

# HABILITATION À DIRIGER DES RECHERCHES

Présentée à

l'Université de Nice-Sophia-Antipolis

Spécialité :

Sciences et technologies de l'information et de la communication

## **La modélisation des systèmes biologiques : Aller-retours le long des fleuves qui circulent entre l'océan du réel et le lac des modèles**

par

Olivier BERNARD



Soutenue le 22 janvier 2004 devant le jury composé de :

Prof. Claude Lobry

Prof. Pierre Bernhard

Dr. Philippe Gros

Prof. Jan Van Impe

Prof. Georges Bastin

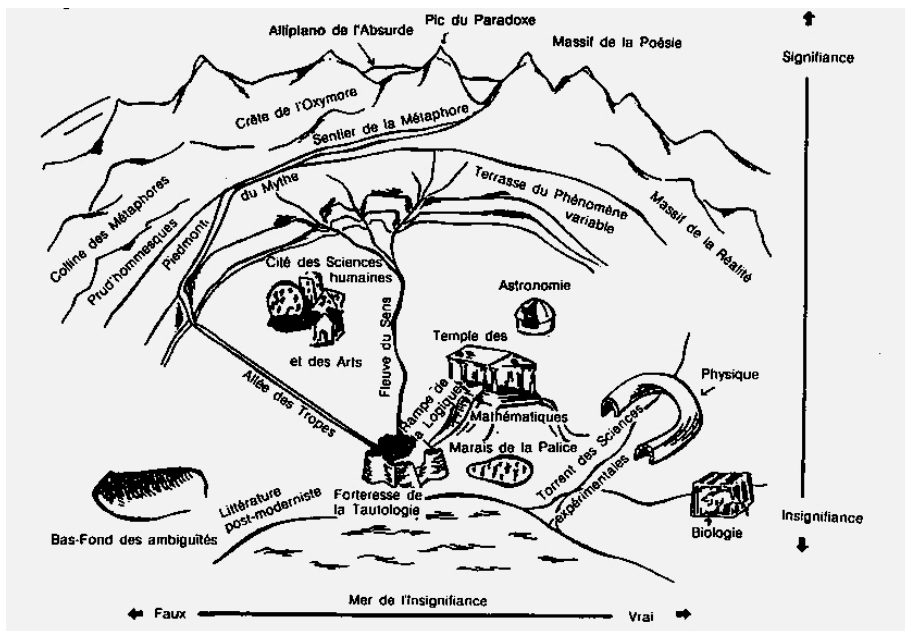
Prof. Michel Loreau

Prof. Jean-Marie Legay





La Carte du Tendre, gravure, 17ème siècle - Paris, B.N.F.



La Carte du Tendre, revue par René Thom "carte du sens". [1]



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>9</b>
1.1	La modélisation des systèmes biologiques : une discipline encore verte . . . . .	9
1.2	Sur la notion de modèle . . . . .	11
1.3	Essai de classification des modèles . . . . .	13
1.4	Vers la “biophysique” . . . . .	16
1.5	Spécificité de mon approche . . . . .	18
1.6	Présentation du mémoire . . . . .	21
<b>2</b>	<b>Les bioréacteurs : des microcosmes simplifiés</b>	<b>23</b>
2.1	Les micro-organismes en bioréacteurs . . . . .	23
2.2	Fonctionnement d’un bioréacteur . . . . .	25
2.3	Les bioréacteurs construits par l’homme . . . . .	26
2.4	Les écosystèmes naturels . . . . .	27
<b>3</b>	<b>La modélisation par bilan de matière</b>	<b>29</b>
3.1	Introduction . . . . .	29
3.2	Le réseau réactionnel . . . . .	30
3.3	Choix du réseau réactionnel . . . . .	30
3.3.1	Exemple : la digestion anaérobie . . . . .	31
3.4	La modélisation des flux de matière . . . . .	32
3.4.1	Introduction . . . . .	32
3.4.2	Raisonnement infinitésimal : étude de l’exemple générique . . . . .	32
3.4.3	Représentation matricielle . . . . .	33
3.4.4	Les flux gazeux . . . . .	34
3.4.5	Electroneutralité et constantes d’affinité . . . . .	34
3.4.6	Exemple : la digestion anaérobie . . . . .	34
3.5	Identification directe du réseau réactionnel à partir d’un jeu de données . . . . .	36
3.5.1	Introduction . . . . .	36
3.5.2	Estimation du nombre de réactions à prendre en compte . . . . .	37
3.5.3	Détermination théorique de $\dim(\mathcal{I}m(K))$ . . . . .	37
3.5.4	Exemple : bioproduction de vanilline . . . . .	39
3.6	Estimation de la matrice pseudo-stoechiométrique à partir d’un jeu de données . . . . .	40

3.6.1	Hypothèses supplémentaires . . . . .	40
3.6.2	Validation . . . . .	41
3.6.3	Exemple : production de lipase à partir d'huile d'olive par <i>Candida rugosa</i> . . . . .	41
3.7	Remarques importantes sur le bilan de matière . . . . .	43
3.7.1	Le bilan ne doit pas nécessairement être fermé . . . . .	43
3.7.2	C'est bien plus que la masse qui est conservée... . . . . .	44
3.8	Conclusion sur le modèle de bilan de masse . . . . .	45
<b>4</b>	<b>Modélisation des cinétiques réactionnelles</b>	<b>47</b>
4.1	Introduction . . . . .	47
4.2	Les contraintes mathématiques . . . . .	47
4.2.1	Positivité des variables . . . . .	47
4.2.2	Variables nécessaires à la réaction . . . . .	48
4.2.3	Exemple : la digestion anaérobie . . . . .	48
4.3	La connaissance phénoménologique . . . . .	49
4.3.1	Recueillir l'information . . . . .	49
4.3.2	Fonctions classiques pour représenter les cinétiques biologiques . . . . .	49
4.3.3	Structure des cinétiques biologiques . . . . .	50
4.4	Représentation des cinétiques au moyen de réseaux de neurones . . . . .	50
4.5	Estimation qualitative directe des cinétiques réactionnelles . . . . .	51
4.5.1	Introduction . . . . .	51
4.5.2	Calcul des taux de réactions . . . . .	52
4.5.3	Définition d'un critère structurel . . . . .	52
4.5.4	Utilisation du critère structurel . . . . .	55
4.6	Conclusion sur la modélisation des cinétiques biologiques . . . . .	57
<b>5</b>	<b>Validation du modèle</b>	<b>59</b>
5.1	Introduction . . . . .	59
5.2	Validation d'un réseau réactionnel . . . . .	60
5.2.1	Position du problème . . . . .	60
5.2.2	Détermination du noyau à gauche de la matrice pseudo-stoechiométrique . . . . .	60
5.2.3	Exemple : bioproduction de vanilline . . . . .	62
5.2.4	Utilisation des données disponibles pour valider le noyau à gauche de $K$ . . . . .	63
5.2.5	Identifiabilité des coefficients pseudo-stoechiométriques . . . . .	63
5.2.6	Identification des coefficients pseudo-stoechiométriques et validation finale . . . . .	65
5.2.7	Comparaison entre plusieurs réseaux réactionnels . . . . .	65
5.2.8	Étude de cas réel . . . . .	65
5.3	Validation qualitative des propriétés asymptotiques du modèle . . . . .	67
5.3.1	Introduction . . . . .	67
5.3.2	La compétition en chémostat . . . . .	68
5.3.3	Évolution des propriétés statiques en fonction des entrées . . . . .	68

5.3.4	Exemple de la modélisation de la croissance phytoplanctonique limitée à la fois par l'azote et la lumière . . . . .	71
5.4	Validation des propriétés dynamiques du modèle . . . . .	74
5.4.1	Introduction . . . . .	74
5.4.2	L'ensemble des événements qualitatifs possibles . . . . .	76
5.4.3	Restriction de l'espace d'état . . . . .	79
5.4.4	Transition entre les domaines $\tilde{\Omega}_{\sigma^q \sigma^p}(x^*)$ . . . . .	79
5.4.5	Barrières dans l'espace d'état . . . . .	80
5.4.6	Théorème de comportement qualitatif . . . . .	80
5.4.7	Représentation graphique . . . . .	80
5.4.8	Comportement asymptotique . . . . .	81
5.4.9	Application à une classe générale de modèles de croissance phytoplanctonique . . . . .	81
5.4.10	Application à la validation qualitative dynamique du modèle . . . . .	85
5.4.11	Autres exemples d'application de l'analyse dynamique qualitative . . . . .	86
5.4.12	Conclusion de l'analyse dynamique qualitative . . . . .	87
5.5	Validation globale du modèle . . . . .	88
5.5.1	Validation semi-quantitative . . . . .	88
5.5.2	Validation quantitative . . . . .	89
5.6	Conclusion sur la validation des modèles . . . . .	89
<b>6</b>	<b>Conclusion</b>	<b>93</b>





# Chapitre 1

## Introduction

*“Ainsi mon dessein n’est pas d’enseigner ici la méthode que chacun doit suivre pour bien conduire sa raison ; mais seulement de faire voir en quelle sorte j’ai tâché de construire la mienne. Ceux qui se mêlent de donner des préceptes se doivent estimer plus habiles que ceux auxquels ils les donnent, et s’ils manquent en la moindre chose, ils en sont blâmables. Mais ne proposant cet écrit que comme une histoire, ou si vous l’aimez mieux que comme comme une fable, en laquelle, parmi quelques exemples qu’on peut imiter on en trouvera peut-être aussi plusieurs autres qu’on aura raison de ne pas suivre, j’espère qu’il sera utile à quelques-uns, sans être nuisible à personne, et que tous me sauront gré de ma franchise.”*

René Descartes, Discours sur la méthode. 1637.

### 1.1 La modélisation des systèmes biologiques : une discipline encore verte

L’apparition depuis les années 1980 de puissants moyens de calcul, leur démocratisation et leur banalisation ont entraîné l’explosion des modèles mathématiques, en particulier dans les sciences du vivant. En quelques jours, il est devenu ”facile” de concevoir le modèle d’un système biologique complexe et de générer une pléiade de courbes colorées et de graphiques séduisants. Bouleau dresse une liste à la Prévert des modèles que l’on peut rencontrer en biologie et il en conclut acerbe :

*”L’impression générale que donnent ces travaux est un ensemble assez hétéroclite. Parfois des schématisations très sommaires sont appliquées à des situations complexes aux implications délicates, parfois au contraire sont employées des méthodes très sophistiquées sur des données très vagues. Des ébauches maladroites avoisinent des effets de style qui évoquent les Précieuses Ridicules. L’emploi de mathématiques trop complexes est un travers fréquent qui nuit gravement à la réputation des mathématiques. Le but peut être*

*d'empêcher la discussion par des considérations savantes ou de tirer avantage d'une tournure à la mode (fractals) mais le plus souvent cela vient de la tendance spontanée du modélisateur à améliorer en perfectionnant et en compliquant."*

Il faut dire que les sciences du vivant qui jusqu'alors avaient été peu mathématisées ont été atteintes de plein fouet par cette vague de "supercomputisme". Ces nouveaux modèles qui prétendent expliquer, quantifier, prédire, ont révolutionné les moeurs et la manière d'aborder les problèmes dans la communauté biologique. De quelle méthodologie relèvent-ils, dans quel cadre peut-on les utiliser, quelles en sont leurs limites, toutes ces questions sur une matière aussi verte (!) sont loin de trouver des réponses simples et consensuelles, nous tacherons d'y réfléchir dans cette introduction. S'il n'est clairement pas question d'apporter une réponse complète à ces interrogations, l'objectif ici sera plutôt de démontrer que des règles du jeu manquent et qu'aucune déontologie de la modélisation en biologie ne fixe vraiment un cadre. Je me baserai sur ce constat pour situer mon approche dans sa globalité par rapport aux grandes tendances émergentes.

Pour commencer je vais cerner la spécificité des systèmes vivants par rapport aux systèmes classiques de la Physique. Notons tout d'abord que les systèmes biologiques sont généralement non linéaires (par exemple ils possèdent souvent plusieurs points d'équilibre), s'opposant ainsi à la majorité des systèmes physiques auxquels nous sommes habitués. Ce sont des systèmes qui sont difficiles à mesurer par manque de capteurs. Les mesures sont souvent constituées d'une chaîne d'étapes manuelles compliquées et engendrent ainsi des données fortement bruitées et à faible fréquence d'échantillonnage. Enfin ces systèmes se caractérisent par leur complexité. Même si le concept de complexité est un peu mou, on peut dire que les systèmes vivants sont fondamentalement hétérogènes et ce pour une très large palette d'échelles. En outre ils incluent un nombre énorme de composés qui interagissent entre eux de l'ångström au kilomètre : molécules, organites, cellules, organes, individus, populations, etc. Cette grande complexité structurelle se double d'une complexité comportementale, guère étonnante en couplant un aussi grand nombre de systèmes non linéaires. Ces systèmes sont en effet bien connus pour leur faculté à générer, dans certaines conditions, des comportements étranges : cycles, chaos, attracteurs étranges, etc. [3].

Cette complexité intrinsèque démarque les systèmes biologiques de la plupart des systèmes physiques, en tout cas des systèmes qui ont été étudiés depuis des siècles et d'où on a pu établir des lois (loi d'Ohm, relation fondamentale de la dynamique, principes de la thermodynamique,...). Ces lois qui sont en général d'excellentes approximations ont largement été validées, admises de tous et sont amplement enseignées et appliquées. Les modèles physiques sont d'une telle précision qu'ils ont renforcé l'idée galiléenne, selon laquelle le monde serait écrit en termes mathématiques.

Toujours est-il que la nature plus compliquée des objets vivants est certainement une des principales raisons expliquant pourquoi ils ne bénéficient pas de modèles aussi précis que ceux que l'on rencontre en physique. Cet "excès" de complexité différencie probablement le monde des physiciens de celui des biologistes. Si dans le premier cas, l'idée qu'il faut identifier ces lois universelles perdure encore, cette vision est beaucoup moins claire dans la communauté biologiste qui argue souvent que le vivant n'est pas réductible à des équations. Mais à bien analyser l'approche de certains biomodélisateurs, ce précepte, lourd de conséquences, semble malgré tout solidement ancré dans les sciences du vivant où la démarche classique consiste finalement à se placer à un niveau de description et de complexité plus proche de ce "langage caché" de la nature.

## 1.2 Sur la notion de modèle

Revenons maintenant sur la notion de modèle, que je vais préciser. Comme le fait remarquer Legay [4], le terme en lui-même est extrêmement ambiguë et recouvre des acceptions assez contradictoires. Il contient notamment à la fois l'idée de maquette, représentation simplifiée de la réalité et la notion d'objet à imiter. Le premier sens va donc vers la simplification de la réalité, alors que le second tendrait à s'en rapprocher en choisissant donc une complexité similaire. Cette ambiguïté sémantique s'amplifie entre communautés issues de disciplines différentes. Ainsi, pour un biologiste, "un modèle" pourra désigner l'organisme étudié, en ce sens qu'il représente à lui seul le comportement d'une population composée de diverses espèces. De la même manière, les microbiologistes étudient comment des substrats "modèles" sont dégradés par des organismes.

Afin de clarifier la notion de modèle, considérons la figure 1.1, inspirée de [5] qui met en correspondance trois entités principales. D'une part le "morceau de réalité" que l'on souhaite représenter (partie du monde réel ou expérience de laboratoire), l'idée que le modélisateur se fait de ce réel et enfin le modèle mathématique associé. Au sens psychanalytique [6], la première entité est le "réel", dans la mesure où il n'a pas été investi par le langage et — *a fortiori* — par l'interprétation du modélisateur, seconde entité. Ce réel, lorsqu'il s'agit par exemple d'une expérience de laboratoire, correspond lui-même à un modèle de la nature, car c'est une sorte de maquette qui n'en reprend que quelques traits.

Pour mieux illustrer cette idée, prenons l'exemple du chémostat, qui sera largement détaillé par la suite. Un chémostat est un dispositif expérimental permettant de faire croître des micro-organismes en milieu ouvert. Examinons la démarche qui consiste à modéliser la croissance du phytoplancton dans la mer. Le premier pas consiste à simplifier le cadre de l'étude et à observer le comportement d'un unique organisme dans un chémostat où un unique facteur limitant (le nitrate) varie. Le chémostat est encore au niveau du réel, mais c'est déjà un modèle de la réalité. Une unique espèce phytoplanctonique s'y développe et n'y est perturbée par aucun prédateur. Elle consomme un unique substrat. Ensuite vient l'idée que le modélisateur a de ce système, caractérisée par un corpus (plus ou moins conscient) d'assertions : "le taux de croissance est proportionnel au taux d'absorption et dépend de la concentration en sels nutritifs extracellulaires". Finalement, sur la base de cette image du réel, un modèle mathématique peut être postulé. Ce modèle est alors associé à un formalisme (équations différentielles ordinaires, équations récurrentes, équations aux dérivées partielles, etc.). En général, le discours sur la réalité ne suffit pas à déterminer un unique modèle et le modélisateur, consciemment ou non, doit rajouter des hypothèses. Des principes universellement employés en physique comme la loi d'action de masse sont souvent utilisés, même si leur application à des particules vivantes en milieu hétérogène reste à démontrer. Il s'agit sans doute là de réminiscences culturelles de ce que Lévy-Leblond [7] appelle la "plurivalence physique des objets mathématiques". Nous sommes habitués à ce que le même modèle décrive avec une précision remarquable des domaines très différents de la physique. Cette idée continue à diffuser dans le monde de la biologie où la spécificité de l'organisme est négligée.

Si je reprends l'exemple de croissance phytoplanctonique, pour écrire le modèle il faut par exemple postuler une expression mathématique reliant taux de croissance et concentration en sels nutritifs. Il y a bien entendu une infinité de choix admissibles, même si l'on restreint le taux de croissance à une fonction croissante et bornée du substrat.

Cette analyse démontre toute l'ambiguïté du terme "modèle", puisque pour arriver à un modèle mathématique, il a fallu passer par 3 types de modélisation, chacune associée à une discipline différente : modélisation



expérimentale, modélisation par le discours et modélisation à l'aide d'un objet mathématique.

## 1.3 Essai de classification des modèles

L'ensemble hétéroclite des modèles en biologie constitue ce que Israel [8] qualifie de véritable "zoo". Devant leur diversité, classer les "animaux" de ce zoo est une tâche délicate et sans doute impossible. Je vais schématiquement séparer les modèles en deux grandes classes, sources de quiproquos entre biomodélisateurs. Les modèles de nature plutôt mécanistes, qui s'appuient sur une connaissance *a priori* des mécanismes pour générer le modèle et les modèles plutôt "boîte noire" qui sont davantage basés sur les données collectées.

L'exemple de la croissance en chémostat se rattache clairement à une approche de type "mécaniste". Dans ce cas, le modèle est construit à partir du discours sur la réalité. Les mesures qui ont pu être recueillies sur le système réel ne servent à aucun moment dans la construction du modèle, seuls les préceptes contenus dans le discours sur la réalité sont utilisés, en faisant donc l'hypothèse forte qu'ils sont suffisants et légitimes. Les modèles "boîte-noire" ne font pas appel à un discours sur le réel, mais se basent sur un ensemble de données disponibles. En fait, dans ce type d'approche, un modèle générique est implicitement utilisé (modèle linéaire de grande dimension, modèle non linéaire de type réseau de neurones, etc.) et la modélisation se résume à l'estimation des paramètres de ce gros modèle. Bien évidemment, comme nous le verrons par la suite, ces modèles "à partir de données" n'ont qu'exceptionnellement les mêmes objectifs que les modèles "à partir du discours".

Les modèles que l'on rencontre en biologie se situent généralement dans une de ces deux classes extrêmes, nous verrons que la démarche que je propose est composite entre ces deux extrêmes. En effet, nous nous aiderons des données disponibles pour mieux formaliser le discours sur la réalité et en limiter les hypothèses fragiles.

Je suggère maintenant de catégoriser les modèles en fonction de leurs objectifs. En effet, des modèles peuvent être construits avec des objectifs, clairement énoncés ou sous-entendus, très différents (je ne discuterai pas ici de la pertinence de ces objectifs).

### Les modèles prédictifs

Ces modèles s'apparentent aux modèles météorologiques dans le sens où leur objectif est de prédire, qualitativement ou quantitativement l'évolution du système entier ou d'un sous-système. Dans certains cas ces modèles peuvent être utilisés simplement pour interpoler une valeur entre deux points de mesures. Dans les communautés plutôt biologiques, c'est souvent cet aspect qui est privilégié et la qualité d'un modèle est implicitement évaluée à l'aide de sa capacité prédictive.

### Les modèles explicatifs

Cette classe de modèle est plus conforme aux attentes des mathématiciens, où le modèle sert à expliquer des phénomènes complexes par des mécanismes plus fondamentaux, plus élémentaires [9]. De tels modèles peuvent être utilisés pour expliquer un comportement observé, par exemple le modèle de Lotka Volterra [10, 11] permet d'expliquer les oscillations décalées d'une population de proies et d'une population de prédateurs. A bien y réfléchir le modèle en lui-même n'est pas explicatif. C'est la théorie issue du modèle qui propose une

explication. C'est un élément clé ; cette théorie doit permettre de faire le lien entre le modèle mathématique et le phénomène que l'on cherche à expliquer. Cela implique en particulier que le modèle doit rester suffisamment simple pour en garder le contrôle et pouvoir mettre en oeuvre un argumentaire mathématique qui donnera le jour à la théorie du modèle.

Ces modèles, en tant que moteurs de réflexion, restent le plus souvent à un niveau qualitatif. En ce sens ils constituent ce qu'Israel [8] qualifie très justement de "sonde conceptuelle que l'on plonge dans la réalité".

A l'extrême, Thom [1] défend une autre classe de modèles explicatifs dont le lien avec le système à modéliser n'est pas direct, mais qui propose des clés de compréhension pour interpréter des observations. Par exemple la théorie des catastrophes prétend davantage attirer l'attention sur un mécanisme possible qu'expliquer *stricto sensu* des observations.

## Les modèles pour stocker la connaissance

Même si ce but est rarement avoué, certains modèles servent surtout de charpente à un dispositif de collecte et de stockage de la connaissance. Le modèle a alors un statut similaire au schéma complexe qu'élabore un systématicien pour représenter comment interviennent et interagissent les différents organes d'un être vivant. Dans ce cadre, améliorer un modèle consiste à décrire plus finement les différents mécanismes et les liens entre les composants. Bien sûr, cette approche va au delà de la simple représentation de la connaissance dans un formalisme structuré. On peut par exemple espérer qu'elle permette de tester la cohérence de cette connaissance et en particulier déterminer si les hypothèses sont contradictoires. Cette vision des modèles a pu prendre corps grâce à la démocratisation des ordinateurs, mais aussi grâce à l'avènement de techniques informatiques permettant plus facilement d'intégrer cette connaissance (par exemple les langages orientés objets ont permis la réalisation de systèmes multi-agents). La complexité mathématique des êtres engendrés par cette approche élimine généralement tout espoir que le modèle puisse être analysé mathématiquement et donc puisse avoir un pouvoir explicatif. De même, le très grand nombre de paramètres introduits et l'incertitude associée à leur estimation amenuise les chances d'utiliser ces modèles dans un but prédictif.

## Les modèles utilisés dans le cadre de problématiques de l'automatique

Ces modèles sont construits pour être ensuite intégrés dans un algorithme dont l'objectif global peut être de nature assez variée : développer une loi de commande pour réguler le système ou l'amener dans un certain état (contrôleur), estimer les variables du système qui ne sont pas mesurées directement (observateur), estimer un des paramètres du procédé (identification), estimer les flux alors que l'on mesure des stocks (observateur), détecter et diagnostiquer une anomalie dans le fonctionnement du procédé (détection de panne et diagnostic), etc. Ces cas d'utilisation, classiques en automatique, font intervenir le modèle ou dans certains cas une sous-partie du modèle. En général, les analyses nécessaires pour élaborer de tels algorithmes ne sont envisageables que pour une complexité du modèle relativement faible. En particulier, en biologie où les modèles sont non linéaires, il est difficile de mettre en oeuvre la plupart des méthodes standards pour des modèles de dimension supérieure à 4. Par exemple, les observateurs à grands gains requièrent des hypothèses assez fines et des calculs compliqués et sont déjà difficiles à mettre en oeuvre en dimension 3 (voir [12] pour un exemple avec des cultures de phytoplancton). Ce type de modèle reste pour l'instant relativement rare en biologie et il est surtout confiné au domaine de la recherche.

## Conclusion : le modèle idéal multiobjectifs existe-t-il ?

Le fantôme classique (et néanmoins compréhensible) du biomodélisateur tient dans l'idée que le modèle est capable simultanément d'atteindre tous ces objectifs. Mais ne jetons pas la pierre aux biomodélisateurs, car cette idée était déjà présente chez les Grecs anciens qui pensaient que les mathématiques constituaient le langage caché de la nature [13]. Cette conception des mathématiques a ensuite largement été relayée par les physiciens [8], où une meilleure connaissance de systèmes plus simples rendait possible tous ces objectifs simultanément.

Illustrons cette idée par un exemple physique simple. Considérons un chariot de masse ( $m$ ) rappelé horizontalement par un ressort de raideur  $k_1$  et freiné par un frottement visqueux (voir figure 1.2). Les frottements sont donc caractérisés par un terme proportionnel à la vitesse de la masse. En utilisant les lois de la mécanique classique, le modèle s'écrit alors sans ambiguïté :

$$\frac{d^2x}{dt^2} = -k_1(x - x_0) - k_2 \frac{dx}{dt}$$

Ce modèle va certainement permettre un bonne **prédiction** de l'évolution du chariot (au moins à court terme), il va **expliquer** pourquoi on observe des oscillations et va permettre de **comprendre** le lien entre la période des oscillations et la masse du chariot. Une fois le modèle posé, la **connaissance** du système est entièrement décrite (et d'ailleurs — via les principes physiques — le modèle est équivalent au discours sur le système). Finalement il ne fait aucun doute qu'un **observateur ou un contrôleur** basé sur ce modèle serait une réussite.

Cet exemple illustre l'idée que, pour des systèmes physiques simples, ces quatre objectifs sont conciliables. En particulier il n'est pas nécessaire de simplifier le modèle pour arriver à un degré de complexité mathématiquement gérable.

Mais les deux conditions que sont la simplicité du système et l'existence d'un corpus de lois constituant d'excellentes approximations ne sont pas satisfaites en biologie, aussi faudra-t-il se focaliser sur l'un de ces quatre objectifs. Dans ce mémoire, les méthodes que je suggère pour élaborer un modèle permettront soit de prédire l'évolution du système soit de traiter des problématiques issues de l'automatique. Je reviendrai dans la conclusion sur l'aspect explicatif des modèles.

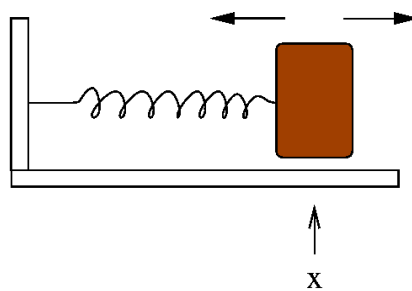


FIG. 1.2 – Le chariot rappelé par un ressort : un système dont un modèle simple peut à la fois prétendre prédire, expliquer et peut également être utilisé pour des problématiques de l'automatique. De plus le modèle contient tout le savoir disponible sur le système.

## 1.4 Vers la “biophysique”

Dans ce paragraphe, en marge des aspects historiques, il sera question de la relation entre la biomodélisation et la physique. Cette réflexion est fondée d’une part sur une note de Lévy-Leblond [7] qui explore les liens entre physique et mathématique et d’autre part sur un article de Lobry [14] qui examine la pertinence des concepts dégagés par Lévy-Leblond dans le cas de la biologie.

Lévy-Leblond défend d’abord le rôle constitutif joué par les mathématiques en physique. En effet, les mathématiques n’ont pas seulement un rôle instrumental de quantification, mais permettent de dégager des concepts, tels que champs, potentiels, etc. Comme le fait remarquer Lobry, la biologie implique également de nombreux concepts qui ne s’entendent qu’à travers une formalisation mathématisée. Par exemple le taux de croissance d’une population fait intervenir la dérivée logarithmique. De manière plus subtile, si l’on veut comprendre pourquoi l’écosystème contenu dans un fermenteur anaérobie peut brutalement dépérir il faudra faire appel à la notion de bassin d’attraction. D’autres exemples de ce rôle constitutif des mathématiques viendront illustrer ce mémoire. Les mathématiques ont donc en physique et en biologie un rôle du même registre.

Le deuxième aspect qui pourrait caractériser la physique est [7] “l’idée que la physique est plus “fondamentale” que les autres sciences de la nature. S’attaquant aux structures plus profondes de la nature, elle mettrait en lumière ses lois les plus générales, implicitement pensées comme “plus simples”, au sens ontologique et donc plus mathématisable.”

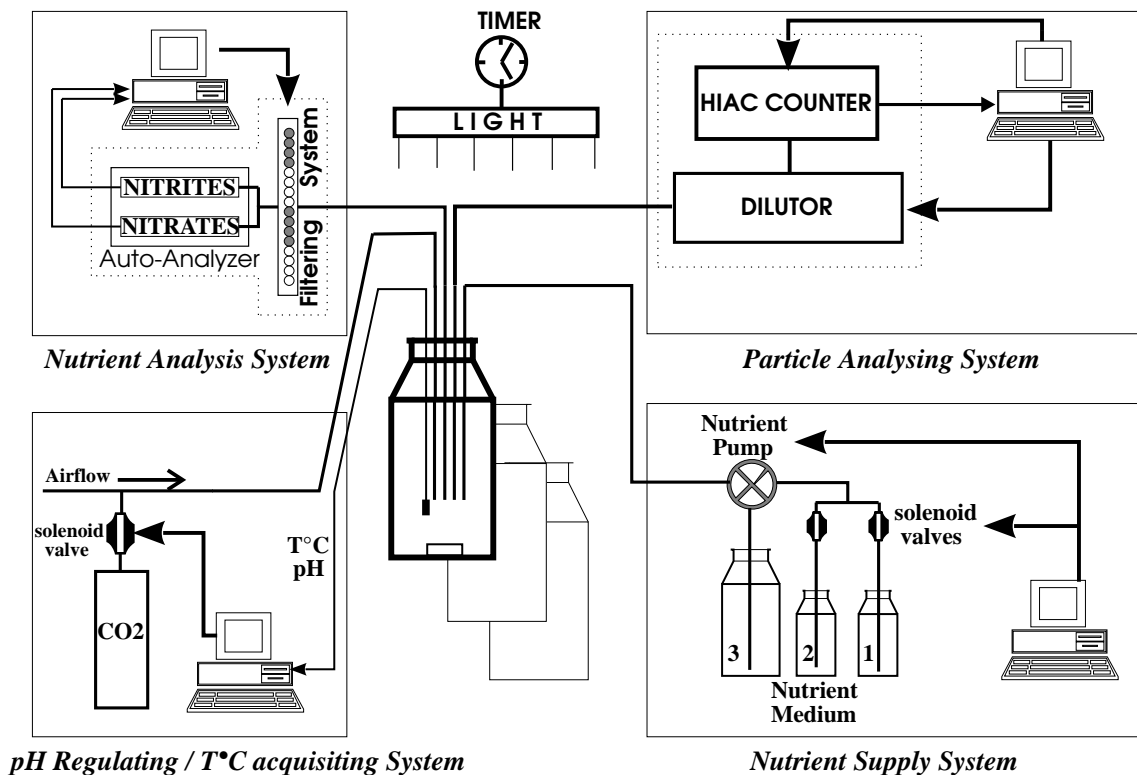


FIG. 1.3 – Schéma synoptique de l’automate de culture de phytoplancton développé au Laboratoire d’Océanographie de Villefranche sur mer (CNRS) sous la direction de A.Sciandra [15], [16].



J’ai insisté sur la complexité des systèmes biologiques, notamment en soulignant l’hétérogénéité et le grand nombre d’éléments impliqués quelle que soit l’échelle considérée. Elle induit *de facto* une complexité expérimentale, comme le fait remarquer Legay [4] : “Plus les hypothèses sont nombreuses, plus le sujet est complexe, plus le contrôle expérimental est nécessaire, mais plus celui-ci devient difficile à organiser et à interpréter”. Mon travail s’est focalisé sur la simplification des biosystèmes en isolant des sous parties (donc en limitant le nombre de composantes) et en les maintenant dans des milieux homogènes. Cette simplification extrême par rapport aux écosystèmes naturels est un pas énorme (c’est déjà un modèle comme je l’ai expliqué), mais cela nous permet de nous rapprocher de l’objet étudié en physique, intrinsèquement plus simple. Je détaillerai dans le chapitre suivant comment une telle approche est réalisable à l’aide de microcosmes que j’appellerai “bioréacteurs”. Cette simplification extrême de l’objet, pour s’apparenter aux dispositifs expérimentaux qu’étudient les physiciens, doit réunir deux conditions.

Tout d’abord, il est nécessaire de pouvoir agir sur le système étudié et de le perturber afin de le placer dans différentes conditions expérimentales. Cette notion d’entrées en automatique donne sens à la notion d’expérience, sans cela il ne s’agit que d’observations [17]. Ensuite, il faut être à même de mesurer précisément l’évolution du système biologique. C’est en général un point délicat car il n’existe que peu de capteurs capables d’évaluer des quantités telles que, par exemple, biomasse ou activité enzymatique. Mais les récents développements de l’automatisation et de l’informatique et la demande industrielle croissante ont permis l’instrumentation lourde de certains microcosmes.

Réunir les trois conditions de simplicité, excitabilité et mesurabilité pour se rapprocher de ce que l’on pourrait qualifier de “biophysique” a constitué une part importante de mes recherches. Le système biophysique ainsi obtenu, que je nommerai “bioréacteur” sera l’objet central de ce mémoire. C’est dans ce cadre simplifié et instrumenté que je présente la modélisation.

J’illustrerai mon propos à l’aide de deux dispositifs expérimentaux pilotés par un ensemble d’ordinateurs. D’une part je présenterai des résultats expérimentaux issus du système de culture de phytoplancton développé au Laboratoire d’Océanographie de Villefranche sur mer (CNRS) par A.Sciandra et G. Malara [15] et que j’ai contribué à améliorer [16] (voir figures 1.3 et 1.4). Ce dispositif contrôlé par un ensemble d’ordinateurs permet d’imposer les conditions de croissance et de suivre avec une très grande précision, grâce à un ensemble d’automates, l’évolution de la biomasse (densité cellulaire, biovolume total, diamètre cellulaire moyen et concentration en chlorophylle *a* et des sels nutritifs (nitrate et nitrite).

J’utiliserai également les données produites au Laboratoire de Biotechnologie de l’Environnement (INRA) sous la direction de J.-P. Steyer [18] qui a développé un bioréacteur anaérobie pour l’épuration (voir figures 1.5 et 1.6) associé à de nombreux capteurs en ligne (demande chimique en oxygène, acides gras volatils, alcalinité, carbone inorganique dissous, débits des gaz, etc). Ce dispositif enregistre avec précision la réponse de l’écosystème à des variations qualitatives et quantitatives du substrat organique.

De même que la complexité [4], la simplification est le résultat d’un choix. Cette simplification peut en particulier amener à des systèmes dont la complexité reste très élevée, dans le sens qui a déjà été défini. Néanmoins, comme nous le verrons dans ce mémoire, des systèmes complexes comme l’écosystème anaérobie de l’INRA se plient parfaitement à cette analyse, même s’ils mettent en jeu plus d’une centaine d’espèces bactériennes [19].

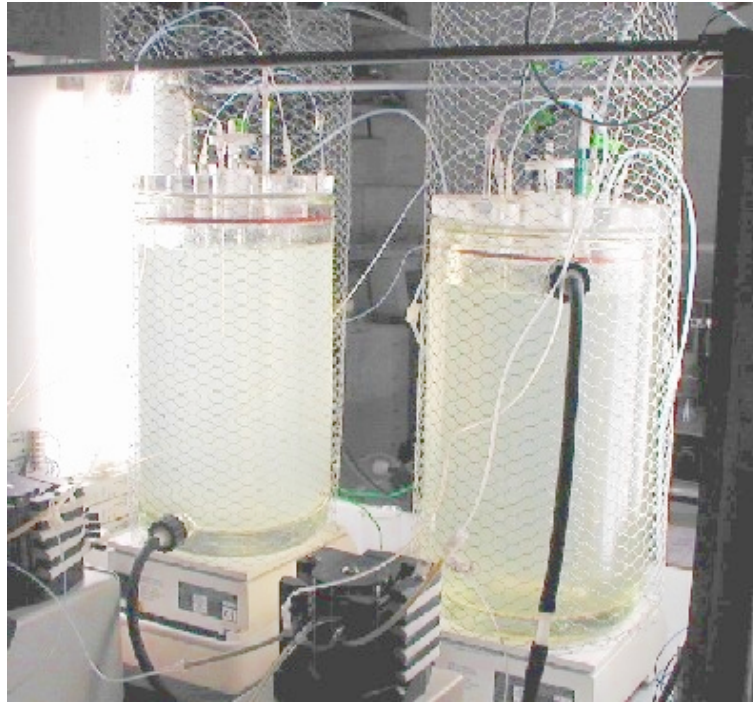


FIG. 1.4 – Photo des chémostats lors d’une expérience avec *Emiliana huxleyi*.

