution est

$$s_\infty(p) = \lambda \exp\left(\frac{2}{\sigma^2} \int_0^p F(r)dr\right) \quad \text{avec} \quad \lambda = \frac{N}{\int_0^K \exp\left(\frac{2}{\sigma^2} \int_0^p F(r)dr\right) dp}.$$

On peut également montrer que cet état stationnaire est atteint avec un taux exponentiel. Pour ce faire, on peut étudier comme dans [56] la fonction

$$G(t) = \|s(t, \cdot) - s_\infty\|_{L^2(0, K; 1/s_\infty)}$$

où la norme pondérée $L^2(0, K; 1/s_\infty)$ est définie par

$$\|f\|_{L^2(0, K; 1/s_\infty)}^2 = \int_0^K f(p)^2 \frac{dp}{s_\infty(p)}.$$

On peut calculer la dérivée temporelle de $G(t)$

$$\begin{aligned} G'(t) &= 2 \int_0^K \partial_t s(t, p) \left(\frac{s(t, p)}{s_\infty(p)} - 1\right) dp \\ &= 2 \int_0^K \left(-\partial_p(F(p)s(t, p)) + \frac{\sigma^2}{2}\partial_{pp}^2 s(t, p)\right) \left(\frac{s(t, p)}{s_\infty(p)} - 1\right) dp \\ &= 2 \int_0^K \left(F(p)s(t, p) - \frac{\sigma^2}{2}\partial_p s(t, p)\right) \partial_p \left(\frac{s(t, p)}{s_\infty(p)} - 1\right) dp \end{aligned}$$

où on a utilisé une intégration par partie et les conditions de bord de l'EDP et la valeur de F en 0 et K .

Comme par définition de s_∞ (voir eq. 5.8) on a $F(p) = \frac{\sigma^2}{2} \frac{\partial_p s_\infty(p)}{s_\infty(p)}$, on obtient :

$$\begin{aligned} G'(t) &= \sigma^2 \int_0^K \left(\frac{\partial_p s_\infty(p) s(t,p)}{s_\infty(p)} - \partial_p s(t,p) \right) \partial_p \left(\frac{s(t,p)}{s_\infty(p)} - 1 \right) dp \\ &= -\sigma^2 \int_0^K s_\infty(p) \partial_p \left(\frac{s(t,p)}{s_\infty(p)} \right) \partial_p \left(\frac{s(t,p)}{s_\infty(p)} - 1 \right) dp = -\sigma^2 \int_0^K s_\infty(p) \left| \partial_p \left(\frac{s(t,p)}{s_\infty(p)} - 1 \right) \right|^2 dp \end{aligned}$$

D'après l'inégalité de Poincaré pour la norme $L^2(0, K; s_\infty)$, il existe une constante $\kappa > 0$ ne dépendant que de K et s_∞ telle que

$$\kappa \int_0^K s_\infty(p) \partial_p \left(\frac{s(t,p)}{s_\infty(p)} - 1 \right)^2 dp \leq \int_0^K s_\infty(p) \left(\frac{s(t,p)}{s_\infty(p)} - 1 \right)^2 dp$$

et ainsi

$$G'(t) \leq -\kappa \sigma^2 G(t)$$

ce qui permet de montrer, en utilisant le lemme de Grönwall, que $G(t) \leq G(0) \exp(-\kappa \sigma^2 t)$, autrement dit que la solution $s(t, p)$ converge vers l'état stationnaire au sens $L^2(0, K; 1/s_\infty)$ au taux exponentiel $\exp(-\kappa \sigma^2 t)$, quelle que soit la condition initiale.

La convergence vers l'état stationnaire pour le modèle (5.5) est ensuite vérifiée numériquement (fig. 5.1, a), pour différentes conditions initiales. Une convergence très rapide est observée, même quand la condition initiale est très éloignée de la distribution stationnaire. Lorsque le réservoir est ajouté (eq. (5.6)-(5.7), fig. 5.1, b), on constate également une convergence rapide vers un état stationnaire, qui est différent de l'état stationnaire sans réservoir. Le réservoir modifie donc la structure de la population d'hôtes en quantité de pathogènes.

Ces résultats ont été valorisés par un article [13] dans les proceedings du CEMRACS 2018.

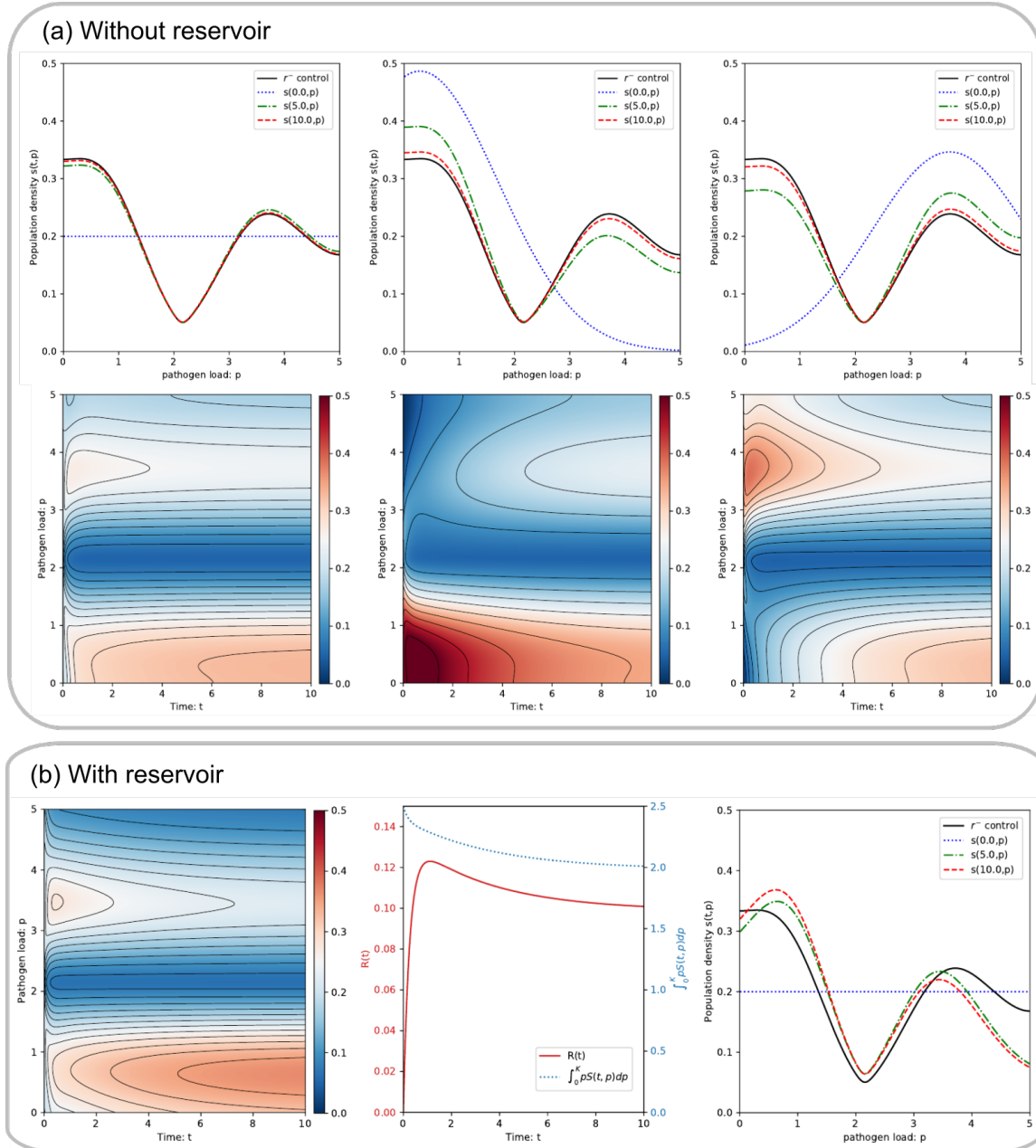


Fig. 5.1

Exploration numérique du modèle sans et avec réservoir. (a) *Sans réservoir, panel supérieure* : les distributions de populations selon p sont calculées pour différentes conditions initiales (ligne bleue pointillée), à 5 (ligne pointillée verte) et 10 (ligne pointillée rouge) unités de temps. En noir, la solution stationnaire $s_{\infty}(p)$ est représentée. La convergence est ainsi vérifiée numériquement. *Panel inférieure* : Les distributions de populations sont représentées en tout temps (abscisses) et en toute densité de pathogènes (ordonnée) pour les différentes conditions initiales. Les isolignes sont représentées en noir. La convergence est également vérifiée, les isolignes étant horizontales en fin de domaine temporel. (b) *Avec réservoir* : *panel de gauche* Les distributions de population sont à nouveau représentées en temps (abscisses) et en densité de pathogènes (ordonnée), avec les isolignes. Un état stationnaire est atteint, les isolignes étant horizontales en $t = 10$. *Panel central* : La densité de pathogènes du réservoir (rouge) ainsi que la quantité totale de pathogènes dans la population d'hôtes (bleue) sont représentées. On constate une forte excrétion (diminution du pathogène total chez les hôtes) tamponnée par le réservoir (augmentation de la courbe rouge). *Panel de droite* : les solutions sont représentées en $t = 5$ (vert) et $t = 10$ (rouge), avec l'état stationnaire sans réservoir. On constate que les états stationnaires avec et sans réservoir diffèrent.

C H A P I T R E
**Conclusion****1 Des axes de recherche à l'interface des mathématiques et de l'écologie microbienne**

Ma recherche en mathématique et modélisation s'intéresse principalement à la modélisation de communautés microbiennes. Elle se focalise notamment sur l'utilisation de données expérimentales pour construire ou calibrer des modèles de dynamique de population microbienne. L'intégration de données omiques par les modèles dynamiques demande de relever différents défis méthodologiques, qui peuvent la plupart du temps se décliner à la fois comme une problématique en écologie microbienne, et comme un verrou méthodologique en mathématiques appliquées. Par exemple, étudier une communauté microbienne naturelle avec des approches holistiques revient, du point de vue de l'écologie microbienne, à identifier des motifs interprétables de présence de taxons microbiens ou de fonctions microbiennes, ce qui se traduit d'un point de vue des mathématiques par une méthode de réduction de dimension identifiant des familles de vecteurs générateurs des données (les motifs) en incluant un volet de modélisation pour permettre leur interprétabilité. De même, dans des approches réductionnistes, les questions d'écologie microbienne consistent à identifier les leviers permettant de piloter la communauté simplifiée, comme des métabolites échangés par des micro-organismes lors de relations trophiques. Cette problématique se traduit dans des modèles mathématiques par des questions d'exploration de réseau métabolique, de choix de variable ou d'exploration numérique. Par ces deux exemples, je souhaite mettre en avant un champ de recherche à l'interface de l'écologie microbienne et des mathématiques, prenant pleinement en compte la question biologique, ainsi que les contraintes inhérentes aux sciences expérimentales, afin d'adapter au mieux les outils mathématiques et les développements méthodologiques pour apporter des éléments de réponse nouveaux à la question biologique. Il faut pour cela développer des méthodes nouvelles dans un continuum de recherche, des données omiques jusqu'au modèle (Fig. 6.1). Ces compétences sont à l'interface de la bio-informatique (métagénomique, réseau métabolique), des statistiques (réduction de dimension, apprentissage) et des systèmes dynamiques (EDO, EDP). Tout en restant spécialisé dans les

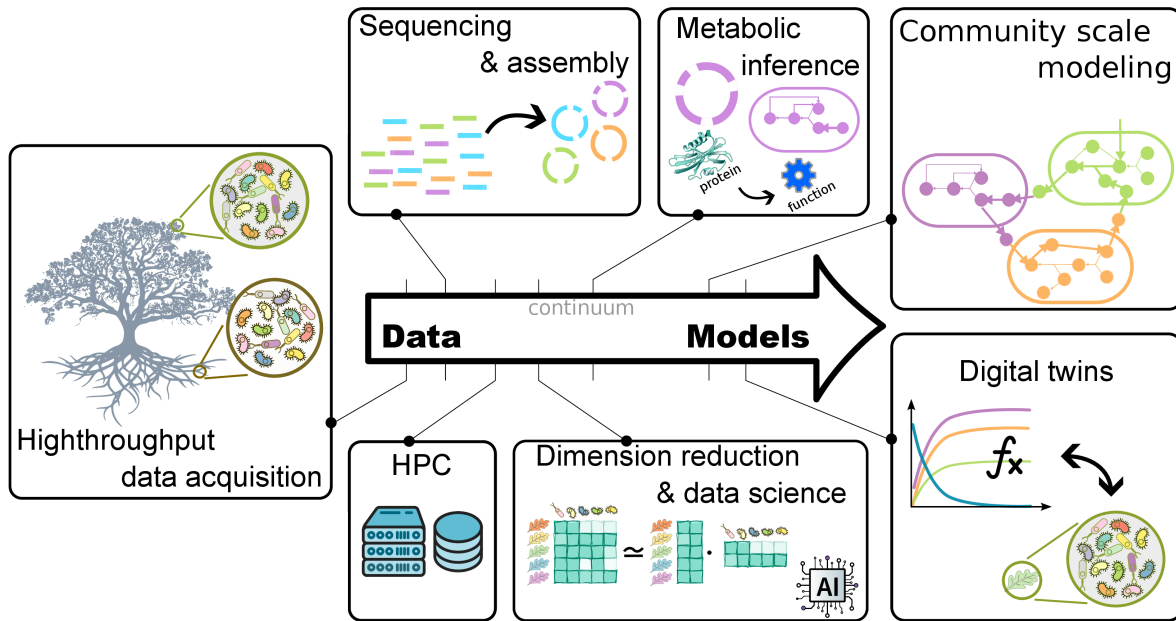


Fig. 6.1 Un continuum de recherche : des séquences jusqu'au modèle. Un continuum de recherche existe en modélisation de communautés microbiennes, allant des données brutes issues des séquenceurs ou appareils de mesure, jusqu'aux modèles de communautés microbiennes. Ce continuum inclut des compétences en bioinformatique, en calcul scientifique, en analyse de données et en mathématiques. Pour ma part, ma recherche se situe plutôt en aval de ce continuum, mais nécessite de fortes interactions avec des chercheurs spécialisés dans les autres compétences.

mathématiques appliquées, il est nécessaire de collaborer étroitement avec des spécialistes des autres thématiques.

J'ai participé à des revues sur la modélisation de communautés microbiennes de manière générale [19], à partir de données métagénomiques [1] ou en particulier chez la vigne [10], ainsi qu'à une publication de diffusion auprès du grand public [6].

2 Un panel étendu d'outils mathématiques pour la modélisation

La démarche de recherche que je mets en œuvre suit souvent le même processus (Fig. 6.2). Tout d'abord, un contexte issu de l'écologie microbienne est formalisé. Je développe alors des modèles mécanistiques à l'échelle des observations, le plus souvent sous forme de systèmes dynamiques déterministes, mais pouvant également inclure des composantes stochastiques. Quand c'est le cas, ces modèles sont étudiés à l'échelle des populations après approximation déterministe. Ils sont ensuite implémentés, en mettant en œuvre des méthodes de calcul scientifiques et des schémas numériques appropriés adossés à des bibliothèques spécialisées, le plus souvent en Python. Ces modèles sont paramétrés à partir de données expérimentales issues de communautés microbiennes, que ce soient des données omiques ou d'imagerie. Pour réduire le coût numérique de ces modèles, j'applique dès que possible des techniques de réduction de modèle (analyse asymptotique) ou d'approximation (par métamodélisation et machine learning). Ces modèles réduits permettent de faciliter leur exploration numérique, et de déployer des méthodes

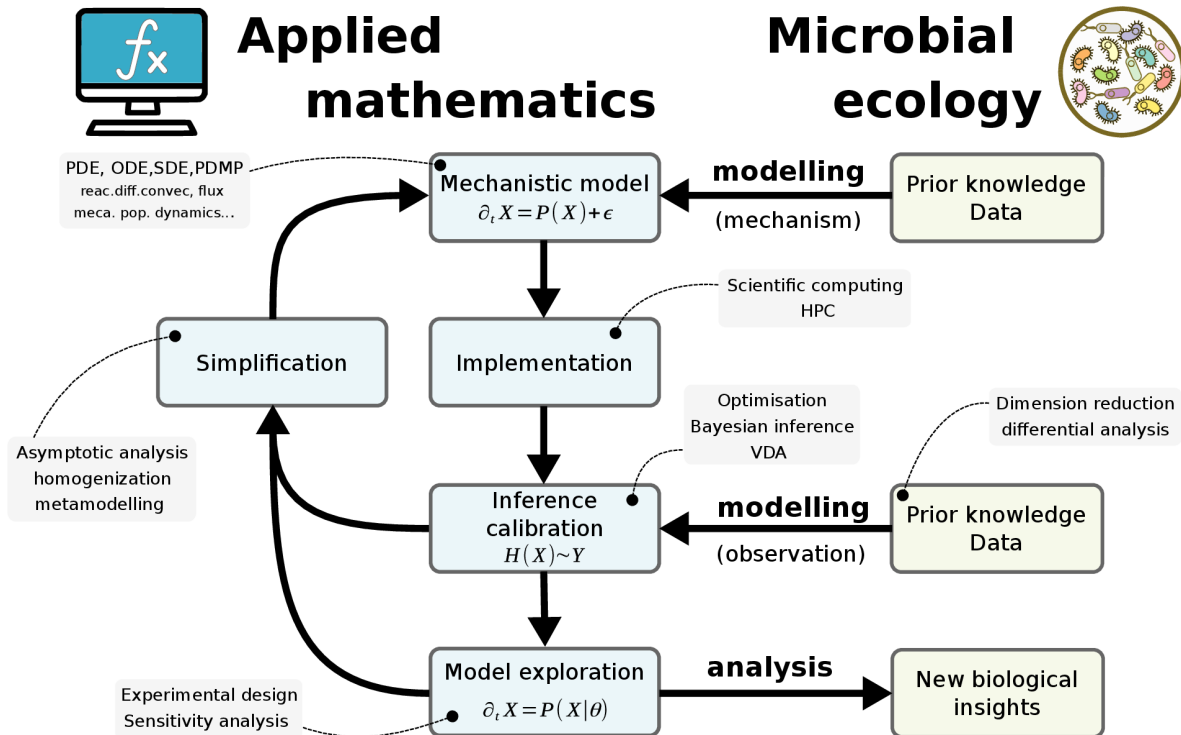


Fig. 6.2 Des outils mathématiques à l'interface entre mathématiques appliquées et écologie microbienne. La démarche de recherche suit le plus souvent ce schéma : à partir d'un contexte issu de l'écologie microbienne, un modèle mécanistique est développé, puis implémenté de manière efficace. Le modèle est ensuite calibré, et ses paramètres sont inférés à partir de données expérimentales. Puis une exploration numérique du modèle est conduite. Enfin, les résultats sont interprétés en vue d'acquies de nouvelles connaissances biologiques. Pour accélérer les calculs, ou pour faciliter l'interprétation des sorties de modèle, ou la paramétrisation, des méthodes de simplification de modèle ou des modèles réduits peuvent être mis en œuvre. La boucle 'implémentation', 'paramétrisation' et 'exploration' est alors actualisée. A chacune de ces étapes, des compétences spécifiques en mathématiques sont utilisées (cases gris clair).

systématiques d'analyse de sensibilité. L'ensemble de ces méthodes permet d'étudier les communautés microbiennes à tous les niveaux de complexité : SynComs, communautés complexes, holobionte et populations d'holobiontes.

3 Un investissement sur les diverses facettes du travail de chercheur

Le manuscrit d'HDR fait la part belle à la production de connaissances et à la description d'une trajectoire de recherche cohérente et structurée. Toutefois, la production de connaissance s'appuie sur d'autres activités du travail de chercheur pouvant le plus souvent la structurer et la nourrir.

3.1 Formation par la recherche

Depuis mon recrutement à l'INRAE, j'ai pu m'investir dans la formation par la recherche à de nombreuses reprises. J'ai d'abord débuté par le co-encadrement puis l'encadrement autonome de stagiaires en master, que ce soit sur des sujets de modélisation comme le microbiote intestinal (T.Goswany 2014, L.Darrigade 2016), la nage dans les biofilms (M.A. Rasendra 2016, A.Braune 2017), l'épidémiologie (C.Mendoza 2019), l'interaction hôte-microbiote (M.Bourgin 2016), le couplage entre modèles métaboliques et systèmes dynamiques (A.Paulay 2019, M.Lecomte 2020) ou l'accélération de modèles métaboliques (A.Worstman 2022, R.Kaempfer 2023). J'ai également participé à l'encadrement de trois projets CEMRACS, qui permettent pendant 6 semaines d'encadrer un groupe de jeunes scientifiques (M2, PhD ou postdoc) sur un sujet de recherche : en 2015, sur la modélisation du microbiote, en 2018, sur l'épidémiologie dans des élevages de volailles et en 2021 sur la métamodélisation de modèles métaboliques. Les trois projets ont été valorisés par des publications [5], [13], [18].

J'ai ensuite participé ou participe actuellement à l'encadrement de doctorantes et doctorants. J'ai d'abord débuté par le co-encadrement, avec B.Laroche, de la thèse de L.Darrigade sur un modèle PDMP du renouvellement de la paroi intestinale financé par l'université Paris Saclay (voir sec. 2 et [4], [9], [34]). J'ai ensuite participé au co-encadrement d'A.Paulay, avec E.Maguin, G.Grimaud et B.Laroche, qui effectuait une thèse CIFRE dans la start up Biomathematica sur la modélisation de la protéolyse avec des modèles genome-scale [3]. Je dirige actuellement la thèse de Sthyve Tatho sur l'intégration de données multi-omiques dans des modèles dynamiques de communautés microbiennes, financée par le PEPR Agroécologie Numérique (Mistic) et co-encadrée par V.Baldazzi. Je participe également au co-encadrement de Sahak Yeghiazaryan sur le couplage de modèles métaboliques avec des modèles thermodynamiques, financée par le métaprogramme Digit-Bio et le département Mica de l'INRAE et dirigée par E.Le Quemener et N.Bernet.

Enfin, j'ai participé à l'encadrement de quatre post-doctorantes et post-doctorants, soit par du co-encadrement comme le post-doc de F.Colas sur l'effet des habitats semi-naturels en agroécologie co-encadré par F.Accatino, financé par le projet API-SMAL (U. Paris Saclay, 12 mois). Mais j'ai également encadré de manière autonome les post-doc de G.Ravel (Financement Mathnum, 12 mois) sur les micro-nageurs dans les biofilms (voir sec. 1 et [11]), de P.Ugalde Salas (financement Aex SLIMMEST Inria, 24 mois) sur la métamodélisation de modèles métaboliques (voir sec.3 et [1], [5]) et celui de T.Malou (financement Pherosensor, 24 mois) sur un problème inverse pour la détection d'insectes ravageurs.

3.2 Montage de projets

Depuis mon recrutement, je me suis investi fortement dans le montage de projets, comme participants ou comme porteur. J'ai d'abord obtenu un financement du métaprogramme MEM/INRA (ModChoCycle, 50k€, co-porté par M.Rhimi, 2016-20018) sur l'impact du microbiote sur le cycle du cholestérol [12]. Je participe également comme responsable ou co-responsable de paquet de travail à deux projets du PPR Cultiver et Protéger Autrement : le projet Pherosensor (porté par P.Lucas, 1.5M€, 2021-2026, WP leader) sur la détection précoce d'insectes ravageurs par les phéromones, ainsi que le projet VITAE (porté par F.Delmotte, 3M€, 2021-2026, co-WP leader) sur la culture de la vigne sans pesticide [10]. Je me suis également fortement impliqué dans le montage du projet Mistic (porté par D.Sherman et P.Abad, 1.5M€, 2023-2028, co-WP leader), sélectionné comme projet phare du PEPR Agroécologie Numérique, centré sur le développement de méthodes pour analyser des communautés microbiennes associées aux plantes [1]. Je me suis

aussi investi dans l'obtention du financement de la thèse de S.Yeghiazaryan, principalement porté par E.Le Quemener, obtenue auprès des départements Mathnum et Mica.

J'ai également obtenu comme porteur un financement du métaprogramme Digit-bio (Artémis, 20k€, 2024-2025) sur les jumeaux numériques en écologie microbienne, l'action exploratoire Inria SLIMMEST (co-porté par C.Frioux, 200k€,2021-2023) sur la métamodélisation de modèles métaboliques [5] et un postdoc (Bioswimmers,60k€,2021) obtenu auprès du département Mathnum [11]. J'ai aussi obtenu des financements pour des mobilités internationales, notamment une bourse Marie-Sklodowska-Curie du programme Agreenskill+ (30k€, 2017-2019) complétée par un projet du fond France-Berkeley (12k€, co-porté par A.Bäumler, 2019-2021) pour financer un séjour à UC Davis, ainsi que des échanges d'étudiants (visite de C.Tiffany en 2020).

3.3 Animation de la science et expertise

Je m'investis fortement dans l'animation scientifique de mon institut, notamment en participant au comité de pilotage du métaprogramme Holoflux. Les métaprogrammes sont des dispositifs de l'INRAE chargés de promouvoir des approches transdisciplinaires sur des fronts de science, par des appels à projet et des actions d'animation et de pilotage. Le métaprogramme Holoflux s'intéresse aux holobiontes microbiens sur tout le périmètre de recherche de l'INRAE.

Je suis également élu à la CSS (commission scientifique spécialisée) MISTI, qui est la commission d'évaluation des chercheurs et chercheuses en mathématiques, informatique et modélisation. Je participe également au bureau de la CSS qui a pour but d'organiser les aspects pratiques de l'évaluation. J'ai participé à plusieurs jury de sélection, à l'INRAE pour des postes de CRCN (2021, 2022, 2024) ou IR (2024, président du jury), mais également pour des postes de maître ou maîtresse de conférence (U.Côte d'Azur, 2020).

J'ai une activité soutenue de relecture d'articles pour des revues de microbiologie (ISME, mSystems, NPJ biofilm), généralistes (Plos ONE) ou de mathématiques appliquées et de modélisation (Journal of Theoretical Biology, Journal of Biological Systems, Computers in Biology and Medicine) pour environ 3 revues par an. Enfin, j'ai évalué plusieurs projets scientifiques pour le Max Plank Institute, le KFG (organisme Autrichien de financement de la recherche, équivalent à l'ANR) ou pour l'université Paris Saclay (école doctorale ABIÉS).