## 1.5 Spécificité de mon approche

La spécificité de mon approche tient d’abord en ce qu’elle s’appuie résolument sur les méthodes de la physique en utilisant plus particulièrement les outils de l’automatique. Les chapitres qui suivent synthétisent ma démarche qui se distingue des approches classiques dans la manière de gérer les deux pôles que sont le modèle et les données. Mon approche ne consiste pas successivement à construire un modèle et à le confronter aux données, ou bien à générer un modèle à partir d’un ensemble de données, mais se nourrit des aller-retours permanents entre ces deux entités. Cela est rendu possible par l’utilisation de dispositifs de laboratoires qui garantissent la reproductibilité des expériences.

De manière plus précise, comme je le détaillerai par la suite, les modèles en biologie reposent sur un fragile échafaudage d’hypothèses. Mon travail a consisté à hiérarchiser ces hypothèses et à en organiser la validation progressive en tâchant de les découpler les unes des autres. Ainsi je montrerai comment tester les hypothèses sur le réseau réactionnel, sur le signe des interactions entre composants, etc. En procédant de la sorte, on voit très vite qu’il existe un niveau à partir duquel certaines parties du modèle semblent fortement spéculatives et en tout cas très difficiles à valider. En fonction des objectifs qui ont été fixés, il faudra n’utiliser que les composants du modèle au delà d’un degré de sûreté fixé. Par exemple, pour construire des observateurs, on peut n’utiliser que la partie du modèle correspondant au bilan de masse [20], les cinétiques réactionnelles — trop peu fiables — étant alors considérées comme des entrées inconnues.

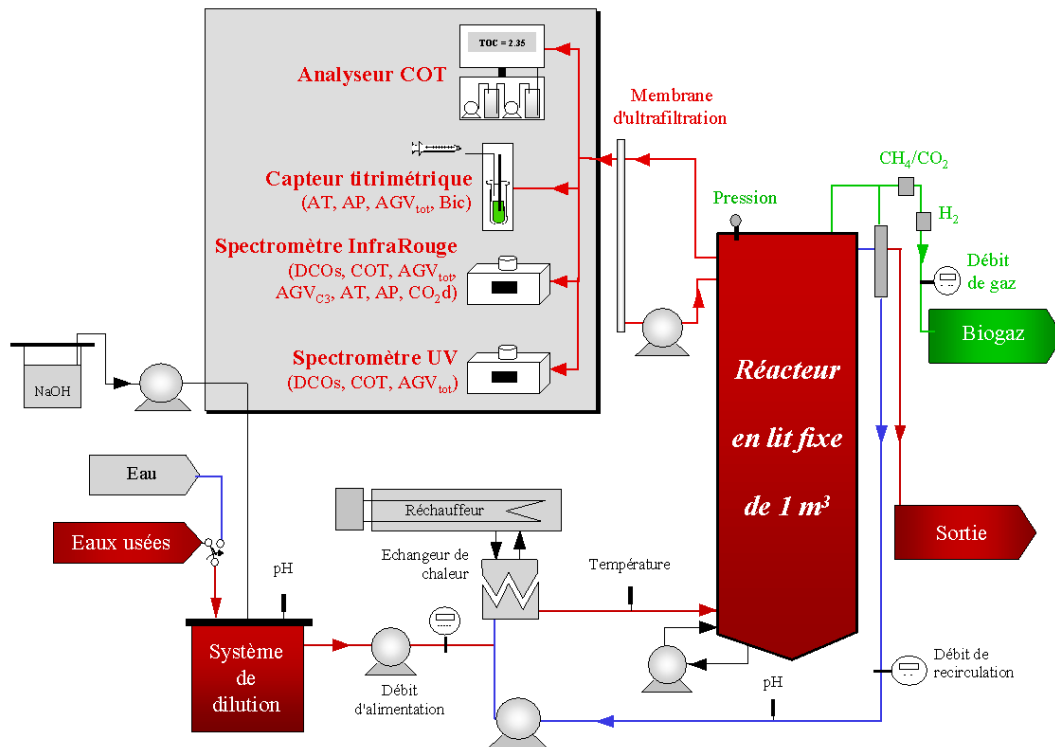


FIG. 1.5 – Schéma synoptique du fermenteur anaérobie fortement instrumenté développé au Laboratoire de Biotechnologie de l'Environnement (INRA) sous la direction de J.-P. Steyer [18].

## Les systèmes biologiques dont il ne sera pas question

Ce mémoire se cantonne à une difficulté assez modeste, puisqu'il se limite aux **systèmes isolés, homogènes et bien mesurés** que sont les bioréacteurs. Dans ce sens ils rassemblent les conditions idéales de la biophysique. Je donnerai malgré tout quelques exemples d'applications *in situ* dans des cadres moins idylliques [21, 22].

Les modèles dont il sera question s'appuient sur la notion de flux de matière entre différents compartiments. En particulier la démarche **ne s'applique pas** à des flux d'information (génomique), d'énergie (même si des liens existent [23, 24]), ou à la propagation d'une épidémie, pour lesquels les lois de conservations sont beaucoup moins claires (si elles existent).

Du fait de ma culture j'ai choisi de considérer un formalisme à base d'équations différentielles, mais la plupart des concepts s'appliquent de manière équivalente avec des équations récurrentes.

## Les travaux de recherches qui ne sont pas dans ce mémoire

Dans ce manuscrit je me suis focalisé sur mes travaux de modélisation, en passant donc sous silence d'autres travaux qui ont pourtant occupé une place importante dans mes recherches. En particulier, mes

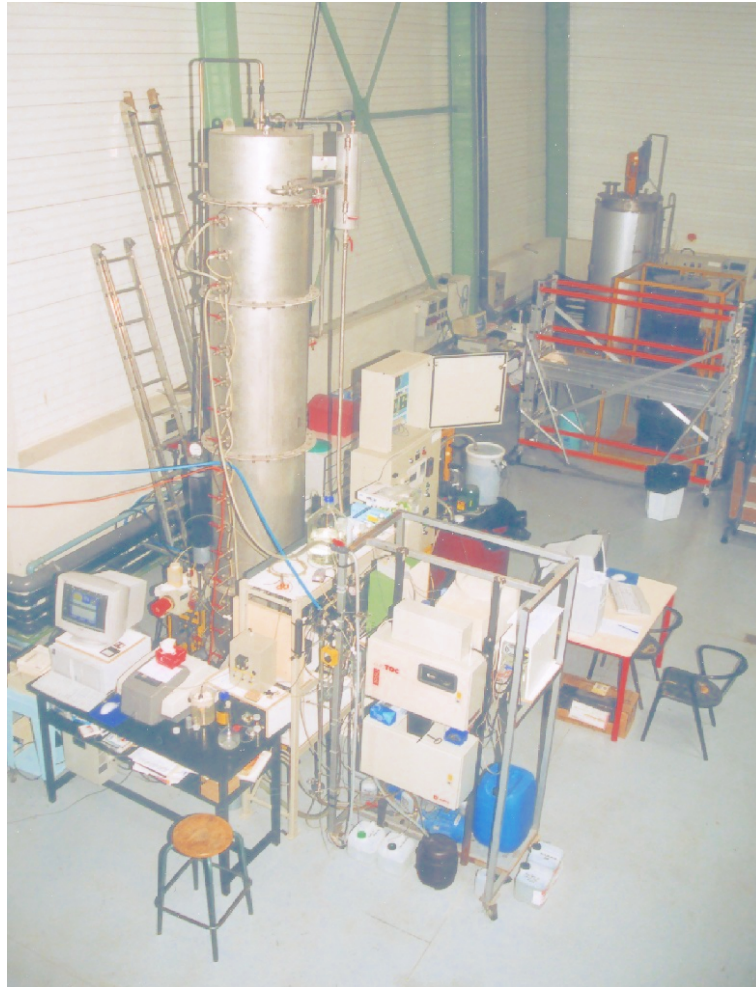


FIG. 1.6 – Fermenteur anaérobie fortement instrumenté développé au Laboratoire de Biotechnologie de l'Environnement (INRA) sous la direction de J.-P. Steyer [18].

travaux sur le développement d'observateurs [25, 26, 12, 27, 28, 29, 30, 31] et de contrôleurs [32, 33, 34, 35, 36, 37] pour les systèmes biologiques. Ces travaux ont joué un rôle déterminant dans mes recherches et m'ont largement motivé pour mettre en place des méthodes propres à la modélisation des systèmes biologiques. En effet, comme tout jeune automaticien pressé de mettre en oeuvre et d'améliorer les magnifiques outils appris à l'école, j'ai commencé par développer des estimateurs d'état (observateurs) pour des étangs de pisciculture, puis plus tard pour des algues phytoplanctoniques se développant en chémostat [26, 12]. Si les résultats de simulation confirmaient la théorie, l'application pratique avec de vraies données expérimentales s'est avérée extrêmement décevante. C'est un fait, les techniques à base de grands gains exacerbent les défauts du modèle et limitent l'applicabilité de nombre d'outils issus de l'automatique non linéaire. Il fallait donc revoir le problème soit en améliorant les modèles, soit en proposant des algorithmes moins sensibles aux incertitudes caractéristiques de la biologie. La première voie est l'objet de ce mémoire, mais j'ai aussi travaillé pour mettre en place des observateurs et des contrôleurs robustes. Ainsi, les observateurs asymptotiques [20] peuvent être robustifiés si l'on prend en compte de nouvelles mesures souvent disponibles mais rarement

exploitées [28]. Il s'avère que ces techniques peuvent se combiner à des observateurs par intervalles [38, 39] et donnent alors lieu à un faisceau d'observateurs par intervalles dont on ne conserve que l'encadrement le plus restreint [40, 31]. J'ai aussi développé une approche alternative qui s'appuie sur la connaissance de la densité de probabilité de la condition initiale et des incertitudes. Il est alors possible, dans certains cas très simples, d'estimer la densité de probabilité des pseudo intégrales premières du système. Cela conduit à la notion d'observateurs probabilistes qui ne fournissent pas uniquement une estimation de l'état, mais qui évaluent également l'incertitude associée à la prédiction [30, 41]. J'ai suivi la même démarche pour le contrôle non linéaire, en mettant en place des contrôleurs largement indépendants des incertitudes sur le système, en particulier en mesurant certaines sorties judicieuses. Cet aspect a en particulier été exploré au cours de la thèse de Ludovic Mailleret [33, 34, 42, 35, 37].

## 1.6 Présentation du mémoire

Ce mémoire s'attache à systématiser l'étude et la modélisation des systèmes biologiques dans un cadre plus rigoureux. La méthodologie qui va être détaillée peut contribuer à élaborer et valider les modèles dans un contexte simplifié, pour un milieu homogène et avec un nombre réduit d'espèces. Nous verrons que même dans ce cadre simplifié les systèmes non linéaires impliqués restent d'une très forte complexité mathématique.

Ce travail, largement pluridisciplinaire a été structuré pour être lisible par des chercheurs de différents domaines. Au prix peut-être d'une trop grande lourdeur, j'ai opté pour un style didactique (d'ailleurs largement inspiré de chapitres de livres auxquels j'ai pu participer [43, 44]). J'ai veillé à ce que ce mémoire soit autosuffisant, même si le lecteur est invité à se référer aux articles pour les détails. Les travaux auxquels j'ai participé sont référencés en caractères gras. La difficulté théorique (même si elle reste très modeste) s'accroît au fil des pages et un lecteur familier des biomathématiques pourra sans doute passer rapidement les premiers chapitres.

Le chapitre de ce mémoire qui suit cette introduction présente dans le détail les microcosmes contrôlés (bioréacteurs) qui serviront de cadre d'application. Le troisième chapitre explicite la première étape de la modélisation par bilan de matière : la définition d'un réseau réactionnel et les équations qui y sont associées. La suite du chapitre est une contribution plus personnelle qui propose des méthodes pour estimer un réseau réactionnel lorsque celui-ci n'est pas connu *a priori*, ou bien lorsqu'il est trop compliqué. Je présente une méthode qui s'appuie sur un jeu de données recueillies pour déterminer le nombre minimal de réactions qui permettraient d'expliquer ces données. En d'autres termes cela revient à déterminer la taille de la matrice pseudo-stoechiométrique. Finalement, je montre que -moyennant des hypothèses biologiques supplémentaires- il est possible d'estimer les coefficients inconnus de cette matrice, déterminant du même coup un réseau réactionnel.

Le quatrième chapitre rappelle quelques règles pour associer une expression mathématique aux cinétiques réactionnelles. Ma contribution apparaît d'une part sur la manière de contraindre un modèle générique (type réseau de neurones) pour que le modèle respecte des contraintes physiques (bilan de masse, positivité des variables, bornitude) et biologiques (monotonie des taux de croissance ou d'absorption). Enfin, je présente une méthode originale pour déterminer par un critère qualitatif les cinétiques biologiques qui sont compatibles avec un jeu de données.

Le cinquième chapitre constitue le coeur de mes recherches. Je propose une méthode pour valider un

modèle à travers une succession d'étapes qualitatives. En m'appuyant sur mes travaux, je montre d'abord comment valider le réseau réactionnel choisi, en validant la conservation de certaines quantités qui s'apparentent à des invariants réactionnels [45, 46, 47]. Je propose ensuite diverses méthodes pour tester les propriétés asymptotiques des modèles (certaines sont classiques). Enfin, et c'est une autre de mes contributions, je détaille une méthode pour tester la validité qualitative des propriétés dynamiques d'un modèle.

Enfin, le dernier chapitre conclut ce mémoire et ouvre des perspectives...

Ce mémoire est agrémenté d'exemples issus des diverses applications sur lesquelles j'ai pu travailler : croissance de cellules phytoplanctoniques limitées par l'azote et la lumière en chémostat, développement de populations de copépodes, interaction de populations de coccinelles et de pucerons en serre de concombres, évolution d'écosystèmes anaérobies en bioréacteur, croissance et bioproduction de champignons filamenteux en bioréacteurs, etc.

## Chapitre 2

# Les bioréacteurs : des microcosmes simplifiés

*“Quant aux infusoires, reprit-il, quant à ces milliards d’animalcules, qui existent par millions dans une gouttelette, et dont il faut huit cent mille pour peser un milligramme, leur rôle n’est pas moins important. Ils absorbent les sels marins, ils assimilent les éléments solides de l’eau, et, véritables faiseurs de continents calcaires, ils fabriquent des coraux et des madrépores ! Et alors la goutte d’eau, privée de son aliment minéral, s’allège, remonte à la surface, y absorbe les sels abandonnés par l’évaporation, s’alourdit, redescend, et rapporte aux animalcules de nouveaux éléments à absorber”*

Jules Verne, *Vingt mille lieues sous les mers*. 1869

### 2.1 Les micro-organismes en bioréacteurs

J’utiliserai le terme de bioréacteur dans une acception assez large, sans une quelconque connotation industrielle. Il s’agira d’un microcosme contrôlé et contenant des organismes vivants, dont les flux de matière à l’entrée et à la sortie sont maîtrisés. Le terme de bioréacteur englobera donc d’un coté des dispositifs de laboratoire de quelques litres ne contenant qu’une espèce en interaction avec un substrat et de l’autre — à l’extrême — des écosystèmes naturels (type lagune ou lac) intégrant tout un réseau trophique.

Néanmoins, je considérerai essentiellement des micro-organismes (bactéries, levures, phytoplancton, etc.) qui consomment des substrats ou se nourrissent d’autres organismes (proies) pour se développer et qui utilisent des précurseurs et des activateurs pour produire des molécules spécifiques.

Dans un bioréacteur on distinguera donc les composants suivants :

- les substrats qui sont nécessaires pour la croissance des micro-organismes, ou bien qui sont des précurseurs d’un composé à produire. Pour soutenir la croissance, le milieu de culture doit contenir tous les éléments

nécessaires (c.a.d N, C, K, P, Fe,...). En général, nombre de ces éléments sont ajoutés en excès de manière qu'ils ne limitent pas la croissance.

- Les organismes qui se développent peuvent être de natures très différentes, (bactéries, phytoplancton, champignons, levures, etc.). La majorité des procédés biotechnologiques développés à l'échelle industrielle utilisent des cultures microbiennes composées d'une seule espèce de micro-organisme (culture pure). Pour les procédés de dépollution on utilise des écosystèmes naturels, qui peuvent donc inclure — c'est le cas des fermenteurs anaérobie — plus d'une centaine d'espèces. Dans le cadre d'une lagune ou d'un lac, on trouvera des organismes plus évolués (zooplancton, poissons, etc.). Notons que certains de ces organismes peuvent être les proies d'autres espèces.
- Les produits des réactions biochimiques. Ces composés synthétisés par les micro-organismes sont utilisés dans divers domaines de l'industrie : agro-alimentaire (huiles, fromages, bières, vins...), industrie chimique (solvants, enzymes, colorants...), pharmaceutique (antibiotiques, hormones, vitamines...) ou pour production d'énergie (éthanol, biogaz...)... Mais cette synthèse se fait naturellement dans les écosystèmes. Par exemple le  $\text{CO}_2$  est un produit très fréquent associé à la respiration des organismes.
- Les catalyseurs, qui ne sont ni produits ni consommés au cours de la réaction, mais qui sont nécessaires pour qu'elle puisse avoir lieu.

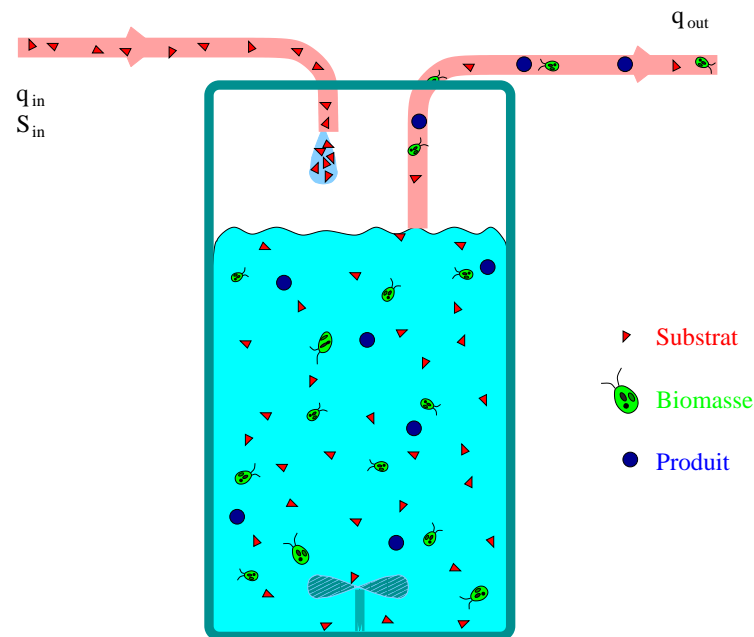


FIG. 2.1 – Principe d'un bioréacteur

Les expériences en bioréacteur relèvent en principe de l'un des 4 objectifs suivants :

- La production de biomasse est un but très courant. C'est le cas, par exemple, des bioréacteurs visant à produire la levure de boulanger, ou bien d'une ferme aquacole.
- On utilise très souvent des bioréacteurs pour faire synthétiser un métabolite (éthanol, pénicilline, ... ) par un micro-organisme.
- La dégradation d'un substrat est l'objectif notamment des procédés de dépollution (traitement biolo-



gique des eaux usées, dégradation de polluants spécifiques, ...).

- Sur des aspects plus fondamentaux (recherche), des bioréacteurs servent également pour étudier des organismes, comprendre la relation entre différentes espèces ou bien observer un écosystème simplifié. Par exemple pour mieux comprendre comment un organisme se développe dans le milieu naturel, ou bien pour observer l'interaction entre un prédateur et sa proie.

## 2.2 Fonctionnement d'un bioréacteur

### Présentation

La figure 2.1 présente un schéma conceptuel du fonctionnement d'un bioréacteur. Il s'agit d'un récipient de volume  $V$  dans lequel se développent les organismes. Ce récipient est rempli par un milieu de renouvellement avec un débit  $Q_{in}$  et se vide avec un débit  $Q_{out}$ . Ces deux débits peuvent être choisis par l'expérimentateur ou résulteront du calcul d'une loi de contrôle (dans ce cas ils dépendront de l'état du système).

On distingue 3 modes d'alimentation classiques des bioréacteurs (figure 2.2).

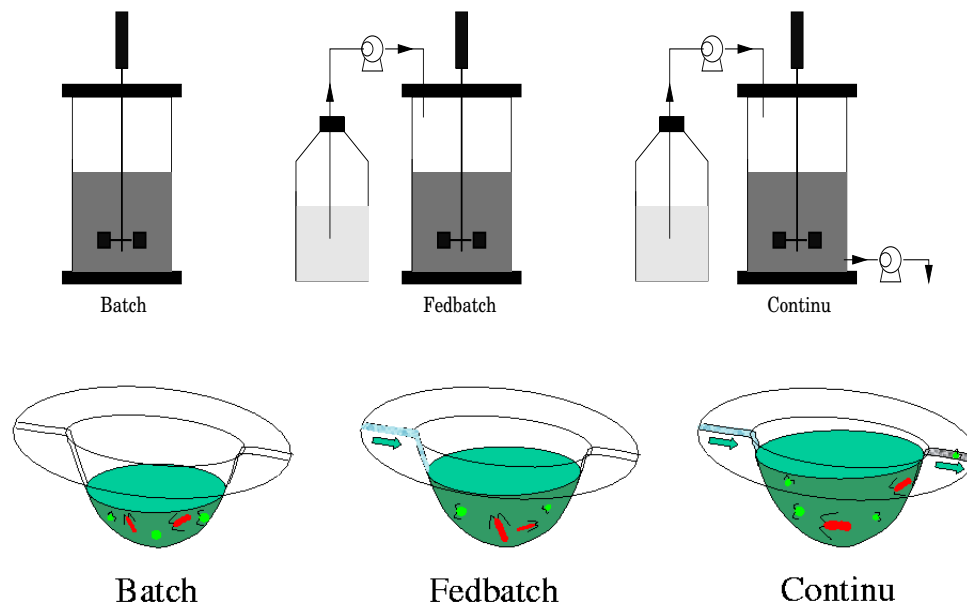


FIG. 2.2 – Les différents modes de fonctionnement d'un bioréacteur construit par l'homme ou d'un bioréacteur naturel (lac).

### Le mode discontinu (ou batch)

Le système est fermé et garde un volume constant. Les espèces sont introduites à l'instant initial (inoculum pour des micro-organismes) ainsi que les nutriments et les précurseurs nécessaires. Lorsque le bioréacteur est utilisé à des fins commerciales, la biomasse et le produit intéressants sont récupérés à l'instant final.

L'avantage de cette approche est qu'elle limite la contamination par d'autres micro-organismes et permet donc de rester plus longtemps dans des conditions axéniques.

En contrepartie les moyens d'action sur de tels systèmes sont limités (aération, température, etc). Un bioréacteur batch présente donc de fortes contraintes qui laissent peu de latitude pour le contrôle. Par ailleurs il est mal adapté à un échantillonnage à haute fréquence qui pourrait significativement modifier son volume. Du fait de sa simplicité, c'est néanmoins le système le plus utilisé dans l'industrie pour produire un métabolite ou une biomasse.

### **Le mode semi-continu (ou fedbatch)**

De même que pour le bioréacteur fermé, la durée d'une expérience semi-continue est finie, la biomasse ou l'un des produits sont récupérés à l'instant final. Mais ici la culture est alimentée de son volume de départ  $V_0$  jusqu'à son volume final  $V_f$  à l'instant final  $t_f$ . Ce mode de fonctionnement permet donc un meilleur contrôle des conditions de croissance, le bioréacteur étant le plus souvent alimenté par un débit contrôlé en boucle fermée.

### **Le mode continu (ou chémostat)**

C'est le mode le plus largement employé dans le domaine de la dépollution biologique de l'eau. Caractérisé par un volume constant, il est soumis à un soutirage de milieu réactionnel égal au flux d'alimentation en matière nutritive. C'est le mode de fonctionnement qui laisse la plus grande latitude pour appliquer les techniques de contrôle. Sa dynamique plus riche (par exemple dans un bioréacteur en batch la plupart des espèces évoluent de manière monotone) en fait aussi un très bon outil pour étudier des phénomènes transitoires liés à la croissance ou à la biotransformation.

## **2.3 Les bioréacteurs construits par l'homme**

Il existe un très grand nombre de réacteurs biologiques construits par l'homme, qui sont adaptés au type de micro-organisme à cultiver. En effet, certaines espèces auront besoin d'un support pour se fixer alors que d'autres se développeront dans le milieu liquide. Elles pourront par ailleurs plus ou moins résister aux contraintes de cisaillement induites par le système d'homogénéisation. Les exigences des organismes détermineront donc le type de bioréacteur à utiliser.

Il y a deux grandes classes de bioréacteurs [48] :

- les bioréacteurs infiniment mélangés (en anglais "Continuous stirred tank reactors", notés CSTR) dans lesquels le milieu est supposé homogène. Plus précisément, nous supposons que les phases solides, liquides et gazeuses sont distribuées de manière homogène dans le bioréacteur.
- les bioréacteurs à gradient spatial de concentration, tels que lits fixes, lits fluidisés, air lifts...

Je considérerai par la suite uniquement la classe des bioréacteurs infiniment mélangés, que je représenterai par des équations différentielles ordinaires.

## 2.4 Les écosystèmes naturels

Quelles sont les différences entre un bioréacteur construit par l'homme et un écosystème naturel? La définition des bioréacteurs s'applique tout aussi bien à un système naturel. La distinction est plus d'ordre pratique que systémique : dans un bioréacteur toutes les entrées sont maîtrisées, alors que pour un écosystème elles sont le plus souvent contraintes par le milieu environnemental (lumière, température, flux liquides, etc.).

La complexité d'un écosystème naturel est en général bien plus forte que celle d'un bioréacteur industriel car un très grand nombre d'espèces sont impliquées, elles sont de natures très différentes (du virus au poisson, en passant par le phytoplancton, etc.). Par ailleurs, contrairement à un bioréacteur où l'on est plus ou moins capable de caractériser les espèces en présence, la succession des espèces en milieu naturel est souvent beaucoup moins bien connue.

Nous verrons que l'homogénéité est un point clé pour la modélisation des biosystèmes. S'il est facile de maintenir une agitation permanente dans un bioréacteur, c'est loin d'être le cas pour un écosystème. Une approche basée sur des équations différentielles fera donc implicitement l'hypothèse d'homogénéité spatiale au sein de l'écosystème, hypothèse certainement beaucoup moins valide que pour un bioréacteur. Ceci explique la tendance actuelle à construire des modèles d'écosystèmes à base d'équations aux dérivées partielles (EDP) résultant du couplage entre un modèle biologique et un modèle physique décrivant le transport et la diffusion des composés biochimiques et des organismes dans le milieu.

Finalement, la limite entre bioréacteur et écosystème est loin d'être claire. Par exemple, pour le traitement biologique de l'eau, un écosystème complet est mobilisé pour dégrader un ensemble de substrats organiques. Dans le cas de l'épuration par lagunage, le "bioréacteur" n'est ni plus ni moins qu'un étang naturel dont le réseau trophique est dopé par un apport de matière organique [49].



## Chapitre 3

# La modélisation par bilan de matière

*“Ainsi, au lieu de ce grand nombre de préceptes dont la logique est composée, je crus que j’aurais assez des quatre suivants (...). Le premier était de ne recevoir jamais aucune chose vraie que je ne connusse évidemment être telle (...). Le second, de diviser chacune des difficultés que j’examinerais en autant de parcelles qu’il se pourrait et qu’il serait requis pour les mieux résoudre. Le troisième de conduire par ordre mes pensées, en commençant par les objets les plus simples et les plus aisés à connaître, pour monter peu à peu comme par degrés jusques à la connaissance des plus composés (...). Et le dernier de faire de partout des dénombrements si entiers et des revues si générales que je fusse assuré de ne rien omettre”.*

René Descartes, Discours sur la méthode. 1637.

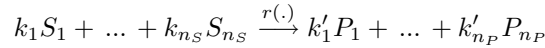
### 3.1 Introduction

La modélisation des systèmes biologiques est une tâche extrêmement délicate car, contrairement à la physique, il n’existe pas de lois admises et reconnues par tous caractérisant l’évolution des micro-organismes. Néanmoins, ces systèmes, au même titre que tous les systèmes physiques, doivent respecter des règles telles que la conservation de la matière, l’électroneutralité des solutions, etc. Nous verrons dans ce chapitre comment charpenter le modèle autour de ces lois physiques, de manière à lui garantir une certaine robustesse.

Ce chapitre, qui s’appuie sur le formalisme introduit par Bastin et Dochain [20], reprend principalement le chapitre du livre auquel j’ai participé [44]. Les exemples qui y sont traités proviennent surtout de travaux sur la modélisation de fermenteurs anaérobies [50] et de bioréacteurs destinés à la production de vanilline par des champignons filamenteux [51]. Le paragraphe 3.6 sur l’estimation du réseau réactionnel à partir d’un jeu de données est détaillé dans la référence [52].

## 3.2 Le réseau réactionnel

Le réseau réactionnel d'un système biochimique est une manière synthétique de représenter au niveau macroscopique l'ensemble des flux de matière entre deux compartiments. On adopte pour cela un formalisme semblable à celui de la chimie [20], en définissant simplement la transformation de  $n_S$  réactants  $S_i$  en  $n_P$  produits  $P_i$  sous la forme suivante :



Dans le cas d'une réaction biologique, l'un des produits correspondra à la biomasse synthétisée. Les coefficients  $k_i$  et  $k'_i$  sont des coefficients stoechiométriques incluant des rendements, ils sont souvent appelés coefficients pseudo-stoechiométriques. La vitesse de la réaction est notée  $r(\cdot)$ . Ainsi, la vitesse de consommation du substrat  $S_i$  sera  $k_i r(\cdot)$ .

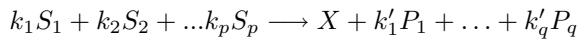
Par convention, la vitesse de la réaction est définie par la vitesse de croissance de la biomasse impliquée, en d'autre terme le coefficient pseudo-stoechiométrique associé à la biomasse est unitaire.

Le réseau réactionnel est en fait une manière synthétique de résumer au niveau macroscopique un ensemble de réactions qui sont supposées déterminer la dynamique du procédé. Le réseau réactionnel est donc le plus souvent fondé sur des hypothèses liées aux connaissances phénoménologiques disponibles.

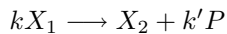
A la différence de la chimie, tous les composés intervenant dans la réaction ne sont pas forcément représentés (et heureusement, car il serait difficile de faire un bilan complet des composés (Fe, Pb, F, ...) nécessaires à la croissance microbienne). Notons également, qu'à la différence de la chimie, les réactions peuvent ne pas être équilibrées en terme de masse ou de charge ionique ; en ce sens les  $k_i$  sont à mi chemin entre des coefficients de rendement et des coefficients stoechiométriques (nous détaillerons cet aspect au paragraphe 3.7).

Les principales réactions de transformation de la matière dans un système biologique sont les suivantes :

- croissance des organismes et biosynthèse (liée au métabolisme secondaire)



- Prédation d'une espèce  $X_1$  par une espèce  $X_2$  et excrétion d'un produit (ammoniac, par exemple) :

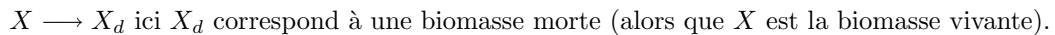


- synthèse d'un produit via le métabolisme primaire



dans ce cas, la fabrication des produits n'est pas directement liée à la croissance, mais elle dépend par exemple d'enzymes fabriquées par les organismes.

- mortalité



Par la suite nous supposerons que le réseau réactionnel est décrit par un ensemble de  $p$  réactions biologiques ou chimiques dans lequel interviennent  $n$  variables (concentrations, biomasses,...).

## 3.3 Choix du réseau réactionnel

Le choix du nombre de réactions à prendre en compte est capital pour la modélisation. Mais il dépend avant tout de la question que l'on se pose, ainsi que de la connaissance dont on dispose sur le système

qu'elle soit quantitative (données disponibles) ou qualitative (connaissance phénoménologique). Le réseau réactionnel, comme nous le verrons par la suite, va conditionner la structure du modèle. Il faudra donc le choisir avec parcimonie, en ayant à l'esprit les objectifs du modèle et la précision qui en est attendue.

Nous pouvons distinguer 3 situations différentes :

- Le réseau réactionnel est parfaitement connu. C'est par exemple le cas de réactions chimiques clairement identifiées.
- Le réseau réactionnel est connu, mais beaucoup trop compliqué par rapport à la connaissance que l'on a du système, ou par rapport aux objectifs que l'on souhaite atteindre.
- Le réseau réactionnel est inconnu, ou connu avec incertitude.

Dans les deux derniers cas, il sera alors nécessaire d'identifier un réseau réactionnel approché qui synthétisera les principaux transferts de matière au sein du système. La complexité retenue impliquera en particulier un choix du nombre d'espèces biologiques et chimiques. En général, ces espèces sont une abstraction dans le sens qu'elles représenteront un agrégat d'espèces. Nous verrons par exemple dans le cas de la digestion anaérobie comment la centaine d'espèces bactériennes se retrouve finalement représentée par deux variables "biomasse".

Nous étudierons comment identifier un tel réseau réactionnel et comment valider les résultats obtenus.

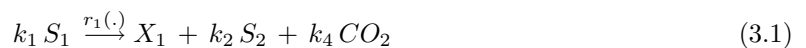
Il est très important de bien remarquer que lorsqu'on suppose un réseau réactionnel, on fait les hypothèses implicites suivantes :

- le réseau réactionnel résume les principaux flux de matière entre les différents compartiments du système ;
- les coefficients pseudo-stoechiométriques sont constants.

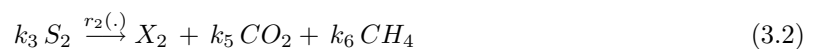
### 3.3.1 Exemple : la digestion anaérobie

Nous allons considérer ici l'exemple de la digestion anaérobie. Ce procédé de traitement biologique de l'eau utilise des bactéries anaérobies pour dégrader la matière organique ( $S_1$ ), qui est un conglomerat de substrats organiques de natures très différentes. C'est en fait un procédé très complexe, dans lequel un grand nombre de populations bactériennes interviennent (plus de 100 espèces sont impliquées [19]). De très nombreux modèles de la digestion anaérobie impliquant de 1 à 9 espèces ont été proposés dans la littérature [53, 54, 55, 56, 57, 58, 59, 60, 61, 62]. Ces modèles rentrent surtout dans la catégorie "modèles pour stocker la connaissance" (cf. classification du paragraphe 1.3), et n'ont été que très peu confrontés à des mesures. Notre objectif est de contrôler cet écosystème, par conséquent le modèle devra être relativement simple. Aussi nous limiterons nous à considérer deux populations bactériennes. Nous supposons alors que la dynamique du système peut être résumée par deux grandes étapes :

- Une étape d'acidogénèse (à vitesse  $r_1(\cdot)$ ), au cours de laquelle le substrat  $S_1$  est dégradé par des bactéries acidogènes ( $X_1$ ) et est transformé en acides gras volatils (AGV) notés  $S_2$ , et en  $CO_2$  :



- Une étape de méthanogénèse (à vitesse  $r_2(\cdot)$ ), où les acides gras volatils sont dégradés en  $CH_4$  et  $CO_2$  par les bactéries méthanogènes ( $X_2$ ).



Les constantes  $k_1, k_2, k_4$ , représentent les coefficients stoechiométriques respectivement associés à la consommation du substrat  $S_1$ , à la production des AGV et de  $\text{CO}_2$  au cours de l'acidogénèse.  $k_3, k_5$  et  $k_6$  représentent respectivement les coefficients stoechiométriques associés à la consommation des AGV et à la production de  $\text{CO}_2$  et de  $\text{CH}_4$  au cours de la méthanogénèse.

Il faut bien noter que ce réseau réactionnel est une vue de l'esprit dans la mesure où les biomasses  $X_1$  et  $X_2$  représentent en fait une flore composée d'espèces différentes. De même pour les substrats  $S_1$  et  $S_2$  qui regroupent un ensemble de composés hétérogènes.