4 perspectives

Dans les prochaines années, je souhaite poursuivre ces différents axes de recherche en me focalisant plus particulièrement sur les microbiotes associés aux plantes. Ce projet comportera trois axes de recherche complémentaires (voir fig. 6.3). Le premier axe consiste à développer des méthodes permettant de diagnostiquer l'état de communautés microbiennes naturelles. Ces méthodes viseront à analyser des données omiques (metabarcoding, métagénomiques ou métabolomiques) pour identifier des idéotypes associés aux communautés et à des états de santé de la plante ou du microbiote. Par idéotype, nous entendons des motifs taxonomiques ou fonctionnels observés en association avec un phénotype particulier de l'hôte. Le deuxième axe consistera à développer des méthodes analytiques permettant d'identifier parmi ces idéotypes des communautés simplifiées de micro-organismes qui soient représentatives de l'idéotype. Ces méthodes pourront se baser sur du criblage de génome, de l'exploration de modèles métaboliques, sur l'analyse de séries temporelles ou des méthodes de réduction de dimension. Enfin, le troisième axe de recherche consistera à développer des méthodes de modélisation de ces communautés

simplifiées, dont la précision puisse permettre de mieux comprendre leur fonctionnement, dans une perspective de jumeaux numériques.

Ces trois axes sont interdépendants : les méthodes d'identification et de modélisation de communautés simplifiées permettent de développer des modèles à l'échelle de communautés complexes, qui peuvent affiner les diagnostics. De plus, les outils diagnostiques et les modèles de SynComs peuvent permettre d'anticiper l'implantation et le maintien de communautés simplifiées porteuses d'un phénotype cible dans une communauté complexe. Ces trois axes représentent un éventail d'outils de modélisation permettant de développer des solutions basées sur l'écologie microbienne en agriculture ou pour les activités forestières. L'objectif final est de disposer d'outils diagnostiques permettant de mesurer des biomarqueurs de santé de l'holobionte. En cas de carence ou de déséquilibre, les outils de sélection de SynCom permettent de développer des communautés simplifiées porteuses d'un phénotype cible. Et les outils de modélisation fournissent des clés de compréhension du fonctionnement de cette communauté réduite, mais également des prédictions de son devenir dans l'holobionte considéré. Se dessine ainsi une chaîne d'outils permettant la personnalisation de solutions basées sur l'écologie microbienne à une situation donnée.

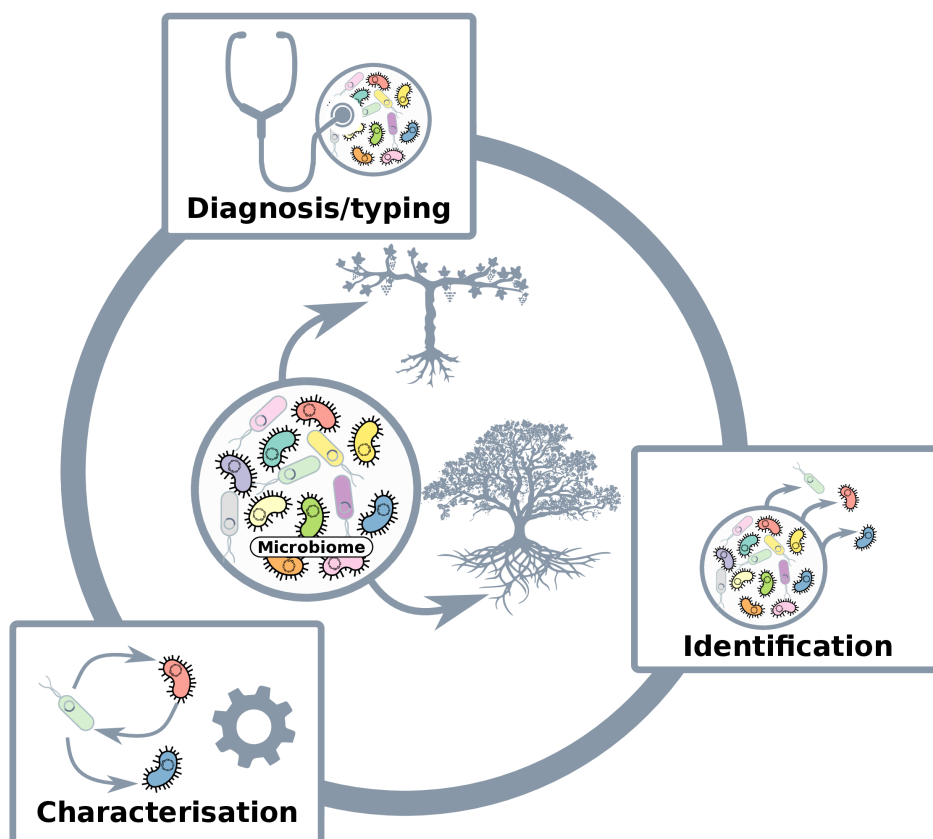


Fig. 6.3 **Projet de recherche futur.** Pour mieux comprendre le fonctionnement des communautés microbiennes associées aux plantes, et l'acquisition de phénotypes bénéfiques pour l'hôte, trois axes complémentaires seront mis en œuvre. Le premier consiste à développer des méthodes permettant de diagnostiquer l'état d'une communauté naturelle. Ceci consiste à identifier des idéotypes associés aux communautés et à des états de santé de la plante. Un deuxième axe consiste à identifier parmi ces idéotypes des communautés simplifiées qui soient représentatives et qui confèrent le phénotype ciblé, par exemple un effet protecteur. Le troisième axe consiste à développer des méthodes de modélisation permettant de mieux comprendre le fonctionnement de ces communautés simplifiées.

Bibliographie personnelle

- [1] K. CERK, P. UGALDE-SALAS, C. G. NEDJAD, M. LECOMTE, C. MULLER, D. J. SHERMAN, F. HILDEBRAND, S. LABARTHE et C. FRIOUX, “Community-scale models of microbiomes: Articulating metabolic modelling and metagenome sequencing”, *Microbial Biotechnology*, t. 17, n° 1, e14396, 2024.
- [2] M. LECOMTE, W. CAO, J. AUBERT, D. J. SHERMAN, H. FALENTIN, C. FRIOUX et S. LABARTHE, “Revealing the dynamics and mechanisms of bacterial interactions in cheese production with metabolic modelling”, *Metabolic Engineering*, S1096717624000302, 2024.
- [3] A. PAULAY, G. M. GRIMAUD, R. CABALLERO, B. LAROCHE, M. LECLERC, S. LABARTHE et E. MAGUIN, “Design of a proteolytic module for improved metabolic modeling of *Bacteroides caccae*”, *mSystems*, N. CHIA, éd., e00153-24, 2024.
- [4] L. DARRIGADE, S. LABARTHE et B. LAROCHE, “Deterministic Limit of a PDMP Model of Epithelial Tissue Interacting with Diffusing Chemicals and Application to the Intestinal Crypt”, 2023.
- [5] C. FRIOUX, S. HUET, S. LABARTHE, J. MARTINELLI, T. MALOU, D. SHERMAN, M.-L. TAUPIN et P. UGALDE-SALAS, “Accelerating Metabolic Models Evaluation with Statistical Metamodels: Application to *Salmonella* Infection Models”, *ESAIM: Proceedings and Surveys*, t. 73, V. EHRLACHER, D. LOMBARDI, O. MULA, F. NOBILE et T. TADDEI, éd., p. 187-217, 2023.
- [6] C. FRIOUX et S. LABARTHE, *Modéliser Les Communautés Bactériennes Pour Mieux Comprendre Leur Fonctionnement*. 2023.
- [7] S. LABARTHE, S. PLANCADE, S. RAGUIDEAU, F. PLAZA OÑATE, E. LE CHATELIER, M. LECLERC et B. LAROCHE, “Four functional profiles for fibre and mucin metabolism in the human gut microbiome”, *Microbiome*, t. 11, n° 1, p. 231, 2023.
- [8] D. MARRA, T. KARAPANTSIOS, S. CASERTA, E. SECCHI, M. HOLYNSKA, S. LABARTHE, B. POLIZZI, S. ORTEGA, M. KOSTOGLU, C. LASSEUR, I. KARAPANAGIOTIS, S. LECUYER, A. BRIDIER, M.-F. NOIROT-GROS et R. BRIANDET, “Migration of surface-associated microbial communities in spaceflight habitats”, *Biofilm*, t. 5, p. 100-109, 2023.
- [9] L. DARRIGADE, M. HAGHEBAERT, C. CHERBUY, S. LABARTHE et B. LAROCHE, “A PDMP model of the epithelial cell turn-over in the intestinal crypt including microbiota-derived regulations”, *Journal of Mathematical Biology*, t. 84, n° 7, p. 60, 2022.
- [10] P. FOURNIER, L. PELLAN, D. BARROSO-BERGADÀ, D. A. BOHAN, T. CANDRESSE, F. DELMOTTE, M.-C. DUFOUR, V. LAUVERGEAT, C. LE MARREC, A. MARAIS, G. MARTINS, I. MASNEUF-POMARÈDE, P. REY, D. SHERMAN, P. THIS, C. FRIOUX, S. LABARTHE et C. VACHER, “The functional microbiome of grapevine throughout plant evolutionary history and lifetime”, in *Advances in Ecological Research*, t. 67, Elsevier, 2022, p. 27-99.

- [11] G. RAVEL, M. BERGMANN, A. TRUBUIL, J. DESCHAMPS, R. BRIANDET et S. LABARTHE, “Inferring characteristics of bacterial swimming in biofilm matrix from time-lapse confocal laser scanning microscopy”, *eLife*, t. 11, e76513, 2022.
- [12] M. BOURGIN, S. LABARTHE, A. KRIAA, M. LHOMME, P. GÉRARD, P. LESNIK, B. LAROCHE, E. MAGUIN et M. RHIMI, “Exploring the Bacterial Impact on Cholesterol Cycle: A Numerical Study”, *Frontiers in Microbiology*, t. 11, p. 1121, 2020.
- [13] S. LABARTHE, B. LAROCHE, T. N. T. NGUYEN, B. POLIZZI, F. PATOUT, M. RIBOT et T. STEGMAIER, “A Multi-Scale Epidemic Model of Salmonella Infection with Heterogeneous Shedding”, *ESAIM: Proceedings and Surveys*, t. 67, p. 261-284, 2020.
- [14] S. LABARTHE, B. POLIZZI, T. PHAN, T. GOUDON, M. RIBOT et B. LAROCHE, “A Mathematical Model to Investigate the Key Drivers of the Biogeography of the Colon Microbiota”, *Journal of theoretical biology*, t. 462, p. 552-581, 2019.
- [15] A. BRIDIER, J.-C. PIARD, C. PANDIN, S. LABARTHE, F. DUBOIS-BRISSENET et R. BRIANDET, “Spatial Organization Plasticity as an Adaptive Driver of Surface Microbial Communities”, *Frontiers in Microbiology*, t. 8, p. 1364, 2017.
- [16] C. CHERBUY, P. VAUGELADE, S. LABARTHE, E. HONVO-HOUETO, B. DARCY-VRILLON, M. WATFORD et P.-H. DUÉE, “The Contribution of Intestinal Gluconeogenesis to Glucose Homeostasis Is Low in 2-Day-Old Pigs”, *The Journal of nutrition*, t. 147, n° 3, p. 361-366, 2017.
- [17] Y. COUDIÈRE, J. HENRY et S. LABARTHE, “An Asymptotic Two-Layer Monodomain Model of Cardiac Electrophysiology in the Atria: Derivation and Convergence”, *SIAM Journal on Applied Mathematics*, t. 77, n° 2, p. 409-429, 2017.
- [18] T. EL BOUTI, T. GOUDON, S. LABARTHE, B. LAROCHE, B. POLIZZI, A. RACHAH, M. RIBOT et R. TESSON, “A Mixture Model for the Dynamic of the Gut Mucus Layer”, *ESAIM: Proceedings and Surveys*, t. 55, E. FRÉNOU, E. MAITRE, A. ROUSSEAU, S. SALMON et M. SZOPOS, éd., p. 111-130, 2016.
- [19] ISAAC NEWTON INSTITUTE FELLOWS, S. WIDDER, R. J. ALLEN, T. PFEIFFER, T. P. CURTIS, C. WIUF, W. T. SLOAN, O. X. CORDERO, S. P. BROWN, B. MOMENI, W. SHOU, H. KETTLE, H. J. FLINT, A. F. HAAS, B. LAROCHE, J.-U. KREFT, P. B. RAINEY, S. FREILICH, S. SCHUSTER, K. MILFERSTEDT, J. R. VAN DER MEER, T. GROSSKOPF, J. HUISMAN, A. FREE, C. PICIOREANU, C. QUINCE, I. KLAPPER, S. LABARTHE, B. F. SMETS, H. WANG et O. S. SOYER, “Challenges in microbial ecology: Building predictive understanding of community function and dynamics”, *The ISME Journal*, t. 10, n° 11, p. 2557-2568, 2016.
- [20] Y. COUDIÈRE, J. HENRY et S. LABARTHE, “A two layers monodomain model of cardiac electrophysiology of the atria”, *Journal of Mathematical Biology*, 2015.
- [21] S. LABARTHE, “Mathématiques et Cardiologie : Modéliser Pour Mieux Comprendre Les Arythmies Auriculaires”, in *5 Jeunes Chercheurs d'avenir 2015*, sér. Promesses de La Science 1, Editions le Pommier, 2015, p. 192.
- [22] S. LABARTHE, J. BAYER, Y. COUDIÈRE, J. HENRY, H. COCHET, P. JAIS et E. VIGMOND, “A Bilayer Model of Human Atria: Mathematical Background, Construction, and Assessment”, *Europace*, t. 16, n° suppl 4, p. iv21-iv29, 2014.
- [23] S. LABARTHE, “Modélisation de l'activité Électrique Des Oreillettes et Des Veines Pulmonaires”, thèse de doct., Université Victor Segalen-Bordeaux II, 2013.

-
- [24] S. LABARTHE, E. VIGMOND, Y. COUDIERE, J. HENRY, H. COCHET et P. JAIS, “Influence of Transmural Slow-Conduction Zones on the Long-Time Behaviour of Atrial Arrhythmia: A Numerical Study with a Human Bilayer Atrial Model”, in *Computing in Cardiology 2013*, IEEE, 2013, p. 1187-1190.
 - [25] S. LABARTHE, E. VIGMOND, Y. COUDIÈRE, J. HENRY, H. COCHET et P. JAIS, “A Computational Bilayer Surface Model of Human Atria”, in *International Conference on Functional Imaging and Modeling of the Heart*, Springer, 2013, p. 27-34.
 - [26] E. VIGMOND, S. LABARTHE, H. COCHET, Y. COUDIERE, J. HENRY et P. JAIS, “A Bilayer Representation of the Human Atria”, in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, p. 1530-1533.
 - [27] R. DUBOIS, S. LABARTHE, Y. COUDIÈRE, M. HOCINI et M. HAISSAGUERRE, “Global and Directional Activation Maps for Cardiac Mapping in Electrophysiology”, in *2012 Computing in Cardiology*, IEEE, 2012, p. 349-352.
 - [28] S. LABARTHE, Y. COUDIERE, J. HENRY et H. COCHET, “A Semi-Automatic Method to Construct Atrial Fibre Structures: A Tool for Atrial Simulations”, in *2012 Computing in Cardiology*, IEEE, 2012, p. 881-884.
 - [29] N. ZEMZEMI, S. LABARTHE, R. DUBOIS et Y. COUDIÈRE, “From Body Surface Potential to Activation Maps on the Atria: A Machine Learning Technique”, in *Computing in Cardiology (CinC), 2012*, 2012, p. 125-128.

Bibliographie générale

- [30] P. T. VAN LEEUWEN, S. BRUL, J. ZHANG et M. T. WORTEL, “Synthetic microbial communities (SynComs) of the human gut: Design, assembly, and applications”, *FEMS Microbiology Reviews*, t. 47, n° 2, fuad012, 2023.
- [31] X. ZHOU, J. WANG, F. LIU, J. LIANG, P. ZHAO, C. K. M. TSUI et L. CAI, “Cross-kingdom synthetic microbiota supports tomato suppression of Fusarium wilt disease”, *Nature Communications*, t. 13, n° 1, p. 7890, 2022.
- [32] A. ALMEIDA, S. NAYFACH, M. BOLAND, F. STROZZI, M. BERACOCHEA, Z. J. SHI, K. S. POLLARD, E. SAKHAROVA, D. H. PARKS, P. HUGENHOLTZ, N. SEGATA, N. C. KYRPIDES et R. D. FINN, “A unified catalog of 204,938 reference genomes from the human gut microbiome”, *Nature Biotechnology*, t. 39, n° 1, p. 105-114, 2021.
- [33] H. A. MABWI, E. KIM, D.-G. SONG, H. S. YOON, C.-H. PAN, E. KOMBA, G. KO et K. H. CHA, “Synthetic gut microbiome: Advances and challenges”, *Computational and Structural Biotechnology Journal*, t. 19, p. 363-371, 2021.
- [34] L. DARRIGADE, “Modélisation Du Dialogue Hôte-Microbiote Au Voisinage de l'épithélium de l'intestin Distal”, thèse de doct., Université Paris-Saclay, 2020.
- [35] C. R. ARMOUR, S. NAYFACH, K. S. POLLARD et T. J. SHARPTON, “A Metagenomic Meta-Analysis Reveals Functional Signatures of Health and Disease in the Human Gut Microbiome”, *MSystems*, t. 4, n° 4, e00332-18, 2019.
- [36] C.-J. DONG, L.-L. WANG, Q. LI et Q.-M. SHANG, “Bacterial communities in the rhizosphere, phyllosphere and endosphere of tomato plants”, *PLOS ONE*, t. 14, n° 11, M.-J. VIROLLE, éd., e0223847, 2019.
- [37] P. KOVATCHEVA-DATCHARY, S. SHOAIE, S. LEE, A. WAHLSTRÖM, I. NOOKAEW, A. HALLEN, R. PERKINS, J. NIELSEN et F. BÄCKHED, “Simplified Intestinal Microbiota to Study Microbe-Diet-Host Interactions in a Mouse Model”, *Cell Reports*, t. 26, n° 13, 3772-3783.e6, 2019.
- [38] J. LLOYD-PRICE, C. ARZE, A. N. ANANTHAKRISHNAN, M. SCHIRMER, J. AVILA-PACHECO, T. W. POON, E. ANDREWS, N. J. AJAMI, K. S. BONHAM, C. J. BRISLAWN et al., “Multi-Omics of the Gut Microbial Ecosystem in Inflammatory Bowel Diseases”, *Nature*, t. 569, n° 7758, p. 655-662, 2019.
- [39] L. M. PROCTOR, H. H. CREASY, J. M. FETTWEIS, J. LLOYD-PRICE, A. MAHURKAR, W. ZHOU, G. A. BUCK, M. P. SNYDER, J. F. STRAUSS, G. M. WEINSTOCK, O. WHITE, C. HUTTENHOWER et THE INTEGRATIVE HMP (iHMP) RESEARCH NETWORK CONSORTIUM, “The Integrative Human Microbiome Project”, *Nature*, t. 569, n° 7758, p. 641-648, 2019.
- [40] J.-C. SIMON, J. R. MARCHESI, C. MOUGEL et M.-A. SELOSSE, “Host-Microbiota Interactions: From Holobiont Theory to Analysis”, *Microbiome*, t. 7, n° 1, p. 1-5, 2019.

- [41] S. L. SALZBERG, “Open questions: How many genes do we have?”, *BMC Biology*, t. 16, n° 1, p. 94, 2018.
- [42] B. J. CALLAHAN, P. J. MCMURDIE et S. P. HOLMES, “Exact Sequence Variants Should Replace Operational Taxonomic Units in Marker-Gene Data Analysis”, *The ISME journal*, t. 11, n° 12, p. 2639-2643, 2017.
- [43] B.-R. KIM, J. SHIN, R. B. GUEVARRA, J. H. LEE, D. W. KIM, K.-H. SEOL, J.-H. LEE, H. B. KIM et R. E. ISAACSON, “Deciphering Diversity Indices for a Better Understanding of Microbial Communities”, *Journal of Microbiology and Biotechnology*, t. 27, n° 12, p. 2089-2093, 2017.
- [44] D. VANDEPUTTE, G. KATHAGEN, K. D’HOE, S. VIEIRA-SILVA, M. VALLES-COLOMER, J. SABINO, J. WANG, R. Y. TITO, L. DE COMMER, Y. DARZI, S. VERMEIRE, G. FALONY et J. RAES, “Quantitative microbiome profiling links gut community variation to microbial load”, *Nature*, t. 551, n° 7681, p. 507-511, 2017.
- [45] A. J. BÄUMLER et V. SPERANDIO, “Interactions between the Microbiota and Pathogenic Bacteria in the Gut”, *Nature*, t. 535, n° 7610, p. 85-93, 2016.
- [46] S. R. BORDENSTEIN et K. R. THEIS, “Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes”, *PLoS biology*, t. 13, n° 8, e1002226, 2015.
- [47] S.-L. E. MARTZ, J. A. K. McDONALD, J. SUN, Y.-g. ZHANG, G. B. GLOOR, C. NOORDHOF, S.-M. HE, T. K. GERBABA, M. BLENNERHASSETT, D. J. HURLBUT, E. ALLEN-VERCOE, E. C. CLAUD et E. O. PETROF, “Administration of defined microbiota is protective in a murine Salmonella infection model”, *Scientific Reports*, t. 5, n° 1, p. 16094, 2015.
- [48] F. RIVERA-CHÁVEZ et A. J. BÄUMLER, “The Pyromaniac inside You: Salmonella Metabolism in the Host Gut”, *Annual review of microbiology*, t. 69, p. 31-48, 2015.
- [49] A. VENKATARAMAN, C. M. BASSIS, J. M. BECK, V. B. YOUNG, J. L. CURTIS, G. B. HUFFNAGLE et T. M. SCHMIDT, “Application of a Neutral Community Model To Assess Structuring of the Human Lung Microbiome”, *mBio*, t. 6, n° 1, M. J. MCFALL-NGAI, éd., e02284-14, 2015.
- [50] J. LI, H. JIA, X. CAI, H. ZHONG, Q. FENG, S. SUNAGAWA, M. ARUMUGAM, J. R. KULTIMA, E. PRIFTI, T. NIELSEN et al., “An Integrated Catalog of Reference Genes in the Human Gut Microbiome”, *Nature biotechnology*, t. 32, n° 8, p. 834-841, 2014.
- [51] R. MARTÍN, S. MIQUEL, P. LANGELLA et L. G. BERMÚDEZ-HUMARÁN, “The role of metagenomics in understanding the human microbiome in health and disease”, *Virulence*, t. 5, n° 3, p. 413-423, 2014.
- [52] A. PILIPENKO, *An Introduction to Stochastic Differential Equations with Reflection*. Universitätsverlag Potsdam, 2014, t. 1.
- [53] M. VAYSSIER-TAUSSAT, E. ALBINA, C. CITTI, J. F. COSSON, M.-A. JACQUES, M.-H. LEBRUN, Y. LE LOIR, M. OGLIASTRO, M.-A. PETIT, P. ROUMAGNAC et al., “Shifting the Paradigm from Pathogens to Pathobiome: New Concepts in the Light of Meta-Omics”, *Frontiers in cellular and infection microbiology*, t. 4, p. 29, 2014.
- [54] K. ATARASHI, T. TANOUE, K. OSHIMA, W. SUDA, Y. NAGANO, H. NISHIKAWA, S. FUKUDA, T. SAITO, S. NARUSHIMA, K. HASE, S. KIM, J. V. FRITZ, P. WILMES, S. UEHA, K. MATSUSHIMA, H. OHNO, B. OLLE, S. SAKAGUCHI, T. TANIGUCHI, H. MORITA, M. HATTORI et K. HONDA, “Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota”, *Nature*, t. 500, n° 7461, p. 232-236, 2013.

-
- [55] B. OKSENDAL, *Stochastic Differential Equations: An Introduction with Applications*. Springer Science & Business Media, 2013.
- [56] F. BOLLEY, I. GENTIL et A. GUILLIN, “Convergence to Equilibrium in Wasserstein Distance for Fokker–Planck Equations”, *Journal of Functional Analysis*, t. 263, n° 8, p. 2430-2457, 2012.
- [57] S. D. EHRLICH, M. CONSORTIUM et al., “MetaHIT: The European Union Project on Metagenomics of the Human Intestinal Tract”, in *Metagenomics of the Human Body*, Springer, 2011, p. 307-316.
- [58] J. RAJENDHRAN et P. GUNASEKARAN, “Microbial Phylogeny and Diversity: Small Subunit Ribosomal RNA Sequence Analysis and Beyond”, *Microbiological research*, t. 166, n° 2, p. 99-110, 2011.
- [59] R. MUÑOZ-TAMAYO, B. LAROCHE, É. WALTER, J. DORÉ et M. LECLERC, “Mathematical modelling of carbohydrate degradation by human colonic microbiota”, *Journal of Theoretical Biology*, t. 266, n° 1, p. 189-201, 2010.
- [60] J. QIN, R. LI, J. RAES, M. ARUMUGAM, K. S. BURGDORF, C. MANICHANH, T. NIELSEN, N. PONS, F. LEVENEZ, T. YAMADA et al., “A Human Gut Microbial Gene Catalogue Established by Metagenomic Sequencing”, *nature*, t. 464, n° 7285, p. 59-65, 2010.
- [61] M. BLAXTER, J. MANN, T. CHAPMAN, F. THOMAS, C. WHITTON, R. FLOYD et E. ABEBE, “Defining Operational Taxonomic Units Using DNA Barcode Data”, *Philosophical Transactions of the Royal Society B: Biological Sciences*, t. 360, n° 1462, p. 1935-1943, 2005.



Annexes

1 Synthèse des travaux de thèse

Je synthétise ici les résultats obtenus durant ma thèse soutenue en décembre 2013 [23].

Contexte : rythmologie et arythmies cardiaques Les tissus cardiaques sont composés de cellules excitables capables de générer, après stimulation, un potentiel d'action qui se propage à travers les tissus dans tout l'organe. La propagation de ce signal électrophysiologique a une double fonction. Localement, ce potentiel d'action génère la contraction du myocarde, mais à l'échelle de l'organe, ce front d'onde assure la bonne coordination des différentes composantes du cœur. Dans des situations pathologiques, la propagation de ce front d'onde peut être perturbée engendrant une contraction désordonnée et inefficace du muscle cardiaque : ce sont les troubles du rythme, ou arythmies. Ma thèse, effectuée dans l'équipe Carmen d'Inria, sous la direction de J.Henry et Y.Coudière et en collaboration avec l'IHU Liryc, s'est intéressée spécifiquement à la modélisation des oreillettes et des veines pulmonaires, afin de mieux comprendre les mécanismes déclencheurs d'arythmies.