## 3.4 La modélisation des flux de matière

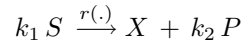
### 3.4.1 Introduction

Dans un réacteur biologique infiniment mélangé, quel que soit le mode de fonctionnement (batch, fedbatch, continu), le comportement dynamique des différents composants de la réaction biologique découle directement de l'expression du bilan de matière, qui traduit que la variation de la quantité d'un composé est égale à la somme de ce qui est produit ou apporté, diminué de ce qui est consommé ou soutiré.

Dans ce cas, à partir du réseau réactionnel et connaissant l'hydrodynamique du bioréacteur (principalement les flux d'entrée et de sortie de la matière), nous allons voir que nous pouvons directement obtenir le modèle de bilan de matière. A des fins didactiques, mais aussi pour mieux cerner où interviennent les hypothèses, nous allons étudier un exemple.

### 3.4.2 Raisonnement infinitésimal : étude de l'exemple générique

Considérons la croissance d'une biomasse sur un substrat, par exemple une population de levures *Saccharomyces cerevisiae* UG5 sur du glucose  $S$  et produisant de l'éthanol  $P$  selon le réseau réactionnel suivant :



La vitesse de croissance des micro-organismes,  $r(\cdot)$  est une fonction *a priori* inconnue des variables du système.

Déterminons tout d'abord l'équation d'évolution du volume  $V$  du bioréacteur. Le débit d'alimentation est  $Q_{in}$ , le débit de soutirage est  $Q_{out}$ . Entre l'instant  $t$  et l'instant  $t + dt$  :

$$V(t + dt) = V(t) + Q_{in} dt - Q_{out} dt$$

soit en passant à la limite  $dt \rightarrow 0$  :

$$\frac{dV}{dt} = Q_{in} - Q_{out} \quad (3.3)$$

Appelons  $n(t) = V(t)X(t)$  la **masse totale** de levure dans le bioréacteur.

Entre  $t$  et  $t + dt$  :

$$n(t + dt) = n(t) + r(\cdot)dt - \frac{Q_{out} dt}{V} n(t)$$

Pour écrire cette équation nous avons dû faire l'**hypothèse capitale** que la concentration de micro-organismes dans le volume  $dV = Q_{out}dt$  était identique à celle du bioréacteur tout entier ( $\frac{n}{V}$ ).



Nous obtenons au final :

$$\frac{dn}{dt} = r(.) - \frac{Q_{out}}{V}n(t)$$

Nous pouvons en déduire l'expression de  $\frac{dX}{dt}$  en prenant en compte l'équation (3.3).

$$\frac{dX}{dt} = r(.) - DX$$

avec  $D = \frac{Q_{in}}{V}$  qui est le taux de dilution.

En appliquant ce bilan de matière à toutes les variables et toujours dans le cadre de l'hypothèse d'homogénéité du milieu, nous obtenons :

$$\left\{ \begin{array}{l} \frac{dX}{dt} = r(.) - DX \\ \frac{dS}{dt} = -k_1 r(.) + D(S_{in} - S) \\ \frac{dP}{dt} = k_2 r(.) - DP \\ \frac{dV}{dt} = Q_{in} - Q_{out} \end{array} \right. \quad \begin{array}{l} (3.4) \\ (3.5) \\ (3.6) \\ (3.7) \end{array}$$

où  $X$  représente la concentration en micro-organismes,  $S$  la concentration en substrat,  $P$  la concentration en produit formé et  $S_{in}$  la concentration de substrat dans l'alimentation. Les coefficients stoechiométriques  $k_1$  et  $k_2$  représentent respectivement le rendement de consommation et de production.

### 3.4.3 Représentation matricielle

Sous les **hypothèses d'homogénéité du milieu**, le réseau réactionnel conduit à des équations qui décrivent **de manière équivalente** la répartition de la matière dans le bioréacteur :

$$\dot{\xi} = Kr(.) + D(\xi_{in} - \xi) - Q(\xi) \quad (3.8)$$

dans laquelle  $\xi$  représente le vecteur d'état du système,  $\xi_{in}$  est la masse entrant dans le bioréacteur et  $r(.)$  le vecteur des vitesses de réaction.  $K$  est la matrice contenant les éléments pseudo-stoechiométriques (combinaison des facteurs de changement d'unité et des rendements associés).  $Q(\xi)$ , sur lequel je reviendrai dans le paragraphe suivant représente les termes d'échange entre la phase gazeuse et la phase liquide. Enfin  $D$  est le taux de dilution, égal au rapport du débit d'alimentation  $Q_{in}$  sur le volume du réacteur  $V$ .

#### Exemple générique : le modèle de Monod

Si l'on considère le modèle (3.4) - (3.7) dans le cas d'un chémostat ( $Q_{out} = Q_{in}$ ,  $D = \frac{Q_{in}}{V}$ ), on peut récrire le modèle sous la forme matricielle (3.8) avec :

$$\xi = \begin{bmatrix} X \\ S \\ P \end{bmatrix}, \quad K = \begin{bmatrix} 1 \\ -k_1 \\ k_2 \end{bmatrix}, \quad \xi_{in} = \begin{bmatrix} 0 \\ S_{in} \\ 0 \end{bmatrix}$$

### 3.4.4 Les flux gazeux

Lors de l'écriture du bilan gazeux, nous devons prendre en compte les espèces qui ont une phase gazeuse. Ces espèces peuvent donc passer dans la phase gazeuse et s'échapper du bioréacteur (elles peuvent aussi passer de la phase gazeuse à la phase liquide).

Il faudra donc ajouter dans le terme de bilan un terme de transfert avec l'extérieur. On utilise pour cela la loi de Henry qui exprime le débit molaire du composé  $C$  de sa phase liquide vers sa phase gazeuse :

$$q_c = K_L a (C - C^*) \quad (3.9)$$

Le coefficient de transfert  $K_L a$  dépend fortement des conditions opératoires et en particulier de l'agitation, de la pression et de la surface d'échange entre les phases liquides et gazeuses [63, 48]. La modélisation de l'évolution de ce paramètre en fonction des conditions opératoires peut s'avérer très délicate.

La quantité  $C^*$  est la concentration de saturation en  $C$  dissout. Cette quantité est reliée à la pression partielle en  $C$  gazeux dans la phase gazeuse ( $P_C$ ) par la constante de Henry :

$$C^* = K_H P_C \quad (3.10)$$

La constante de Henry peut elle aussi varier, mais dans une moindre mesure, en fonction du milieu de culture et de la température.

Par ailleurs, lorsque plusieurs espèces gazeuses se trouvent simultanément dans la phase liquide, elles doivent suivre la loi des gaz parfaits, qui va se traduire par un rapport constant entre débit molaire et pression partielle. Ainsi, pour  $m$  espèces gazeuses  $C_1 \dots C_m$  :

$$\frac{P_{c1}}{q_{c1}} = \frac{P_{c2}}{q_{c2}} = \dots = \frac{P_{cm}}{q_{cm}} \quad (3.11)$$

### 3.4.5 Electroneutralité et constantes d'affinité

L'électroneutralité des solutions est une seconde règle à laquelle les systèmes biologiques ne dérogent pas : la concentration en anions pondérée par le nombre de charges doit égaler la concentration en cations pondérée de la même façon.

Les réactions purement chimiques sont souvent bien connues et une constante d'affinité leur est en général associée. Cette constante est souvent reliée à la concentration en protons  $H^+$ , donc au pH.

### 3.4.6 Exemple : la digestion anaérobie

#### Bilan de matière

Reprenons l'exemple de la digestion anaérobie dont le réseau réactionnel est donné par (3.1) et (3.2). Nous supposons que le méthane est très peu soluble et donc qu'il n'est pas stocké dans la phase liquide et passe très rapidement dans la phase gazeuse. Le  $CO_2$  quant à lui est stocké dans la phase liquide où il entre dans le compartiment de carbone inorganique ( $C$ ).

$$\frac{dX_1}{dt} = r_1(\cdot) - DX_1 \quad (3.12)$$

$$\frac{dX_2}{dt} = r_2(\cdot) - DX_2 \quad (3.13)$$

$$\frac{dS_1}{dt} = D(S_{1in} - S_1) - k_1 r_1(\cdot) \quad (3.14)$$

$$\frac{dS_2}{dt} = D(S_{2in} - S_2) + k_2 r_1(\cdot) - k_3 r_2(\cdot) \quad (3.15)$$

$$\frac{dC}{dt} = D(C_{in} - C) - q_C(\xi) + k_4 r_1(\cdot) + k_5 r_2(\cdot) \quad (3.16)$$

$S_{1in}$ ,  $S_{2in}$  et  $C_{in}$  sont respectivement les concentrations dans l'influent en substrat, AGV et carbone inorganique dissout. Le terme  $q_C(\xi)$  représente le flux de carbone inorganique (sous forme de  $CO_2$ ) de la phase liquide à la phase gazeuse.

### Flux gazeux

Le débit de méthane est directement relié à la vitesse de la méthanogénèse :

$$q_M = k_6 r_2(\cdot) \quad (3.17)$$

Le flux de  $CO_2$  suit la loi de Henry :

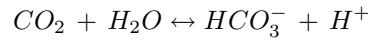
$$q_C(\xi) = K_L a(CO_2 - K_H P_C) \quad (3.18)$$

où  $P_C$  est la pression partielle de  $CO_2$ .

### Constantes d'affinité

Dans l'exemple de la digestion anaérobie, nous allons exploiter l'électroneutralité de la solution et les équilibres chimiques.

Nous considérerons que dans la gamme de pH usuelle pour ce type de procédé ( $6 \leq pH \leq 8$ ) les AGV sont sous leur forme ionisée. Le  $CO_2$  dissout est en équilibre avec le bicarbonate :



La constante d'affinité de la réaction est alors

$$K_b = \frac{HCO_3^- H^+}{CO_2} \quad (3.19)$$

### Electroneutralité de la solution

Les cations ( $Z$ ), sont principalement des ions qui ne sont pas affectés par les réactions biochimiques ( $Na^+$ , ...). De ce fait, la dynamique des cations va simplement suivre, sans modification, la concentration en cations  $Z_{in}$  dans l'entrée, si bien que :

$$\frac{dZ}{dt} = D(Z_{in} - Z) \quad (3.20)$$

Les anions quant à eux sont représentés majoritairement par les AGV et le bicarbonate, l'électroneutralité nous assure alors que

$$Z = S_2 + HCO_3^- \quad (3.21)$$

## Conclusion

En rajoutant l'équation (3.20), le modèle s'exprime finalement sous la forme matricielle (3.8), avec :

$$\xi = \begin{bmatrix} X_1 \\ X_2 \\ Z \\ S_1 \\ S_2 \\ C \end{bmatrix}, r(\cdot) = \begin{bmatrix} r_1(\cdot) \\ r_2(\cdot) \end{bmatrix}, K = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ -k_1 & 0 \\ k_2 & -k_3 \\ k_4 & k_5 \end{bmatrix} \quad (3.22)$$

$$\xi_{in} = \begin{bmatrix} 0 \\ 0 \\ Z_{in} \\ S_{1in} \\ S_{2in} \\ C_{in} \end{bmatrix}, Q = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ q_C(\xi) \end{bmatrix}, \quad (3.23)$$

L'élimination des variables  $\text{HCO}_3^-$ ,  $\text{CO}_2$ , et  $P_C$  en utilisant les équations (3.18), (3.11) (3.19) et (3.21), conduit à l'expression suivante pour  $P_C(\xi)$  (cf [64]) :

$$P_C(\xi) = \frac{\phi - \sqrt{\phi^2 - 4K_H P_T (C + S_2 - Z)}}{2K_H} \quad (3.24)$$

en posant :  $\phi = C + S_2 - Z + K_H P_T + \frac{k_6}{k_L a} r_2(\cdot)$

Finalement, cela nous donne :

$$q_C(\xi) = k_L a (C + S_2 - Z - K_H P_C(\xi)) \quad (3.25)$$

## 3.5 Identification directe du réseau réactionnel à partir d'un jeu de données

### 3.5.1 Introduction

Jusqu'ici nous avons adopté une démarche mécaniste. Nous supposons les mécanismes et les flux de matière au sein du système connus et nous en déduisons la structure du modèle la plus adaptée, basée sur un réseau réactionnel. Dans ce chapitre, nous allons nous placer dans une optique complètement différente. Nous supposerons que les mécanismes responsables de la dynamique du système sont inconnus (ou mal connus) et nous tacherons de les estimer sur la base d'un ensemble d'observations disponibles. La première étape consistera donc à déterminer le nombre de réactions nécessaires pour représenter les flux de matière au sein du système. Nous verrons ensuite comment identifier ces réactions ou, autrement dit, comment déterminer la matrice pseudo-stoechiométrique  $K$ .

### 3.5.2 Estimation du nombre de réactions à prendre en compte

Dans cette section, nous proposons une méthode pour déterminer  $p$  le nombre minimum de réactions nécessaires pour expliquer la dynamique observée expérimentalement. Nous supposons que nous mesurons un sous-ensemble  $\zeta$  de  $n$  composantes de  $\xi$  qui interviennent dans le système (*c.-à-d.*  $n$  variables qui présentent des variations significatives au cours du temps). En effet les mesures des autres composantes de l'état (appelées  $\tilde{\zeta}$ ) peuvent ne pas être disponibles. Nous supposons cependant que nous mesurons plus de variables qu'il y a de réactions :  $n > p$  (comme  $p$  est *a priori* inconnu, nous vérifierons cette hypothèse *a posteriori*). Si ces composantes ont une phase gazeuse, nous supposons que les débits gazeux associés  $Q(\zeta, \tilde{\zeta})$  sont mesurés.

Les équations associées à  $\zeta$  sont alors :

$$\frac{d\zeta}{dt} = K r(\zeta, \tilde{\zeta}) + D(\zeta_{in} - \zeta) - Q(\zeta, \tilde{\zeta}), \quad (3.26)$$

**Remarque 1**  $K$  est en fait la sous-matrice d'une matrice pseudo-stoechiométrique décrivant les flux de matière au sein d'un système plus grand. Par abus de notation nous gardons les notations  $K$  et  $Q$  pour ce sous-système.

### 3.5.3 Détermination théorique de $\dim(\mathcal{I}m(K))$

Intégrons l'équation (3.26) entre 2 instants :

$$\zeta(t) - \zeta(0) - \int_0^t D(\zeta_{in}(\tau) - \zeta(\tau)) + Q(\xi(\tau)) d\tau = K \int_0^t r(\xi(\tau)) d\tau, \quad (3.27)$$

Posons :

$$\eta(t) = \zeta(t) - \zeta(0) - \int_0^t D(\zeta_{in}(\tau) - \zeta(\tau)) + Q(\xi(\tau)) d\tau \quad (3.28)$$

et

$$\epsilon(t) = \int_0^t r(\xi(\tau)) d\tau$$

Les équations (3.27) peuvent alors être reformulées :

$$\eta(t) = K \epsilon(t) \quad (3.29)$$

Le vecteur  $\eta(t)$  peut être estimé au cours du temps en utilisant des mesures disponibles. La valeur de l'intégrale dans (3.28) peut être calculée par exemple par une méthode des trapèzes. Plus généralement, afin d'améliorer le nettoyage des données (réduction du bruit et diminution de l'autocorrélation) n'importe quelle application différentielle scalaire linéaire peut être appliquée à  $\eta(t)$ .

En effet, si  $G(s, \theta)$  est une fonction de transfert comprenant une combinaison linéaire d'intégrations, de différentiations et de retards, nous avons :

$$\mathcal{Y}(s) = G(s, \theta) \mathcal{U}(s) = K G(s, \theta) \mathcal{W}(s)$$

où  $\mathcal{U}(s)$  et  $\mathcal{W}(s)$  représentent respectivement les transformées de Laplace de  $\eta(t)$  et de  $\epsilon(t)$ .  $\mathcal{Y}(s)$  est la transformée de Laplace du signal après filtrage. Si nous appelons  $u(t)$  et  $w(t)$  les signaux issus respectivement de  $\eta(t)$  et de  $\epsilon(t)$  après filtrage, nous obtenons alors :

$$u(t) = K w(t) \quad (3.30)$$

Par exemple, la moyenne mobile est un filtre simple :

$$u(t) = \frac{1}{T} \int_t^{t+T} \eta(\tau) d\tau \quad (3.31)$$

$$w(t) = \frac{1}{T} \int_t^{t+T} \epsilon(\tau) d\tau \quad (3.32)$$

Par la suite, nous ne précisons pas le type de filtre appliqué aux données (même si nous supposons que les données ont été lissées par le filtre approprié).

Pour déterminer la dimension de la matrice  $K$  nous allons maintenant estimer la dimension de l'image de  $K$ , en d'autres termes, la dimension de l'espace où vit  $u(t)$ . Notons que nous supposons implicitement que  $K$  est une matrice de plein rang. En effet, si cela n'était pas le cas, cela signifierait que le même comportement dynamique pourrait être obtenu avec une matrice  $K$  de dimension inférieure.

La détermination de la dimension de l'espace dans lequel se trouve  $u(t)$  est un problème classique en analyse de données. Cela correspond à l'analyse en composantes principales, qui détermine la dimension de l'espace vectoriel engendré par les vecteurs  $k_i$ , colonnes de  $K$ .

Nous considérerons pour cela la matrice  $n \times N$ ,  $U$ , obtenue à partir d'un ensemble de  $N$  valeurs de  $u$  calculées aux temps  $t_i$  (nous choisirons plus de mesures que de variables mesurées :  $N > n$ ) :

$$U = (u(t_1), \dots, u(t_N))$$

Nous construisons également la matrice des taux de réaction associés, qui est inconnue :

$$W = (w(t_1), \dots, w(t_N))$$

Nous supposons que la matrice  $W$  est de plein rang. En d'autres termes que les réactions sont indépendantes. Si ce n'était pas le cas, le comportement observé pourrait être reproduit avec moins de réactions.

**Propriété 1** Soit une matrice  $K$  de rang  $p$ , si  $W$  est de plein rang, alors la matrice  $n \times n$  :

$$M = UU^T = KWW^TK^T$$

est de rang  $p$ . Comme c'est une matrice symétrique elle peut s'écrire :

$$M = P^T \Sigma P$$

où  $P$  est une matrice orthogonale ( $P^T P = I$ ) et

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & & \dots & & 0 \\ 0 & \sigma_2 & 0 & & & 0 \\ \vdots & & \ddots & & & \\ & & & \sigma_p & & \\ & & & & 0 & \\ & & & & & \ddots & \vdots \\ 0 & & \dots & & & & 0 \end{pmatrix}$$

avec  $\sigma_{i-1} \geq \sigma_i > 0$  for  $i \in \{2, \dots, p\}$ .

De plus les vecteurs propres associés à  $\sigma_i$  forment une base orthonormale de  $\text{Im } K$ .

C'est une application directe du théorème de décomposition en valeurs singulières [65]. Comme  $\text{rang}(M) = \text{rang}(KW) = \text{rang}(K) = \text{rang}(\Sigma) = p$ , on obtient le résultat.

Le nombre de réactions nécessaires pour reproduire les données mesurées correspond au nombre de valeurs singulières non nulles de  $UU^T$ .

En pratique le cas idéal que nous venons de présenter est perturbé pour quatre raisons principales :

- le réseau réactionnel que nous recherchons est une première approximation de l'ensemble des réactions chimiques ou biochimiques qui peuvent être très complexes. La "vraie" matrice de  $K$  est probablement une matrice beaucoup plus compliquée. L'ensemble des réactions qui sont rapides ou de faible amplitude peut être considéré comme perturbation d'un réseau réactionnel principal.
- les mesures sont corrompues par du bruit. Ce bruit peut être très important, particulièrement dans le cas des mesures de quantités biologiques qui souffrent d'un manque de capteurs fiables.
- les mesures sont rarement toutes disponibles exactement au même instant  $t_i$  et donc elles doivent être interpolées à  $t_i$ .
- Pour estimer  $u(t)$  dans l'équation (3.28) il faut estimer la valeur d'une intégrale en passant donc par une approximation.

En conséquence, dans le cas réel, la matrice  $M = UU^T$  n'aura aucune valeur propre rigoureusement nulle. Il faudra donc déterminer le nombre de vecteurs propres à considérer afin de représenter une approximation raisonnable de l'espace vectoriel engendré par  $u(t)$ . Pour résoudre ce problème, remarquons que les valeurs propres  $\sigma_i$  de  $M$ , une fois le problème normalisé, correspondent à la variance associée au vecteur propre  $\lambda_i$  (axe d'inertie). La méthode consiste alors à choisir les  $p$  premiers axes principaux qui représentent une variance totale supérieure à un seuil fixé. Par exemple, dans l'exemple suivant, nous avons fixé un seuil (selon l'information disponible sur des mesures de bruit) à 95% de la variance, ceci a mené au choix de 5 axes et donc  $p = 5$ .

**Remarque 2** si  $\text{rang}(M) = n$  cela signifie que  $\text{rank}(W) \geq n$ . Dans ce cas nous ne pouvons pas estimer  $p$  et des mesures de variables supplémentaires sont nécessaires pour appliquer les méthodes présentées.

### 3.5.4 Exemple : bioproduction de vanilline

Nous considérons ici la production de vanilline à partir d'acide vanillique par le champignon filamenteux *Pycnoporus cinnabarinus* [66]. Cette biotransformation résulte d'un ensemble complexe de réactions [67], la plupart d'entre elles étant mal connues. Dans une première étape (qui dure généralement pendant les 3 premiers jours), le mycélium utilise les substrats disponibles (azote, maltose et glucose) pour se développer. La croissance est aérobie et donc l'oxygène est consommé et du  $\text{CO}_2$  est produit. Dans une deuxième étape, la biosynthèse est déclenchée par l'addition de la cellobiose 2 heures avant l'addition continue d'acide vanillique. Le mycélium transforme alors l'acide vanillique en methoxyhydroquinone ou en vanilline. Dans ce dernier cas, la vanilline peut également être dégradée en alcool vanillique. Les variables liées à cette bioproduction sont les sources de carbone (maltose et glucose), la source d'azote (ammonium), l'oxygène, le dioxyde de carbone, la biomasse fongique et les composés phénoliques (acide vanillique, vanilline, alcool vanillique et methoxyhydroquinone). Il résulte que 10 variables évoluent pendant le processus de croissance.

Les vecteurs  $u(t_i)$  (équation 3.31) ont été calculés à partir de données expérimentales. La figure 3.1

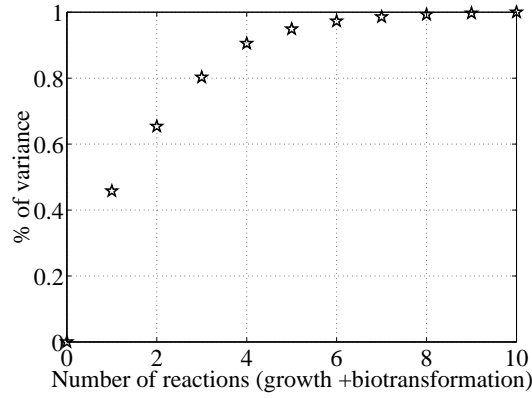


FIG. 3.1 – Variance totale expliquée en fonction du nombre de réactions pour la bioproduction de vanilline par le champignon filamenteux *Pycnoporus cinnabarinus*.

représente la variance cumulée associée au nombre d'axes d'inertie considérés. Quatre réactions sont suffisantes pour expliquer 80% de la variance observée. Cinq réactions expliquent 95% de la variance. Cette analyse a motivé la structure du modèle exposé dans [51].

### 3.6 Estimation de la matrice pseudo-stoechiométrique à partir d'un jeu de données

Utilisons la propriété 1 qui dit que  $\mathcal{Im} K$  est engendré par les vecteurs propres  $\rho_i$  associés aux valeurs propres non nulles de  $UU^T$ . A partir des données expérimentales collectées dans la matrice  $UU^T$  nous pouvons calculer les  $p$  vecteurs propres  $\rho_i$  qui engendrent  $K$ . Cela signifie que chaque colonne  $k_i$  de  $K$  est une combinaison linéaire des  $\rho_i$ . En d'autres termes, il existe une  $p \times p$  matrice  $G$  telle que

$$K = \rho G$$

où les colonnes de la matrice  $\rho$  sont les vecteurs propres  $\rho_j$ .

Autrement dit, la famille des matrices pseudo-stoechiométriques possibles  $K$  est paramétrisée par  $G$ .

**Remarque :** En général, puisque les taux de réaction sont inconnus, la matrice  $G$  (et donc la matrice  $K$ ) n'est pas identifiable. Ceci est illustré par l'exemple simple qui suit. Si  $r_1(\xi)$  et  $r_2(\xi)$  sont deux taux de réaction, le terme  $Kr(\xi)$  peut être écrit :

$$\begin{aligned} Kr(\xi) &= k_1 r_1(\xi) + k_2 r_2(\xi) \\ &= \frac{k_1 + k_2}{2} (r_1(\xi) + r_2(\xi)) + \frac{k_1 - k_2}{2} (r_1(\xi) - r_2(\xi)) \end{aligned}$$

et donc les matrices  $K = [k_1 \ k_2]$  et  $\bar{K} = [\frac{k_1 + k_2}{2} \ \frac{k_1 - k_2}{2}]$  peuvent toutes les deux produire le même résultat. Les taux de réaction associés à la deuxième matrice sont alors :  $\bar{r}_1(\xi) = r_1(\xi) + r_2(\xi)$  et  $\bar{r}_2(\xi) = r_1(\xi) - r_2(\xi)$ .

#### 3.6.1 Hypothèses supplémentaires

Afin de rendre la matrice  $G$  (et donc  $K$ ) identifiable de manière unique, nous devons ajouter des contraintes structurales.



Tout d'abord, nous imposerons (sans perte de généralité) que chaque taux de réaction est normalisé pour une espèce biochimique et donc que chaque colonne de la matrice  $K$  contient un +1 ou un -1. Ceci induit des contraintes supplémentaires sur l'ensemble des matrices possibles  $G$ . Parfois nous pouvons ne pas connaître le signe de l'élément : les deux cas possibles (+1 ou un -1) doivent alors être considérés.

Si d'autres contraintes sont nécessaires, nous devons utiliser des hypothèses biologiques. Par exemple, nous pouvons supposer qu'un élément spécifique n'est pas impliqué dans l'une des  $p$  réactions (signifiant qu'il y a un zéro dans  $K$ ). Il est clair par exemple que la première réaction impliquera seulement les substrats qui étaient présents au début de la fermentation. Nous pouvons également imposer le respect d'une stoechiométrie, ou n'accepter qu'une perte de masse dans le système. On peut également essayer de trouver une matrice  $K$  impliquant un nombre minimum de composants dans chaque réaction (c.-à-d. contenant le nombre maximum de zéros). Si ces hypothèses ne sont pas suffisantes, plusieurs matrices  $K$  peuvent alors être identifiées et peuvent dépendre de paramètres inconnus. La signification biologique des réactions correspondantes et leur légitimité d'un point de vue chimique doivent alors être discutées.

### 3.6.2 Validation

Le résultat principal de l'analyse précédente est la détermination des variables qui sont des substrats ou des produits dans les réactions ou, en d'autres termes, les signes des éléments de  $K$ .

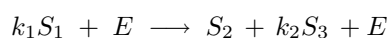
Un autre résultat important est la caractérisation des variables qui ne sont pas impliquées dans une réaction et qui correspondent aux éléments nuls dans la matrice  $K$ . Cependant, en pratique il est très peu probable que l'analyse fournisse une valeur exactement nulle de ces éléments de  $K$ . L'idée consiste alors à remplacer les éléments très petits par des zéros, puis à valider le réseau réactionnel correspondant en utilisant les techniques qui seront présentées au paragraphe 5.2.

### 3.6.3 Exemple : production de lipase à partir d'huile d'olive par *Candida rugosa*

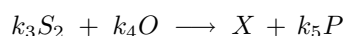
Considérons l'exemple d'une croissance compétitive sur deux substrats [68], ce qui pourrait représenter, par exemple, la production de lipase à partir d'huile d'olive par *Candida rugosa*. La levure se développe sur deux substrats qui sont produits par l'hydrolyse d'un substrat organique complexe primaire.

Le réseau réactionnel suivant en 3 étapes a été proposé dans la littérature :

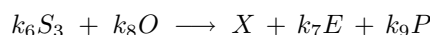
– Hydrolyse :



– Croissance à partir de  $S_2$  :



– Croissance à partir de  $S_3$  :



où  $S_1$  est le substrat primaire (huile d'olive),  $S_2$  (glycérol) et  $S_3$  (acide gras) sont les substrats secondaires.  $E$  est l'enzyme (lipase) produite,  $X$  la biomasse,  $O$  l'oxygène dissout et  $P$  le dioxyde de carbone.

La matrice pseudo-stoechiométrique associée est donc :

$$K = \begin{pmatrix} -k_1 & 0 & 0 \\ 1 & -k_3 & 0 \\ k_2 & 0 & -k_6 \\ 0 & 0 & k_7 \\ 0 & 1 & 1 \\ 0 & -k_4 & -k_8 \\ 0 & k_5 & k_9 \end{pmatrix}$$

△

Nous supposons que ce réseau réactionnel est inconnu et doit être découvert à partir des mesures des diverses concentrations. Ici les données seront simulées par un modèle (voir [52]) mais naturellement dans la pratique les données sont obtenues à partir des expériences. Nous supposons que les variables d'état,  $E$ ,  $X$ ,  $O$  et les débits gazeux  $q_{O_2}$  et  $q_{CO_2}$  ont été mesurés. Nous supposons ici que la variable d'état  $S_1$  n'a pas été enregistrée, afin de montrer que l'approche est applicable même si l'ensemble des variables d'état n'est pas mesuré. Le taux de dilution et le taux d'apport en substrat ont été choisis afin de garantir que le système reste suffisamment excité. Cela assure que les signaux enregistrés auront un contenu assez riche pour obtenir de bons résultats d'identification.

Les vecteurs  $u(t_i)$  ont été calculés à partir de ces données simulées puis normalisés. Finalement, les vecteurs propres de  $UU^T$  ont été calculés. Cette analyse permet tout d'abord de conclure que seules 3 réactions sont nécessaires pour reproduire le comportement dynamique de ce système.

Nous supposons ici que la première réaction est connue et donc nous nous concentrons sur les deux autres réactions.

<b><math>K</math> utilisée pour la simulation</b>	<b><math>K</math> identifiée</b>
$\begin{pmatrix} -5 & 0 \\ 0 & -0.5 \\ 0 & 0.2 \\ 1 & 1 \\ -2 & -1 \\ 0.3 & 1.5 \end{pmatrix}$	$\begin{pmatrix} -3.54 & 0 \\ 0 & -0.51 \\ 0.01 & 0.22 \\ 1 & 1 \\ -1.34 & -0.87 \\ 0.18 & 1.51 \end{pmatrix}$

TAB. 3.1 – Coefficients de la matrice  $K$  utilisés pour la simulation et valeurs identifiées.

Nous pouvons maintenant calculer les quantités  $U_i$  associées aux 6 variables mesurées en utilisant une moyenne mobile. Puis, nous construisons la matrice  $M = U^T U$ . Les vecteurs propres  $\rho_i$  liés aux deux plus grandes valeurs propres constituent alors la base de  $\mathcal{Im} K$ . Comme  $G$  est une matrice  $2 \times 2$ , les colonnes  $k_1$  et  $k_2$  de  $K$  peuvent être écrites :

$$k_1 = \alpha_{11}\rho_1 + \alpha_{12}\rho_2 \text{ et } k_2 = \alpha_{21}\rho_1 + \alpha_{22}\rho_2 \quad (3.33)$$

Nous procédons maintenant en deux étapes successives :

*i. normalisation .*

Les coefficients pseudo-stoechiométriques associés à la croissance de la biomasse sont normalisés :  $k_{14} = 1$  et  $k_{24} = 1$ . Nous obtenons alors :

$$\begin{aligned} k_{14} = 1 &= \alpha_{11}\rho_{14} + \alpha_{12}\rho_{24} \\ k_{24} = 1 &= \alpha_{21}\rho_{14} + \alpha_{22}\rho_{24} \end{aligned} \quad (3.34)$$

En utilisant les équations 3.33 et 3.34 avec les valeurs obtenues de  $\rho_1$  et de  $\rho_2$ , nous pouvons maintenant écrire la matrice  $K$  paramétrisée par  $\alpha_{11}$  et  $\alpha_{22}$  comme suit :

$$K = \begin{pmatrix} -1.42\alpha_{11} - 2.65 & -1.2\alpha_{22} + 1.12 \\ 0.2\alpha_{11} - 0.13 & 0.17\alpha_{22} - 0.67 \\ -0.08\alpha_{11} + 0.062 & -0.071\alpha_{22} + 0.28 \\ 1 & 1 \\ -0.19\alpha_{11} - 1.2 & -0.16\alpha_{22} - 0.72 \\ -0.53\alpha_{11} + 0.51 & -0.45\alpha_{22} + 1.93 \end{pmatrix}$$

*ii. Hypothèses supplémentaires .*

Pour déterminer de manière unique la matrice  $K$  nous ajoutons deux hypothèses :

**Hypothèse 1** *La réaction a encore lieu si  $S_2$  [resp.  $S_3$ ] n'est pas présent au temps initial. De plus  $S_3$  [resp.  $S_2$ ] n'est pas produit au cours de cette réaction.*

En d'autres termes cela signifie que  $S_2$  est le (seul) substrat d'une réaction et que  $S_3$  est le (seul) substrat de l'autre. Ainsi nous imposerons  $k_{12} = 0$  et  $k_{21} = 0$ .

Ces contraintes nous permettent de calculer  $\alpha_{11}$  (0,621) et  $\alpha_{22}$  (0,93).

Finalement nous pouvons maintenant évaluer la matrice  $K$  (voir le Tableau 3.1). La matrice identifiée est très proche de la vraie matrice  $K$ . Le coefficient  $k_{13}$ (théoriquement nul) vaut 0,01 et peut être négligé par rapport aux autres coefficients de  $K$ .

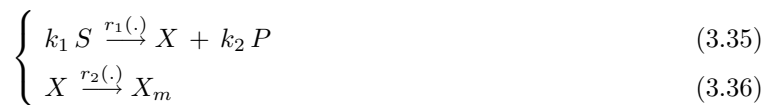
En définitive, la partie inconnue de la matrice  $K$  a été correctement identifiée. Par ailleurs les estimations des autres termes de la matrice  $K$  sont tout à fait précises.

## 3.7 Remarques importantes sur le bilan de matière

### 3.7.1 Le bilan ne doit pas nécessairement être fermé

Tout d'abord et c'est souvent ambiguë, le bilan de masse ne doit pas nécessairement être fermé *stricto sensu*. En effet, il n'est pas nécessaire que le modèle de bilan de matière conserve la masse et une partie de la matière peut se perdre (mortalité, fuite, etc). Dans ce cas le modèle considéré est en fait le sous modèle d'un modèle qui, lui, respecte la conservation de la masse. Nous allons illustrer ce point sur un exemple.

Reprenons l'exemple générique d'un micro-organisme se développant en consommant un substrat  $S$  et en produisant un métabolite  $P$  (cf. paragraphe 3.4.2) et ajoutons-y la mortalité de la biomasse. Cela revient à compléter le réseau réactionnel par une deuxième réaction :



où  $X_m$  représente la biomasse morte. Si nous écrivons maintenant le modèle de bilan de matière associé, nous obtenons :

$$\begin{cases} \frac{dX}{dt} = r_1(\cdot) - r_2(\cdot) - DX & (3.37) \\ \frac{dS}{dt} = -k_1 r_1(\cdot) + D(S_{in} - S) & (3.38) \\ \frac{dP}{dt} = k_2 r_1(\cdot) - DP & (3.39) \\ \frac{dX_m}{dt} = r_2(\cdot) - DX_m & (3.40) \end{cases}$$

Le bilan est fermé, et cela se vérifie aisément car la quantité  $z = \frac{S}{k_1} + \frac{P}{k_2} + X + X_m$  est (asymptotiquement) conservée. Considérons maintenant la formulation matricielle associée à la structure générale (3.8) et omettons l'équation de la biomasse morte (3.40). Le modèle pour  $X, S$  et  $P$ , dans lequel il y a une **fuite de masse** s'écrit :

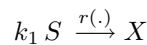
$$\xi = \begin{bmatrix} X \\ S \\ P \end{bmatrix}, \quad K = \begin{bmatrix} 1 & -1 \\ -k_1 & 0 \\ k_2 & 0 \end{bmatrix}, \quad r(\cdot) = \begin{bmatrix} r_1(\cdot) \\ r_2(\cdot) \end{bmatrix}, \quad \xi_{in} = \begin{bmatrix} 0 \\ S_{in} \\ 0 \end{bmatrix}$$

Maintenant, il est aisé de remarquer que la quantité  $k_1 P + k_2 S$  est encore (asymptotiquement) conservée, même si la masse totale du système ne l'est plus.