Axe de recherche : modélisation de l'activité électrique dans les oreillettes et veines pulmonaires Le modèle d'électrophysiologie cardiaque sur lequel j'ai principalement travaillé, appelé modèle monodomaine, est un modèle macroscopique homogénéisé décrivant l'activité électrique à l'échelle de l'organe par une équation de réaction diffusion, couplée à un système d'équations différentielles ordinaires non linéaires modélisant les échanges ioniques générant le potentiel d'action. Le modèle s'écrit

$$\begin{cases} \beta(C_m \partial_t u - I_{ion}(u, v)) = \operatorname{div}(\sigma \nabla u) + I_{app} \\ \partial v + g(u, v) = 0 \end{cases} \quad (\text{A.1})$$

où les variables u et v sont respectivement le potentiel transmembranaire et un vecteur de variables d'état décrivant l'évolution du potentiel d'action. Les fonctions sources I_{ion} et g

reproduisent l'évolution des courants transmembranaires et des variables d'état. Les paramètres β et C_m sont des caractéristiques électrophysiologiques des cellules, et le tenseur σ modélise l'anisotropie de la diffusion induite par l'organisation des myocytes en fibres musculaires. La fonction I_{app} représente une stimulation extérieure. Des conditions initiales et des conditions frontières sont enfin ajoutées.

Un premier volet théorique de ma thèse a consisté à tirer profit de la faible épaisseur des oreillettes afin de dériver un modèle bisurfacique constitué de deux EDP couplées, l'une modélisant l'endocarde, et l'autre l'épicarde. Nous avons pu démontrer rigoureusement [20] que les solutions $(\hat{u}_\epsilon^{(k)}, \hat{v}_\epsilon^{(k)})$, $k = 1, 2$, du problème surfacique couplé

$$\begin{cases} \beta \left(C_m \partial_t \hat{u}_\epsilon^{(k)} - I_{ion}(\hat{u}_\epsilon^{(k)}, \hat{v}_\epsilon^{(k)}) \right) = \operatorname{div}_{x,y}(\sigma^{(k)} \nabla_{x,y} \hat{u}_\epsilon^{(k)}) + (-1)^k \gamma_\epsilon \frac{\hat{u}_\epsilon^{(1)} - \hat{u}_\epsilon^{(2)}}{\epsilon^2}, \\ \partial \hat{v}_\epsilon^{(k)} + g(\hat{u}_\epsilon^{(k)}, \hat{u}_\epsilon^{(k)}) = 0 \end{cases}, \quad (\text{A.2})$$

convergent à l'ordre 3 vers la moyenne dans l'épaisseur des solutions tridimensionnelles du problème A.1 quand l'épaisseur du tissu ϵ tend vers 0.

Après avoir développé une méthode permettant de modéliser l'orientation complexe des fibres cardiaques dans l'oreillette [28] et prouvé la convergence rigoureusement [17], nous avons proposé un modèle précis des oreillettes humaines sur une géométrie issue d'imagerie médicale et des orientations de fibres issues d'histologie [22], [25], [26] qui ont permis d'identifier l'impact de zones de faible conduction sur des arythmies [24]. J'ai également contribué durant ma thèse au développement d'une méthode de résolution du problème inverse utilisant des RKHS [29] ainsi qu'à des développements de méthode de mesures électrophysiologiques durant des interventions [27]. Ma thèse a été récompensée par différents prix, notamment par le prix Le Monde de la recherche universitaire, aboutissant à la publication d'un résumé "grand public" de mes travaux [21].



Curriculum Vitae

Simon Labarthe

Junior Researcher INRAE, Biogeco/Pléiade - Cestas/Talence, (France)

Mathematical modeling applied to microbial ecology

Education

- U. Bordeaux** Bordeaux, FR
PhD in applied mathematics and scientific computing 2010–2013
– Title : mathematical modeling of the electrical activity of the atria and pulmonary veins.
– Phd supervisor : Yves Coudière et Jacques Henry .
- Agrégation de mathématiques** 2013
National competitive examination for Secondary and University teachers.
- UPMC (cohabitation Univ. Paris Nord, ENPC, ENS, Ecole Polytechnique)** Paris, FR
Master of Applied Mathematics : Numerical analysis and PDEs 2008-2010
– Internship : reaction diffusion equation with Allee's effect: geometry influence.
– Phd supervisor : Jean-Michel Roquejoffre and Henry Berestycki.
- Concours de Recrutement de Professeur des écoles** 2004
Competitive examination for Primary teachers.

Experience

- Permanent researcher (CRCN)** Cestas/Talence, FR
INRAE (Biogeco) & Inria (Pléiade) & University of Bordeaux 2021 –...
- Permanent researcher (CRCN)** Jouy-en-Josas, FR
INRAE (MaIAGE) & University of Paris Saclay 2014 –2021
- Visiting scholar** Davis, US
UC Davis. Visiting scholar (10 months) at A.Bäumler's lab 2017 –2019
- Temporary research assistant (ATER)** Bordeaux, FR
U. Bordeaux & INP Bordeaux in applied mathematics 2013 –2014
- PhD student** Bordeaux, FR
U. Bordeaux & Inria BSO & IHU Liryc. 2010 –2013
– applied mathematics and scientific computing : study of heart arrhythmias
- Primary teacher** Paris, FR
Primary teachers in different regions of France 2004 –2010

📍 Biogeco/Pléiade, Pierroton, 33610 Jouy-en-Josas, France

✉ simon.labarthe@inrae.fr

🌐 Personal website

🆔 0000-0002-5463-7256

 **HAL** [simon-labarthe](https://hal.inrae.fr/simon-labarthe)

- [1] K. Cerk *et al.*, “Community-scale models of microbiomes: Articulating metabolic modelling and metagenome sequencing”, *Microbial Biotechnology*, vol. 17, no. 1, e14396, 2024.
- [2] M. Lecomte *et al.*, “Revealing the dynamics and mechanisms of bacterial interactions in cheese production with metabolic modelling”, *Metabolic Engineering*, S1096717624000302, 2024.
- [3] A. Paulay *et al.*, “Design of a proteolytic module for improved metabolic modeling of *Bacteroides caccae*”, *mSystems*, N. Chia, Ed., e00153–24, 2024.
- [4] L. Darrigade *et al.*, “Deterministic limit of a PDMP model of epithelial tissue interacting with diffusing chemicals and application to the intestinal crypt”, 2023.
- [5] C. Frioux *et al.*, “Accelerating metabolic models evaluation with statistical metamodelling: Application to *Salmonella* infection models”, *ESAIM: Proceedings and Surveys*, vol. 73, V. Ehrlacher *et al.*, Eds., pp. 187–217, 2023.
- [6] C. Frioux *et al.*, *Modéliser les communautés bactériennes pour mieux comprendre leur fonctionnement*. 2023.
- [7] S. Labarthe *et al.*, “Four functional profiles for fibre and mucin metabolism in the human gut microbiome”, *Microbiome*, vol. 11, no. 1, p. 231, 2023.
- [8] D. Marra *et al.*, “Migration of surface-associated microbial communities in spaceflight habitats”, *Biofilm*, vol. 5, p. 100 109, 2023.
- [9] L. Darrigade *et al.*, “A PDMP model of the epithelial cell turn-over in the intestinal crypt including microbiota-derived regulations”, *Journal of Mathematical Biology*, vol. 84, no. 7, p. 60, 2022.
- [10] P. Fournier *et al.*, “The functional microbiome of grapevine throughout plant evolutionary history and lifetime”, in *Advances in Ecological Research*, vol. 67, Elsevier, 2022, pp. 27–99.
- [11] G. Ravel *et al.*, “Inferring characteristics of bacterial swimming in biofilm matrix from time-lapse confocal laser scanning microscopy”, *eLife*, vol. 11, e76513, 2022.
- [12] M. Bourgin *et al.*, “Exploring the Bacterial Impact on Cholesterol Cycle: A Numerical Study”, *Frontiers in Microbiology*, vol. 11, p. 1121, 2020.
- [13] S. Labarthe *et al.*, “A multi-scale epidemic model of salmonella infection with heterogeneous shedding”, *ESAIM: Proceedings and Surveys*, vol. 67, pp. 261–284, 2020.
- [14] S. Labarthe *et al.*, “A mathematical model to investigate the key drivers of the biogeography of the colon microbiota”, *Journal of theoretical biology*, vol. 462, pp. 552–581, 2019.
- [15] A. Bridier *et al.*, “Spatial Organization Plasticity as an Adaptive Driver of Surface Microbial Communities”, *Frontiers in Microbiology*, vol. 8, p. 1364, 2017.
- [16] C. Cherbuy *et al.*, “The contribution of intestinal gluconeogenesis to glucose homeostasis is low in 2-Day-Old pigs”, *The Journal of nutrition*, vol. 147, no. 3, pp. 361–366, 2017.
- [17] Y. Coudière *et al.*, “An Asymptotic Two-Layer Monodomain Model of Cardiac Electrophysiology in the Atria: Derivation and Convergence”, *SIAM Journal on Applied Mathematics*, vol. 77, no. 2, pp. 409–429, 2017.
- [18] T. El Bouti *et al.*, “A mixture model for the dynamic of the gut mucus layer”, *ESAIM: Proceedings and Surveys*, vol. 55, E. Frénod *et al.*, Eds., pp. 111–130, 2016.
- [19] Isaac Newton Institute Fellows *et al.*, “Challenges in microbial ecology: Building predictive understanding of community function and dynamics”, *The ISME Journal*, vol. 10, no. 11, pp. 2557–2568, 2016.
- [20] Y. Coudière *et al.*, “A two layers monodomain model of cardiac electrophysiology of the atria”, *Journal of Mathematical Biology*, 2015.
- [21] S. Labarthe, “Mathématiques et cardiologie : Modéliser pour mieux comprendre les arythmies auriculaires”, in *5 Jeunes Chercheurs d’avenir 2015*, ser. Promesses de La Science 1, Editions le Pommier, 2015, p. 192.
- [22] S. Labarthe *et al.*, “A bilayer model of human atria: Mathematical background, construction, and assessment”, *Europace*, vol. 16, no. suppl 4, pp. iv21–iv29, 2014.
- [23] S. Labarthe, “Modélisation de l’activité électrique des oreillettes et des veines pulmonaires”, Ph.D. dissertation, Université Victor Segalen-Bordeaux II, 2013.
- [24] S. Labarthe *et al.*, “A Computational Bilayer Surface Model of Human Atria”, in *FIMH 2013 - 7th International Conference on Functional Imaging and Modeling of the Heart*, ser. Lecture Notes In Computer Sciences, London, United Kingdom: Springer, 2013.
- [25] S. Labarthe *et al.*, “Influence of transmural slow-conduction zones on the long-time behaviour of atrial arrhythmia: A numerical study with a human bilayer atrial model”, in *Computing in Cardiology 2013*, IEEE, 2013, pp. 1187–1190.
- [26] E. Vigmond *et al.*, “A bilayer representation of the human atria”, in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 1530–1533.
- [27] R. Dubois *et al.*, “Global and directional activation maps for cardiac mapping in electrophysiology”, in *2012 Computing in Cardiology*, IEEE, 2012, pp. 349–352.
- [28] S. Labarthe *et al.*, “A semi-automatic method to construct atrial fibre structures: A tool for atrial simulations”, in *2012 Computing in Cardiology*, IEEE, 2012, pp. 881–884.
- [29] N. Zemzemi *et al.*, “From body surface potential to activation maps on the atria: A machine learning technique”, in *Computing in Cardiology (CinC), 2012*, 2012, pp. 125–128.

📍 Biogeco/Pléiade, Pierroton, 33610 Jouy-en-Josas, France

✉ simon.labarthe@inrae.fr

Awards

- Poster award, CANUM 2012
- Poster award, “Printemps de la Cardiologie” (national conference on cardiology) 2012
- Thesis award, “Thes’AQT Region Aquitaine” 2013
- Thesis award, “le Monde de la recherche universitaire” du journal “Le Monde” 2014

Grants & Projects

- MEM/INRA, MODCHOCYCLE. Amount : 50 K€. co-PI (with M.Rhimi). Bacterial impact on cholesterol cycle. 2016-2018
- Agreenskills+ fellowship . (Marie-Sklodowska-Curie program). Amount: 30K€, PI. Mathematical modeling of Salmonella infection. 2017-2019
- France-Berkeley Fund, Articulate. Amount : 12 K\$. co-PI (with A.Bäumler). articulation of mathematical and experimental models to study gut microbiota ecology 2019-2021
- ANR (PPR-PCA), Pherosensor. WP leader. Early detection of insect pests using pheromone olfactory sensors 2021-2026. Amount: 1,5M€
- ANR (PPR-PCA), VITAE, WP co-leader. Cultivating the grapevine without pesticides : towards agroecological wine- producing socio-ecosystems. 2021-2026. Amount: 3M€
- Mathnum/INRAE, Bioswimmers, PI. 2021. Post-doc fellowship. Modeling microbial swimmers in biofilms. Amount: 60K€
- Action Exploratoire Inria, SLIMMEST, co-PI (with C.Frioux). Surrogate modeling of metabolic models. 2021-2023. Amount: 200K€
- ANR (PEPR Agroécologie numérique), Mistic. WP leader. Computational models of crop plant microbial biodiver- sity 2023-2028. Amount: 1,5M€
- Digit-bio+MICA/INRAE, MODOSAO, co-PI. 2023-2026. PhD fellowship. Modeling bioprocesses integrating metabolic modeling. Amount: 110K€
- Digit-bio/INRAE, Artemis. PI. Digital twins of microbial communities 2024-2025. Amount: 20K€