### 3.7.2 C'est bien plus que la masse qui est conservée...

L'exemple précédent suggère déjà clairement que ce type de modélisation fonctionne y compris pour des systèmes dans lesquels la masse n'est pas conservée (parce qu'on ne considère pas tous les compartiments). Nous allons voir que les flux dont il est question —certes reliés à la masse— sont en fait bien plus abstraits.

Considérons à nouveau l'exemple de la croissance d'une population de bactéries, dont la biomasse est estimée par la masse en carbone. Supposons que le substrat considéré soit l'ammonium  $NH_4^+$ . Le substrat peut donc s'exprimer par une concentration en azote, alors que la biomasse est associée à une concentration en carbone. Le sens physique de la réaction suivante paraîtrait sans doute aberrant à un chimiste :



Cette réaction est associée au modèle :

$$\xi = \begin{bmatrix} X \\ S \end{bmatrix}, \quad K = \begin{bmatrix} 1 \\ -k_1 \end{bmatrix}, \quad \xi_{in} = \begin{bmatrix} 0 \\ S_{in} \end{bmatrix}$$

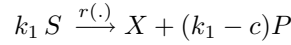
Cela nous amène tout naturellement à la conservation de la quantité  $z = k_1 X + S$ . Pour des raisons d'homogénéité  $k_1$  est fondamentalement un coefficient qui intègre un changement d'unité pour passer d'une concentration de carbone à une concentration d'azote.

Nous allons maintenant montrer que  $z$  ne correspond pas à la masse totale, et nous allons calculer celle-ci.

Supposons que la concentration  $c$  d'azote par unité de biomasse soit constante et connue. La quantité d'azote contenue dans la biomasse est donc  $cX$  et la quantité totale d'azote dans le bioréacteur est donc  $cX + S$ , qui devrait être conservée. Si le bilan de masse était fermé, lors de l'identification de  $k_1$  sur base

de données expérimentales (en suivant par exemple la méthode proposée au paragraphe 5.2.4) on devrait trouver une valeur de  $k_1$  très proche de celle de  $c$ . Dans nombre de cas, on obtient en fait  $k_1 \geq c$ .

Cela s'explique simplement si l'on admet qu'une partie de la masse d'azote s'échappe, par exemple sous la forme d'un produit excrété (et non mesuré)  $P$ . Le "vrai" réseau réactionnel est alors :



ou  $P$  est un composé (exprimé en concentration d'azote) qui échappe à la mesure et qui représente donc la perte d'azote. Le nouveau modèle est maintenant :

$$\xi = \begin{bmatrix} X \\ S \\ P \end{bmatrix}, \quad K = \begin{bmatrix} 1 \\ -k_1 \\ k_1 - c \end{bmatrix}, \quad \xi_{in} = \begin{bmatrix} 0 \\ S_{in} \\ 0 \end{bmatrix}$$

Les équations de conservation sont alors données par la conservation des quantités

$$\begin{cases} z_1 = k_1 X + S & (3.41) \\ z_2 = cX + S + P & (3.42) \end{cases}$$

La quantité  $z_2$  correspond maintenant à la "vraie" conservation de la quantité totale d'azote.

On consultera [9] pour un autre exemple qui démontre que la masse n'est qu'un cas très particulier parmi l'ensemble des quantités conservées.

En conclusion, ce qui pourrait donner l'impression d'un manque de rigueur est plutôt lié à un abus de langage ; le terme de bilan de masse étant clairement usurpé. La notion fondamentale dont il est question est à relier au principe d'invariant asymptotique encore appelé invariant réactionnel [45, 46, 47], qui n'est pas forcément directement relié à la masse.

Par contre, cette approche ne se limitant pas à considérer des systèmes où la masse est conservée *stricto sensu*, son champ d'application s'élargit *de facto*. Cette souplesse est sans doute la raison clé qui explique le succès de cette méthode.

### 3.8 Conclusion sur le modèle de bilan de masse

A ce stade, nous arrivons donc à un modèle reposant sur les principes physiques et chimiques suivants :

- Bilan de masse
- Bilan ionique
- Constante d'affinité
- Loi des gaz parfaits
- Loi de Henry

La plus importante des hypothèses (en terme de fiabilité) est celle qui propose le bilan de matière.

Cette première étape de la modélisation est capitale, car les hypothèses impliquées peuvent être testées indépendamment du reste du modèle. Elles peuvent donc être consolidées et enrichies au fil des nouvelles expériences.

Mon approche s'est d'abord appuyée sur la connaissance phénoménologique pour générer le modèle. Puis, j'ai montré comment obtenir le modèle directement à partir d'une analyse des données disponibles.

En pratique, un aller-retour entre ces deux approches doit permettre de converger vers une (ou plusieurs) structure(s).

Ce modèle, pourra être utilisé tel quel, par exemple pour développer des observateurs où les vitesses de réactions sont considérées comme des entrées inconnues (observateurs asymptotiques [20]). Le modèle de bilan de matière peut également être utilisé comme base à des lois de contrôle [37].

Néanmoins, si l'objectif défini au départ nécessite une formulation complète du modèle (simulation, contrôle optimal, observateurs à convergence aussi rapide que voulu, etc) il faudra alors expliciter les vitesses de réactions  $r_i(\cdot)$  en fonction des variables d'état du système et des entrées (variables environnementales). Cette étape est beaucoup plus délicate et nous devons faire de nombreuses hypothèses difficiles à vérifier comme nous allons le voir au chapitre suivant.

## Chapitre 4

# Modélisation des cinétiques réactionnelles

*“Je me plaisais surtout aux mathématiques, à cause de la certitude et de l’évidence de leurs raisons, mais je ne remarquais point encore leur vrai usage, et, pensant qu’elles ne servaient qu’aux arts mécaniques, je m’étonnais de ce que, leurs fondements étant si fermes et si solides, on n’avait rien bâti dessus de plus relevé.”*

René Descartes, Discours sur la méthode. 1637.

### 4.1 Introduction

Pour certains des objectifs qui ont motivé la modélisation (contrôle optimal, simulation, prédiction,...) il est nécessaire de disposer d’une expression reliant les vitesses de réactions aux variables du système. Il faut toutefois bien garder à l’esprit que ce sont le plus souvent des relations spéculatives et très approximatives, et que l’on s’éloigne ainsi de la modélisation d’ordre physique présentée dans le paragraphe précédent.

Dans ce chapitre, nous allons voir comment hiérarchiser les hypothèses sur les cinétiques pour aboutir à nouveau à une description structurée en niveaux de fiabilité.

Ce chapitre est associé au chapitre du livre [44], les aspects liés à la modélisation hybride par réseau de neurones (paragraphe 4.4 ) sont traités plus en détail dans [69, 70]. La méthode pour l’estimation directe de la structure des cinétiques (paragraphe 4.5) correspond à [71].

### 4.2 Les contraintes mathématiques

#### 4.2.1 Positivité des variables

*A priori*, on connaît un certain nombre de contraintes physiques que le modèle doit respecter : les variables doivent rester positives et elles devront être bornées si la quantité de matière entrant dans le bioréacteur est bornée. Ces contraintes physiques vont imposer des contraintes sur la structure des  $r_i(\cdot)$ . Certaines quantités

(pourcentages, ratios,...) doivent rester entre deux bornes connues. Pour garantir que le modèle conserve bien cette propriété il doit vérifier la propriété suivante :

**Propriété 2 (H1)** *On souhaite que  $\forall t \geq 0 \xi_i(t) \in [L_{i,min}, L_{i,max}]$ . Pour chaque variable d'état  $\xi_i$ , le champ  $\dot{\xi}_i$  sur l'axe  $\xi_i = 0$  doit donc pointer vers la partie de l'espace admissible. En d'autres termes, la propriété suivante doit être vérifiée :*

$$\begin{aligned}\xi_i = L_{i,min} &\Rightarrow \dot{\xi}_i \geq 0 \\ \xi_i = L_{i,max} &\Rightarrow \dot{\xi}_i \leq 0\end{aligned}$$

**Cas particulier :** pour qu'une variable  $\xi_i$  reste positive, il faut s'assurer que  $\xi_i = 0 \Rightarrow \dot{\xi}_i \geq 0$ .

### 4.2.2 Variables nécessaires à la réaction

La deuxième contrainte que devront respecter les cinétiques biochimiques est liée au réseau réactionnel. La réaction ne peut pas avoir lieu si l'un des réactants nécessaires à la réaction est manquant.

On utilise en fait la propriété mathématique classique :

**Lemme 1 (Lemme d'Hadamard)** *Soit  $f$  une fonction de classe  $C^k$  d'un ouvert convexe de  $\mathbf{R}^n$  dans  $\mathbf{R}$ . Si  $x_i = 0 \Rightarrow f(x) = 0$ , alors il existe une fonction  $g$  de classe  $C^{k-1}$  telle que  $f(x) = x_i g(x)$ .*

Cela explique la propriété suivante :

**Propriété 3** *Si  $\xi_j$  est un réactant de la réaction  $i$ , alors  $\xi_j$  peut être mis en facteur dans  $r_i$  :*

$$r_i(\xi, u) = \xi_j \nu_{ij}(\xi, u) \quad (4.1)$$

On vérifie alors aisément que  $\xi_j = 0 \Rightarrow r_i(\xi, u) = 0$ .

Nous avons la même propriété pour les réactions associées à une biomasse  $X$ . Ainsi, une réaction de croissance va s'écrire :

$$r_i(\xi, u) = \mu_i(\xi, u)X$$

Le terme  $\mu_i(\cdot)$  est alors appelé le taux de croissance.

### 4.2.3 Exemple : la digestion anaérobie

Considérons le modèle donné par les équations (3.12) à (3.15). Et appliquons le principe de positivité des variables :

$$X_1 = 0 \Rightarrow r_1(\cdot) \geq 0 \quad (4.2)$$

$$X_2 = 0 \Rightarrow r_2(\cdot) \geq 0 \quad (4.3)$$

$$S_1 = 0 \Rightarrow D(S_{1in} - S_1) - k_1 r_1(\cdot) \geq 0 \quad (4.4)$$

$$S_2 = 0 \Rightarrow D(S_{2in} - S_2) + k_2 r_1(\cdot) - k_3 r_2(\cdot) \geq 0 \quad (4.5)$$

Les équations (4.2) et (4.3) n'apportent aucune information car les  $r_i$  sont positifs. Pour que (4.4) et (4.5) soient respectées quelles que soient les conditions expérimentales, en particulier pour  $D=0$ , il faut que :

$$r_1(\cdot) = S_1 \phi_1(\cdot) \text{ et } r_2(\cdot) = S_2 \phi_2(\cdot)$$



Par ailleurs, les biomasses  $X_1$  et  $X_2$  sont nécessaires, respectivement pour les réactions d'acidogénèse et de méthanogénèse, de ce fait :

$$r_1(\cdot) = \mu_1(\cdot)X_1 \text{ et } r_2(\cdot) = \mu_2(\cdot)X_2$$

Donc au final, on doit avoir :

$$r_1(\cdot) = S_1X_1\nu_1(\cdot) \quad (4.6)$$

$$r_2(\cdot) = S_2X_1\nu_2(\cdot) \quad (4.7)$$

## 4.3 La connaissance phénoménologique

### 4.3.1 Recueillir l'information

Afin de proposer une expression pour les cinétiques réactionnelles, nous allons exploiter la connaissance phénoménologique disponible, bien que celle-ci soit, en général critiquable.

Tout d'abord, les expériences en laboratoire permettent souvent de déterminer les variables qui agissent sur la vitesse des réactions. Nous avons vu que parmi ces variables nous devons trouver (au moins) les réactants et éventuellement la biomasse.

Ensuite, il faut déterminer si la vitesse de la réaction est activée ou inhibée par ces variables. Il arrive fréquemment qu'une variable soit activatrice et qu'elle devienne inhibitrice à très haute concentration (toxicité).

Il reste maintenant à proposer une expression analytique qui va permettre de prendre en compte les contraintes mathématiques et la connaissance phénoménologique sur le procédé. Dans les travaux de modélisation, les auteurs s'appuient le plus souvent sur des observations expérimentales (lorsqu'elles existent) ou bien sur des travaux antérieurs de modélisation de processus similaires. Dans d'autres cas les modèles sont basés sur des analogies, soit avec la mécanique (taux de rencontre, loi d'action de masse), la chimie (cinétiques enzymatiques) ou bien la biologie (par comparaison à des expressions standards, type Monod).

Dans tous les cas, on privilégie le principe de parcimonie, afin que les modèles obtenus puissent être identifiés et validés par la suite.

Les paragraphes suivants présentent les modèles les plus souvent retenus pour décrire la croissance limitée par un substrat.

### 4.3.2 Fonctions classiques pour représenter les cinétiques biologiques

Même si le taux de croissance dépend fortement des conditions opératoires (température, pH...), du milieu réactionnel (concentrations en composés carbonés, azotés, phosphorés, en sels minéraux, en oxygène...), l'expression la plus couramment utilisée est le modèle empirique de Monod [72]. Cette expression décrivait à l'origine la croissance bactérienne, en se basant sur la loi introduite au début du 20<sup>ième</sup> siècle par Michaelis-Menten pour représenter une cinétique enzymatique :

$$\mu(S) = \mu_{max} \frac{S}{K_s + S} \quad (4.8)$$

Cette expression, dans laquelle  $\mu_{max}$  est le taux de croissance maximal et  $K_s$  la constante de demi-saturation, permet de décrire le phénomène de limitation de la croissance par manque de substrat et l'arrêt complet

lorsque le substrat n'est plus disponible. Notons que les analogies avec les cinétiques enzymatiques ont souvent été utilisées pour déterminer les modèles de croissance ou de prédation [73, 74].

Par ailleurs, les phénomènes d'inhibition par excès de substrat sont généralement modélisés par l'expression de Haldane, introduite dans le cas des réactions enzymatiques et reprise par Andrews [75] dans le cas des réactions biologiques :

$$\mu(S) = \mu_{max} \frac{S}{K_s + S + \frac{S^2}{K_i}} \quad (4.9)$$

avec  $K_i$  la constante d'inhibition (g/l).

Il convient de noter que de nombreuses autres relations algébriques ont été établies pour décrire ces cinétiques. Certains modèles prennent en compte l'influence de la concentration en micro-organismes, en co-métabolite, de la température du pH, etc. (voir [20, 76, 48, 77] pour une liste d'une cinquantaine de modèles).

### 4.3.3 Structure des cinétiques biologiques

En examinant les cinétiques habituellement utilisées par les modélisateurs, il apparaît qu'une grande majorité des fonctions retenues sont de la forme :

$$r(\xi) = \frac{\prod_{\xi_j \in \mathcal{N}} \xi_j}{\psi(\xi)} \quad (4.10)$$

Cette équation signifie que la vitesse de réaction est proportionnelle aux concentrations de l'ensemble  $\mathcal{N}$  des réactants  $\xi_j$  (incluant la biomasse ou les catalyseurs) nécessaires à la réaction.

Remarquons que dans la plupart des modèles la fonction  $\psi(\xi)$  est **positive** et **strictement croissante** [71, 52]. En effet, les fonctions  $\psi$  sont souvent le produit de termes du type :

- $(K + S)$  pour les cinétiques de type Michaelis-Menten
- $(K_1 + S + \frac{S^2}{K_2})$  pour les cinétiques de type Haldane ( inhibition par une forte concentration de  $S$ )
- $(K + P)$  inhibition hyperbolique par le produit  $P$
- $e^{kP}$  inhibition exponentielle par le produit  $P$
- $(K + X)$  limitation de la croissance de la biomasse  $X$  par surpeuplement
- ...

Nous utiliserons cette propriété pour tacher d'identifier  $\psi(\xi)$  à l'aide de jeux de données.

## 4.4 Représentation des cinétiques au moyen de réseaux de neurones

De manière alternative, on peut modéliser les cinétiques en utilisant un réseau de neurones (modèles hybrides). On ne fait pas d'hypothèse *a priori* sur la cinétique (ou simplement les contraintes qui garantiront que les trajectoires du système gardent un sens biologique acceptable) et on identifie directement les cinétiques au cours de l'apprentissage du réseau.

Néanmoins, il faut décider des variables qui influencent les cinétiques et qui seront les entrées du réseau de neurones.

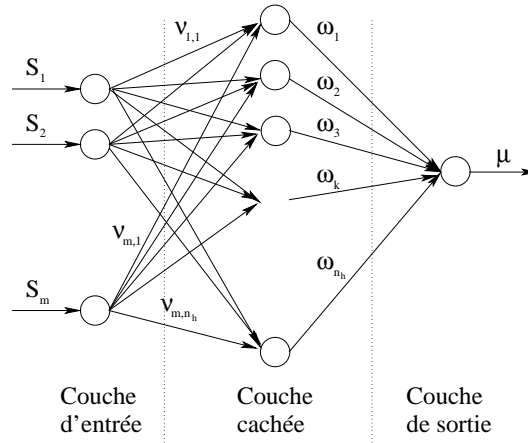


FIG. 4.1 – Schéma d'un réseau de neurones "feedforward" incluant une couche cachée.

Une vue schématique d'un réseau "feedforward" est représentée figure 4.1 pour une seule couche cachée. L'expression de la sortie du réseau en fonction des entrées s'exprime de la manière suivante :

$$\mu(S_1, \dots, S_m) = \sum_{k=1}^{n_h} \omega_k \phi\left(\sum_{i=1}^m v_{ki} S_i\right) \quad (4.11)$$

où  $n_h$  représente le nombre de neurones dans la couche cachée. Les  $\omega_k$  et les  $v_{ki}$  sont respectivement les poids des couches d'entrée et de sortie. La fonction  $\phi$  représente la fonction d'activation associée à chaque neurone (sigmoïde, tangente hyperbolique, gaussienne, etc).

Le choix du type de réseau et du nombre de neurones est une question relativement classique et le lecteur pourra se référer à [78] pour davantage de précisions.

Notons qu'il est souvent avantageux de ne pas utiliser un réseau de neurones pour modéliser une vitesse de réaction  $r_i(\xi)$ , mais pour modéliser le terme  $\nu_{ij}(\xi)$  dans l'équation 4.1. Ceci garantit, non seulement le respect de la conservation de la masse, mais aussi de la bornitude et de la positivité des variables. On se référera à [79, 70] pour davantage de détails ainsi que pour un exemple de mise en place de réseau de neurones dans le cadre de la digestion anaérobie.

Une fois la structure du réseau choisie, il faut ensuite identifier les poids du réseau au cours de la phase d'apprentissage. Pour davantage de précision, se référer à [69, 79, 70].

Cette approche revient en fait à transformer le problème de modélisation des fonctions cinétiques (ou plus précisément des  $\nu_{ij}(\xi)$ ) en un problème d'identification paramétrique des poids du réseau.

## 4.5 Estimation qualitative directe des cinétiques réactionnelles

### 4.5.1 Introduction

L'objectif de ce chapitre est de déterminer des structures possibles pour les cinétiques réactionnelles. Nous nous fonderons sur un jeu de données disponibles et nous utiliserons des arguments de nature **qualitative** pour déterminer d'une part les variables qui influencent la réaction (en plus de l'ensemble des réactants

nécessaires à la réactions) ainsi que leur effet sur la réaction. En d'autres termes, en nous appuyant sur la remarque du paragraphe 4.3.3, nous tâcherons de déterminer **qualitativement**  $\psi$ .

Avant de détailler la démarche, nous allons expliciter le mode de calcul des taux de réaction au cours du temps (cf. [69]).

### 4.5.2 Calcul des taux de réactions

Dans le cadre des hypothèses que nous avons formulées au paragraphe 3.5.2, nous avons la propriété triviale suivante :

**Proposition 1** *Considérons le système (3.8) :*

$$\dot{\xi} = Kr(\cdot) + D(\xi_{in} - \xi) - Q(\xi) \quad (4.12)$$

où  $K$  est une matrice de plein rang et supposons qu'il y a plus de variables que de réactions ( $n > p$ ), il existe une  $p \times n$  matrice  $L$  telle que

$$LK = I_{p \times p}$$

En conséquence, en posant  $\chi = L\xi$  et  $\chi_{in} = L\xi_{in}$ , nous avons :

$$\frac{d\chi}{dt} = r(\cdot) + D(\chi_{in} - \chi) - LQ(\xi) \quad (4.13)$$

Rappelons qu'à ce stade la matrice  $K$  est connue et ses coefficients identifiés (on se référera au paragraphe 5.2.6 pour davantage de détails sur l'identification). Nous pouvons donc calculer la matrice  $L$  et donc les  $p$  variables  $\chi_i$ .

En lissant les données ainsi obtenues (par exemple à l'aide de fonctions splines), on pourra estimer  $\frac{d\chi}{dt}$  et par suite évaluer la valeur de  $r(\cdot)$  à différents instants associés à différentes valeurs de  $\xi$ . Pour un exemple d'application de cette méthode le lecteur se référera à [51].

Notons enfin que la matrice  $L$  n'est pas définie de manière unique. Il conviendra de choisir la matrice  $L$  la moins sensible aux incertitudes sur les coefficients de  $K$ .

L'issue de ce calcul nous amènera à un ensemble de points de données obtenus à différents instants  $t_i$ , sous la forme de couples  $(\xi(t_i), r(t_i))$ .

### 4.5.3 Définition d'un critère structurel

A partir des différentes mesures  $(\xi(t_i), r(t_i))$ , nous pouvons estimer les valeurs de  $\psi$  définies par l'équation (4.10) :

$$\psi(t_i) = \frac{\prod_{\xi_j \in \mathcal{N}} \xi_j(t_i)}{r(t_i)} \quad (4.14)$$

Nous supposons donc qu'un ensemble  $\mathcal{D}$  de couples de données est disponible, sous la forme :

$$\mathcal{D} = \{d^i = (\xi^i, \psi^i) \quad i = 1, \dots, N\}$$

où  $\xi^i$  et  $\psi^i$  représentent respectivement les estimées de  $\xi$  et de  $\psi(\xi)$  au temps  $t_i$ .

L'objectif de ce paragraphe va être de proposer un critère pour identifier des fonctions  $\psi$ , croissantes et de structure minimale. En effet, nous avons vu que la grande majorité des modèles classiques de cinétiques réactionnelles mettent en jeu une fonction  $\psi(\xi)$  croissante (voir paragraphe 4.3.3). Plus précisément, nous proposons une méthode qui permet de tester si  $\psi(\xi)$  peut s'écrire sous la forme suivante :

$$\psi(\xi) = (\xi_i + a)\phi(\tilde{\xi})$$

où  $\tilde{\xi}$  représente un sous vecteur extrait de  $\xi$ , et  $\phi(\tilde{\xi})$  est elle même une fonction strictement croissante.

Pour simplifier les notations, nous noterons  $x = \xi_i \in \mathbf{R}$  la composante de  $\xi$  qui peut se mettre en facteur (après ajout de la quantité  $a$ ). Nous noterons  $y = \tilde{\xi} \in \mathbf{R}^p$  ( $p \leq n - 1$ ) le vecteur des autres variables d'état dont dépend la fonction  $\psi$ .

Nous avons bien évidemment la même partition pour les mesures, avec les notations correspondantes  $(x^i, y^j)$ ,  $i = 1, \dots, N$ .

Pour tout réel  $\lambda > 0$  et tout entier  $i \in \{1, \dots, N\}$ , nous définissons le sous-ensemble  $\mathcal{S}_i(\lambda)$  de  $\mathcal{D}$  associé à la mesure  $d^i$  :

$$\mathcal{S}_i(\lambda) = \left\{ d^j \in \mathcal{D} : y^j < y^i, \frac{\psi^j}{x^j + \lambda} \geq \frac{\psi^i}{x^i + \lambda} \right\} \quad (4.15)$$

où l'inégalité  $y^i < y^j$  signifie que  $y^j - y^i$  est un vecteur strictement positif de  $\mathbf{R}^p$ .

**Principe de l'approche :** supposons que  $\psi(\xi) = \psi(x, y) = (x + a)\phi(y)$  où  $\phi$  est strictement croissante. Alors il est clair que l'ensemble  $\mathcal{S}_i(a) = \emptyset$ , puisque  $\frac{\psi}{x+a} = \phi(y)$  est une fonction strictement croissante de  $y$ . L'idée que nous allons développer consiste donc à balayer l'ensemble  $\lambda \in \mathbf{R}_+$ , et à observer comment varie  $\text{card}(\mathcal{S}_i(\lambda))$ .

Nous allons donc construire à partir de  $\mathcal{S}_i(\lambda)$  les critères structurels suivants, calculés sur l'ensemble  $\mathcal{D}$  :

$$\mathcal{I}_i(\lambda) = \text{card}(\mathcal{S}_i(\lambda)) \quad (4.16)$$

$$\mathcal{J}(\lambda) = \sum_{i=1}^N \mathcal{I}_i(\lambda) \quad (4.17)$$

Comme nous le montrerons ultérieurement (voir par exemple la figure 4.4), les fonctions  $\mathcal{I}_i(\lambda)$  et  $\mathcal{J}(\lambda)$  sont continues par morceaux. L'ensemble des discontinuités de  $\mathcal{I}_i(\lambda)$  peut-être calculé comme suit :

$$\Lambda_i = \left\{ 0 < \sigma_{ij} < +\infty : \sigma_{ij} = \frac{\psi^j x^i - \psi^i x^j}{\psi^i - \psi^j}; j = 1, N \right\} \quad (4.18)$$

Pour  $\mathcal{J}(\lambda)$ , cet ensemble est la réunion des  $\Lambda_i$  :

$$\Lambda = \left\{ \sigma_{ij} > 0 : \sigma_{ij} = \frac{\psi^j x^i - \psi^i x^j}{\psi^i - \psi^j}; i = 1, N; j > i \right\} \quad (4.19)$$

Cet ensemble de discontinuités a un nombre fini d'éléments qui sont classés en ordre croissant et dénommés (par abus de notation)  $\sigma_q$  (où  $q \in \mathbf{N}$  est un indice de classement) :  $\sigma_{q+1} > \sigma_q$ .

**Definition 1** Nous dirons que l'application  $\phi$  est strictement croissante sur  $\mathcal{D}$  si pour tout couple  $(d^i, d^j) \in \mathcal{D}^2$ ,  $y^j < y^i \implies \phi(y^j) < \phi(y^i)$

En d'autres termes, nous supposons que  $\phi$  est croissante à partir de la connaissance de l'ensemble des  $\mathcal{D}$ .

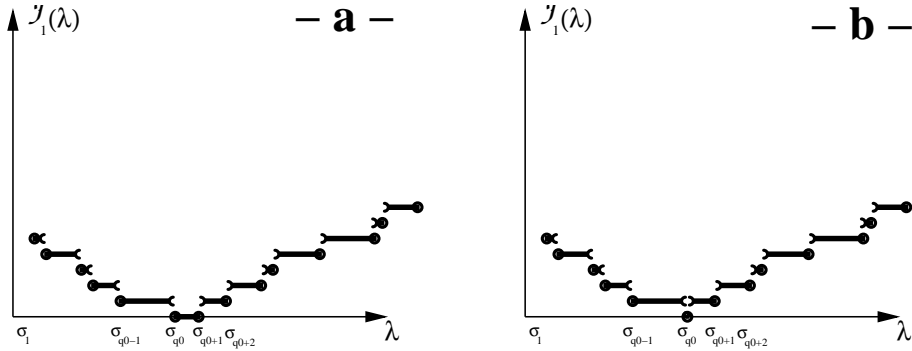


FIG. 4.2 – Forme du critère  $\mathcal{J}$  lorsque  $\psi(\xi) = \psi(x, y) = (x + a)\phi(y)$  pour les 2 cas possibles, a)  $a$  appartient à  $\Lambda$  b)  $a$  n'appartient pas à  $\Lambda$ .

**Théorème 1** Il existe un réel  $a > 0$  tel que  $\psi(x, y)$  peut être écrit pour tout  $d^i \in \mathcal{D}$  :

$$\psi^i = \psi(x^i, y^i) = (x^i + a)\phi(y^i) \quad (4.20)$$

où  $\phi$  est croissant sur  $\mathcal{D}$ , si et seulement si  $\mathcal{J}(\lambda)$  est une application constante par morceaux sur chaque intervalle  $]\sigma_q, \sigma_{q+1}[$  satisfaisant l'une des assertions suivantes :

- i. Il existe  $a > 0$  tel que  $\mathcal{J}(a) = 0$  et si  $a \notin \Lambda$ ,  $\mathcal{J}(\lambda) = 0 \quad \forall \lambda \in [\sigma_{q0}, \sigma_{q0+1}]$ , avec :  $\sigma_{q0} < a < \sigma_{q0+1}$
  - ii. Il existe  $a > 0$  tel que pour tous les  $\sigma_q \in \Lambda$  (excepté  $\sigma_q = a$  si  $a \in \Lambda$ ) :
    - si  $0 < \lambda < a$  :  $\lim_{\lambda \rightarrow \sigma_q^-} \mathcal{J}(\lambda) = \lim_{\lambda \rightarrow \sigma_q^+} \mathcal{J}(\lambda) + 1$
    - si  $\lambda > a$  :  $\lim_{\lambda \rightarrow \sigma_q^+} \mathcal{J}(\lambda) = \lim_{\lambda \rightarrow \sigma_q^-} \mathcal{J}(\lambda) + 1$
- Et, de plus,  $d^j \notin \mathcal{S}_i(\lambda)$  si  $\sigma_{ij} < 0$  ou si  $\sigma_{ij} = +\infty$ .

**Preuve :** voir [80].

Remarquons que cela signifie que  $\mathcal{J}$  est décroissant avant  $]\sigma_{q0}, \sigma_{q0+1}[$  et croissant après, et vaut zéro sur un intervalle (qui peut être réduit au point  $a$ ) contenant  $a$  (cf figure 4.2).

**Remarque :** La valeur de  $a$  n'est pas unique si  $a \notin \Lambda$  ; tout  $a \in [\sigma_{q0}, \sigma_{q0+1}]$  satisfait cette propriété.

Ce théorème va donc permettre de tester si, pour un choix de  $x$  et de  $y$ , la fonction  $\psi$  peut s'écrire sous la forme (4.20). En effet, si c'est le cas, le calcul de  $\mathcal{J}(\lambda)$  doit conduire à un critère en fonction de  $\lambda$  de la forme présentée sur la figure 4.2.

Lorsque  $\psi(x, y)$  ne dépend en fait pas de  $x$  (autrement dit  $a = +\infty$ ), le théorème suivant montre que  $\mathcal{J}(\lambda)$  est décroissant :

**Théorème 2**  $\psi(x, y)$  peut être écrit comme une fonction de  $y$  (indépendant de  $x$ ) :

$$\psi(x, y) = \phi(y)$$

où  $\phi(y)$  est une fonction positive strictement croissante sur  $\mathcal{D}$ , si et seulement si, pour tout  $\sigma_q \in \Lambda$

$$\lim_{\lambda \rightarrow \sigma_q^-} \mathcal{J}(\lambda) = \lim_{\lambda \rightarrow \sigma_q^+} \mathcal{J}(\lambda) + 1$$

*c-à-d que  $\mathcal{J}$  est décroissante.*

**Preuve :** voir [80].

#### 4.5.4 Utilisation du critère structurel

Comme nous le verrons dans l'exemple, le critère structurel va permettre de tester la vraisemblance de différentes structures. L'idée va donc consister à tester automatiquement toutes les structures possibles (au moins parmi un sous ensemble vraisemblable *a priori*).

Après qu'une structure  $\psi(x, y)$  a été retenue, le même raisonnement peut être appliqué à  $\frac{\phi(\bar{x}, \bar{y})}{\bar{x}+a}$  pour déterminer par récurrence la structure complète de  $r(\xi)$ . Une fois que nous nous sommes ramenés à un sous-ensemble  $y$  de  $\xi$  qui est réel, il devient alors possible d'estimer un modèle simple pour  $\phi(y)$ . L'exemple suivant éclairera ces explications.

Ce critère est testé dans [71] pour des simulations de la croissance d'une biomasse bactérienne  $X$ , consommant un substrat  $S$  et produisant un métabolite  $P$  (voir figure 4.3).

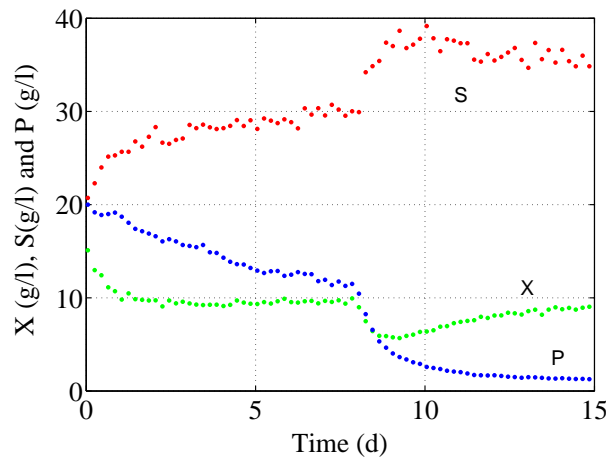


FIG. 4.3 – Expérience simulée à partir d'un modèle (supposé inconnu) rééchantillonnée à 5 points par jour et perturbée par un bruit blanc multiplicatif.

Deux réactions interviennent dans ce système, et l'on s'intéresse ici à la cinétique de la première réaction qui s'écrit nécessairement

$$r_1(\xi) = \frac{SX}{\psi_1(\xi)} \quad (4.21)$$

On souhaite tester les hypothèses suivantes sur la structure de  $\psi_1$  :

$$(H1) \quad \psi_1 \stackrel{?}{=} (S + K_S)\phi_1(X)$$

$$(H2) \quad \psi_1 \stackrel{?}{=} (S + K_S)\phi_1(P)$$

$$(H3) \quad \psi_1 \stackrel{?}{=} (P + K_P)\phi_1(S)$$

$$(H4) \quad \psi_1 \stackrel{?}{=} (B + K_B)\phi_1(S)$$

où  $\phi_1$  est une fonction strictement croissante (l'hypothèse (H1) a servi à générer les données).

On peut construire les critères  $\mathcal{J}_i$  associés aux hypothèses (Hi). Les résultats obtenus sont présentés sur les figures 4.4 et 4.5.

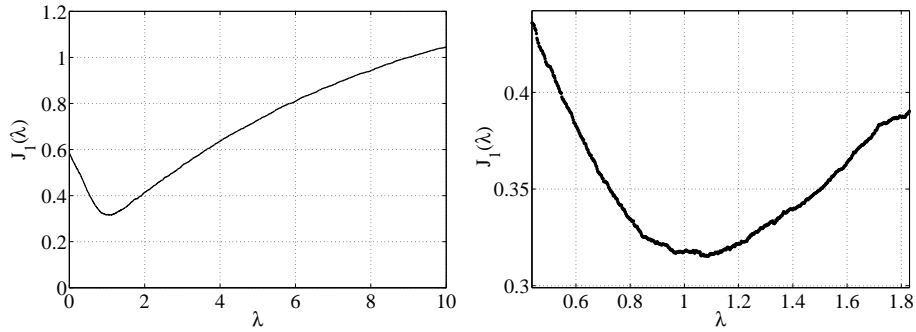


FIG. 4.4 – Application du critère structurel  $\mathcal{J}_1(\lambda)$  pour tester la structure de la cinétique réactionnelle  $r_1$ . La figure de droite est un zoom de  $\tilde{\mathcal{J}}_1(\lambda)$  autour de son minimum. Le critère est normalisé par le nombre de points discriminants.

Il est intéressant de noter que l'ajout de bruit perturbe quantitativement le critère, mais pas qualitativement. En effet, la fonction  $\mathcal{J}_1(\lambda)$  conserve un minimum global (Figure 4.4), toutefois il existe d'autres minima et le minimum global ne correspond plus à un critère nul.

Il est assez surprenant de voir que les critères  $\mathcal{J}_2(\lambda)$ ,  $\mathcal{J}_3(\lambda)$  et  $\mathcal{J}_4(\lambda)$  présentent également un minimum global (figure 4.5). Le nombre de points qui contredisent l'hypothèse (H1) est toutefois 10 fois inférieur au nombre de points qui contredisent les hypothèses (H2), (H3) et (H4). Davantage de travail permettrait d'affiner ces critères en les associant à des tests statistiques de monotonie [81].

L'exemple démontre aussi que, malgré l'aspect fondamentalement qualitatif de ce critère, il permet d'estimer la valeur du paramètre  $a$  avec une très bonne précision.