Internship mentoring

Internships

- Tagathy Goswany, 2014, stage M2
population dynamic model of gut microbiota. With B.Laroche, M.Ribot, T.Goudon
- CEMRACS 2015: 6 weeks internship with 4 PhD/post-doc
Modeling the fluid mechanics of the gut mucus layer. With B.Laroche, M.Ribot, T.Goudon
- Leo Darrigade, 2016, stage M2
modeling the gut microbiota in its environment. With B.Laroche
- Melanie Bourgin, 2016, stage M2
Whole-body model of cholesterol metabolism, including gut microbiota. With M.Rhimi
- Marie-Ange Rasendra, 2016, stage M1
Modeling diffusion in biofilms
- Arthur Braune, 2017, stage M1
Modeling bacterial swimmers in biofilms
- CEMRACS 2018: 6 weeks internship with 3 PhD/post-doc
Modeling Salmonella epidemics in poultry farms. With B.Laroche, M.Ribot, B.Polizzi
- Cloé Mendoza, 2019, stage M1
Modeling the impact of diffusion in epidemic spreading
- Amandine Paulay, 2019, stage M2
Modeling protein metabolism in intestinal bacteria. With B.Laroche, G.Grimaud, M.Leclerc
- Maxime Lecomte, 2020, stage M2
Modeling lactic bacteria for cheese organoleptic qualities improvement. With C.Frioux
- CEMRACS 2021: 6 weeks internship with 2 PhD/post-doc
Surrogate modelling of metabolic models. With D.Sherman, C.Frioux
- Arie Worstman, 2022, stage M2
Accelerating metabolic models with surrogate modelling. With C.Frioux, P.Ugalde
- Rafael Kaempfer Danin, 2023, stage M2
Surrogate modelling for a host-microbiota model. With C.Frioux, P.Ugalde

📍 Biogeco/Pléiade, Pierroton, 33610 Jouy-en-Josas, France

✉ simon.labarthe@inrae.fr

🌐 Personal website

🆔 0000-0002-5463-7256

🏠 HAL simon-labarthe

PhD supervision

- Leo Darrigade, 2017-2020, EDMH Paris Saclay, *Host-microbiota cross-talk near distal intestinal epithelium, with B.Laroche et C.Cherbuy*
- Amandine Paulay, 2019-2022, Abies UPS, CIFRE *Modeling protein degradation in the gut microbiota, with B.Laroche, M.Leclerc, E.Maguin, G.Grimaud*
- Sahak Yeghiazaryan, 2023-2026, U.Montpellier *Coupling genome-based models with a thermodynamic framework of microbial syntrophy in anaerobic digestion, with E.Le Quemener, N.Bernet*
- Sthyye Tatho, 2024-2027, EDMI U.Bordeaux *Integration multi-omics time-series in dynamic models of microbial community, with V.Baldazzi*

Post-doc supervision

- Floriane Colas, 2019-2020, API-SMAL (U.Saclay BASC Labex project) *Multimodel analysis of the effects of semi-natural habitat increase in agricultural landscapes, with F.Accatino*
- Guillaume Ravel, 2021, MathNum fellowship *Modeling of bacterial swimmers to enhance diffusion in biofilms, with M.Bergmann, A.Iollo, A.Trubuil, R.Briandet*
- Pablo Ugalde Salas, 2021-2023, SLIMMEST *RKHS metamodeling for metabolic models fast approximation*
- Thibault Malou, 2022-2024, Pherosensor *Locating insect pests by solving an inverse problem for pheromon dispersion*

Teaching

- U. Paris Saclay, master 2 MSV, 2017 *system dynamics, PDEs, model reduction. TD (6h)*
- U. Evry Val-d'Essonne, L2, 2015-2017 *analysis, probability. TD (56h/an)*
- U. Bordeaux, L1-L2, 2014 *analysis. TD (96h)*
- IPB, Bordeaux., L2, 2013 *Scientific computing in Fortran 90. TD (96h)*
- IUT HSE, Bordeaux, L2, 2012-2013 *Probability and statistics. TD (128h)*
- Ecole primaire, MEN, 2004-2010 *Primary teacher. Full time*

Responsibilities, Expertise & Collective duties

INRAE level

- Elected to CSS MISTI (Commission Scientifique Spécialisée: INRAE evaluation comity), 2020-2024. Nominated to the steering comity (Bureau).
- Steering comity of Metaprogram Holoflux (since 2022).
- Jury: CRCN competitive selection (2021, 2022, 2024). President of competitive selection: IR (2024).
- Representation of INRAE Bordeaux Center in R3IA Regional network (2022-).

Other institutes

- Jury: MCF, Univ. Côte d'Azur (2020).
- Scientific comity for GDR MathSav Conference (2022)

Reviews

- Journals: ISME, Plos ONE, mSystems, NPJ Biofilm, Journal of Theoretical Biology, Journal of Biological Systems, Computers in Biology and Medecine
- Expertise for grants: Max Plank Institute (2023), KFG (Austrian science funding agency, 2024), U.Paris Saclay (ABIENS, doctoral grant, 2024), annual AMI of Holoflux metaprogram.

Other

- Biogeco Unit: "Groupe animation". Seminary organization comity.
- Member of Jobim 2025 organization comity.

📍 Biogeco/Pléiade, Pierroton, 33610 Jouy-en-Josas, France

✉️ simon.labarthe@inrae.fr

Posters

- Journées MathSav, 10/2022, Besançon, France
- JOBIM, 07/2022, Rennes, France
- ECMTB, 07/2018, Lisbon, Portugal
- Agreenskills meeting, 06/2018, Edimburgh.
- Workshop Liryc, 10/2013, Pessac.
- Computing in Cardiology, 09/2012, Krakow, Poland.
- Mayneord Phillips Summer School Cardiac Imaging & Modelling, 07/2012, Oxford, UK.
- National congress of numerical analysis (CANUM 2012), 05/2012, Super Besse.
- National congress of cardiology « Printemps de la Cardiologie » , 04/2012, Bordeaux.

Presentations

- Workshop Metabolisme, 03/2024, Toulouse, France, **invited**
- Ferment'IA, 09/2023, Saclay, France, **invited**
- Mini-Symposium Micalis, 06/2023, Jouy-en-Josas, France
- Séminaire EGFV, 02/2023, Bordeaux, France
- ModStatsAP, 12/2022, Paris, France, **invited**
- Journées Françaises de la Nutrition, 11/2022, Toulouse, France, **invited**
- IBSB, 09/2022, Jena, Germany (hybrid)
- Workshop Gut models, 11/2021, Besançon, France, **invited**
- Séminaire LABRI, 03/2022, Bordeaux, France
- Séminaire MIAT, 03/2021, Toulouse, France
- Workshop ENS, 10/2020, on-line
- Séminaire IMB (with C.Tiffany), 03/2020, Bordeaux, France

- Séminaire IMB, 02/2020, Bordeaux, France
- Séminaire MaIAGE, 01/2020, Jouy, France
- Séminaire M2 MathSV, Ecole Polytechnique, 09/2019, Palaiseau, France
- Séminaire Biogeco, 09/2019, Cestas, France
- Bäumlér's lab meeting, 03/2019, Davis, USA
- CEMRACS, 08/2018, Marseille France
- Workshop, Fields institute, 05/2018, Toronto, Canada.
- ECMTB, 07/2016, Nottingham, UK.
- CEMRACS, 08/2015, Marseille, France.
- Séminaire MaIAGE, 10/2014, Jouy, France.
- Franco-Romanian symposium of Applied Mathematics, 08/2014, Lyon, France.
- PhD defense, 12/2013, Talence, France.
- CINC 2013, 09/2013, Zaragoza, Espagne.
- EMBC 2013, 07/2013, Osaka, Japon.
- FIMH 2013, 06/2013, London, UK.
- symposium "AF: Clinical challenges for biophysical modelling". 06/2013, London .
- PlaFrim day. 04/2013 ,LABRI.
- Print. de la Cardiologie, 04/2013, Marseille.
- Séminaire Etudiant. LMAP. 02/2013. Pau.
- Journal Club IHU Liryc, 01/2013. Bordeaux.
- Summer School on Cardiac Imaging & Modelling. 07/2012, Oxford, UK.
- Print. de la Cardiologie, 04/2012, Bordeaux.
- Séminaire IHU Liryc, 04/2012, Bordeaux.
- Séminaire Etudiant, IMB, 01/2012, Bordeaux.



Résumé étendu

Après une thèse soutenue en décembre 2013 portant sur la modélisation mathématique de l'électro-physiologie cardiaque, j'ai été recruté comme chargé de recherche à l'INRAe en septembre 2014 dans l'unité MaIAGE afin de travailler sur la modélisation mathématique de communautés microbiennes. J'ai ensuite bénéficié d'une mobilité vers l'unité Biogeco (Cestas) et je fais depuis partie de l'équipe Pléiade (INRAe/Inria). J'ai fait le choix de ne pas présenter mes travaux de thèse dans cette synthèse, dont une brève description peut être trouvée en annexe, mais d'exposer les différents axes de recherche menés depuis mon recrutement à l'INRAe pour la modélisation de communautés microbiennes. Je présente d'abord une brève introduction sur les communautés microbiennes et sur les enjeux de modélisation, puis ensuite les différents axes de recherche que j'ai suivis. Pour chaque axe, une contextualisation succincte de la problématique biologique est proposée, puis l'axe de recherche en mathématique ou modélisation est décrit.

1 Introduction : modélisation de communautés microbiennes

Les communautés microbiennes sont des écosystèmes extrêmement complexes, tant par le nombre d'individus qu'ils contiennent, que par leurs diversités taxonomiques ou fonctionnelles. Par exemple, pour le microbiote intestinal, il peut y avoir jusqu'à 10^{11} bactéries par gramme de fèces issues de plusieurs centaines d'espèces, dont le métagénome, c'est à dire l'ensemble des gènes microbiens de la communauté, contient un nombre de gènes deux ordres de grandeur supérieur au nombre de gènes du génome humain. Ces communautés peuvent être étudiées de manière holistique, en analysant des données métagénomiques à l'échelle de la communauté, ou réductionniste en analysant des communautés microbiennes réduites. De plus, ces communautés vivent en interaction avec un environnement physico-chimique qui peut être régulé par un organisme hôte vivant en symbiose avec la communauté. Ces liens intriqués entre hôte et microbes ont amené le concept d'holobionte, l'ensemble symbiotique formé par l'hôte et son microbiote résidant. Enfin, différents hôtes peuvent être en contact, engendrant des flux microbiens entre leurs communautés symbiotiques ou des flux de pathogènes (voir Fig. 1.1 pour un résumé de ces

différents niveaux d'étude des communautés microbiennes).

J'étudie les communautés microbiennes à ces différentes échelles de complexité avec des équations de dynamique de population pour une population microbienne $i \in [0, N_b]$ et un métabolite $j \in [0, N_m]$

$$\partial_t b_i + \text{div}(F_{b,i}(b, m)) = G_{b,i}(b, m) \quad (\text{C.1})$$

$$\partial_t m_j + \text{div}(F_{m,j}(b, m)) = G_{m,j}(b, m) \quad (\text{C.2})$$

Dans ce système, b et m sont les vecteurs de dimension N_b et N_m regroupant les densités de microbes et métabolites considérés. $F_{b,i}$ et $F_{m,j}$ désignent les opérateurs différentiels décrivant les flux locaux de microbes ou métabolites. Les fonctions $G_{b,i}$ et $G_{m,j}$ décrivent les taux de croissance microbienne et de consommation ou de production de métabolites. Du point de vue de la modélisation, les différentes échelles d'études des communautés microbiennes impliquent différentes problématiques mathématiques. Les approches holistiques visent à analyser des données omiques par des approches d'apprentissage pour identifier la structure du système (C.1)-(C.2), c'est à dire les nombres N_b , N_m de populations microbiennes et de métabolites et les fonctions G_b et G_m (sec. 2). Dans les approches réductionnistes, le défi consiste à profiter du petit nombre d'espèces microbiennes pour bâtir les fonctions G_b et G_m à partir de données génomiques. Les modèles résultants sont précis mais gourmands numériquement, ce qui amène à des problématiques d'approximation de modèle (sec. 3). Considérer les interactions physico-chimiques avec l'environnement consiste principalement à caractériser les fonctions $F_{b,i}$ et $F_{m,j}$ ou à ajouter un nouveau compartiment "hôte". Ceci m'a amené à considérer des modèles réduits de mécanique des fluides (sec. 4.1), à étudier des modèles stochastiques de crypte intestinale (sec. 4.2) ou à inférer les interactions de micro-nageurs avec leur environnement (sec. 4.3).

Je note par ailleurs que j'ai participé à des revues sur la modélisation de communautés microbiennes de manière générale [19] ou en particulier chez la vigne [10].

2 Approches holistiques : analyse des liens structures-fonctions

Contexte : données métagénomiques pour la caractérisation fonctionnelle de communautés Les données métagénomiques permettent de caractériser le potentiel fonctionnel d'une communauté microbienne prise dans son ensemble. Produire des données métagénomiques consiste à détruire les parois cellulaires microbiennes d'un échantillon de communauté pour accéder au matériel génomique microbien, qui est découpé et mélangé : on parle alors de métagénome. Un séquenceur tire alors au hasard des fragments génomiques du métagénome et les séquence. Des pipelines bioinformatiques visent alors à regrouper ces fragments en gènes (contigage) puis en génome. Ces gènes peuvent être ensuite annotés fonctionnellement pour constituer des catalogues de gènes. Enfin, les fragments initiaux peuvent être associés aux gènes du catalogue pour évaluer leur fréquence dans le métagénome, ce qui caractérise fonctionnellement l'échantillon. Le produit final de ce pipeline est donc une matrice X , de N_s (nombre d'échantillons) lignes et N_g (nombre de gènes du catalogue) colonnes, où $X_{i,j}$ est la fréquence du gène j dans l'échantillon i . Ces matrices sont grandes, N_s étant de l'ordre du millier et N_g de l'ordre de la dizaine de millions.