Cette piste mérite d'être creusée et d'être testée en vraie grandeur avec des données expérimentales.

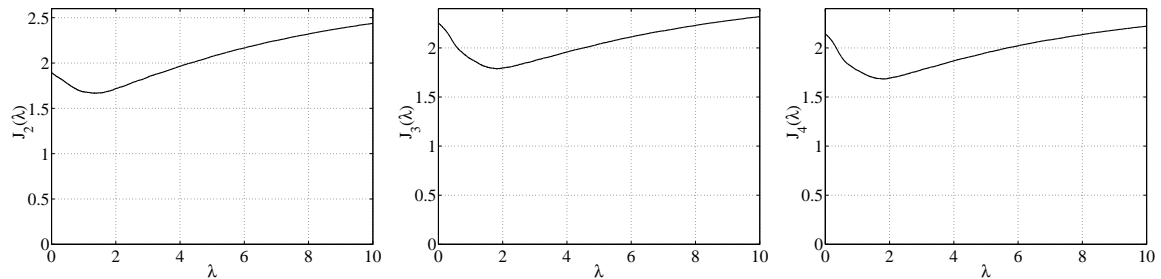


FIG. 4.5 – Critères structurels  $\mathcal{J}_2(\lambda)$ ,  $\mathcal{J}_3(\lambda)$  et  $\mathcal{J}_4(\lambda)$  pour tester les respectivement les hypothèses (H2), (H3) et (H4). Les critères sont normalisés par le nombre de points discriminants.



## 4.6 Conclusion sur la modélisation des cinétiques biologiques

Il ressort de ce chapitre que la représentation des cinétiques réactionnelles est un sujet épineux. C'est tout particulièrement à ce niveau que réapparaît la variabilité caractéristique de la biologie.

Néanmoins, il reste possible d'ordonner les hypothèses sur la structure des cinétiques car certaines contraintes sont nécessaires pour garantir que le modèle conserve des propriétés physiques. Au delà de ces hypothèses relativement fiables, les autres conjonctures nécessaires pour finaliser le modèle sont extrêmement difficiles à vérifier : l'étape de validation jouera donc un rôle fondamental.



## Chapitre 5

# Validation du modèle

*”De ma cellule, il m’est difficile de dire comment est fait ce château d’If où depuis tant d’années je me trouve prisonnier (...). Les images que Faria et moi nous faisons de la forteresse deviennent de plus en plus dissemblables : parti d’une figure simple, Faria la complique à l’extrême afin d’y inclure chacun des détails imprévus qu’il rencontre sur son chemin ; moi, partant du désordre de ces données, je vois en chaque obstacle pris à part l’indice d’un système d’obstacles, je développe chaque segment selon une figure régulière, je mets ensemble ces figures en qualité de faces d’un même solide, polyèdre ou hyper-polyèdre, j’inscris ces polyèdres dans des sphères ou des hyper-sphères, et ainsi dans la mesure même où je fais le tour de la forteresse je la simplifie, j’en donne une définition sous l’espèce d’un rapport numérique ou encore dans une formule algébrique.*

*Mais pour penser de cette façon une forteresse j’ai besoin que l’abbé Faria ne cesse pas de se battre avec les éboulements de terre, boulons en acier, écoulements d’égouts, guérites de sentinelles, sauts dans le vide, angles rentrants des murs de fondation, parce que la seule façon de renforcer la forteresse pensée est de la mettre continuellement à l’épreuve de la véritable.”*

Alexandre Dumas, Le Comte de Monte-Cristo. 1845-1846.

### 5.1 Introduction

La validation est la dernière phase de la modélisation. C’est sans doute la phase la plus importante, mais c’est aussi celle qui est le plus souvent négligée. Elle est d’autant plus cruciale que nous avons vu qu’il était nécessaire pour élaborer un modèle de faire un grand nombre d’hypothèses, souvent spéculatives. Avant d’utiliser le modèle, il conviendra donc de le valider correctement. Cette phase fait normalement suite à la

phase d'identification que nous n'aborderons que très partiellement et indirectement dans les paragraphes 4.5.3 et 5.2.6. Pour davantage de détails sur les aspects de l'identification paramétrique, le lecteur consultera [82].

L'objectif général de la validation est de vérifier que le modèle répond bien aux objectifs qui lui ont été fixés. Plus précisément, nous allons voir comment tester **isolément** les différentes hypothèses qui ont pu être supposées au cours de la construction du modèle. Pour cela nous procéderons en 3 étapes, au cours desquelles nous testerons :

- Le réseau réactionnel
- Les prédictions qualitatives du modèle
- Le modèle dans sa globalité (réseau réactionnel+cinétiques+paramètres)

La validation doit se fonder sur un jeu de données n'ayant pas servi à construire ou à identifier le modèle et correspondant à des conditions expérimentales différentes (sinon on teste plutôt la reproductibilité expérimentale). Si ces hypothèses ne sont pas satisfaites, le modèle ne peut pas prétendre à la validation.

Insistons sur le fait qu'un modèle ne peut jamais être validé. Il ne peut qu'être "falsifié", au sens où des données peuvent se révéler en contradiction avec le modèle. Néanmoins un modèle est dit "validé" s'il n'est pas "falsifié" par un jeu de données incluant un large spectre de conditions expérimentales. Dans certains cas, nous pourrions même être plus précis, et proposer un test de validité, associé à un risque  $0 < \alpha < 1$  que cette hypothèse soit fausse.

Ce chapitre qui se situe au coeur de mes travaux de recherches était au stade embryonnaire dans [44]. Il reprend tout d'abord un article sur la validation du réseau réactionnel [80] (paragraphe 5.2). L'exemple illustrant la validation des propriétés asymptotiques du modèle (paragraphe 5.3) provient de [83]. L'étude des propriétés qualitatives dynamiques (paragraphe 5.4) est le fruit d'une longue série de travaux [84, 85, 86, 87, 21, 88]. Je mentionne en toute fin un travail plus récent sur la validation semi-quantitative des modèles [89, 90] (paragraphe 5.5.1).

## 5.2 Validation d'un réseau réactionnel

### 5.2.1 Position du problème

Dans ce chapitre nous verrons comment comparer un réseau réactionnel, par l'intermédiaire de sa matrice pseudo-stoechiométrique  $K$ , à un ensemble de données afin de le valider. La difficulté tient ici à ce que certains coefficients pseudo-stoechiométriques (et même tous!) peuvent être inconnus.

Nous proposons une méthode qui permettra en même temps de tester la validité du réseau réactionnel (sous un seuil  $\alpha$  fixé à l'avance) et d'identifier les coefficients pseudo-stoechiométriques. L'idée consiste à vérifier si les vecteurs du noyau à gauche de  $K$  vérifient bien certaines relations. Cette méthode repose sur les mêmes fondements que la démarche présentée au paragraphe 3.5. Elle peut en particulier être appliquée pour valider les résultats obtenus lors de la phase d'identification du réseau réactionnel.

### 5.2.2 Détermination du noyau à gauche de la matrice pseudo-stoechiométrique

Considérons un vecteur  $\lambda \in \text{Ker } K^T$ , cela signifie :

$$\lambda^T K = 0$$

nous choisirons  $\lambda$  tel que sa composante  $i_0$  soit unitaire :  $\lambda_{i_0} = 1$

Nous reprendrons ici les notations introduites au paragraphe 3.5.3. Considérons la quantité scalaire  $\lambda^T u(t)$ , d'après l'équation (3.30), elle satisfait à chaque instant  $t$  :

$$\lambda^T u(t) = 0$$

En d'autres termes, nous avons :

$$u_{i_0}(t) = - \sum_{j \neq i_0} \lambda_j u_j(t) \quad (5.1)$$

Cela signifie que les  $u_j$  sont liés par une relation linéaire. L'idée immédiate que l'on pourrait avoir serait de vérifier si la relation (5.1) est en adéquation avec les données. Ceci peut être fait en effectuant une régression linéaire entre  $u_{i_0}$  et les  $u_j$ .

**Remarque 1 :** naturellement la même relation (5.1) est également satisfaite pour  $\delta u_j(t)$ , ou pour n'importe quelle transformation linéaire  $u(t)$  qui utilise des intégrations, différentiations, retards, etc, cf. paragraphe 3.5.3.

**Remarque 2 :** Les systèmes basés sur un bilan de matière fermé (tous les transferts de masse sont représentés), satisfont la propriété importante qu'il existe un vecteur positif  $\omega$  dans  $\text{Ker } K^T$  [91]. En fait, les relations linéaires de type (5.1) traduisent de manière générale la conservation de certaines quantités au sein du système. Cette notion est liée aux invariants réactionnels [47].

Nous devons toutefois garder à l'esprit que les  $u_j$  sont *a priori* non indépendants, puisqu'ils peuvent être reliés par d'autres relations liés à d'autres vecteurs du noyau à gauche de  $K$ . En particulier, nous avons vu que  $\text{rang}(U) = p$  et une régression (5.1) ne saurait donc impliquer plus de  $p + 1$  variables  $u_j$  indépendantes.

Nous choisirons donc les vecteurs  $\lambda$  du noyau à gauche tels qu'il seront associés à des  $u_j$  indépendants.

**Definition 2** Nous dirons qu'un ensemble de  $k$  données transformées  $\{u_{i_1} \dots u_{i_k}\}$  est associé à un vecteur du noyau à gauche  $\lambda$  si  $\lambda_j = 0$  pour tous les indices  $j \notin \{i_1 \dots i_k\}$ . Nous dirons que  $\lambda$  est associé à la  $k \times p$  matrice  $\tilde{K}$ , qui est la sous-matrice constituée des lignes  $i_1$  à  $i_k$  de  $K$ . Enfin nous appelons  $\tilde{\lambda}$  le vecteur réduit obtenu en enlevant tous les éléments nuls dans  $\lambda$ .

Nous avons donc  $\tilde{\lambda}^T \tilde{K} = 0$  et  $\tilde{\lambda}$  n'a aucune composante nulle. Par conséquent,  $\sum_{i_k} \lambda_{i_k} u_{i_k}(t) = 0$ .

**Definition 3** Nous notons  $d(\lambda) = \dim(\tilde{\lambda})$  que nous appelons la dimension de régression associée à  $\lambda$ . Le nombre de composantes inconnues de  $\lambda$  sera noté  $d_u(\lambda)$

$d(\lambda)$  correspond donc au nombre d'éléments non nuls dans  $\lambda$ .

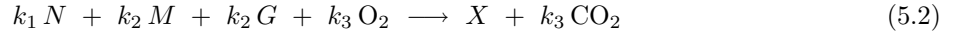
**Definition 4** nous dirons qu'un vecteur propre à gauche  $\lambda$  est **sain** si sa  $d(\lambda) \times n_r$  matrice associée  $\tilde{K}$  ne contient que des sous matrices  $k \times n_r$  (avec  $k < d(\lambda)$ ) de plein rang ou — de manière équivalente — si  $\dim(\text{Ker } \tilde{K}) = 1$ .

**Propriété 4** Le vecteur  $\lambda$  associé à un ensemble de données transformées  $u_i$  est sain si et seulement si les  $u_i$  ne sont reliés par aucune autre relation que celle définie par  $\lambda$ .

**Preuves :** voir [80]

### 5.2.3 Exemple : bioproduction de vanilline

La croissance aérobie d'une biomasse de champignon filamenteux ( $X$ ) à partir d'une source de carbone et d'une source d'azote ( $N$ ) est représentée par le réseau réactionnel suivant :



où  $k_1$ ,  $k_2$  et  $k_3$  représentent les coefficients pseudo-stoechiométriques associés respectivement à la consommation d'azote, de sucres et à la respiration de l'oxygène. La source de carbone se compose des sucres (glucose  $G$  et maltose  $M$ ) et la source d'azote est l'ammonium. Il a été établi [66, 67] que la réaction de conversion de l'acide vanillique ( $A$ ) en vanilline ( $V$ ) par *P. cinnabarinus* se déroule généralement en compétition avec deux autres réactions : une oxydation de l'acide vanillique en methoxyhydroquinone ( $H$ ) et une réduction de la vanilline en alcool vanillique ( $L$ ).

Le réseau réactionnel est alors le suivant :



La première réaction inclut le glucose ( $G$ ) comme source de carbone (avec un coefficient pseudo-stoechiométrique  $k_4$ ). Finalement, nous considérons une réaction associée à la mortalité du champignon :



où  $X_d$  représente la biomasse morte.

$$\xi = \begin{pmatrix} X \\ X_d \\ M \\ G \\ N \\ A \\ V \\ H \\ L \end{pmatrix}, \quad K = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -k_2 & 0 & 0 & 0 & 0 \\ -k_2 & 0 & -k_4 & 0 & 0 \\ -k_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (5.7)$$

$$r = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{pmatrix}, \quad \xi_{in} = \begin{pmatrix} 0 \\ 0 \\ M_{in} \\ G_{in} \\ N_{in} \\ A_{in} \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (5.8)$$

On vérifiera aisément que les vecteurs suivants sont des vecteurs sains du noyau à gauche de  $K$  :

$$\begin{aligned}\lambda_1 &= (k_2, k_2, 1, 0, 0, 0, 0, 0)^T \\ \lambda_2 &= (k_1, k_1, 0, 0, 1, 0, 0, 0)^T \\ \lambda_3 &= (0, 0, 0, 0, 0, 1, 1, 1)^T \\ \lambda_4 &= \left(0, 0, 0, 1, -\frac{k_2}{k_1}, 1, k_4 + 1, 1, k_4 + 1\right)^T\end{aligned}$$

Par exemple,  $\lambda_1$  est associé à la régression :

$$k_2 u_1(t) + k_2 u_2(t) + u_3(t) = 0$$

Notons aussi que la relation associée à  $\lambda_3$  correspond à la conservation de la masse des composés phénolés.

#### 5.2.4 Utilisation des données disponibles pour valider le noyau à gauche de $K$

**Propriété 5** Soit  $\lambda$  un vecteur sain du noyau de  $K^T$ , associé à  $\tilde{\lambda}$  et à un ensemble de  $d(\lambda)$  données transformées  $u_{i_j}$  ( $i_j \in \{i_1..i_{d(\lambda)}\}$ ). Notons  $\mathcal{S}$  l'ensemble des indices  $j$  tels que  $\tilde{\lambda}_j$  est connu. Alors le critère de coût suivant :

$$J(\alpha) = \sum_{t=t_1}^{t_N} \left( \sum_{j \in \mathcal{S}} \tilde{\lambda}_j u_{i_j}(t) - \sum_{j \notin \mathcal{S}} \alpha_j u_{i_j}(t) \right)^2 \quad (5.9)$$

admet un unique minimum, de valeur nulle, obtenu pour  $\alpha_j = \tilde{\lambda}_j$  (pour tous les  $j \notin \mathcal{S}$ ).

Notons que minimiser  $J(\alpha)$  est exactement un problème de régression **linéaire**. La validation consistera donc à vérifier que  $J(\alpha)$  peut être correctement minimisé, ou en d'autres termes, que la régression entre  $v = \sum_{j \in \mathcal{S}} \tilde{\lambda}_j u_{i_j}$  et les  $u_{i_j}$  (pour tous les  $j \notin \mathcal{S}$ ) est significative (au seuil  $\alpha$  fixé). Cette analyse doit être exécutée sur tous les vecteurs sains du noyau de  $K^T$ .

#### 5.2.5 Identifiabilité des coefficients pseudo-stoechiométriques

Nous allons examiner dans ce paragraphe si les valeurs de  $\lambda_i$  identifiées à l'aide de l'ensemble des régressions données par les équations (5.1) permettent de déterminer tous les coefficients pseudo-stoechiométriques  $k_i$ .

Cette propriété d'identifiabilité quand les taux de réaction  $r(\zeta)$  sont inconnus est appelée C-identifiabilité dans [92], qui proposent un test à partir d'une condition nécessaire et suffisante.

Une version de ce théorème est proposée ici dans le cadre simplifié des matrices  $K$  de plein rang :

**Théorème 3 (Chen et Bastin 1995)** Soit  $K$  une  $n \times p$  matrice de plein rang, avec  $n > p$ . La  $j$  ième colonne de  $K$  est dite C-identifiable si et seulement si il existe une partition non singulière  $(K_a K_b)$ , où  $K_a$  est une sous-matrice  $p \times p$  de plein rang qui ne contient aucun élément inconnu dans sa  $j$  ième colonne.

Nous proposons ici une condition nécessaire plus précise sur la C-identifiabilité de chaque élément de la matrice :

**Théorème 4** Soit  $K$  une  $n \times p$  matrice de plein rang, avec  $n > p$ . L'élément (inconnu)  $k_{ij}$  de  $K$  est C-identifiable s'il existe une sous-matrice  $K_a$  de plein rang  $k \times p$ , avec  $k \leq p$ , qui ne contient aucun élément inconnu sur sa  $i$  ième colonne, telle que la  $(k + 1) \times p$  sous-matrice de  $K$  :

$$\Xi = \begin{pmatrix} K_a \\ K_{bj} \end{pmatrix}$$

vérifie  $\text{rang}(\Xi) < k + 1$ , où  $K_{bj}$  est la  $j$  ième ligne de  $K$ .

**Preuve :** voir [80]

**Remarque :** Ce critère, bien qu'il soit plus compliqué que celui proposé par [92], caractérise la C-identifiabilité pour chaque élément de  $K$  et non pas uniquement pour les colonnes entières.

**Exemple :** Considérons la matrice  $K$  suivante :

$$K = \begin{pmatrix} k_{11} & 1 \\ 1 & 0 \\ k_{31} & 0 \end{pmatrix} \quad (5.10)$$

Le théorème 3 conclut que la première colonne de  $K$  n'est pas C-identifiable, puisqu'il n'est pas possible de trouver une  $2 \times 2$  sous-matrice  $K_a$  qui ne contienne aucun élément inconnu dans sa première colonne. Maintenant si nous appliquons le théorème 4, nous pouvons considérer les sous-matrices suivantes :

$$K_a = \begin{pmatrix} 1 & 0 \end{pmatrix}, K_b = \begin{pmatrix} k_{31} & 0 \end{pmatrix}$$

alors  $\Xi = \begin{pmatrix} 1 & 0 \\ k_{31} & 0 \end{pmatrix}$  est de rang 1 et vérifie la condition  $k + 1 = 2 > \text{rang}(\Xi)$ , il s'en suit que  $k_{31}$  est C-identifiable. Il est maintenant clair que  $k_{11}$  n'est pas C-identifiable, autrement la première colonne de  $K$  serait C-identifiable.

Remarquons que l'analyse du noyau de la matrice  $K^T$  fournit également un critère pour examiner l'identifiabilité de  $k_{ij}$ . Même si ce critère est moins commode, il donne des idées sur l'identifiabilité pratique, comme nous le verrons dans la propriété suivante.

**Définition 5** Une base constituée des  $p$  premiers vecteurs sains indépendants  $\lambda_i$  sera appelée base saine du noyau.

**Propriété 6** les coefficients pseudo-stoechiométriques  $k_{ij}$  sont C-identifiables si et seulement s'il peuvent être calculés à l'aide de n'importe quelle combinaison des composantes des vecteurs  $\lambda_i$  qui forment une base saine du noyau

Dans l'exemple de l'équation (5.10), la base saine du noyau de  $K^T$  était :

$$\tilde{\lambda} = (0, -k_{31}, 1)^T$$

Il s'ensuit que  $k_{31}$  est C-identifiable et que  $k_{11}$  n'est pas C-identifiable.



### 5.2.6 Identification des coefficients pseudo-stoechiométriques et validation finale

Une fois que nous savons que les coefficients pseudo-stoechiométriques sont identifiables, nous pouvons estimer leur valeur à partir des données expérimentales. Pour ceci, nous emploierons la régression associée à la base saine du noyau  $K^T$  donnée par l'équation (5.1). La significativité statistique (au seuil  $\alpha$ ) de la corrélation permettra de tester si les vecteurs  $\tilde{\lambda}_i$  sont bien dans le noyau  $K^T$ . La validation finale consistera à contrôler que les coefficients ont le bon signe.

Ce test doit être réalisé en rapport avec l'incertitude sur  $\lambda_i$  obtenue par la régression linéaire (5.1). En effet, il se peut que  $k_i$  ait une valeur négative, mais avec un intervalle de confiance intersectant le domaine positif.

### 5.2.7 Comparaison entre plusieurs réseaux réactionnels

Ici nous considérerons le problème classique en modélisation, où l'on se demande quel est le meilleur réseau réactionnel parmi un ensemble de  $q$  réseaux *a priori* possibles.

Le but de cette section est de déterminer comment choisir parmi ces  $q$  hypothèses celles qui fournissent une matrice pseudo-stoechiométrique en accord avec l'ensemble de données. Parmi cet ensemble admissible, nous verrons comment déterminer le meilleur. Remarquons cependant que, dans la plupart des cas,  $q$  est un nombre petit puisqu'il y a seulement quelques réseaux réactionnels possibles. La méthode consiste donc à examiner chaque matrice  $K_i$  suivant la méthodologie exposée dans la section (5.2.6), puis à choisir le modèle présentant la meilleure adéquation. Naturellement il est souvent difficile de choisir entre 2 modèles qui présentent chacun de bons résultats de validation. D'autres critères peuvent alors intervenir, tels que l'identifiabilité de la matrice associée, ou sa simplicité.

### 5.2.8 Étude de cas réel

Nous étudierons des données de la croissance et de la biotransformation du champignon filamentueux *Pycnoporus cinnabarinus*.

Nous considérerons ici plus particulièrement la phase de croissance. La croissance aérobie de la biomasse de mycélium ( $X$ ) à partir d'une source de carbone (glucose  $G$  et maltose  $M$ ) et d'une source d'azote ( $N$ ) peut *a priori* être raisonnablement représentée par les 3 réseaux réactionnels :

- Réseau 1 : le champignon se développe sur le maltose, le glucose et l'azote ; il peut toutefois transformer le maltose en glucose au cours d'une première étape.



- Réseau 2 : le champignon se développe sur le glucose et l'azote ; il transforme le maltose en glucose au cours d'une première étape.



- Réseau 3 : le champignon se développe soit sur le glucose et l'azote soit sur le maltose et l'azote ; dans ce second cas du glucose est excrété.



Les matrices pseudo-stoechiométriques liées à (5.11), (5.12) et (5.13) sont alors respectivement :

$$K_1 = \begin{pmatrix} 0 & -k_1 \\ -1 & -k_3 \\ 2 & -k_2 \\ 0 & 1 \end{pmatrix}, K_2 = \begin{pmatrix} 0 & -k_1 \\ -1 & 0 \\ 2 & -k_2 \\ 0 & 1 \end{pmatrix}, K_3 = \begin{pmatrix} -k_1 & -k_4 \\ -k_2 & 0 \\ k_3 & -k_5 \\ 1 & 1 \end{pmatrix} \quad (5.14)$$

Les vecteurs sains des noyaux de ces 3 matrices pseudo-stoechiométriques (PS), sont présentés dans le tableau 5.1.

Matrice PS	Base de noyaux sains de $K^T$	Régressions	Équilibrage
$K_1$	$\lambda_1^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ k_1 \end{pmatrix}, \lambda_2^1 = \begin{pmatrix} 0 \\ -2 \\ 1 \\ -2k_3 + k_2 \end{pmatrix}$	$u_1 = -c_1^{1+} u_4$ $2u_2 + u_3 = c_3^1 u_4$	$\mathcal{B}(\lambda_1^1) = 1$ $\mathcal{B}(\lambda_2^1) = 1$
$K_2$	$\lambda_1^2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ k_1 \end{pmatrix}, \lambda_2^2 = \begin{pmatrix} 0 \\ 2 \\ 1 \\ k_2 \end{pmatrix}$	$u_1 = -c_1^{2+} u_4$ $2u_2 + u_3 = c_3^2 u_4$	$\mathcal{B}(\lambda_1^2) = 1$ $\mathcal{B}(\lambda_2^2) = 1$
$K_3$	$\lambda_1^3 = \begin{pmatrix} 0 \\ \frac{k_5 + k_3}{k_2} \\ 1 \\ k_5 \end{pmatrix}, \lambda_2^3 = \begin{pmatrix} 1 \\ \frac{k_4 - k_1}{k_2} \\ 0 \\ k_4 \end{pmatrix}$	$u_3 = -c_1^{3+} u_2 - c_2^{3+} u_4$ $u_1 = c_3^3 u_2 - c_4^{3+} u_4$	$\mathcal{B}(\lambda_1^3) = 4.48$ $\mathcal{B}(\lambda_2^3) = 6.29$

TAB. 5.1 – Vecteurs des noyaux et régressions associées aux matrices pseudo-stoechiométriques pour chacun des réseaux considérés pour la croissance de *Pycnoporus cinnabarinus* sur l'ammonium, le maltose et le glucose. Les réels  $c_i^{j+}$  sont positifs, les  $c_i^j$  peuvent être de signe quelconque.

Les régressions associées sont présentés dans le Tableau 5.2 pour les 3 matrices pseudo-stoechiométriques. Elles ont été calculées en utilisant 70 points de données issus de 9 expériences différentes. L'intervalle de confiance pour les paramètres a été estimé en utilisant une distribution de Student avec un seuil  $\alpha$  de 5% et la significativité de la régression a été examinée sous l'hypothèse  $H_0$  : les coefficients sont différents de zéro.

Dans ce cas, les réseaux 1 et 2 sont acceptables, on privilégiera alors le réseau 2 qui est plus simple et dont les coefficients pseudo-stoechiométriques sont tous identifiables.

Paramètre	min	max	Positivité	Significativité	Conclusion
$c_1^{1+}$	0.41	0.79	Oui	Oui	$k_1 \in [0.41, 0.79]$
$c_3^1$	1.4	1.77	/	Oui	$k_2 - 2k_3 \in [1.4, 1.77]$
$c_1^{2+}$	0.41	0.79	Oui	Oui	$k_1 \in [0.41, 0.79]$
$c_3^{2+}$	1.4	1.78	Oui	Oui	$k_2 \in [1.4, 1.78]$
$c_1^{3+}$	0.72	1.1	Oui	Oui	$\frac{k_5+k_3}{k_2} \in [0.72, 1.1]$
$c_2^{3+}$	1.40	1.78	Oui	Oui	$k_5 \in [1.40, 1.78]$
$c_3^3$	0.93	1.28	/	Non	$\frac{k_4-k_1}{k_2} \in [0.93, 1.28]$
$c_4^{3+}$	-0.45	-0.11	Non	Non	$k_4 \in [-0.45, -0.11]$

TAB. 5.2 – Estimation des intervalles de confiance pour les valeurs de paramètre et significativité des régressions (seuil  $\alpha=5\%$ ).

## 5.3 Validation qualitative des propriétés asymptotiques du modèle

### 5.3.1 Introduction

A ce stade, nous supposons que le réseau réactionnel et donc le modèle de bilan de matière a été validé. Nous considérerons donc le modèle de simulation incluant les cinétiques réactionnelles. Néanmoins nous pouvons discuter de la validité du modèle lorsque seules des hypothèses qualitatives sur les cinétiques (dépendance fonctionnelle, monotonie, etc.) ont été formulées et nous supposerons que les paramètres ne sont pas connus (ou dans certains cas qu'il appartient à un domaine connu).

La première chose à faire est de vérifier si les propriétés qualitatives du modèle en temps infini correspondent à des observations expérimentales.

Lorsque les entrées du modèle sont constantes, le modèle prédit-il bien un équilibre, ou bien un comportement plus complexe (cycles limites, chaos,...) en accord avec les observations ?

Ces propriétés qualitatives sont souvent dépendantes des entrées (supposées constantes) appliquées (taux de dilution, concentrations dans l'alimentation, etc.). Observe-t-on un changement qualitatif (l'équilibre qui était stable devient instable) lorsque l'on fait varier ces entrées ?

Cette manière qualitative de comparer un modèle à des résultats expérimentaux est classique en biologie. Bien souvent c'est d'ailleurs le principal critère qui est retenu, en particulier en écologie.

Par exemple la *première loi de Volterra* stipulait que la coexistence entre un prédateur et une proie présente un caractère oscillatoire périodique. C'est typiquement une propriété qualitative issue du modèle qu'il resterait à confronter à des données et Volterra a cherché de longues années, avec son beau-fils D'Ancona, des séries de données qui vérifieraient cette propriété des modèles proies-prédateurs [8]. La célèbre série des lynx et des lièvres (figure 5.1), obtenues à partir des données de fourrures collectées par la compagnie de la baie d'Hudson (plus de un siècle de données, depuis le milieu du XIX ième siècle) corrobore la première loi de Volterra.

Le paragraphe suivant expose un remarquable exemple de validation des propriétés asymptotiques du modèle.

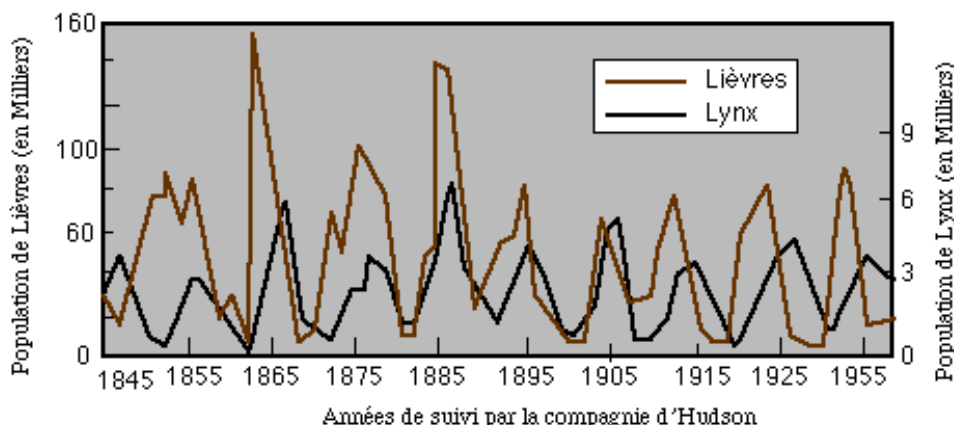


FIG. 5.1 – Exemple de validation de propriétés qualitatives d'un modèle : l'oscillation de populations de lynx et de lièvres corrobore la première loi de Volterra.

### 5.3.2 La compétition en chémostat

Hansen et Hubell présentent une très belle illustration de la validation asymptotique d'un modèle. Ils étudient la compétition entre deux espèces bactériennes en chémostat, en supposant une cinétique de croissance de type Monod :

$$\mu_i(S) = \mu_{max\ i} \frac{S}{S + K_{s\ i}}$$

L'étude du modèle montre qu'il y a (dans le cas générique) trois équilibres, un seul d'entre eux étant globalement attractif. Le vainqueur de la compétition dépend du taux de dilution ; c'est l'espèce avec le plus petit ratio  $J_i$  :

$$J_i = \frac{K_{s\ i}}{\mu_{max\ i} - D}$$

Hansen et Hubell ont pu comparer ces prédictions avec des résultats expérimentaux pour différentes espèces en compétition (une souche de *Escherichia coli* et une souche de *Pseudomonas aeruginosa*, ou 2 souches de *Escherichia coli* qui diffèrent par leur taux de croissance maximal). Dans le deuxième cas et en jouant sur la concentration en acide nalidixique qui inhibe le taux de croissance maximal, il ont ainsi pu obtenir, conformément à la théorie, coexistence (en tout cas sur la courte durée de l'expérience) des deux populations bactériennes (voir Figure 5.2).

Nous verrons ultérieurement une technique qui permet de montrer que le modèle n'est néanmoins pas entièrement satisfaisant [87] : les oscillations observées ne peuvent pas être reproduites avec la structure choisie pour le modèle.

### 5.3.3 Évolution des propriétés statiques en fonction des entrées

Cette manière de caractériser la validité d'un modèle, même si elle n'est qu'exceptionnellement utilisée dans le monde de l'Automatique, n'est pas nouvelle. Pour preuve, l'histoire relatée par [8] sur le problème

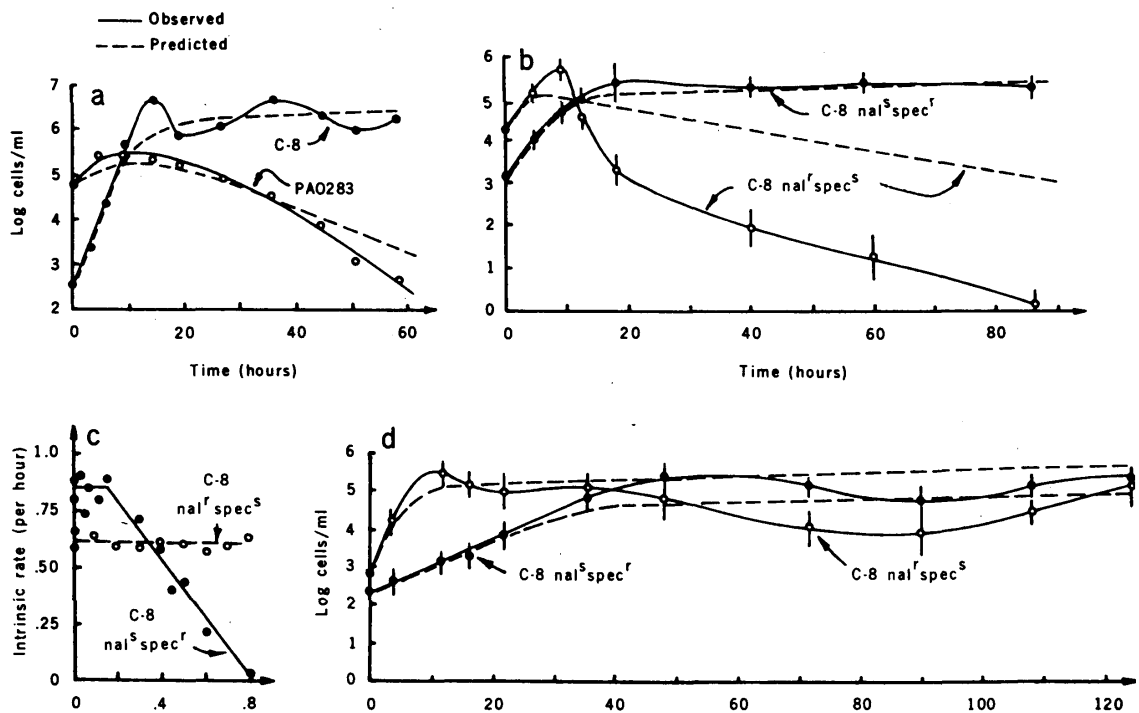


FIG. 5.2 – Validation qualitative expérimentale du comportement du modèle. Les prédictions qualitatives du modèle sont vérifiées pour : a) 2 espèces (*Escherichia coli*, souche C-8 et *Pseudomonas aeruginosa*, souche PA0283) qui diffèrent par leur constante de demi-saturation. b) 2 souches de *Escherichia coli* qui diffèrent par leur taux de croissance maximal. d) Coexistence obtenue avec 2 souches de *Escherichia coli* qui ont le même paramètre  $J_i$ . La figure c) représente l'effet de l'acide nalidixique sur le taux de croissance maximal pour les souches considérées C-8. D'après Hansen et Hubbell (1980).

rencontré par d'Ancona, beau-fils de Volterra. D'Ancona étudiait une série de données sur la pêche en mer Adriatique de poissons prédateurs (sélaciens). Il apparaissait que la proportion des sélaciens restait stable jusqu'aux dernières années de la guerre où elle se mettait alors à croître. Volterra modifia son modèle Proies-Prédateurs en y ajoutant un terme de capture par pêche (la mortalité par pêche, identique pour les deux espèces est appelée  $\epsilon$ ) :

$$\begin{aligned}\frac{dx}{dt} &= Ax - Bxy - \epsilon x \\ \frac{dy}{dt} &= -Cy + Dxy - \epsilon y\end{aligned}$$

Il apparaissait alors qu'une diminution de l'effort de pêche lié à la guerre entraînait un déplacement des valeurs moyennes des deux populations sur une période donnée par  $(\frac{C+\epsilon}{D}, \frac{A-\epsilon}{B})$ . Volterra énonça ainsi sa *troisième loi*

*"A cause de la pêche, la valeur moyenne de la population de proies  $x$  s'accroît de la quantité  $\frac{\epsilon}{D}$ , tandis que la valeur moyenne de la population des prédateurs  $y$  diminue de la quantité  $\frac{\epsilon}{B}$ ".*

Cette manière de valider les modèles, ou du moins d'en extraire des éléments qualitatifs confrontables à des données expérimentales est utilisée en écologie. Pour preuve la longue série de travaux d'Arditi et coll. [94, 95, 96, 97, 98, 99, 100] visant à montrer que des cinétiques "ratio dépendantes" peuvent s'avérer plus pertinentes que des cinétiques "proie dépendantes". Nous ne détaillerons pas cette problématique mais simplement les arguments utilisés par leurs auteurs dans le cadre de l'étude de chaînes trophiques. L'idée consiste [94] à calculer les populations à l'équilibre pour deux modèles concurrents. L'un se base sur une "réponse fonctionnelle" proie-dépendante et l'autre sur une hypothèse de ratio dépendance. Les auteurs considèrent ensuite différents taux d'enrichissement du milieu et analysent comment les différents niveaux trophiques y répondent. Dans le cas de la ratio dépendance tous les niveaux trophiques répondent positivement, alors que pour un modèle proie dépendant selon la distance au niveau trophique le plus élevé, on observe une augmentation, une diminution ou bien l'absence de changement. Ils confrontent ensuite ces arguments théoriques à des mesures de terrain provenant d'une succession de lac d'eau douce associés à un gradient de nutriments. Des relevés expérimentaux ont clairement montré une augmentation de tous les niveaux trophiques (phytoplancton, zooplancton herbivore, zooplancton carnivore, poissons) le long du gradient de nutriment, validant donc les prédictions du modèle densité dépendant et invalidant un modèle proie-dépendant.

L'idée sous-jacente consiste donc à analyser le comportement asymptotique du modèle en fonction des entrées  $u$ . Dans le cas le plus simple, où il existe un unique équilibre globalement attractif  $y^*$ , cela revient donc à étudier l'application :  $u \rightarrow y^*$ . Dans des cas plus complexes, tel que celui du modèle Volterra-D'Ancona, l'application est en fait la moyennisation sur un cycle des trajectoires en régime établi.

Notons que  $y^*$  représente l'ensemble des variables qui peuvent être mesurées (sorties). Pour davantage filtrer les modèles, il est également intéressant de vérifier aussi l'adéquation du modèle pour toute fonction simple (non paramétrée) faisant intervenir les composantes de  $y$ , par exemple  $\frac{y_1}{y_3}$  ou bien  $y_1 e^{-y_1}$ . Ainsi, dans l'exemple qui suit, nous n'inspecterons pas uniquement les concentrations mesurées correspondant aux variables d'état du modèle, nous analyserons également les rapports de ces concentrations (quota).

Cette analyse très simple est néanmoins extrêmement instructive et elle contraint fortement le modèle, comme nous allons le voir dans l'exemple.

Soulignons que si la théorie sous-jacente est triviale, elle implique un protocole expérimental particulier. Ce protocole consiste à balayer (si possible de manière intelligente) l'espace (statique) des entrées et à mesurer

avec précision les valeurs obtenues en régime permanent. Une telle approche est illustrée dans l'exemple ci-dessous.

### 5.3.4 Exemple de la modélisation de la croissance phytoplanctonique limitée à la fois par l'azote et la lumière

Ces travaux, réalisés d'abord avec D. Eveillard (2000), puis avec L. Pawlowski (2001), ont conduit à un modèle que je vais brièvement exposer [83].

Des études expérimentales [101] ont démontré l'interaction entre la limitation de la croissance par la lumière et par l'azote. Ce phénomène d'adaptation physiologique est connu sous le nom de phénomène de compensation. La cellule peut compenser une diminution d'intensité lumineuse en augmentant sa capacité à exploiter les photons par une augmentation de son contenu pigmentaire [102] et en régulant le Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco). Rubisco et chlorophylle représentent d'importantes réserves d'azote dans la cellule [103].

Des expériences ont permis d'observer expérimentalement le comportement entrées-sorties obtenu pour des variations du taux de dilution et de l'intensité lumineuse (cf. tableau 5.3).

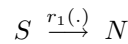
Par exemple la figure 5.3 représente en fonction du taux de dilution les valeurs obtenues à l'équilibre pour le carbone particulaire ( $C$ ), la chlorophylle  $a$  ( $L$ ) et le rapport  $L/C$  pour des cultures en chémostat de *Rhodomonas salina*.

Quantité	$I$	$D$	Bibliographie
$C^*$	+	-	[104]
$L^*$	-	+	[105]
$\frac{N^*}{C^*}$	-	+	[106, 107, 108]
$\frac{L^*}{C^*}$	- & +	+	[108, 104]

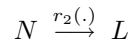
TAB. 5.3 – Comportement entrées-sorties qualitatif **expérimental** : évolution des états d'équilibres en fonction de  $I$  et  $D$  (+ : croissant, - : décroissant, - & + : décroissant puis croissant).

Le modèle que nous avons construit suppose que la croissance de phytoplancton se produit selon deux voies métaboliques distinctes : prise et assimilation de l'azote d'une part et fixation de carbone par la photosynthèse d'autre part. Le réseau réactionnel pour la voie de l'azote est donc :

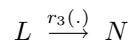
- La source d'azote (nitrate) est absorbée par la cellule et entre ainsi dans la réserve d'azote intracellulaire ( $N$ ) :



- dans une deuxième étape, la réserve d'azote est utilisée pour produire les protéines chlorophylliennes ( $L$ ) associées à la photosynthèse :



- La chlorophylle se dégrade naturellement et retourne dans le stock d'azote intracellulaire :



Nous supposons que la voie d'assimilation du carbone résulte principalement de deux réactions :

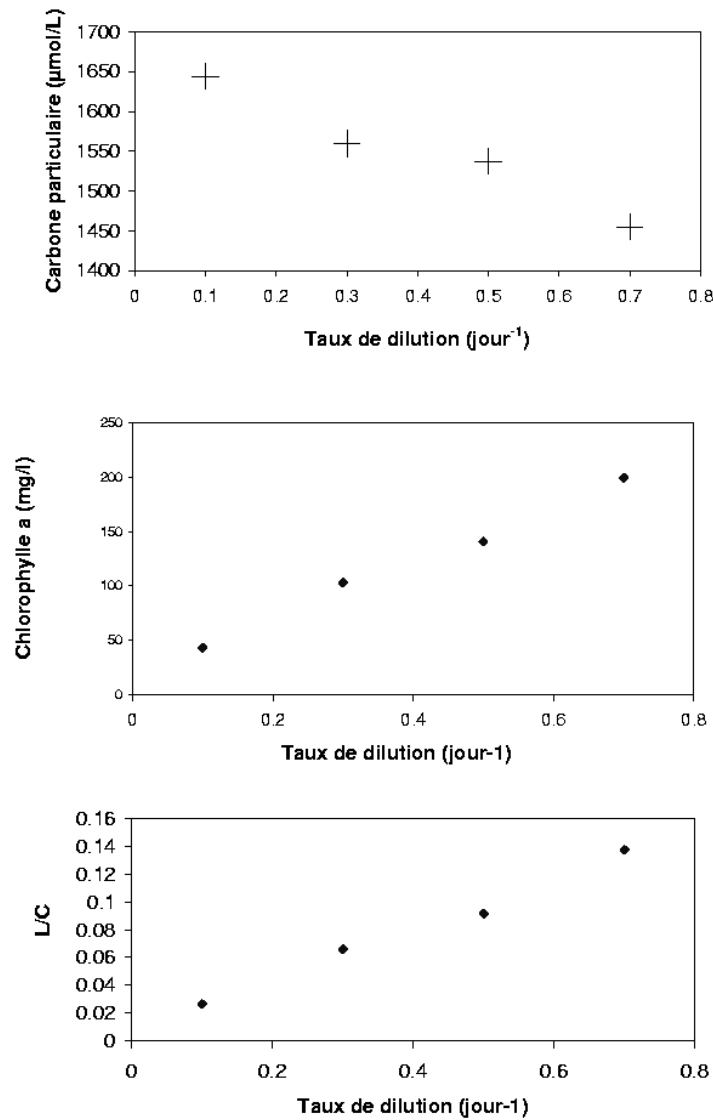
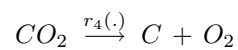
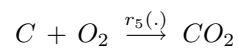


FIG. 5.3 – Exemple de comportement entrées-sorties à l'équilibre. Valeurs à l'équilibre pour le carbone particulaire, la chlorophylle *a* et le rapport L/C en fonction du taux de dilution pour des cultures de *Rhodomonas salina*.

- le carbone inorganique ( $CO_2$ ) est incorporé dans la cellule par la photosynthèse pour former le carbone particulaire ( $C$ ).



- Une partie du carbone est perdu par la respiration :



Le modèle a donc la forme générale (3.8) des systèmes biologiques en bioréacteur, avec :



$$\xi = \begin{pmatrix} S \\ N \\ L \\ C \\ CO_2 \\ O_2 \end{pmatrix}, \xi_{in} = \begin{pmatrix} S_{in} \\ 0 \\ 0 \\ 0 \\ CO_{2in} \\ O_{2in} \end{pmatrix}, r(\xi) = \begin{pmatrix} r_1(\xi) \\ r_2(\xi) \\ r_3(\xi) \\ r_4(\xi) \\ r_5(\xi) \end{pmatrix},$$

$$K = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}, Q(\xi) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ Q_{CO_2} \\ Q_{O_2} \end{pmatrix} \quad (5.15)$$

$CO_{2in}$  et  $O_{in}$  sont respectivement la concentration de  $CO_2$  et d' $O_2$  dans le milieu de renouvellement. Le débit gazeux du dioxyde de carbone et de l'oxygène sont notés  $Q_{CO_2}$  (qui est généralement négatif en raison de la photosynthèse) et  $Q_{O_2}$ .

Les hypothèses qui ont été faites pour obtenir les cinétiques et leur dépendance par rapport à l'intensité lumineuse, ont abouti aux expressions suivantes :

$$r(\xi) = \begin{pmatrix} \rho_m \frac{S}{S(t)+K_S} C \\ \gamma(I) N \frac{L}{C} \\ \beta L \\ a(I) L \\ \lambda C \end{pmatrix}$$

Il est maintenant possible d'analyser les propriétés qualitatives du modèle à l'équilibre indépendamment des valeurs des paramètres afin de les comparer aux observations.

Des calculs élémentaires permettent de décrire l'évolution des différentes quantités à l'équilibre lorsque  $I$  et  $D$  varient. Notons qu'*a priori* il n'est pas du tout évident d'obtenir des variations monotones. Ceci n'est d'ailleurs vrai le plus souvent que dans une certaine plage de fonctionnement (loin du point de lessivage) et nous avons donc supposé que  $S^* \ll S_{in}$ .

Nous obtenons par exemple :

$$\frac{L^*}{C^*} = \frac{D + \lambda}{a(I)} \quad (5.16)$$

$$\frac{N^*}{C^*} = \frac{D + \beta}{\gamma(I)} \quad (5.17)$$

$$S^* = \frac{K_S D [k(I)(D + \lambda) + (D + \beta)]}{\rho_m \gamma(I) - D [k(I)(D + \lambda) + (D + \beta)]} \quad (5.18)$$

L'étude des variables d'état à l'équilibre en fonction des entrées ( $D$  et  $I$ ) conduit au tableau 5.4.

Le comportement entrées-sorties à l'équilibre synthétisé dans le tableau 5.4 respecte les observations expérimentales à l'équilibre (cf. tableau 5.3). Ce n'est, par exemple, pas le cas pour le modèle proposé par

Quantité	$I$	$D$
$C^*$	+	-
$L^*$	-	+
$N^*$	+	-
$\frac{N^*}{C^*}$	-	+
$\frac{L^*}{C^*}$	- & +	+

TAB. 5.4 – Comportement entrées-sorties qualitatif **théorique** : évolution des états d'équilibres en fonction de  $I$  et  $D$  (+ : croissant, - : décroissant, - & + : décroissant puis croissant).

[104] qui prévoit (ce qui est plus intuitif) une diminution de la concentration en chlorophylle avec le taux de dilution, alors que le contraire est observé.

## 5.4 Validation des propriétés dynamiques du modèle

### 5.4.1 Introduction

Les objectifs de cette étude sont plus ambitieux car il s'agit maintenant de caractériser le comportement **transitoire** du modèle afin de le comparer à des observations expérimentales. Comme nous souhaitons toujours pouvoir caractériser la validité du modèle indépendamment des paramètres utilisés et même des fonctions retenues, nous avons choisi de décrire de manière qualitative ce comportement transitoire. Nous voulons par exemple déterminer quelles suites de minima et de maxima devraient être observées pour les différentes variables du système.

L'analyse du comportement d'un système dynamique non linéaire est une tâche difficile, pour laquelle il n'existe que très peu d'outils. En effet, il est bien connu qu'un modèle dynamique non-linéaire peut produire des comportements complexes [109], même pour de faibles dimensions. Lorsque le modèle n'est pas complètement connu (des paramètres sont inconnus ou même certaines fonctions ne sont pas explicitées), le problème est bien plus compliqué. Le problème que nous étudierons ici consiste à décrire le comportement dynamique qualitatif d'un modèle décrit de manière incomplète.

La description dynamique d'un modèle qualitatif est un des buts du raisonnement qualitatif (QR) [110, 111]. Si le modèle est suffisamment connu (c.-à-d. avec des incertitudes bornées sur les paramètres), alors les méthodes semi-quantitatives peuvent être utilisées [110, 112],[90]. Dans certains cas (en dimension 2) les algorithmes de la simulation qualitative peuvent s'interpréter simplement [113, 114, 115]. Nous verrons au paragraphe 5.5.1 comment ces techniques peuvent également intervenir pour comparer des modèles et discuter de leur validité [90].

Nous présentons ici une approche qui décrit le comportement transitoire d'un système dynamique défini seulement par des propriétés de signe. Cette méthode convient particulièrement à une classe de systèmes (systèmes structurés en boucle avec des interactions monotones) qui contient de nombreux modèles en biologie dans les domaines de la régulation génique, [116], des systèmes compartimentaux [117], de la croissance cellulaire [118] et du développement des populations structurées en stades [119]. Néanmoins l'analyse peut également être appliquée à d'autres modèles, à condition qu'il y ait assez de zéros dans la matrice jacobienne,

nous le verrons plus en détail par la suite.

**Notations.** Pour  $y \in \mathbf{R}$  nous considérons la fonction “signe” :

$$\text{signe}(y) = \begin{cases} -1 & \text{si } y < 0 \\ 0 & \text{si } y = 0 \\ 1 & \text{si } y > 0 \end{cases}$$

Pour  $x \in \mathbf{R}^n$ ,  $\text{signe}(x)$  est le vecteur de composantes  $\text{signe}(x_i)$ .

La matrice  $\text{diag}(x)$  est la matrice diagonale ayant  $x \in \mathbf{R}^n$  sur sa diagonale principale.

Soit  $\Omega$  un domaine ouvert convexe de  $\mathbf{R}^n$  et  $f$  une application  $C^1$  de  $\Omega$  sur  $\mathbf{R}^n$ . Nous considérons le système différentiel autonome sur  $\Omega$  :

$$(\Sigma) \left\{ \dot{x} = f(x) \right.$$

**Definition 6** Le système  $(\Sigma)$  a une structure de boucle si  $f_i(x) = f_i(x_i, x_{i+1}) \forall i \in \{1, \dots, n\}$ .

La vitesse de chacune des variables dépend seulement de la variable elle-même et de la suivante (les indices sont comptés modulo  $n$ ).

**Definition 7** Le système  $(\Sigma)$  est à interactions monotones sur  $\Omega$  si chaque dérivée partielle  $\partial f_i / \partial x_j(x)$  pour  $i \neq j$  ne s'annule jamais sur  $\Omega$ .

Les éléments hors diagonales de la matrice jacobienne sont de ce fait de signe fixe sur  $\Omega$ . Les signes des éléments définissent ce que nous appellerons la structure du système.

Nous considérerons l'ensemble  $\mathcal{S}_n$  qui contient  $2^n$  éléments :

$$\mathcal{S}_n \stackrel{\text{def}}{=} \{ \sigma^i = {}^T(\sigma_1^i, \dots, \sigma_n^i) / \sigma_j^i \in \{-1, 1\} \}$$

Cet ensemble  $\mathcal{S}_n$  représente tous les vecteurs de  $\mathbf{R}^n$  dont les composantes sont  $+1$  ou  $-1$ . Nous choisirons d'ordonner  $\mathcal{S}_n$ , de sorte que  $\sigma^q$  sera le  $q^{\text{ieme}}$  élément. Conventionnellement, nous choisirons  $\sigma^1 = {}^T(1, \dots, 1)$ .

**Definition 8** Pour  $x^* \in \Omega$  le système à interactions monotones  $(\Sigma)$  est diagonalement  $x^*$ -monotone si pour tout  $j$  dans  $\{1, \dots, n\}$ , pour tout  $q$  dans  $\{1, \dots, 2^n\}$ , le signe de la dérivée partielle  $\partial f_j / \partial x_j(x)$  est fixe sur chaque domaine :

$$W_{\sigma^q}(x^*) \stackrel{\text{def}}{=} \{x \in \Omega; \text{diag}(\sigma^q)(x - x^*) > 0\}$$

Nous disons que les domaines  $W_{\sigma^q}(x^*)$  sont les  $x^*$ -orthants. Ils sont délimités par les hyperplans liés à  $x^*$ . Nous notons  $V_i(x^*)$  le  $i^{\text{eme}}$  hyperplan associé à  $x^*$  :

$$V_i(x^*) \stackrel{\text{def}}{=} \{x \in \Omega; x_i = x_i^*\}$$

De la même manière, nous définissons les orthants de l'espace des vitesses ou  $z$ -orthants :

$$Z_{\sigma^p} \stackrel{\text{def}}{=} \{x \in \Omega; \text{diag}(\sigma^p)f(x) > 0\}$$

Les  $z$ -orthants sont délimités par les isoclines nulles  $U_i$  :

$$U_i \stackrel{\text{def}}{=} \{x \in \Omega; f_i(x) = 0\}$$

Nous considérons également les ensembles suivants qui sont des réunions d'ensembles précédemment définis :

- La réunion des hyperplans associé à  $x^*$  :  
 $V(x^*) \stackrel{\text{def}}{=} \bigcup_{i=1}^n V_i(x^*)$ .
- La réunion des isoclines nulles :  $U \stackrel{\text{def}}{=} \bigcup_{i=1}^n U_i$

### 5.4.2 L'ensemble des événements qualitatifs possibles

Dans cette section nous considérerons un système à structure de boucle et à interaction monotone  $(\Sigma)$  et nous supposerons qu'il existe (au moins) un point d'équilibre  $x^* \in \Omega$  pour lequel  $(\Sigma)$  est diagonalement  $x^*$ -monotone.

Nous déterminerons alors l'ensemble des orthants de l'espace des vitesses  $Z_{\sigma^q}$  compatibles avec un  $W_{\sigma^q}(x^*)$  donné de l'espace des déviations. En d'autres termes nous recherchons tous les signes de  $f(x)$  possibles pour  $x$  dans un  $x^*$ -orthant donné, nous voulons donc caractériser l'ensemble des signes suivants :

$$\mathcal{F}^q \stackrel{\text{def}}{=} \{\text{signe}(f(x)), x \in W_{\sigma^q}(x^*) \setminus U\} \quad (5.19)$$

#### Approche locale

Dans un premier temps, nous allons déterminer cet ensemble pour le linéarisé de  $(\Sigma)$  autour du point  $x^*$  :

$$\Delta x = Df(x^*) \Delta x \quad (5.20)$$

où  $Df(x)$  représente la matrice jacobienne au point  $x$  et  $\Delta x \stackrel{\text{def}}{=} x - x^*$ . Notons que  $\text{diag}(\sigma^q)\Delta x$  est un vecteur positif pour  $x \in W_{\sigma^q}(x^*)$ . Naturellement, dans le cas non-linéaire, nous excluons les isoclines nulles du système linéaire (linéarisé d'un  $U_i$ ).

Nous considérerons la matrice  $\mathcal{M}^q \stackrel{\text{def}}{=} Df(x^*)\text{diag}(\sigma^q)$ , dont les éléments sont notés  $m_{kl}^q$  de signes  $s_{kl}^q = \text{signe}(m_{kl}^q)$ .

Le problème de la détermination des signes possibles pour  $Df(x^*)\Delta x$  lorsque  $x \in W_{\sigma^q}(x^*)$  est donc équivalent à la détermination de l'ensemble :

$$\mathcal{L}^q \stackrel{\text{def}}{=} \{\text{signe}(\mathcal{M}^q \xi), \xi > 0, (\xi + x^*) \in \Omega\} \quad (5.21)$$

Nous avons naturellement :  $\mathcal{L}^q \subset \mathcal{F}^q$  : les signes possibles localement au voisinage de  $x^*$  sont inclus dans l'ensemble de tous les signes possibles pour le système non-linéaire original.

Nous déterminerons l'ensemble  $\mathcal{L}^q$ , en utilisant deux lemmes complémentaires ( $s_{kl}^q$  prend ses valeurs dans  $\{-1, 0, 1\}$ ).

**Lemme 2** *Considérons le linéarisé d'un système diagonalement  $x^*$ -monotone structuré en boucle  $(\Sigma)$ , au point  $x^* \in \Omega$ . Pour un  $\sigma^q$  donné, s'il existe un indice  $k$  tel que l'une des deux conditions suivantes est satisfaite :*

1.  $s_{k,k}^q = 0$
2.  $s_{k,k}^q = s_{k,k+1}^q$

alors

$$\mathcal{L}^q = \{l^q = {}^T(l_1^q, \dots, l_n^q); l_j^q \in \{s_{j,j}^q \oplus s_{j,j+1}^q\}\} \quad (5.22)$$

La table de l'opérateur  $\oplus$  (somme qualitative "générique ") est donnée dans le tableau 5.5.

$1 \oplus 1$	$\{1\}$
$-1 \oplus 1$	$\{-1, 1\}$
$-1 \oplus -1$	$\{-1\}$
$-1 \oplus 0$	$\{-1\}$
$1 \oplus 0$	$\{1\}$
$0 \oplus 0$	$\{0\}$

TAB. 5.5 – Table de la somme qualitative (commutative)  $\oplus$ .

**Preuve :** voir [84]

Le lemme 2 dit que, dans les conditions données, tous les signes *a priori* admissibles sont obtenus. Le lemme suivant couvre les cas restants et il s'avère que la situation est un peu plus compliquée :

**Lemme 3** *Considérons le linéarisé d'un système diagonalement  $x^*$ -monotone à structure de boucle  $(\Sigma)$ , avec  $x^* \in \Omega$ .*

*Si pour tout indice  $k$  :  $s_{k,k}^q = -s_{k,k+1}^q$ , alors*

$$\mathcal{L}^q = \mathcal{S}_n - \mathcal{C}^q \quad (5.23)$$

*où l'ensemble  $\mathcal{C}^q$  est obtenu comme suit à partir de  $D^q \stackrel{\text{def}}{=} {}^T(s_{1,1}^q, s_{2,2}^q, \dots, s_{n,n}^q)$  :*

- si  $\det [Df(x^*)] < 0$ ,  $\mathcal{C}^q = \{\prod_{j=1}^n \sigma_j^q s_{j,j}^q D^q\}$
- si  $\det [Df(x^*)] = 0$ ,  $\mathcal{C}^q = \{-\prod_{j=1}^n \sigma_j^q s_{j,j}^q D^q, \prod_{j=1}^n \sigma_j^q s_{j,j}^q D^q\}$
- si  $\det [Df(x^*)] > 0$ ,  $\mathcal{C}^q = \{-\prod_{j=1}^n \sigma_j^q s_{j,j}^q D^q\}$

**Preuve :** voir [84].

### Approche globale

Afin de déterminer l'ensemble des signes possibles pour  $f(x)$ , nous récrivons le système  $(\Sigma)$  sous une autre forme. En utilisant le fait que  $\mathcal{L}^q \subset \mathcal{F}^q$ , nous déterminerons les divers cas pour lesquels cette inclusion est stricte.

**Lemme 4** *Si  $x^*$  est un point d'équilibre, le système  $(\Sigma)$  peut être récrit*

$$\dot{\Delta}x = A(x, x^*) \cdot \Delta x$$

*Si  $(\Sigma)$  a des interactions monotones, alors la matrice  $A(x, x^*)$  a les mêmes signes hors-diagonaux que la matrice jacobienne  $Df(x^*)$  de  $(\Sigma)$ . Si de plus  $(\Sigma)$  est diagonalement  $x^*$ -monotone, les éléments diagonaux*

de  $A(x, x^*)$  sont de signes fixes dans le domaine  $W_{\sigma^q}(x^*)$  et sont de même signe que ceux de  $Df(x^*)$  (excepté pour les éléments  $Df(x^*)$  qui sont nuls).

**Preuve :** voir [85, 84].

Nous emploierons les mêmes notations pour  $A(x, x^*)$  que pour  $Df(x^*)$ , c.-à-d. que nous noterons  $\mathcal{M}^q(x) = A(x, x^*) \text{diag}(\sigma^q)$  et  $t_{kl}^q$  le signe (fixé) de ses éléments  $m_{kl}^q(x)$ .

Nous considérons d'abord le cas simple où les éléments diagonaux de la matrice jacobienne sont non nuls (et donc les matrice  $A(x, x^*)$  et  $Df(x^*)$  sont de même signe). Dans ce cas le cadre du Lemme 4 est satisfait :

**Lemme 5** *considérons un système en boucle diagonalement  $x^*$ -monotone ( $\Sigma$ ) où  $x^*$  est un point d'équilibre. Si les deux conditions suivantes sont vérifiées dans  $W_{\sigma^q}(x^*)$  :*

1.  $\forall k, t_{k,k}^q = s_{k,k}^q \neq 0$
2.  $\exists k, t_{k,k}^q = t_{k,k+1}^q$

alors nous avons  $\mathcal{F}^q = \mathcal{L}^q$ .

Nous allons maintenant considérer l'autre cas correspondant au lemme 3.

**Lemme 6** *Considérons un système en boucle ( $\Sigma$ ) diagonalement  $x^*$ -monotone, où  $x^*$  est un point d'équilibre. Supposons que pour tout  $k$  nous avons  $t_{k,k}^q = -t_{k,k+1}^q$ .*

*Si  $\det(A(x, x^*))$  s'annule et change de signe sur  $W_{\sigma^q}(x^*)$ , alors l'ensemble  $\mathcal{F}^q$  contient tous les orthants possibles :*

$$\mathcal{F}^q = \mathcal{S}_n \tag{5.24}$$

si ce n'est pas le cas, alors

$$\mathcal{F}^q = \mathcal{S}_n - \mathcal{C}^q \tag{5.25}$$

où l'ensemble  $\mathcal{C}^q$  est obtenu à partir de  $D^q \stackrel{\text{def}}{=} T(t_{1,1}^q, t_{2,2}^q, \dots, t_{n,n}^q)$  comme suit :

- si  $\det[A(x^\dagger, x^*)] < 0$ ,  $\mathcal{C}^q = \{\prod_{j=1}^n \sigma_j^q s_{j,j}^q D^q\}$
- si  $\det[A(x^\dagger, x^*)] > 0$ ,  $\mathcal{C}^q = \{-\prod_{j=1}^n \sigma_j^q s_{j,j}^q D^q\}$

$x^\dagger \in W_{\sigma^q}(x^*)$  étant un point quelconque où le déterminant de  $A(x^\dagger, x^*)$  ne s'annule pas.

**Remarque :** s'il existe un autre point d'équilibre  $x^\dagger \in W_{\sigma^q}(x^*)$ , alors  $A(x^\dagger, x^*)(x^\dagger - x^*) = 0$  et par conséquent  $\det(A(x^\dagger, x^*)) = 0$ . Nous sommes alors dans le cas du lemme ci-dessus.

### Partition de l'espace d'état en régions admissibles

**Definition 9** *Pour  $\sigma^p \in \mathcal{F}_q$  nous définissons l'ensemble ouvert suivant :*

$$\Omega_{\sigma^q \sigma^p}(x^*) \stackrel{\text{def}}{=} W_{\sigma^q}(x^*) \cap Z_{\sigma^p} = \{x \in W_{\sigma^q}(x^*); \text{diag}(\sigma^p)f(x) > 0\} \tag{5.26}$$

Remarquons qu'un grand nombre de  $\Omega_{\sigma^q \sigma^p}(x^*)$  sont vides. Les  $\Omega_{\sigma^q \sigma^p}(x^*)$  non vides représentent donc les situations qualitatives permises par le modèle et seront appelées régions admissibles. Remarquons également que ces régions n'intersectent pas les isoclines nulles  $U_i$  et l'hyperplan  $V_i(x^*)$ .

Nous pouvons maintenant considérer la partition suivante de l'espace d'état  $\Omega$  :

$$\Omega = \left( \bigcup_{\sigma^q \in \mathcal{S}_n, \sigma^p \in \mathcal{F}^q} \Omega_{\sigma^q \sigma^p}(x^*) \right) \cup V(x^*) \cup U \quad (5.27)$$

### 5.4.3 Restriction de l'espace d'état

Dans l'analyse que nous proposons, nous enlevons un ensemble de trajectoires qui peuvent avoir des comportements marginaux, nous démontrons en particulier que cet ensemble est de mesure nulle, moyennant des hypothèses qui sont généralement vérifiées [84]. L'espace final de phase sera appelé  $\tilde{\Omega}$ .

Maintenant nous pouvons définir l'ensemble ouvert  $\tilde{\Omega}$ , qui est  $\Omega$  privé de ces ensembles de mesure nulle.

Dans le cas particulier (que nous appellerons le cas E) où les deux surfaces  $U_i$  et  $V_{i-1}(x^*)$  coïncident sur un ensemble ouvert, nous n'enlèverons pas cet ensemble qui n'est pas associé à un ensemble de trajectoires de mesure nulle. Dorénavant, tout se passe dans cet espace restreint  $\tilde{\Omega}$ . Pour une région admissible  $\Omega_{\sigma^q \sigma^p}(x^*)$ , nous définissons l'ensemble non vide :

$$\tilde{\Omega}_{\sigma^q \sigma^p}(x^*) \stackrel{\text{def}}{=} \Omega_{\sigma^q \sigma^p}(x^*) \cap \tilde{\Omega}$$

Deux régions de l'espace de phase seront dites voisines si elles ne diffèrent que par un signe (de la déviation ou de la vitesse). Remarquons que nous avons supprimé (en restreignant  $\Omega$ ) la possibilité d'aller d'une région  $\tilde{\Omega}_{\sigma^q \sigma^p}(x^*)$  à l'autre si elles diffèrent de plus d'un signe (à l'exception du cas E qui est un peu dégénéré).

**Definition 10** Deux régions  $\tilde{\Omega}_{\sigma^{q_1} \sigma^{p_1}}(x^*)$  et  $\tilde{\Omega}_{\sigma^{q_2} \sigma^{p_2}}(x^*)$  sont appelées :

- stricts  $U$ -voisins si  $\sigma^{q_1} = \sigma^{q_2}$  et s'il existe un unique  $k \in \{1, ..n\}$  tel que  $\sigma_k^{p_1} = -\sigma_k^{p_2}$ .
- stricts  $V$ -voisins si  $\sigma^{p_1} = \sigma^{p_2}$  et s'il existe un unique  $k \in \{1, ..n\}$  tel que  $\sigma_k^{q_1} = -\sigma_k^{q_2}$ .

**Definition 11** Dans le cas E, nous dirons que  $\tilde{\Omega}_{\sigma^{q_1} \sigma^{p_1}}(x^*)$  et  $\tilde{\Omega}_{\sigma^{q_2} \sigma^{p_2}}(x^*)$  sont des UV-voisins stricts si pour tout  $i \neq k$ ,  $\sigma_i^{p_1} = \sigma_i^{p_2}$ ,  $\sigma_{i+1}^{q_1} = \sigma_{i+1}^{q_2}$  et  $\sigma_k^{p_1} = -\sigma_k^{p_2}$ ,  $\sigma_{k+1}^{q_1} = -\sigma_{k+1}^{q_2}$ .

### 5.4.4 Transition entre les domaines $\tilde{\Omega}_{\sigma^q \sigma^p}(x^*)$

Nous pouvons maintenant considérer l'espace restreint  $\tilde{\Omega}$  partitionné en domaines ouverts  $\tilde{\Omega}_{\sigma^q \sigma^p}(x^*)$ . Nous prouverons ici que la transition entre ces domaines obéit à des règles déterminées par le signe des éléments extradiagonaux de la matrice jacobienne.

**Théorème 5** (Théorème de transition entre les régions)

Considérons un système  $(\Sigma)$  en boucle et à interactions monotones et un point d'équilibre  $x^*$ . Supposons que  $\tilde{\Omega}_{\sigma^{q_1} \sigma^{p_1}}(x^*)$  et  $\tilde{\Omega}_{\sigma^{q_2} \sigma^{p_2}}(x^*)$  sont deux voisins stricts.

Rappelons que  $t_{k,k+1}$  est le signe de l'élément  $(k, k+1)$  de la matrice jacobienne.

- Traversée d'un  $U_k$  : Supposons qu'ils sont stricts  $U$  (ou UV)-voisins. Si  $t_{k,k+1} \sigma_{k+1}^{p_1} = \sigma_k^{p_1}$  [resp.  $-\sigma_k^{p_1}$ ], alors la traversée de  $U_k$  est possible seulement de  $\tilde{\Omega}_{\sigma^{q_2} \sigma^{p_2}}(x^*)$  vers  $\tilde{\Omega}_{\sigma^{q_1} \sigma^{p_1}}(x^*)$  [ resp  $\tilde{\Omega}_{\sigma^{q_1} \sigma^{p_1}}(x^*)$  vers  $\tilde{\Omega}_{\sigma^{q_2} \sigma^{p_2}}(x^*)$  ] et elle correspond à un minimum [resp. un maximum] de la variable  $x_k$ .
- traversée d'un  $V_i$  : Supposons qu'ils sont stricts  $V$  (ou UV)-voisins. Si  $t_{k,k+1} \sigma_{k+1}^{q_1} = \sigma_k^{q_1}$  [resp.  $-\sigma_k^{q_1}$ ], alors la traversée de  $V_k(x^*)$  est possible seulement de  $\tilde{\Omega}_{\sigma^{q_2} \sigma^{p_2}}(x^*)$  vers  $\tilde{\Omega}_{\sigma^{q_1} \sigma^{p_1}}(x^*)$  [ resp  $\tilde{\Omega}_{\sigma^{q_1} \sigma^{p_1}}(x^*)$  vers  $\tilde{\Omega}_{\sigma^{q_2} \sigma^{p_2}}(x^*)$  ] et elle correspond pour  $x_k$  à une traversée de bas en haut [ resp de haut en bas ] de son équilibre  $x_k^*$ .

Nous dirons que  $\tilde{\Omega}_{\sigma^{q_1}\sigma^{p_1}}(x^*)$  [ resp  $\tilde{\Omega}_{\sigma^{q_2}\sigma^{p_2}}(x^*)$  ] est accessible depuis  $\tilde{\Omega}_{\sigma^{q_2}\sigma^{p_2}}(x^*)$  [ resp  $\tilde{\Omega}_{\sigma^{q_1}\sigma^{p_1}}(x^*)$  ].

Les preuves de ces théorèmes sont très semblables et peuvent être consultées dans [88, 87].

En dimension 2 ces transitions pourraient avoir été obtenues à partir de l'analyse plus traditionnelle du plan de phase.

### 5.4.5 Barrières dans l'espace d'état

**Lemme 7** Supposons qu'il existe deux applications  $C^1$   $\Phi : x \in \Omega \longrightarrow \Phi(x) \in \mathbf{R}$   $\psi : u \in \mathbf{R} \longrightarrow \psi(u) \in \mathbf{R}$  qui vérifient :

$$D\Phi(x).f(x) = \psi(\Phi(x))$$

alors  $\mathcal{R} \stackrel{\text{def}}{=} \{x \in \Omega; \psi(\Phi(x)) = 0\}$  séparent  $\Omega$  en régions positivement invariantes.

$$\mathcal{R}^- \stackrel{\text{def}}{=} \{x \in \Omega; \psi(\Phi(x)) < 0\} \text{ et } \mathcal{R}^+ \stackrel{\text{def}}{=} \{x \in \Omega; \psi(\Phi(x)) > 0\}$$

**Corollaire 1** S'il existe une région de l'espace d'état  $\tilde{\Omega}_{\sigma^q\sigma^p}(x^*)$  telle que :  $\tilde{\Omega}_{\sigma^q\sigma^p}(x^*) \cap \mathcal{R}^- = \emptyset$  [resp  $\tilde{\Omega}_{\sigma^q\sigma^p}(x^*) \cap \mathcal{R}^+ = \emptyset$ ], alors toute trajectoire initialisée dans  $\mathcal{R}^-$  [resp dans  $\mathcal{R}^+$ ] n'atteindra jamais la région  $\tilde{\Omega}_{\sigma^q\sigma^p}(x^*)$ .