Axe de recherche : NMF pour l'apprentissage de populations fonctionnelles Afin d'utiliser la matrice X pour inférer N_b , N_m , G_b et G_m du système (C.1)-(C.2), deux étapes de simplification ont été menées. D'abord une méthode basée sur une connaissance biologique *a priori* a été développée : des gènes spécifiques du métabolisme des fibres dans l'intestin humain ont été

sélectionnés et regroupés par fonctions à différents niveau du réseau métabolique de dégradation des fibres en $N_{AFT} = 101$ traits fonctionnels, résultant en une matrice X_{AFT} de dimension $N_s \times N_{AFT}$. Ensuite, une méthode de réduction de dimension (Non-négative Matrix Factorization, ou NMF) a été mise en œuvre afin d'interpréter la matrice X_{AFT} comme le mélange pondéré de k profils fonctionnels "universels". Plus précisément, nous avons résolu le problème : trouver la matrice de poids W^* et la matrice de profils universels H^* tels que

$$W^*, H^* = \underset{\substack{W \in \mathbb{R}^{N_s, k}, H \in \mathbb{R}^{k, N_{AFT}} \\ W \geq 0, H \geq 0 \\ F \cdot H \leq 0}}{\operatorname{argmin}} \left(\| (X_{AFT} - W \cdot H) \cdot D^{-1} \|_2^2 + \mu \left(\|W\|_2^2 + \|H \cdot D^{-1}\|_{1,2}^2 \right) \right). \quad (\text{C.3})$$

où D est une matrice diagonale de renormalisation et F est une matrice définissant des contraintes linéaires entre gènes telles qu'observées dans des génomes représentatifs du microbiote intestinal. Les hyper-paramètres k et μ sont sélectionnés selon différents critères (bi-cross validation, erreur de reconstruction et stabilité d'identification de H). Comme les profils contenus dans la matrice H sont directement reliés au réseau métabolique de dégradation des fibres et donc interprétables biologiquement, ils peuvent être utilisés pour bâtir les fonctions sources G_b et G_m d'un système (C.1)-(C.2) contenant $N_b = k$ populations fonctionnelles. Ce travail est publié récemment [7]. Cet axe de recherche est inclus dans le projet Mistic coordonné par Pléiade et financé par le PePR Agroécologie Numérique.

3 Approches réductionnistes pour l'analyse des communautés

Contexte : jumeaux numériques de SynCom. Une approche réductionniste pour analyser des interactions microbiennes consiste à sélectionner parmi la communauté un petit nombre d'individus dont le potentiel fonctionnel est représentatif de la communauté. Une fois isolés, ces N_b organismes peuvent être cultivés en communautés synthétiques (SynCom) de manière contrôlée et répétable. Du point de vue de la modélisation, l'avantage est de pouvoir séquencer complètement les individus de la SynCom et dériver après annotation fonctionnelle du génome un réseau métaboliques représentatif de chaque espèce. Ce réseau métabolique peut être représenté par une matrice de réaction R contenant autant de lignes n_m que de métabolites et autant de colonnes n_r que de réactions dans le réseau, chaque élément de la matrice étant le coefficient stœchiométrique pour le métabolite m dans la réaction r . Il s'agit ensuite de prédire les flux métaboliques de métabolites transformés par le réseau pour dériver les fonctions $G_{b,i}$ et $G_{m,j}$ avec les relations

$$G_{b,i} = \nu_b^{(i)} b_i \quad G_{b,j} = \sum_{i=1}^{N_b} \nu_{m,j}^{(i)} b_i$$

où $\nu_b^{(i)}$ est le flux de production de biomasse pour la bactérie i et $\nu_{m,j}^{(i)}$ est le flux de production (ou de consommation) du métabolite j par la bactérie i . Une méthode classique pour prédire les flux $\nu_b^{(i)}$ et $\nu_{m,j}^{(i)}$ est la méthode de Flux Balance Analysis (FBA) résolvant le problème de programmation linéaire

$$\text{Trouver } \nu_b^*, \nu^* \text{ tels que } \nu_b^*, \nu^* = \underset{\substack{R \cdot \nu = 0 \\ \nu_l(b, m) \leq \nu \leq \nu_u(b, m)}}{\operatorname{argmax}} \nu_b. \quad (\text{C.4})$$

Les contraintes $\nu_l(b, m)$ et $\nu_u(b, m)$ sont des fonctions dépendant des variables d'état du système (C.1)-(C.2) permettant de coupler fortement les équations (C.1)-(C.2) et 3.2 via la fonction $\mathcal{F} : (\nu_l(b, m), \nu_u(b, m)) \mapsto \nu_b^*, \nu^*$. L'avantage de cet environnement de modélisation, connu sous la dénomination de *dynamic FBA*, est qu'il prend en compte l'ensemble des réactions métaboliques contenues dans le génome de chaque espèce. Il permet, en résolvant les équations (C.1)-(C.2) couplées pour le calcul de la fonction source au problème (C.4), de constituer des jumeaux numériques de SynCom cultivées expérimentalement, et de comparer certains paramètres comme le suivi dynamique de la biomasse et de la production de métabolites.

Axe de recherche : approximation de modèle par *machine learning*. La résolution du problème (C.1)-(C.2) couplé à (C.4) nécessite la résolution du problème d'optimisation (C.4) à chaque pas d'espace et de temps du système d'EDP (C.1)-(C.2). Bien que des solveurs de problèmes de programmation linéaires particulièrement efficaces soient disponibles, ce couplage représente une charge computationnelle prohibitive pour la résolution d'EDP. Nous avons donc développé une approche de métamodélisation permettant de calculer une solution approchée du problème (C.4), et donc de la fonction \mathcal{F} , pour un coût numérique bien moindre. Concrètement, nous avons assemblé une base d'apprentissage composée d'une matrice $\mathbf{X} = (\nu_l(b_s, m_s), \nu_u(b_s, m_s))_{1 \leq s \leq N_s}$ de contraintes du problème (C.4) calculées pour N_s points de l'espace d'état (b, m) du système (C.1)-(C.2), ainsi que de la matrice $\mathbf{Y} = \mathcal{F}(X)$. Ensuite, nous avons approché \mathcal{F} dans un espace de RKHS particulier, appelé ANOVA-RKHS. L'ANOVA-RKHS permet de discrétiser le problème d'approximation de la fonction \mathcal{F} , via le *Representer theorem*, mais aussi de réaliser une sélection de variables sur les colonnes de X . Plus précisément, il s'agit de résoudre le problème d'inférence, pour une colonne j de \mathbf{Y} ,

$$\hat{\theta}_{0,j}, (\hat{\theta}_{p,j})_{p \in \mathcal{P}} := \underset{\substack{\theta_{0,j} \in \mathbb{R} \\ \theta_{p,j} \in \mathbb{R}^{N_s}, \forall p \in \mathcal{P}}}{\operatorname{arg\,min}} \quad \|\mathbf{Y}_j - (\theta_{0,j} \mathbb{1} + \sum_{p \in \mathcal{P}} K_p \theta_{p,j})\|_2^2 + \lambda \sum_{p \in \mathcal{P}} \|\theta_{p,j}\|_2. \quad (\text{C.5})$$

Dans cette équation, \mathcal{P} est l'ensemble des multi-indices sur les colonnes de \mathbf{X} , K_p est une matrice de Gram d'un noyau reproduisant k_p défini sur les colonnes de \mathbf{X} correspondant au multi-indice p et λ est un hyperparamètre de régularisation pour le terme groupe-lasso $\sum_{p \in \mathcal{P}} \|\theta_{p,j}\|_2$. Ce terme group-lasso permet la sélection de variables. Après inférence, une approximation de \mathcal{F} en un nouveau point (ν_l, ν_u) externe à la base d'apprentissage peut être obtenue avec la formule explicite

$$\hat{\mathcal{F}}_j(\nu_l, \nu_u) := \hat{\theta}_{0,j} + \sum_{p \in \mathcal{P}} F_p(\nu_l, \nu_u) \cdot \hat{\theta}_{p,j} \quad (\text{C.6})$$

où $F_p(\nu_l, \nu_u)$ est le vecteur de dimension N_s obtenu par l'évaluation du noyau $k_p((\nu_l, \nu_u), \cdot)$ en les N_s points de \mathbf{X} .

Dans l'article [5], nous avons mis en œuvre cette méthode de métamodélisation pour une SynCom de deux bactéries dans un système décrivant une infection à la *Salmonelle* et nous avons obtenu une accélération par un facteur 50 du temps de calcul pour une précision acceptable dans l'approximation de la dynamique du système. Cet axe de recherche bénéficie d'une action Exploratoire de l'Inria (Aex SLIMMEST) qui finance le post-doc de Pablo Ugalde Salas que j'encadre. Par ailleurs, la modélisation de SynCom est aussi l'objet de la thèse d'Amandine Paulay (bourse CIFRE) que je co-supervise.

4 Interaction du microbiote avec son environnement

4.1 Modèle du microbiote intestinal

Contexte : environnement biophysique du microbiote intestinal. Comme indiqué dans la section 2, l'environnement biochimique du microbiote intestinal peut être approché par les métabolites et les réactions impliquées dans le réseau métabolique de dégradation des fibres, principal substrat pour les populations microbiennes. Cependant, l'environnement physique est également particulièrement spécifique : le transit intestinal tend à lessiver le microbiote intestinal. Par ailleurs, une couche de mucus, un fluide visqueux enveloppant les cellules épithéliales de l'intestin, permet à certaines bactéries de se protéger du transit intestinal. Enfin, certaines bactéries sont capables de se déplacer activement vers un métabolite cible par chimiotactisme. Nous avons ainsi un milieu multi-phasique de différentes viscosités dont la dynamique détermine un champ d'advection des populations microbiennes, qui peuvent elles-mêmes se mouvoir activement.

Axe de recherche : microbiote intestinal en interaction avec son environnement multi-phasique, simplification de modèle. Il s'agit dans ce contexte de définir les opérateurs $F_{b,i}$ et $F_{m,j}$ du système (C.1)-(C.2). Nous utilisons un modèle de mélange pour modéliser ce système : en notant $\mathcal{I} := \{l_{muc}, l_{lum}, b_1, \dots, b_{N_b}\}$ l'ensemble des différentes phases du fluide multi-phasique, i.e. respectivement le mucus, le liquide luminal les N_b populations bactériennes, et en notant $f_k, k \in \mathcal{I}$ la fraction volumique correspondante, nous supposons que $\sum_{k \in \mathcal{I}} f_k = 1$ (saturation du milieu).

Nous supposons ensuite que chaque phase est transportée par une vitesse commune u provenant du flux intestinal, et que chaque phase bactérienne a additionnellement sa propre vitesse active \mathcal{V}_{b_i} . On suppose que la dynamique de u est décrite par l'équation de Stokes

$$\nabla p - \operatorname{div}(\mu(f_{l_{muc}}, f_{l_{lum}})D(u)) = 0 \quad (\text{C.7})$$

où p est un champ de pression, $D(u) = \frac{1}{2}(\nabla u + \nabla u^T)$, et μ est la viscosité locale dépendant de la composition locale du fluide en liquide luminal et mucus. En sommant les équations du système (C.1)-(C.2) et en utilisant la condition de saturation, on obtient la condition de compressibilité

$$\operatorname{div}(u) = -\operatorname{div}\left(\sum_{i=1}^{N_b} f_{b_i} \mathcal{V}_{b_i}\right).$$

En approchant la géométrie de l'intestin par un cylindre dont on suppose que le rayon est négligeable devant la longueur, on peut dériver une approximation analytique des champs de vitesse u et \mathcal{V}_{b_i} . Cette dérivation est une adaptation de la dérivation des écoulements laminaires à partir des équations de Stokes au cas où la viscosité est variable. Dans ce régime asymptotique, la résolution du système (C.1)-(C.2) est simplifiée par le calcul analytique du champ de vitesse, ce qui nous a permis de dériver un modèle du microbiote intestinal incluant des écoulements multi-phasiques pour un coût numérique réduit. Ces modèles ont été valorisés par la publication [18] pour le modèle d'écoulement, et par [14] pour le modèle du microbiote intestinal. Cette publication contient également une analyse de sensibilité du modèle aux différents paramètres permettant d'identifier les paramètres déterminant la biogéographie des populations bactériennes dans l'intestin. Un autre exemple d'interaction hôte-microbiote est un modèle compartimentaire modélisant l'impact du microbiote intestinal sur le cycle du cholestérol dans les différents organes de l'hôte [12].

4.2 Modélisation des régulations de la crypte intestinale par le microbiote

Contexte : renouvellement de la crypte intestinale influencé par le microbiote. Les cryptes intestinales sont des villosités composées de quelques centaines de cellules épithéliales composant la surface du colon. C'est également le lieu de l'essentiel des mécanismes déterminant l'impact du microbiote sur l'hôte. En effet, les cellules épithéliales métabolisent principalement des composés produits par le microbiote, les acides gras à chaîne courte (AGCC). De plus, ces AGCC régulent également la différenciation des cellules souches en cellules différenciées. A l'inverse, les cellules de l'hôte régulent les niveaux d'oxygène dans la crypte, ce qui impacte les populations bactériennes.