**Corollaire 2** Toute trajectoire initialisée dans une région  $\tilde{\Omega}_{\sigma^{q_1}\sigma^{p_1}}(x^*) \subset \mathcal{R}^+$  [resp  $\tilde{\Omega}_{\sigma^{q_1}\sigma^{p_1}}(x^*) \subset \mathcal{R}^-$ ], n'atteindra jamais les régions  $\tilde{\Omega}_{\sigma^{q_2}\sigma^{p_2}}(x^*) \subset \mathcal{R}^-$  [resp  $\tilde{\Omega}_{\sigma^{q_2}\sigma^{p_2}}(x^*) \subset \mathcal{R}^+$ ].

#### Remarques

- Une telle propriété est fréquente dans les modèles à bilan de matière, où  $u$  est lié à la conservation, de la masse et  $\psi$  est linéaire [20], cf. paragraphe 5.2.2.
- Comme  $u$  vérifie une équation scalaire, les états limites non vides sont les équilibres ; la propriété E est donc valide.

### 5.4.6 Théorème de comportement qualitatif

Le théorème suivant décrit le comportement des trajectoires d'un système différentiel  $(\Sigma)$  à interactions monotones, à structure de boucle et à diagonale monotone : le domaine  $\Omega$ , limité à  $\tilde{\Omega}$  est divisé en régions admissibles  $\tilde{\Omega}_{\sigma^{q_1}\sigma^{p_1}}(x^*)$  ; les transitions possibles entre ces régions sont données par le Théorème 5.

**Théorème 6** (Comportement qualitatif global) Chaque trajectoire de  $(\Sigma)$  dans un domaine  $\tilde{\Omega}_{\sigma^q\sigma^p}(x^*)$  soit :

- Reste dans  $\tilde{\Omega}_{\sigma^q\sigma^p}(x^*)$  et va à l'infini
- Reste dans  $\tilde{\Omega}_{\sigma^q\sigma^p}(x^*)$  et va vers un équilibre  $x^\dagger$  dans la fermeture de  $\tilde{\Omega}_{\sigma^q\sigma^p}(x^*)$
- Va vers un des voisins stricts  $\tilde{\Omega}_{\sigma^{q'}\sigma^{p'}}(x^*)$  qui sont accessibles.

### 5.4.7 Représentation graphique

Nous représenterons chaque région possible  $\tilde{\Omega}_{\sigma^q\sigma^p}(x^*)$  par une matrice de deux colonnes de signes : la première colonne représente  $\sigma^q$  et la seconde  $\sigma^p$ . Par exemple, la région  $\{x \in \Omega; x_1 > x_1^*, x_2 < x_2^*, x_3 >$



$x_3^*, \dot{x}_1 < 0, \dot{x}_2 < 0, \dot{x}_3 > 0$  est représentée par la matrice :

$$\begin{pmatrix} + & - \\ - & - \\ + & + \end{pmatrix}$$

### 5.4.8 Comportement asymptotique

Notons que cette analyse permet de fournir des résultats sur le comportement asymptotique du système [88, 84]. Nous donnerons simplement le Lemme suivant :

**Lemme 8** *Si, dans le graphe de transition d'un système  $\Sigma$ , il n'y a aucun cycle contenant un extremum de la variable  $x_k$ , alors pour presque toute trajectoire,  $x_k$  va soit vers un équilibre dans la fermeture de  $\Omega$  soit part à l'infini.*

**Corollaire 3** *S'il n'y a aucun cycle dans le graphe de transition d'un système  $\Sigma$ , presque toute les trajectoires vont vers un équilibre dans la fermeture de  $\Omega$  ou divergent.*

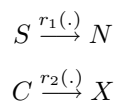
En d'autres termes, il n'y a aucun comportement périodique ou récurrent, ni chaos ou tout autre comportement complexe.

### 5.4.9 Application à une classe générale de modèles de croissance phytoplantonique

#### Présentation des modèles

Pour illustrer l'analyse qualitative des systèmes à structure de boucle nous considérerons une culture de micro algues en photobioréacteur continu. La croissance des algues est limitée par une source d'azote ( $\text{NO}_3$ ) notée  $S$ , et utilise le carbone inorganique dissout ( $C$ ), principalement sous forme de  $\text{CO}_2$ . La biomasse algale ( $X$ ) correspondra donc également à une quantité d'azote particulaire ( $N$ ).

Pour décrire simultanément l'absorption de carbone et d'azote dans la cellule, nous considérerons donc le schéma réactionnel suivant :



En posant  $\xi = (X, N, S, C)^t$ , on peut alors écrire le modèle à bilan de matière de type (3.8) avec :

$$K = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ -1 & 0 \\ 0 & -k_1 \end{pmatrix}, \xi_{in} = \begin{pmatrix} 0 \\ 0 \\ S_{in} \\ C_{in} \end{pmatrix} Q(\xi) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ Q_c(\xi) \end{pmatrix} \quad (5.28)$$

Nous remarquons ici que le rendement d'assimilation de l'azote et du carbone sont supposés unitaires.

Après un changement de variables (voir [88]) pour adimensionnaliser le système (5.28) par rapport aux entrées, le modèle décrivant la croissance de la biomasse phytoplantonique ( $x_2 = \frac{X}{DS_{in}}$ ) sur le substrat ( $x_1 = \frac{S}{S_{in}}$ ) en chémostat peut alors se récrire (nous ne considérons plus la variable  $C$ ) :

$$(\Sigma_{PGM}) \begin{cases} \dot{x}_1 = 1 - x_1 - \rho(x_1)x_2 \\ \dot{x}_2 = (\mu(x_3) - 1)x_2 \\ \dot{x}_3 = \rho(x_1) - \mu(x_3)x_3 \end{cases} \quad (5.29)$$

La variable  $x_3$  est le quota interne des cellules c.-à-d. la quantité de nutriment intracellulaire par unité de biomasse ( $x_3 = \frac{N}{X}$ ). Les fonctions  $\rho$  et  $\mu$  représentent respectivement le taux d'absorption du substrat et le taux de croissance.

Parmi les modèles  $(\Sigma_{PGM})$ , le modèle de Droop [106, 118] est très populaire. Pour ce modèle particulier nous avons :

$$\rho(x_1) = a_1 \frac{x_1}{a_2 + x_1}; \quad \mu(x_3) = a_3 \left(1 - \frac{a_4}{x_3}\right)$$

Les fonctions utilisées pour  $\rho$  et  $\mu$  sont des conjectures et ne sont pas justifiées par une validation *stricto sensu* en conditions transitoires.

L'étude suivante s'applique dans un cadre plus général et nous faisons seulement des hypothèses qualitatives, de sorte que l'analyse peut être généralisée à n'importe quelles fonctions  $\rho$  et  $\mu$  raisonnables.

Les hypothèses suivantes, corroborées par les expériences, sont généralement formulées par les biologistes afin de représenter la croissance du phytoplancton [120].

**Hypothèses :** dans le domaine physique considéré,

$$\Omega = \{x \in \mathbf{R}_+^3; x_1 > 0, x_2 > 0, x_3 > 0\} :$$

- **(H1)** : Le taux d'absorption  $\rho$  est une fonction positive croissante de  $x_1$ . Elle vérifie  $\rho(0) = 0$ .
- **(H2)** : Le taux de croissance  $\mu$  est une fonction positive strictement croissante de  $x_3$ .
- **(H3)** : Il existe un équilibre dans le domaine ouvert  $\Omega$ .

La classe des modèles  $(\Sigma_{PGM})$  qui vérifient les hypothèses (H1)-(H2)-(H3) sera appelée la classe des modèles de croissance phytoplanctonique (PGM).

Ce sont des modèles en boucle et à interaction monotone. La matrice Jacobienne a les signes suivants sur  $\Omega$  :

$$\begin{pmatrix} -1 & -1 & 0 \\ 0 & t_{22}(x) & +1 \\ +1 & 0 & -1 \end{pmatrix} \quad (5.30)$$

avec

$$t_{22}(x) = \text{signe}(\dot{x}_2) = \text{signe}(\mu(x_3) - 1) \quad (5.31)$$

**Propriété 7** Les PGM ont deux équilibres :

-  $x^* \in \Omega$  :

$$x_3^* = \mu^{-1}(1); \quad x_1^* = \rho^{-1}(\mu^{-1}(1)); \quad x_2^* = \frac{1 - \rho^{-1}(\mu^{-1}(1))}{\mu^{-1}(1)}$$

-  $x^b$ , un équilibre instable sur la frontière de  $\Omega$  :

$$x_2^b = 0; \quad x_1^b = 1; \quad x_3^b \text{ unique solution de : } \mu(x_3^b)x_3^b = \rho(x_1^b)$$

Nous considérerons donc les 3 hyperplans  $V_i(x^*) = \{x \in \Omega; x = x_i^*\}$  qui séparent l'espace  $\Omega$  en 8 régions  $W_{\sigma^j}(x^*)$ .

### Étude des situations qualitatives possibles

Pour obtenir l'ensemble des domaines compatibles avec la classe des modèles ( $\Sigma_{PGM}$ ), nous appliquons les résultats des lemmes 5 et 6 pour les 8 orthants  $W_{\sigma^q}(x^*)$ .

L'ensemble des domaines qualitatifs possibles  $\tilde{\Omega}_{\sigma^q\sigma^p}(x^*)$  se déduit des résultats de la section 5.4.2. Le calcul du signe des 8 matrices  $\mathcal{M}^q$  est résumé dans le Tableau 5.6.

**Propriété 8** Il y a 18 régions qualitatives possibles  $\tilde{\Omega}_{\sigma^q\sigma^p}(x^*)$  pour la classe de modèles PGM.

$\sigma^q$	-1 +1 +1	-1 -1 +1	+1 -1 +1	+1 +1 +1	-1 +1 -1	-1 -1 -1	+1 -1 -1	+1 +1 -1
$\mathcal{M}^q$	$\begin{pmatrix} +1 & -1 & 0 \\ 0 & 0 & +1 \\ -1 & 0 & -1 \end{pmatrix}$	$\begin{pmatrix} +1 & +1 & 0 \\ 0 & 0 & +1 \\ -1 & 0 & -1 \end{pmatrix}$	$\begin{pmatrix} -1 & +1 & 0 \\ 0 & 0 & +1 \\ +1 & 0 & -1 \end{pmatrix}$	$\begin{pmatrix} -1 & -1 & 0 \\ 0 & 0 & +1 \\ +1 & 0 & -1 \end{pmatrix}$	$\begin{pmatrix} +1 & -1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & +1 \end{pmatrix}$	$\begin{pmatrix} +1 & +1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & +1 \end{pmatrix}$	$\begin{pmatrix} -1 & +1 & 0 \\ 0 & 0 & -1 \\ +1 & 0 & +1 \end{pmatrix}$	$\begin{pmatrix} -1 & -1 & 0 \\ 0 & 0 & -1 \\ +1 & 0 & +1 \end{pmatrix}$
$\mathcal{F}^q$	*	+1	*	-1	*	+1	*	-1
	+1	+1	+1	+1	-1	-1	-1	-1
	-1	-1	*	*	*	*	+1	+1
Graphical representation of the $\tilde{\Omega}_{\sigma^q\sigma^p}(x^*)$	$\begin{pmatrix} - & - \\ + & + \\ + & - \end{pmatrix}$	$\begin{pmatrix} - & + \\ - & + \\ + & - \end{pmatrix}$	$\begin{pmatrix} + & + \\ - & + \\ + & - \end{pmatrix}$	$\begin{pmatrix} + & - \\ + & + \\ + & + \end{pmatrix}$	$\begin{pmatrix} - & - \\ + & - \\ - & + \end{pmatrix}$	$\begin{pmatrix} - & + \\ - & - \\ - & + \end{pmatrix}$	$\begin{pmatrix} + & + \\ - & - \\ - & + \end{pmatrix}$	$\begin{pmatrix} + & - \\ + & - \\ - & + \end{pmatrix}$

TAB. 5.6 – Matrices  $\mathcal{M}^q$  et ensembles  $\mathcal{F}^q$  associés au modèle de Droop en fonction de chaque  $\sigma^q$ . Chaque domaine possible  $\tilde{\Omega}_{\sigma^q\sigma^p}(x^*)$  est représenté par une matrice de signe : la première colonne contient le signe de l'écart à l'équilibre  $x^*$ , la seconde contient le signe de la tendance des variables. Le symbole  $*$  signifie soit +1 soit -1.

Ces situations qualitatives sont présentées dans le tableau 5.6. Remarquons qu'en dimension 3 il y a *a priori*  $2^6 = 64$  situations qualitatives possibles. Cette étude montre que seuls 14 de ces cas subsistent pour les modèles PGM. La détermination de cet ensemble de domaines possibles peut être utilisé comme un premier filtre pour examiner la structure du modèle. Si l'on peut observer expérimentalement un événement qualitatif n'appartenant pas à cet ensemble, cela signifie que la structure du modèle n'est pas correcte.

Pour aller plus loin dans la description du comportement qualitatif du modèle, nous pouvons maintenant construire le graphe de transition brut en appliquant les résultats du théorème 5. Nous obtenons le graphe

de transition (fig. 5.4) associé au modèle  $(\Sigma_{PGM})$ . Ce graphe récapitule la succession possible des extrema ou des traversées d'équilibre.

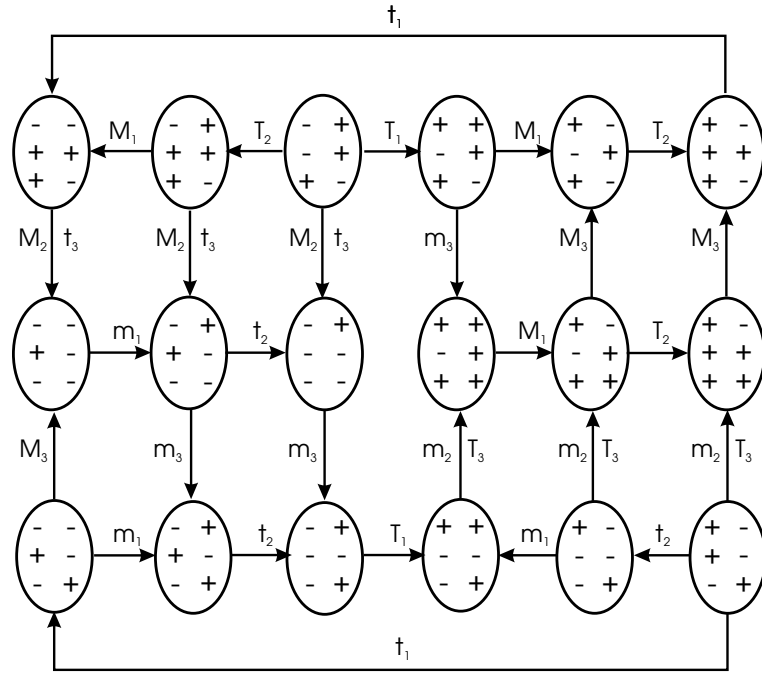


FIG. 5.4 – Graphe de transition brut.  $M_i$  [resp.  $m_i$ ] représente un maximum [resp. un minimum] pour la variable  $x_i$ .  $T_i$  [resp.  $t_i$ ] est associé à la traversée de la valeur d'équilibre du bas vers le haut [resp. du haut vers le bas] de la variable  $x_i$ .

En utilisant les résultats du paragraphe 5.4.5 et en particulier du lemme 7, il est ensuite possible de simplifier ce graphe brut.

**Propriété 9** Selon la condition initiale certaines régions de  $\Omega$  sont inaccessibles (Le signe  $\star$  représente un  $+$  ou un  $-$ ) :

- si  $u(0) < 1$  alors les régions  $\begin{pmatrix} + & \star \\ + & \star \\ + & \star \end{pmatrix}$  et les régions  $\begin{pmatrix} \star & - \\ \star & - \\ \star & - \end{pmatrix}$  sont inaccessibles
- si  $u(0) > 1$  alors les régions  $\begin{pmatrix} - & \star \\ - & \star \\ - & \star \end{pmatrix}$  et les régions  $\begin{pmatrix} \star & + \\ \star & + \\ \star & + \end{pmatrix}$  sont inaccessibles.

La propriété 9 permet maintenant de simplifier le graphe de transition en fonction des conditions initiales de  $u$  (Figures 5.6 et 5.5).

### Propriétés du modèle

L'analyse des graphes de transitions permet en particulier de démontrer la stabilité de l'équilibre non trivial  $x^*$  :

**Lemme 9** Presque toute trajectoire converge vers l'équilibre  $x^*$ .

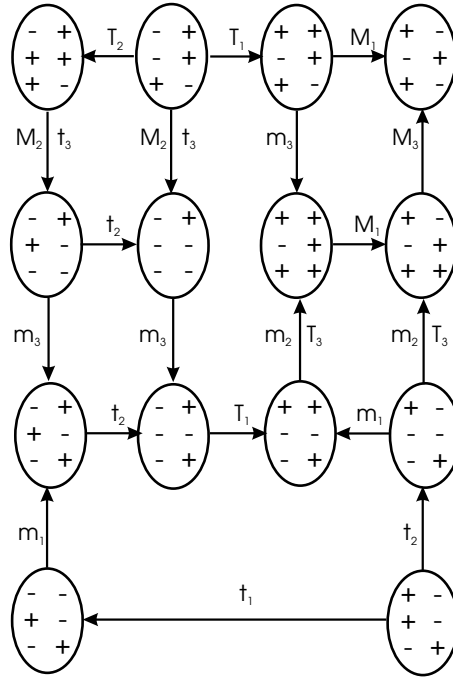


FIG. 5.5 – Graphe de transition simplifié pour  $u(0) < 1$ .

La preuve est une conséquence du fait que les graphes 5.6 et 5.5 ne contiennent pas de cycle. Cela signifie que toutes les variables finissent par être monotones. Comme le système est borné, elles doivent donc converger vers un équilibre (cf. [87]).

### 5.4.10 Application à la validation qualitative dynamique du modèle

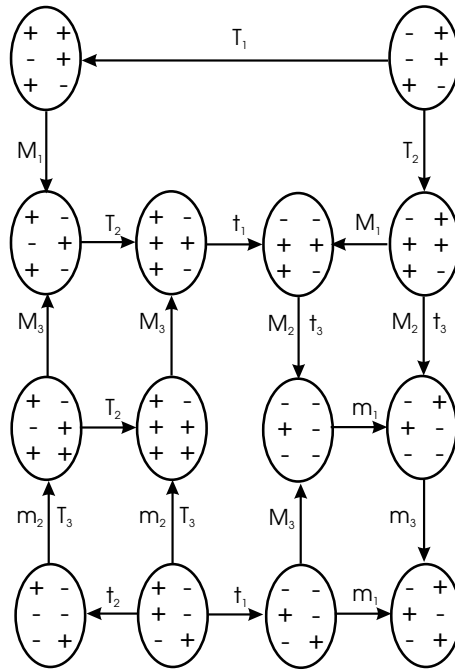
#### Principe

Cette méthodologie peut maintenant être employée pour comparer qualitativement le comportement réel d'un système (obtenu par les mesures) et le comportement théorique contenu dans le graphe de transition. Notez que la valeur de l'équilibre  $x^*$  est obtenue à partir des données expérimentales.

Si l'ordre des événements qualitatifs observés expérimentalement ne correspond pas à une des successions contenue dans le graphe, c'est qu'il y a un conflit entre le modèle et les données.

Le graphe de transition nous fournira même un ensemble de critères pour diagnostiquer un défaut du modèle. Le diagnostic découlera de la localisation de la contrainte cachée dans le graphe qui est violée. La raison du conflit entre les comportements réel et théorique est essentiellement de deux types :

- Une transition de  $x_i$  ne respecte pas la direction d'une flèche. Ceci indique que le modèle utilisé pour la variable  $x_i$  n'est pas conforme aux données. Plus précisément cela veut dire que le signe des interactions avec la variable  $x_{i+1}$  n'est pas correct, ou bien qu'il y a une autre interaction qui n'est pas représentée. Dans tous les cas cela touche les signes extradiagonaux de la matrice jacobienne.
- Un domaine qualitatif qui n'existe pas dans le graphe est observé expérimentalement. Si la transition était compatible avec les règles de transition imposées par les éléments hors-diagonaux de la matrice

FIG. 5.6 – Graphe de transition simplifié pour  $u(0) > 1$ .

jacobienne, ceci peut indiquer que la partition de l'espace par les  $\Omega_{\sigma^q \sigma^p}(x^*)$  est erronée. Ceci indique probablement une erreur dans le signe d'un terme diagonal de la matrice jacobienne.

#### 5.4.11 Autres exemples d'application de l'analyse dynamique qualitative

Ces techniques de validation qualitatives ont été appliquées à de nombreux exemples qui ne présentent pas tous la structure de boucle étudiée ici [87, 21, 86, 85, 22].

Dans [87], nous montrons que le modèle de compétition entre deux espèces (cf. chapitre 5.3.2), s'il est pertinent dans ses prédictions à l'équilibre, ne peut néanmoins pas prétendre expliquer les oscillations des populations bactériennes observées expérimentalement [93] (voir figure 5.2).

Dans [86, 22] nous étudions des populations de copépodes (microcrustacés marins dont le développement est structuré en stades). Ce type de structure induit pour certaines variables une dynamique de type  $\dot{x}_i = f_i(x_i, x_{i+1})$ . Certaines variables doivent donc respecter des règles de transition. Dans la plupart des cas, ces règles sont en fait intuitives et bien connues des biologistes qui les considèrent triviales. Ces règles ont pu être vérifiées dans de nombreuses expériences, y compris sur des séries de données assez longues.

Néanmoins, certaines "anomalies" avaient pu être remarquées dans les données de laboratoire. Par exemple des pics d'abondance des jeunes stades naupliens du copépode *Acartia clausi* ne respectent pas les règles de transition. Nous montrons que cela n'est pas explicable avec un modèle où la dynamique de la variable  $x_i$  est une fonction monotone de  $x_{i+1}$  (et dépend d'elle même). La prédation des nauplii par les adultes semblait une explication possible. L'hypothèse de perturbation du développement liée à une fluctuation de la nourriture est à exclure car ces stades naupliens ne se nourrissent pas. Par la suite cette conjecture sur le cannibalisme d'*A. clausi* a été vérifiée par des biologistes. Le même type d'analyse a été appliqué sur

des données *in situ* en mer du Japon, où le développement d'une cohorte de *C. abdominalis* a pu être observé. On montre également que les jeunes nauplii sont certainement consommés, soit par les adultes de la même espèce, soit par ceux de *Acartia omorii* également présents.

Dans cette analyse nous n'avons pas utilisé une modélisation complète du développement de cette espèce, notamment parce qu'il existe de nombreuses incertitudes quant aux effets (non mesurés) d'une fluctuation de la nourriture de ces copépodes. Nous avons donc focalisé notre raisonnement sur les stades naupliens qui ne se nourrissent pas.

Dans [85] nous considérons une population de cellules phytoplanctoniques soumis à la fluctuation périodique d'une source d'azote. Nous montrons que les règles de transition établies pour des entrées constantes par morceaux (et valables pour les intervalles où ces entrées sont constantes) peuvent être étendues aux systèmes forcés périodiquement. Il faut pour cela réaliser une moyenne mobile sur l'état. La tendance des variables moyennées sur une fenêtre correspondant à la période du forçage doit alors suivre un graphe de transition.

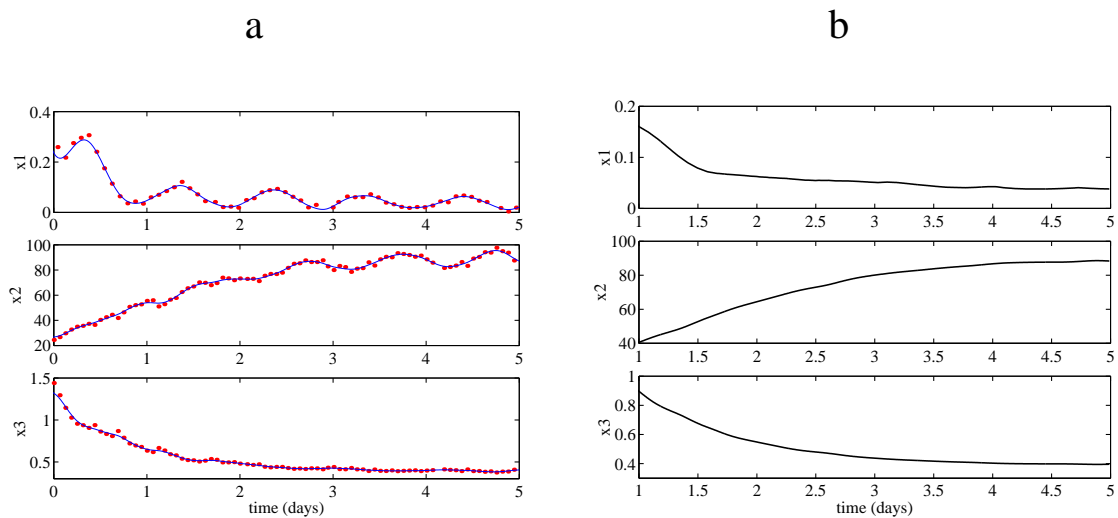


FIG. 5.7 – a. *Dunaliella tertiolecta* croissant en chémostat limité par une source de nitrate variant lentement (période : 24 heures). b. Données traitées par une moyenne mobile.

En utilisant ces techniques, nous avons montré qu'un changement qualitatif du comportement de la population s'opérait pour les fréquences de l'apport en nitrate élevées. En effet, à faible fréquence d'apport en nitrate (figure 5.7) le graphe de transition décrit correctement les données. Pour une fréquence plus élevée (figure 5.8 avec une période de 8h), le graphe ne permet plus de décrire les observations. Dans ce second cas, il semble en effet que la division des cellules se synchronise sur la source d'azote. La structure du modèle de Droop s'avère alors inappropriée.

#### 5.4.12 Conclusion de l'analyse dynamique qualitative

Nous avons vu qu'il était possible d'obtenir une description qualitative globale du comportement transitoire aussi bien que du comportement asymptotique d'une classe de modèles biologiques. Ces résultats sont globaux c.-à-d. qu'ils ne résultent pas de considérations linéaires locales. Insistons sur l'étonnante propriété à la base du comportement qualitatif de ces systèmes : seuls les signes de la matrice Jacobienne conditionnent

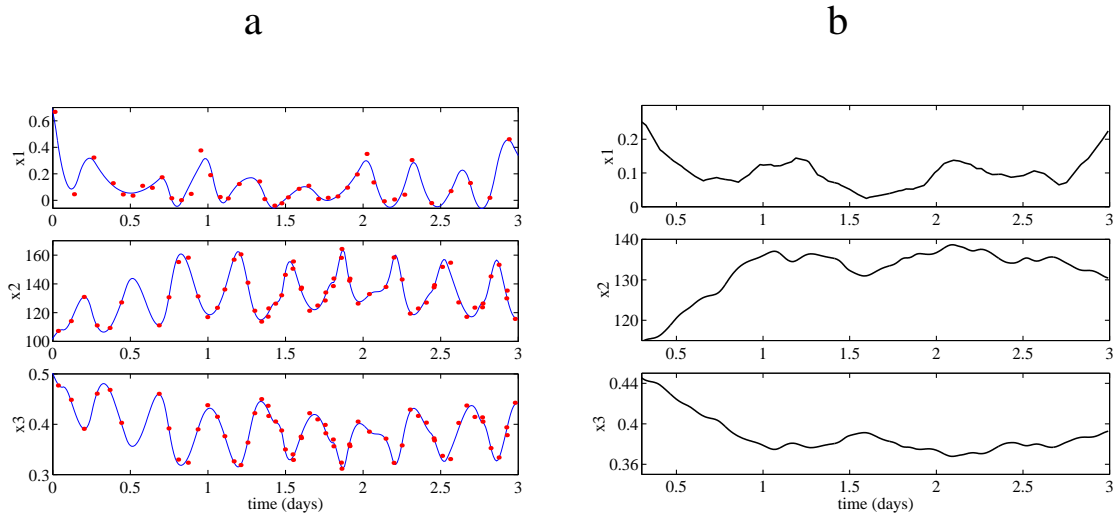


FIG. 5.8 – a. *Dunaliella tertiolecta* croissant en chiméostat limité par une source de nitrate variant rapidement. b. Données traitées par une moyenne mobile.

la dynamique qualitative. Cette analyse est donc particulièrement bien adaptée au contexte biologique, où la description exacte des fonctions utilisées ou bien la valeur des paramètres sont sujets à controverse (voir le paragraphe 4.5 sur la modélisation des cinétiques biologiques).

Nous avons aussi démontré, à travers un grand nombre d'exemples que cette démarche restait efficace dès lors que le système ne s'éloignait pas trop de la structure de boucle. Moins il y aura de zéros dans la matrice jacobienne et moins l'approche présentée générera de contraintes. Le graphe de transition comportera alors davantage d'état possibles ainsi que de transitions à double sens entre ces états.

Cette méthodologie peut également être employée dans le contexte des problèmes inverses où l'on essaye d'identifier un modèle. Comment trouver la structure d'un modèle à partir d'un ensemble de données ? S'il y a un grand nombre d'expériences disponibles qui contiennent une quantité suffisante d'information qualitative (transitions qualitatives), on peut rapporter ces observations dans un graphe et observer les transitions qui sont toujours exécutées de la même manière, ainsi que les domaines qualitatifs qui sont atteints. Ceci devrait permettre d'identifier un graphe de transition expérimental. La prochaine étape consistera alors à trouver une structure de modèle qui produit le graphe de transition observé.

## 5.5 Validation globale du modèle

### 5.5.1 Validation semi-quantitative

Cette approche fait suite à l'étape d'identification, mais prend toutefois en compte l'incertitude caractéristique de la biologie. Les paramètres appartiennent alors à des intervalles.

Des méthodes semi-quantitatives peuvent alors être utilisées [110, 112, 111], [90] pour simuler ces modèles, en prédisant les traits qualitatifs tels que les extrema et en en caractérisant l'amplitude par des intervalles.

L'idée proposée dans [89, 90] consiste à sélectionner parmi un ensemble de modèles ceux dont les simulations semi-quantitatives restent en adéquation avec les données, d'une part parce que les traits qualitatifs



correspondent aux observations, d'autre part parce que les intervalles pour les extrema, valeurs d'équilibre, etc. et les mesures concordent. Si c'est le cas, les estimations des intervalles paramétriques sont affinés par des techniques de type estimation robuste [82]. Si tous les modèles en lice satisfont les critères qualitatifs et quantitatifs, la méthode de discrimination calcule un critère d'entropie qui permet de concevoir l'expérience la plus informative. Cette approche aboutit donc à un protocole expérimental redéfini après chaque nouvelle expérience et qui propose les conditions les plus discriminantes pour les modèles.

En appliquant cette approche [89, 90], nous avons montré que les modèles de Monod [72] et Contois [121] ne permettent pas de décrire la croissance du phytoplancton. Les modèles de Droop [106, 118] et de Caperon-Meyer [122], quant à eux, produisent de (bons) résultats sensiblement équivalents.

### 5.5.2 Validation quantitative

C'est la manière la plus classique de valider un modèle, en comparant quantitativement les résultats obtenus par le modèle avec des données expérimentales. Dans cette approche, le modèle est considéré dans sa globalité et en particulier, les valeurs des paramètres sont fixées. Si l'adéquation entre modèle et expérience n'est pas bonne, il sera difficile de déterminer si le problème est plutôt structurel (mauvaises hypothèses sur le réseau réactionnel), lié aux expressions mathématiques utilisées pour les cinétiques, ou bien si les paramètres du modèle ne sont pas correctement estimés.

Si les précédentes étapes de la validation ont été franchies avec succès, on peut penser qu'une inadéquation provient avant tout d'une mauvaise estimation des paramètres.

En théorie, la validation d'un modèle consiste à tester des hypothèses statistiques sur les résidus obtenus en soustrayant les prédictions du modèle et les valeurs mesurées (homogénéité, stationnarité, indépendance et normalité)[82]. En pratique et compte tenu de la difficulté pour valider *stricto sensu* les modèles biotechnologiques, les biomodélisateurs se contentent souvent d'une bonne adéquation visuelle. Il conviendrait au moins de renforcer ce critère subjectif par une analyse de la corrélation entre les prédictions et les mesures.

Obtenir un modèle biologique dont la validation quantitative n'est toutefois pas mauvaise n'est cependant pas impossible, si l'on en juge par exemple par les résultats obtenus pour le modèle de fermentation anaérobie [50] (Figures 5.9 et 5.10). Ce modèle, même 6 ans après son développement (et sa calibration) présente une adéquation avec les données étonnement bonne [123].

## 5.6 Conclusion sur la validation des modèles

Sans doute parce que l'on nous apprend dès notre plus jeune âge à élaborer des modèles physiques en se basant sur des lois validées (relation fondamentale de la dynamique, loi d'ohm, etc.), nous sommes habitués à construire des modèles qui sont d'excellentes approximations du monde physique. De par notre culture, la validation d'un modèle ne nous apparaît pas être une nécessité. Ceci explique certainement pourquoi si peu de modèles en biologie peuvent se targuer d'une vraie validation. Cette faiblesse confère à la modélisation en biologie une image peu rigoureuse. Mais comme j'ai essayé de le démontrer au travers de mes travaux, ce n'est pas une fatalité et il est temps d'élaborer un cadre plus strict pour caractériser plus précisément le lien entre modèle et réalité.

Les différentes étapes de validation que j'ai mises en avant dans ce mémoire ne sont en réalité pas purement séquentielles, comme je l'ai présenté pour des raisons didactiques. Le réseau réactionnel et les interactions

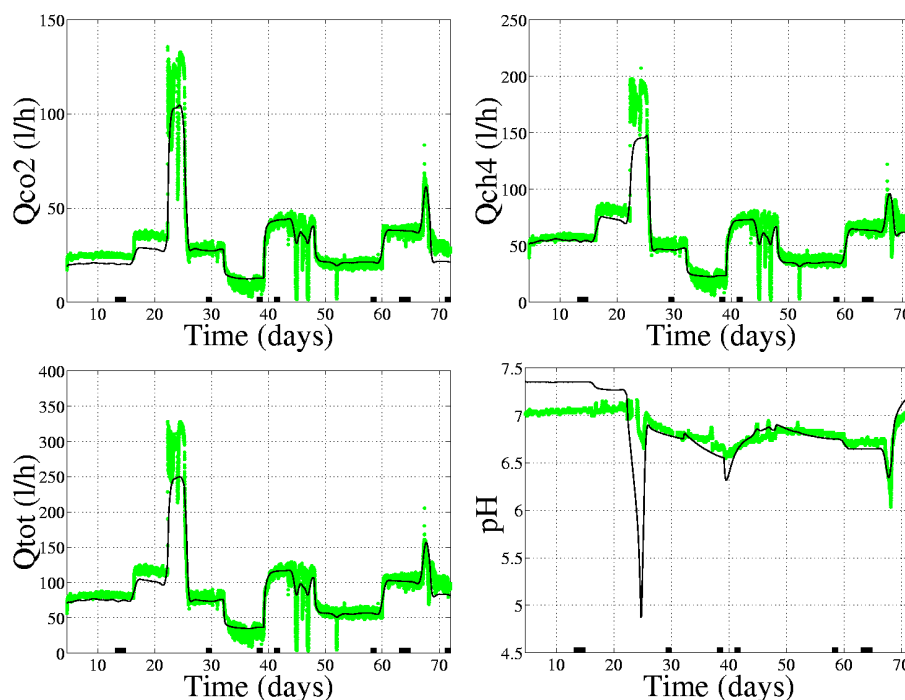


FIG. 5.9 – Comparaison entre mesures et résultats de simulation pour les débits gazeux et le pH. Données INRA-LBE (Narbonne, France).

entre variables ne sont pas vraiment complémentaires, mais chacune de ces approches propose un éclairage différent sur la validité du modèle. Par ailleurs l'analyse de ses points défailants permet en retour de modifier le modèle. Le nouveau modèle doit à son tour être testé, ce qui fait que la démarche de validation est une approche itérative qui accompagne l'élaboration du modèle.

Comme je l'ai rappelé, il n'est pas possible de valider un modèle. On peut l'invalider (le falsifier), mais le valider nécessiterait un nombre infini d'expériences. On considère que le modèle est valide si dans le panel de données disponibles il n'y a pas de conflit notoire avec les données. Ou du moins, si les zones conflictuelles peuvent être expliquées (stress des organismes, expérience douteuse, défaillance probable d'un capteur, etc.). En définitive et au delà de la question manichéenne de la validité, il convient de déterminer une marge de validité associée au modèle c'est-à-dire de dégager la frontière entre les conditions dans lesquelles le modèle est acceptable et celles dans lesquelles il devient insuffisant.