Axe de recherche : un modèle PDMP de la dynamique de la crypte intestinale incluant le microbiote. Afin d'intégrer les observations biologiques réalisées à l'échelle cellulaire, nous avons formalisé un modèle individu-centré de la crypte intestinale : chaque cellule est modélisée individuellement par des variables d'état décrivant sa position dans l'espace et son type cellulaire. Afin de coupler des mécanismes stochastiques tels que les événements de différenciation, de division ou de mort, avec des mécanismes déterministes comme les efforts mécaniques des cellules les unes sur les autres, nous avons opté pour un modèle de type *Piecewise-Deterministic Markov Process* (PDMP) : entre les événements stochastiques décrits par des sauts dans l'espace d'état, l'évolution du système est formalisé par des systèmes dynamiques déterministes. Le couplage avec les populations microbiennes est effectué dans la paramétrisation des taux et des intensités de saut qui dépendent des densités locales en AGCC. Nous avons ensuite dérivé rigoureusement, sous hypothèse de grande population, une approximation déterministe du système de PDMP sous la forme d'une EDP non locale. Enfin, sous hypothèse de petit rayon d'interaction entre cellules, nous avons formellement dérivé une deuxième approximation sous forme d'une équation de réaction-diffusion non linéaire. Ce travail a été réalisé dans le cadre de la thèse de Léo Darrigade, que j'ai co-encadrée avec Béatrice Laroche. Le modèle PDMP a été valorisé dans [9] où le modèle PDMP a été formalisé et exploré numériquement. Puis l'approximation déterministe a été utilisée pour réaliser une analyse de sensibilité du modèle.

4.3 Interactions de micro-nageurs avec leur environnement dans un biofilm

Contexte : biofilms et micro-nageurs comme facilitateurs de contrôle Certaines bactéries ont la capacité de former des biofilms : les bactéries sécrètent de larges protéines qui s'entremêlent pour former un mélange visqueux, la matrice extracellulaire, dans lequel elles s'agglomèrent. Sous forme de biofilm, la communauté bactérienne est beaucoup plus résistante aux stress mécaniques ou chimiques (comme un biocide ou un désinfectant). Lorsque les bactéries sont pathogènes, la formation de biofilm pose d'importants problèmes sanitaires, comme par exemple dans des applications médicales ou agroalimentaires. Il y a quelques années, des collaborateurs spécialisés dans l'étude des biofilms par microscopie ont fait l'observation que des bactéries nageuses avaient la possibilité de nager dans des biofilms de *Staphylococcus aureus*, laissant derrière eux des pores ouvertes dans le biofilm et la matrice extracellulaire. Ils ont alors démontré par une preuve de concept que ces pores amélioreraient la diffusivité et la pénétrabilité, et donc l'efficacité de biocides permettant d'éradiquer le biofilm. Ils ont alors criblé une centaine de bactéries nageuses pour déterminer leur impact comme facilitateur de traitement de biofilm.

Axe de recherche : caractérisation de bactéries nageuses dans un biofilm. Afin d'identifier des cocktails de bactéries nageuses optimisant le traitement des biofilms par biocide, il est devenu primordial de caractériser de manière quantifiable leur nage, notamment leurs interactions avec

le biofilm. Dans le cadre du post-doc de Guillaume Ravel, financé par le département Mathnum de l'INRAe et que j'ai co-encadré avec Angelo Iollo et Michel Bergmann, nous avons développé un modèle de marche aléatoire de bactéries nageuses, en interaction avec le biofilm sous-jacent. L'inférence de ce modèle à partir de données de microscopie confocale a permis de caractériser l'interaction de ces nageuses avec le biofilm. Plus précisément, le modèle s'écrit

$$d\mathbf{v} = \underbrace{\gamma(\alpha(b) - \|\mathbf{v}\|) \frac{\mathbf{v}}{\|\mathbf{v}\|}}_{\text{sélection de vitesse}} dt + \underbrace{\beta \frac{\nabla b}{\|\nabla b\|}}_{\text{sélection de direction}} dt + \underbrace{\eta dt}_{\text{terme aléatoire}}. \quad (\text{C.8})$$

C'est un modèle de conservation du mouvement supposant que les variations de vitesses \mathbf{v} des micro-nageurs sont issues de deux forces déterministes perturbées par un terme aléatoire gaussien d'intensité η . La première force sélectionne le module de la vitesse de nage, l'amenant vers la valeur $\alpha(b)$ dépendant de la densité locale de biofilm b avec un temps de relaxation $1/\gamma$. Ce terme suppose donc que la densité locale de biofilm détermine une vitesse "atteignable" que le nageur est capable de détecter et d'atteindre. La fonction α est une fonction affine. La seconde force influe sur la direction : elle suppose que le micro-nageur est capable de détecter les gradients de densité de biofilm et d'orienter sa nage vers des gradients favorable avec intensité β . L'inférence a été effectuée par inférence Bayésienne sur 3 populations différentes de *Bacillus* et a permis de relier les caractéristiques d'interaction avec le biofilm (c'est à dire les différents paramètres de l'équation) avec des caractéristiques physiologique des espèces de micro-nageurs, notamment la forme du complexe de flagelles. Le modèle et la méthode d'inférence ont été valorisés par un article [11] et plusieurs présentations en conférence (IBSB2022, CEMRACS 2021, ECMTB2018). J'ai également participé à un article de revue sur les mécanismes spatiaux dans les biofilms [15].

5 Flux microbiens entre holobiontes et épidémiologie

Contexte : des flux microbiens entre holobiontes. Jusqu'ici, nous nous sommes intéressés à la caractérisation d'holobiontes : une communauté microbienne en interaction avec un hôte. Nous avons pu étudier la diversité inter-individuelle des microbiotes, notamment des microbiotes intestinaux (cf Sec. 2). Si la structure des communautés microbiennes symbiotiques dépendent pour une large part des caractéristiques génétiques de l'hôte (qui participe à la sélection des communautés microbiennes par son système immunitaire ou par la régulation de niches nutritionnelles favorables aux symbiotes), elle dépend également de flux microbiens provenant de l'extérieur, notamment des interactions avec d'autres hôtes. Ces interactions deviennent cruciales dans le cadre de microbes pathogènes, notamment pour comprendre leur diffusion et leur épidémiologie. En particulier, il est important de comprendre la diffusion de pathogènes entériques comme des *Salmonelles* dans des élevages, dans le but de contrôler la présence de ces pathogènes dans la chaîne alimentaire.

Axe de recherche : modèle populationnel incluant la dynamique d'infection intra-hôte. Dans le cadre d'un projet européen One Health (Momir-PPC), nous avons proposé un projet CEMRACS en 2018 s'intéressant à la modélisation de propagation de *Salmonelles* dans les élevages de volailles. Un modèle intra-hôte, déterminant la dynamique interne du pathogène, a d'abord été défini sous forme d'une SDE :

$$dP = \underbrace{\left(P(K - P)(A - \alpha P + C \frac{P^n}{P^n + p_*^n}) \right)}_{=F(p): \text{drift déterministe}} dt + \sigma dB \quad (\text{C.9})$$

Cette équation considère une perturbation gaussienne autour d'un drift déterministe de la densité intra-hôte en pathogène P . Ce drift prend en compte différents mécanismes de pathogénicité comme l'inflammation (regroupée dans le terme $C \frac{p^n}{p^n + p_*^n}$, où p_* et C sont des paramètres), ou l'interaction avec le microbiote résident (regroupée dans les autres paramètres K , A et α). Selon les valeurs de paramètre, ce drift peut avoir 2 à 5 états stationnaires, et la valeur $P = 0$ est instable si et seulement si $A > 0$.

Cette SDE permet d'écrire une équation à l'échelle de la population d'hôte, structurée en densité intra-hôte de pathogène :

$$\partial_t s(t, p) = -\partial_p (F(p) s(t, p)) + \frac{\sigma^2}{2} \partial_{pp}^2 s(t, p) \quad (\text{C.10})$$

Dans cette équation, $s(., p)$ est la densité d'hôte dans l'élevage ayant une densité interne de pathogène p . Les variations en niveau de pathogène suivent le drift déterministe de l'infection F ou évoluent aléatoirement via le terme de diffusion. Cette équation peut être couplée à un "réservoir" décrivant le niveau environnemental de pathogène dans l'élevage, ce qui permet de prendre explicitement en considération les flux de pathogène inter-hôtes

$$\begin{aligned} \partial_t s(t, p) &= -\partial_p ((F(p) + \beta_{in}(p)r(t) - \beta_{ex}(p)) s(t, p)) + \frac{\sigma^2}{2} \partial_{pp}^2 s(t, p) \\ \partial_t r(t) &= -\left(\gamma^* + \int_0^K s(t, p) \beta_{in}(p) dp \right) r(t) + \int_0^K s(t, p) \beta_{ex}(p) dp \end{aligned}$$

Dans cette équation, r est la densité en pathogène du réservoir, β_{in} et β_{ex} sont des fonctions décrivant l'ingestion et l'excrétion de pathogènes et γ^* est un terme de mort de pathogène en milieu extérieur.

Ces équations ont permis d'étudier les distributions stationnaires de population d'hôtes et d'étudier différentes modalités de contrôle comme le nettoyage ou l'application d'antibiotiques. Ce travail a été valorisé par un article [13].

6 Conclusion

Ma recherche en mathématique et modélisation s'intéresse principalement à la modélisation de communautés microbiennes. Elle se focalise notamment sur l'utilisation de données expérimentales pour construire ou calibrer des modèles. Il faut pour cela développer des méthodes au carrefour de la bioinformatique (métagénomique, réseau métabolique), des statistiques (réduction de dimension, apprentissage) et des EDP. Je développe des modèles à l'échelle des observations, pouvant inclure des composantes stochastiques, que j'étudie ensuite à l'échelle des populations après approximation déterministe. Soucieux de réduire le coût numérique des modèles que je manipule, j'applique dès que possible des techniques de réduction de modèle (analyse asymptotique) ou d'approximation (par machine learning). Ces modèles réduits permettent de faciliter leur exploration numérique, et de déployer des méthodes systématiques d'analyse de sensibilité. L'ensemble de ces méthodes permet d'étudier les communautés microbiennes à tous les niveaux de complexité : SynComs, communautés complexes, holobionte et populations d'holobiontes.

Résumé Contributions à la modélisation mathématiques en écologie microbienne, des données aux modèles

Les micro-organismes forment des communautés complexes en interaction avec leur environnement direct. L'étude de ces communautés fait appel à des concepts d'écologie adaptés aux spécificités microbiennes, auxquels la modélisation mathématique apporte des méthodes analytiques et quantitatives permettant d'intégrer des données et décrire des mécanismes écologiques. Dans ce manuscrit, je fais un tour d'horizon des principaux travaux de modélisation en écologie microbienne que j'ai pu effectuer ces dix dernières années.

Tout d'abord, une méthode permettant d'introduire des a priori biologiques à des techniques de réduction de dimension est présentée, afin de produire des analyses interprétables de données omiques de grande dimension. Ensuite, des modèles de communautés microbiennes simplifiées sont introduits : ces modèles permettent de prendre en compte de manière fine le métabolisme des différents micro-organismes de la communauté simplifiée en utilisant des modèles métaboliques. Toutefois, ces modèles ne sont pas spatialisables en raison de leur temps de calculs prohibitifs : une technique de méta-modélisation permet de produire des approximations rapides de ces modèles pour les coupler à des EDP. Après quoi, je m'intéresse à la modélisation des interactions entre le microbiote et son environnement au travers de trois modèles : 1) un modèle bayésien des capacités de nage de populations bactériennes dans des biofilms estimé à partir d'images de microscopie confocale ; 2) un modèle stochastique des régulations des cellules de l'épithélium de l'intestin, en interaction avec le microbiote intestinal, dont une limite déterministe est obtenue sous hypothèse de grande population ; 3) un modèle du microbiote intestinal dans son environnement prenant en compte la mécanique des différents fluides contenus dans l'intestin. Enfin, je m'intéresse au couplage de plusieurs microbiotes, avec un modèle épidémiologique d'infection à une bactérie entérique, mettant à jour différentes typologie d'hôtes lors de l'épidémie.

Mots-clés : Modélisation, EDP/EDO, écologie microbienne, métamodélisation, simplification de modèle, réduction de dimension.

Abstract (in english) Contributions to mathematical modelling in microbial ecology, from data to models

Microorganisms form complex communities interacting with their direct environment. The study of these communities involves ecology concepts adapted to microbial specificities. Mathematical modelling provides these approaches with analytical and quantitative methods for integrating data and describing ecological mechanisms. In this manuscript, I sketch the main microbial ecology modelling studies that I have carried out over the last ten years.

First, a method for introducing biological prior to dimension reduction techniques is presented, in order to produce interpretable analyses of high-dimensional omics data. Secondly, models of simplified microbial communities are introduced: these models allow the metabolism of the various micro-organisms in the simplified community to be taken into account in a detailed manner using metabolic models. However, these models cannot be spatialised due to their prohibitive computation times: surrogate modelling techniques are used to produce fast approximations of these models, so that they can be used in PDEs. Next, the focus is on how to model the interactions between the microbiota and its environment, by introducing three different models: 1) a Bayesian model of the swimming capacity of bacterial populations in biofilms inferred with confocal microscopy data; 2) a stochastic model to study the regulation of intestinal epithelial cells in interaction with the intestinal microbiota, with a deterministic limit obtained under a large population assumption; 3) a model of the intestinal microbiota in its environment, taking into account the mechanics of the various fluids contained in the intestinal lumen. Finally, I am interested in the coupling of several microbiota in an epidemiological model of infection by an enteric bacterium which reveals the population-wide distribution of different host typologies during an epidemic.

Keywords: Modelling, PDE/ODE, microbial ecology, surrogate modelling, model simplification, dimension reduction.