Gardons à l'esprit que la validation du modèle est relative aux objectifs que l'on s'était fixé au départ. Suivant le degré de précision que l'on attendait du modèle, on acceptera tel modèle qui serait refusé pour une précision supérieure. La validation peut alors prendre des sens fondamentalement différents. Valider un modèle explicatif pourrait consister à vérifier expérimentalement (pour autant que cela soit possible) que les mécanismes postulés se produisent, par exemple en suggérant de nouvelles expériences. Pour valider un modèle développé pour établir un contrôleur, il faut en fait vérifier que le contrôleur fonctionne en pratique, comme la théorie le prédit.

Ces réflexions nous ramènent à un point central de la validation : la possibilité d'organiser les expériences

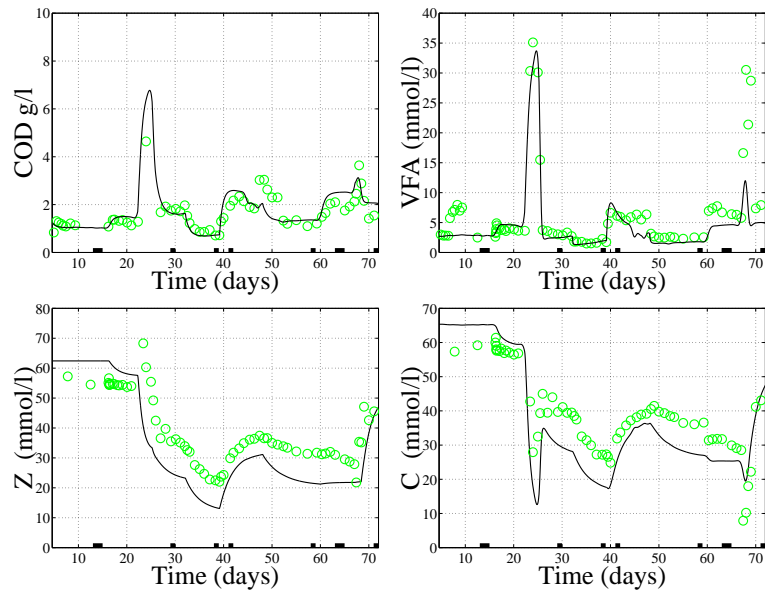


FIG. 5.10 – Comparaison entre mesures et résultats de simulation pour la demande chimique en oxygène (COD) et les acides gras volatiles (VFA), l'alcalinité (Z) et le carbone inorganique dissout (TIC). Données INRA-LBE (Narbonne, France).

inspirées de l'analyse du modèle. Cette condition *sine qua non* à la validation du modèle n'est possible que dans la mesure où nous nous sommes placés dans le cadre "biophysique" que j'ai mentionné en introduction. Ne pas se donner les moyens d'éprouver la théorie générée par le modèle ruinerait toute possibilité de caractériser la pertinence du modèle. Cette notion toute popperienne selon laquelle la science se caractérise par sa réfutabilité est capitale et trop souvent oubliée des biomodélisateurs.



## Chapitre 6

# Conclusion

*“Sebbene di un interesse di giorno in giorno crescente le applicazioni delle matematiche alle scienze biologiche ci appaiono esse pure al loro inizio”*

*“Malgré un intérêt croissant et continu, l'application des mathématiques aux sciences biologiques semble être encore à l'étape initiale”*

Extrait du discours à l'académie de Rome de Vito Volterra lors du cours d'ouverture de l'année académique 1901-1902.

Dans ce mémoire, j'ai organisé mes travaux sur la modélisation de manière à élaborer une méthode constructive et systématique pour développer des modèles de biosystèmes à partir de l'analyse des flux de masse. Cette approche consiste fondamentalement à organiser progressivement la confrontation entre un modèle et les mesures réalisées sur un système biologique.

La première étape de la modélisation consiste à rassembler les principes physiques et chimiques et à faire l'hypothèse d'un réseau réactionnel pour arriver à un modèle de bilan de matière. Mes travaux ont plus particulièrement porté sur l'estimation du nombre de réactions nécessaires pour prétendre reproduire un ensemble de données, sur l'identification des coefficients pseudo-stoechiométriques inconnus, ou bien directement sur la détermination de la matrice pseudo-stoechiométrique.

Dans une deuxième étape, il faut tirer profit des contraintes que le modèle doit respecter et utiliser des relations empiriques pour essayer de déterminer une expression pour les cinétiques. J'ai notamment utilisé ces contraintes pour élaborer des modèles hybrides (utilisant des réseaux de neurones) qui satisfont ces critères. Par ailleurs, j'ai mis en place une méthode pour identifier qualitativement la structure de ces cinétiques réactionnelles.

La validation du modèle, ultime étape de la modélisation — au coeur de mes recherches — est trop souvent négligée. Au cours de cette phase, la qualité du modèle doit être évaluée le plus objectivement possible, et les domaines de fonctionnement pour lesquels il atteint ses limites doivent être identifiés. La

méthode que je présente consiste à élaborer une théorie issue de ce modèle et à vérifier la consistance de cette théorie avec des observations. Mes travaux se sont notamment focalisés d'une part sur la validation du réseau réactionnel et d'autre part sur la caractérisation qualitative des transitoires de certaines classes de systèmes biologiques.

Les méthodes que j'ai présentées sont fortement liées à la vision par bilan de masse des systèmes biologiques qui fait émerger une structure linéaire (les non-linéarités étant regroupées dans les termes de cinétiques réactionnelles). Cette structure permet alors de découpler les deux niveaux d'hypothèses (réseau réactionnel et cinétiques). J'insiste à nouveau sur le mauvais choix du terme "bilan de masse", qui est sans doute usurpé dans le sens où il pourrait sous-entendre que l'on s'intéresse à des systèmes où le bilan est fermé. En fait l'approche sur laquelle je fonde ma démarche [20] s'appuie sur la notion d'invariant réactionnel [45, 46, 47], principe plus général que celui de conservation de la masse. Comme je l'ai expliqué en détail (voir paragraphe 3.7) il est possible d'obtenir des équations de conservation au sein d'un système qui n'est pas conservatif au sens de la masse.

Ces méthodes ont été élaborées sous l'impulsion de domaines très différents tels que l'océanographie (modélisation de croissances phyto- et zooplanctoniques), la lutte biologique (modélisation du développement de populations de coccinelles et de leur prédation de pucerons), l'épuration de l'eau (digestion anaérobie, boues activées, lagunage), ou bien l'agroalimentaire (croissance et bioproduction de champignons filamenteux).

J'insiste pour qu'il soit bien clair que l'approche de la modélisation des systèmes biologiques que j'ai choisie est un parti pris et qu'elle ne saurait prétendre à une quelconque universalité. En effet, le risque de cette approche, qui pourrait s'apparenter à une suite de "recettes", serait de faire émerger un modèle absolu, résultat d'une approche itérative induisant de nombreux retours à l'expérience. Un modèle unique pour décrire un phénomène est certainement dangereux, comme le souligne Bouleau [2]. De manière assez subtile, cet unique modèle risquerait de s'immiscer au niveau de la représentation que se fait le modélisateur de la réalité et de s'y substituer. Il devient alors extrêmement difficile de s'échapper de l'image de la réalité formée par le modèle. On rencontre souvent cette confusion chez les modélisateurs entre le modèle et le système réel, où la réponse à une question sur le réel se situe au niveau du modèle. Cet effet ornière [2] est très pernicieux car il semble engendrer et autoriser une vision biaisée du monde par l'entremise du modèle. Il est donc primordial de souligner que "les procédures de validation fournissent une acceptabilité au modèle et non une légitimité" [2].

Dès lors, que ce soit pour prendre du recul par rapport à **son** modèle ou pour sortir des ornières, il est essentiel d'adopter une approche multimodèles en proposant un ensemble de modèles pour représenter le système biologique. Ces modèles peuvent d'ailleurs avoir des objectifs distincts et apporter chacun un éclairage très différent, voire contradictoire. C'est une de mes problématiques actuelles, je m'attache à gérer simultanément un ensemble de modèles de complexité croissante et à profiter du point de vue fourni par chacun des modèles. Dans le cadre de modèles mécanistes, une autre vertu de l'approche multimodèles est de proposer des explications alternatives. En effet, des modèles basés sur des mécanismes radicalement différents peuvent produire les mêmes phénomènes (cycles limites, excitabilité, etc.). Cette non injectivité de l'ensemble des modèles vers l'ensemble des effets sera sans doute mieux appréhendée et les conclusions plus nuancées si l'on considère une panoplie de modèles explicatifs.

Se détacher d'un modèle particulier, c'est aussi rendre au modèle son vrai statut, celui d'outil. Comme

le dit Legay [4], on ne concevrait pas une boîte à outils avec un unique ustensile ; de même doit on aborder la modélisation. Il vaudrait d'ailleurs mieux parler de "méthode des modèles" [4] que de modélisation pour limiter cette ambiguïté.

A l'extrême, pour bénéficier encore de davantage de recul faudrait-il utiliser simultanément des modèles élaborés par des auteurs indépendants, voire même développés avec des outils et des formalismes différents. A la condition bien sûr que chacun de ces modèles ait pu passer les tests d'acceptabilité que j'ai largement mentionnés.

Je me suis efforcé de démontrer, en reprenant les arguments de Lobry [14] que les nouveaux outils expérimentaux et théoriques nous conduisent naturellement à l'ère de la biophysique. C'est dans ce cadre que nous pouvons envisager sereinement les liens que le modèle entretient avec le système réel. L'idée maîtresse est le très fort couplage entre une théorie sous-tendue par des modèles et un dispositif expérimental qui assure une très bonne reproductibilité, permet d'exciter largement le système et d'en mesurer les réponses. Ces principes sont notamment très bien illustrés par l'ensemble des travaux de l'équipe COMORE sur le dispositif de culture de phytoplancton piloté par ordinateurs, ainsi que par les travaux sur l'écosystème anaérobie très finement contrôlé et mesuré, en collaboration avec Jean-Philippe Steyer (LBE, INRA Narbonne).

Pour être mise en oeuvre, l'approche biophysique nécessite une importante multidisciplinarité qui engage des compétences pointues dans les domaines allant de la biologie aux mathématiques. L'interaction entre ces pôles de compétences est primordiale car elle assure une circulation entre le réel et la théorie. Ce passage obligé est souvent périlleux car le langage naturel est ambiguë et se comprendre est souvent l'étape la plus critique, mais aussi la plus passionnante. L'autre difficulté tient à mon avis à ce qu'il faut tout à la fois respecter et solliciter les expertises de chacun. Pour développer un modèle il faut donc se situer au centre de la chaîne qui lie le biologiste à un bout et le mathématicien à l'autre. Posséder en même temps ces trois expertises est un rêve schizophrène dangereux, même s'il correspond davantage à l'image que nous avons des grands savants.

Sans se substituer à ces experts (qui entrent néanmoins tous dans la catégorie des "biophysiciens"), le biomodélisateur doit donc s'impliquer tant du côté biologique que du côté purement mathématique, condition nécessaire pour organiser correctement la comparaison entre modèle et théorie. De cet équilibre délicat naissent les méthodes propres à la modélisation et c'est l'essence de mes recherches. Ces méthodes s'appuient sur les concepts forts de l'automatique (notion de système, d'entrée, de sortie). Mais il faut bien noter que les outils proposés dans le cadre de l'automatique classique n'apportent pas fondamentalement d'aide pour la modélisation des systèmes non linéaires. C'est assez explicable en remarquant que, pour les systèmes linéaires, modéliser est fondamentalement une question d'identification paramétrique (estimation de la dimension du système, puis de ses paramètres).

Le chemin pour apprendre à élaborer et à utiliser les modèles biologiques est encore long et semé d'embûches et j'ai laissé de très nombreuses questions en suspens. En particulier, les aspects statistiques n'ont pas été développés. Est-il possible, à l'instar des tests statistiques classiques, de proposer des tests d'acceptation des modèles, avec un risque  $\alpha$  de se tromper ? Comment estimer la variabilité associée aux prédictions d'un modèle ? Comment caractériser plus rigoureusement le domaine de validité d'un modèle ? D'autres problèmes cruciaux restent ouverts, comme l'estimation des paramètres de modèles non linéaires lorsque le système est mal mesuré. J'ai abordé très superficiellement certaines de ces questions dans ce mémoire, mais de nombreuses thèses seront encore nécessaires pour progresser dans ces domaines.

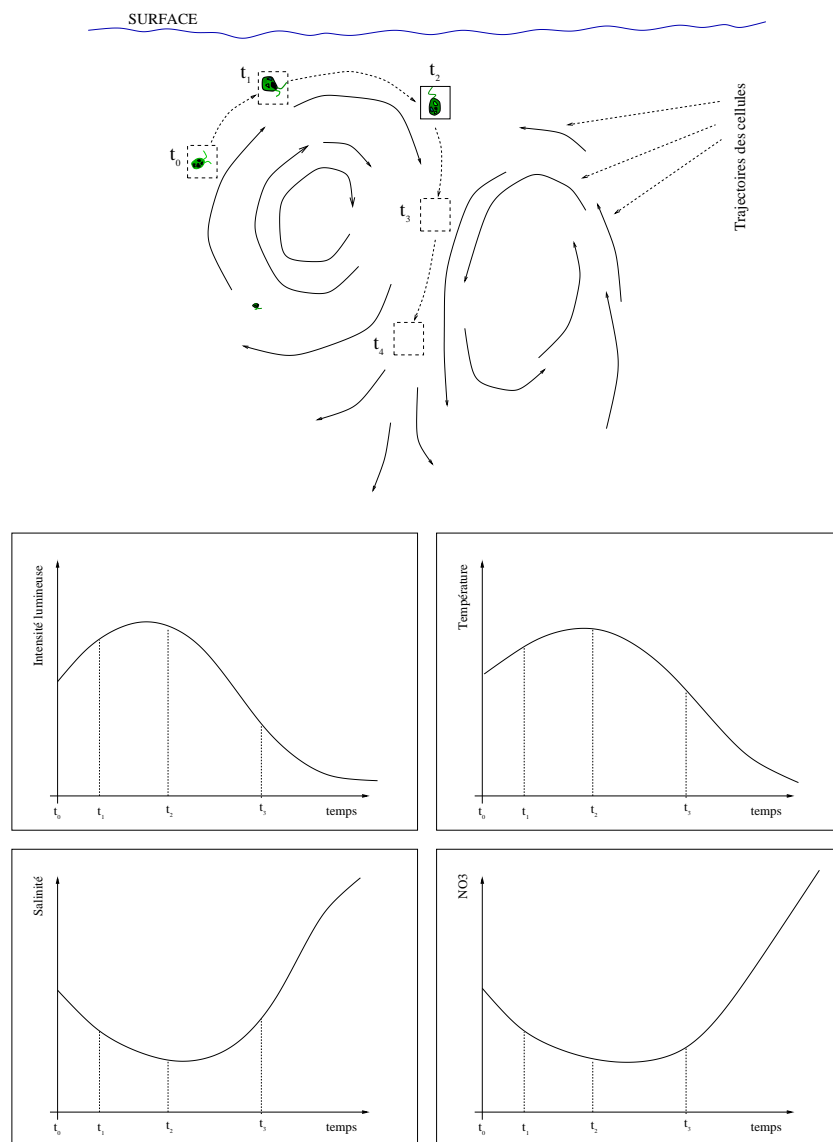


FIG. 6.1 – Schéma caricatural de la trajectoire d’une cellule phytoplanctonique dans une zone frontale engendrant un mélange de deux masses d’eau, et représentation de l’environnement résultant perçu par la cellule.

Si j’ai présenté une approche embryonnaire pour la modélisation systématique des bioréacteurs, cadre de laboratoire extrêmement simplifié, il reste encore un travail considérable pour espérer décrire des écosystèmes *in situ*, qui mettent en jeu un grand nombre d’espèces dans un environnement non contrôlé. La spatialisation y joue un rôle fondamental et ne peut bien souvent plus être négligée. Sans doute faudra-t-il en laboratoire se rapprocher progressivement de ces systèmes en considérant progressivement des bioréacteurs plus complexes.

La synthèse de bioréacteurs de complexité intermédiaire entre un écosystème *in situ* et les bioréacteurs relativement élémentaires dont il a été question dans ce mémoire est un compromis pour s’affranchir des difficultés d’une étude *in situ* tout en restant “plus proche” du système naturel. Ainsi l’Ecotron développé en



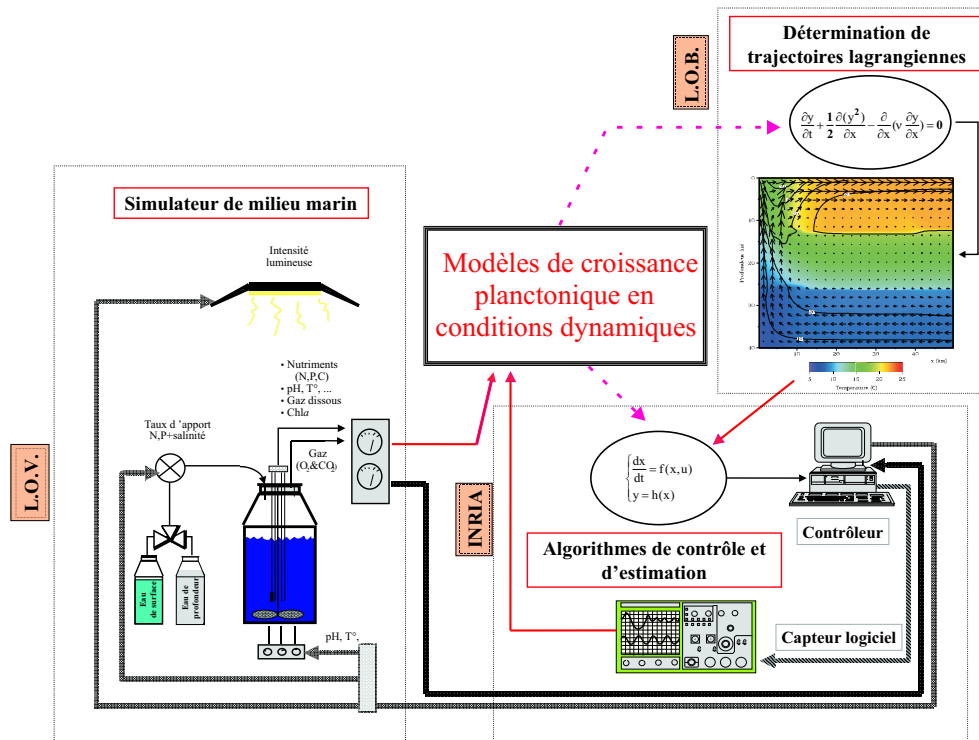


FIG. 6.2 – Schéma synoptique du projet SEMPO.

Grande-Bretagne [124, 125] reproduit en chambres contrôlées des écosystèmes terrestres de taille réduite mais complets, c'est-à-dire comprenant plantes, animaux, micro-organismes, sol et atmosphère. En m'inspirant de cette idée, j'ai mis en place le projet SEMPO (simulateur d'environnement marin piloté par ordinateurs) afin de reproduire *in vitro* l'environnement dynamique perçu par une population de cellules phytoplanctoniques au cours de leur trajet dans la colonne d'eau. En collaboration avec des physiciens (Y. Leredde, COM, Marseille et M. Levy, Lodyc, Paris) nous utilisons des codes de calculs hydrodynamiques pour estimer les trajectoires lagrangiennes suivies par les particules phytoplanctoniques. Ces trajectoires seront ensuite reproduites expérimentalement, en soumettant les cellules de phytoplancton dans le système de culture de Villefanche-sur-mer [16] aux mêmes conditions que celles perçues par les particules phytoplanctoniques dans le modèle hydrodynamique (voir Figure 6.1). Pour réaliser ce SEMPO, cela implique bien évidemment de mettre en place des algorithmes de contrôle et d'estimation d'état (voir Figure 6.2), qui constituent une autre facette de mes recherches [12, 26, 36]. Par la suite, en appliquant les méthodes développées dans ce mémoire, plus traditionnellement développées en dehors du champs de l'écologie, on peut espérer améliorer la représentation et la compréhension de tels systèmes.

Cette progression dans le domaine expérimental devra bien sûr aller de pair avec une progression théorique : la plupart des méthodes proposées s'appliquent à des classes de systèmes simples (modèles en boucle, modèles coopératifs, etc.) et les concepts devront être progressivement élargis. Certaines idées peuvent facilement être transposées à des classes de modèles plus larges, par exemple des systèmes représentés par des équations aux dérivées partielles qui seraient établies à partir de bilans de masse.



# Bibliographie

- [1] R. Thom, *Prédire n'est pas expliquer*. Flammarion, 1993.
- [2] N. Bouleau, *Philosophie des mathématiques et de la modélisation*. l'Harmattan, 1999.
- [3] R. May, "Simple mathematical models with very complicated dynamics," *Nature*, vol. 261, p. 459, 1976.
- [4] J.-M. Legay, *L'expérience et le modèle*. INRA Editions, 199.
- [5] C. Lobry, *Basic principles of Mathematical Modeling*. Encyclopedia Of Life Support Systems, UNESCO Publishing-Eolss Publishers, Oxford, à paraître.
- [6] G. Diatkine, *Jacques Lacan*. Puf, 1997.
- [7] J.-M. Lévy-Leblond, *Physique et mathématique*, pp. 195–210. Penser les mathématiques, Points, 1982.
- [8] G. Israel, *La mathématisation du réel*. Seuil, 1996.
- [9] J.-L. Gouzé, "Analyse mathématique qualitative de modèles biologiques." Mémoire d'Habilitation à diriger les recherches, Université de Nice, 1995.
- [10] A. Lotka, *Elements of Physical Biology*. Williams & Wilkins Co., Baltimore, 1925.
- [11] V. Volterra, "Variations and fluctuations in the numbers of coexisting animal species," in *The golden age of theoretical ecology : 1923-1940* (F. Scudo and J. Ziegler, eds.), vol. 22 of *Lecture notes in biomathematics*, Springer-Verlag, 1978, 1927.
- [12] O. Bernard, G. Sallet, and A. Sciandra, "Nonlinear observers for a class of biological systems. Application to validation of a phytoplanktonic growth model.," *IEEE Trans. Aut. Cont.*, vol. 43, pp. 1056–1065, 1998.
- [13] G. L. Lloyd, *Une histoire de la science grecque*. La découverte, 1990.
- [14] C. Lobry, "La modélisation en écologie. du laboratoire à l'écosystème," in *Actes de la journée annuelle de la SMF - "Mathématiques et biologie"*, pp. 47–65, paris, 2002.
- [15] G. Malara and A. Sciandra, "A multiparameter phytoplanktonic culture system driven by microcomputer," *J. Applied Phycol*, no. 3, pp. 235–241, 1991.
- [16] O. Bernard, G. Malara, and A. Sciandra, "The effects of a controlled fluctuating nutrient environment on continuous cultures of phytoplankton monitored by a computer," *J. Exp. Mar. Biol. Ecol*, no. 197, pp. 263–278, 1996.
- [17] J.-C. Poggiale, "Influence de l'hétérogénéité spatiale sur la structure et le fonctionnement des écosystèmes marins. Méthodes mathématiques et modélisation." Mémoire d'Habilitation à diriger les recherches, Université de Marseille, 2002.

- [18] J.-P. Steyer, “Modélisation, commande et diagnostic des procédés biologiques de dépollution.” Mémoire d’Habilitation à diriger les recherches, Université de Perpignan, 1998.
- [19] J. Godon, E. Zumstein, P. Dabert, F. Habouzit, and R. Moletta, “Molecular microbial diversity of an anaerobic digester as determined by small-subunit rDNA sequence analysis,” *Applied and Environmental Microbiology*, vol. 63, pp. 2802–2813, 1997.
- [20] G. Bastin and D. Dochain, *On-line estimation and adaptive control of bioreactors*. Elsevier, 1990.
- [21] O. Bernard and J.-L. Gouzé, “Transient behavior of biological models as a tool of qualitative validation - Application to the Droop model and to a N-P-Z model,” *J.Biol.Syst.*, vol. 4, no. 3, pp. 303–314, 1996.
- [22] S. Souissi and O. Bernard, *Identification of interactions in copepod populations using a qualitative study of stage-structured population models*, pp. 365–383. Handbook of Scaling Methods in Aquatic Ecology : Measurement, Analysis, Simulation, CRC Press, 2003.
- [23] S. A. L. M. Kooijman, *Dynamic Energy and Mass Budgets in Biological Systems*. Cambridge University Press, Great Britain, 2000.
- [24] M. Loreau, “Mass and energy flow in closed ecosystems : do ecological or mathematical constraints prevail ?,” *J. Theor. Biology*, no. 168, pp. 237–243, 1994.
- [25] O. Bernard and J.-L. Gouzé, *Estimation d’état*, ch. 4, pp. 87–120. Automatique des bioprocédés, Paris : Hermes Science, 2001.
- [26] O. Bernard, A. Sciandra, and G. Sallet, “A non-linear software sensor to monitor the internal nitrogen quota of phytoplanktonic cells,” *Oceanologica Acta*, vol. 24, pp. 435–442, 2001.
- [27] O. Bernard and J.-L. Gouzé, *State estimation for bioprocesses*, pp. 813–855. Trieste : ICTP, 2002.
- [28] O. Bernard, Z. Hadj-Sadok, and J.-L. Gouzé, “Observers for the biotechnological processes with unknown kinetics. application to wastewater treatment,” in *Proceedings of CDC 2000*, Sydney, Australia, 2000.
- [29] O. Bernard, Z. Hadj-Sadok, and D. Dochain, “Software sensors to monitor the dynamics of microbial communities : application to anaerobic digestion,” *Acta Biotheo.*, vol. 48, pp. 197–205, 2000.
- [30] M. Fruchard, O. Bernard, and J.-L. Gouzé, “Interval observers with confidence levels. application to the activated sludge process,” in *Proceedings of the IFAC World Congress*, pp. CD-ROM, Barcelona, Spain, 2002.
- [31] O. Bernard and J.-L. Gouzé, “Closed loop observers bundle for uncertain biotechnological models,” *J. Process. Contr.* à paraître.
- [32] O. Bernard, M. Polit, Z. Hadj-Sadok, M. Pengov, D. Dochain, M. Estaben, and P. Labat, “Advanced monitoring and control of anaerobic wastewater treatment plants : software sensors and controllers for an anaerobic digester,” *Wat. Sci. Technol.*, vol. 43, no. 7, pp. 175–182, 2001.
- [33] L. Mailleret, O. Bernard, and J.-P. Steyer, “Contrôle asymptotique non linéaire des fermenteurs anaérobies,” in *Proceedings of the CIFA Congress*, pp. 301–306, Nantes, France, 2002.
- [34] L. Mailleret and O. Bernard, “A simple robust controller to stabilise an anaerobic digestion process,” in *Proceedings of the computer applications in biotechnology-CAB8*, pp. 213–218, Quebec, 2001.

- [35] L. Mailleret, O. Bernard, and J.-P. Steyer, "Robust regulation of anaerobic digestion processes," *Wat. Sci. Technol.*, vol. 48, pp. 87–94, 2003.
- [36] L. Mailleret, J.-L. Gouzé, and O. Bernard, "Nonlinear control for algae growth models in the chemostat," in *Proceedings of ECC03*, Cambridge, UK, 2003.
- [37] L. Mailleret, O. Bernard, and J.-P. Steyer, "Robust nonlinear adaptive control for bioreactors with unknown kinetics," *Automatica*. accepté.
- [38] J. L. Gouzé, A. Rapaport, and Z. Hadj-Sadok, "Interval observers for uncertain biological systems," *Ecological modelling*, vol. 133, pp. 45–56, 2000.
- [39] M. Z. Hadj-Sadok and J. L. Gouzé, "Bounds estimation for uncertain models of wastewater treatment," *IEEE International Conf. on Control and Applications, Trieste, Italy*, pp. 336–340, 1998.
- [40] O. Bernard and J.-L. Gouzé, "Closed loop observers bundle for uncertain biotechnological models," in *Proceedings of CAB9*, Nancy, F, 2004. à paraître.
- [41] B. Chachuat and O. Bernard, "Design of probabilistic observers for mass-balance based bioprocess models," in *Proceedings of CAB9*, Nancy, F, 2004. à paraître.
- [42] L. Mailleret, O. Bernard, and J.-P. Steyer, "Contrôle asymptotique non-linéaire des fermenteurs anaérobie," *Journal Européen des systèmes automatisés*, vol. 37, pp. 275–286, 2003.
- [43] O. Bernard and I. Queinnec, *Modèles dynamiques de procédés biochimiques. Propriétés des modèles*, ch. 2, pp. 23–52. Automatique des bioprocédés, Paris : Hermes Science, 2001.
- [44] O. Bernard, *Mass balance modelling of bioprocesses*, pp. 769–812. Trieste : ICTP, 2002.
- [45] M. Fjeld, "On a pitfall in stability analysis of chemical reactions," *Chem. Engin. Science*, vol. 23, pp. 565–573, 1968.
- [46] O. Asbjørnsen and M. Fjeld, "Response modes of continuous stirred tank reactors," *Chem. Engin. Science*, vol. 25, pp. 1627–1636, 1970.
- [47] O. Asbjørnsen, "Reaction Invariants in the Control of Continuous Chemical Reactors," *Chem. Engin. Science*, vol. 27, pp. 709–717, 1972.
- [48] J. E. Bailey and D. F. Ollis, *Biochemical engineering fundamentals*. McGraw-Hill, 1986.
- [49] O. Bernard, C. Bonvillain, D. Benyamina, M. Schaegger, A. Pauss, and D. Dochain, "Modelling of extensive wastewater treatment plants," in *Proceedings of the International Symposium on Environmental Biotechnology*, pp. 373–376, Oostende, Belgium, 1997.
- [50] O. Bernard, Z. Hadj-Sadok, D. Dochain, A. Genovesi, and J.-P. Steyer, "Dynamical model development and parameter identification for an anaerobic wastewater treatment process," *Biotech. Bioeng.*, no. 75, pp. 424–438, 2001.
- [51] O. Bernard, G. Bastin, C. Stentelaire, L. Lesage-Meessen, and M. Asther, "Mass balance modelling of vanillin production from vanillic acid by cultures of the fungus *Pycnoporus cinnabarinus* in bioreactors," *Biotech. Bioeng.*, pp. 558–571, 1999.
- [52] O. Bernard and G. Bastin, "Identification of reaction schemes for bioprocesses : determination of an incompletely known yield matrix," in *Proceedings of ECC03*, Cambridge, UK, 2003.

- [53] D. Hill and C. Barth, "A dynamic model for simulation of animal waste digestion," *Journal of the Water Pollution Control Association*, vol. 10, pp. 2129–2143, 1977.
- [54] F. Mosey, "Mathematical modelling of the anaerobic digestion process : regulatory mechanisms for the formation of short-chain volatile acids from glucose," *Water Science and Technology*, vol. 15, pp. 209–232, 1983.
- [55] R. Moletta, D. Verrier, and G. Albagnac, "Dynamic modelling of anaerobic digestion," *Wat.Res.*, vol. 20, pp. 427–434, 1986.
- [56] D. Costello, P. Greenfield, and P. Lee, "Dynamic modelling of a single-stage high-rate anaerobic reactor - I. Model derivation," *Water Research*, vol. 25, pp. 847–858, 1991.
- [57] D. Costello, P. Greenfield, and P. Lee, "Dynamic modelling of a single-stage high-rate anaerobic reactor - II. Model verification," *Water Research*, vol. 25, pp. 859–871, 1991.
- [58] L. Fernandes, K. Kennedy, and Z. Ning, "Dynamic modelling of substrate degradation in sequencing batch anaerobic reactors (SBAR)," *Water Research*, vol. 27, pp. 1619–1628, 1993.
- [59] G. Kiely, G. Tayfur, C. Dolan, and K. Tanji, "Physical and mathematical modelling of anaerobic digestion of organic wastes," *Water Research*, vol. 31, pp. 534–540, 1997.
- [60] D. Batstone, J. Keller, B. Newell, and M. Newland, "Model development and full scale validation for anaerobic treatment of protein and fat based wastewater," *Water Science and Technology*, vol. 36, pp. 423–431, 1997.
- [61] D. Batstone, J. Keller, R. Angelidaki, S. Kalyuzhnyi, S. Pavlostathis, A. Rozzi, W. S. H. Siegrist, and V. Vavilin, "The IWA anaerobic model no 1," in *Proceedings of the 9th IWA World Congress on Anaerobic digestion*, Antwerp, Belgium, 2001.
- [62] IWA Task Group, *Anaerobic Digestion Model No. 1 (ADM1)*. London : IWA Publishing, 2002.
- [63] J. Merchuk, "Further considerations on the enhancement factor for oxygen absorption into fermentation broth," *Biotechnol. & Bioeng.*, vol. 19, pp. 1885–1889, 1977.
- [64] O. Bernard, Z. Hadj-Sadok, D. Dochain, A. Genovesi, and J.-P. Steyer, "Dynamical model development and parameter identification for an anaerobic wastewater treatment process," *Biotech.Bioeng.*, to appear.
- [65] R. Horn and C. Johnson, *Matrix analysis*. Cambridge University Press, 1992.
- [66] B. Falconnier, C. Lapiere, L. Lesage-Meessen, G. Yonnet, P. Brunerie, B. Colonna-Ceccaldi, G. Corrieu, and M. Asther, "Vanillin as a product of ferulic acid biotransformation by the white-rot fungus *Pycnoporus cinnabarinus* I-937 : identification of metabolic pathways," *J. Biotechnol*, vol. 37, pp. 123–132, 1994.
- [67] L. Lesage-Meessen, M. Delattre, M. Haon, J. Thibault, B. Colonna-Ceccaldi, P. Brunerie, and M. Asther, "A two-step bioconversion process for vanillin production from ferulic acid combining *Aspergillus niger* and *Pycnoporus cinnabarinus*," *J. Biotechnol*, vol. 50, pp. 107–113, 1996.
- [68] P. Serra, J. del Rio, J. Robusté, M. Poch, C. Sola, and A. Cheruy, "A model for lipase production by candida rugosa," *Bioprocess Engineering*, vol. 8, pp. 145–150, 1992.
- [69] L. Chen, O. Bernard, G. Bastin, and P. Angelov, "Hybrid modelling of biotechnological processes using neural networks," *Contr. Eng. Practice*, vol. 8, pp. 821–827, 2000.

- [70] A. Karama, O. Bernard, A. Genovesi, D. Dochain, A. Benhammou, and J.-P. Steyer, "Hybrid modelling of anaerobic wastewater treatment processes," *Wat. Sci. Technol.*, vol. 43, no. 1, pp. 43–50, 2001.
- [71] O. Bernard and G. Bastin, "Modelling of biological kinetics using a new structural criterion," in *Proceedings of the computer applications in biotechnology-CAB7*, pp. 95–100, Osaka, May 31–June 4, 1998, 1998.
- [72] J. Monod, *Recherches sur la Croissance des Cultures Bactériennes*. Paris : Hermann, 1942.
- [73] L. A. Segel, *Modeling Dynamic Phenomena in Molecular and Cellular Biology*. Cambridge : Cambridge University Press, 1984.
- [74] L. Edelstein, *Mathematical Models in Biology*. New York : Random House, 1988.
- [75] J. Andrews, "A mathematical model for the continuous culture of microorganisms utilizing inhibitory substrate," *Biotechnol. & Bioeng.*, vol. 10, pp. 707–723, 1968.
- [76] D. Dochain, *On line parameter estimation, adaptive state estimation and control of fermentation processes*. PhD thesis, Université Catholique de Louvain-la-Neuve, Belgique, 1986.
- [77] G. Stephanopoulos, A. Aristidou, and J. Nielsen, *Metabolic Engineering*. Elsevier Science, 1998.
- [78] J. Hertz, A. Krogh, and R. Palmer, *Introduction to the Theory of Neural Computation*. Addison-Wesley, 1991.
- [79] A. Karama, O. Bernard, J. Gouzé, A. Benhammou, and D. Dochain, "Hybrid neural modelling of an anaerobic digester with respect to biological constraints," *Wat. Sci. Technol.*, vol. 43, no. 7, pp. 1–8, 2001.
- [80] O. Bernard and G. Bastin, "On the estimation of the pseudo-stoichiometric matrix for mass balance modeling of biotechnological processes," *Math. Biosciences*, submitted.
- [81] I. Gijbels, P. Hall, M. Jones, and I. Koch, "Tests for monotonicity of a regression mean with guaranteed level," *Biometrika*, no. 87, pp. 663–673, 2000.
- [82] E. Walter and L. Pronzato, *Identification of parametric models from experimental data*. Springer, 1997.
- [83] L. Pawlowski, O. Bernard, E. L. Floc'h, and A. Sciandra, "Qualitative behaviour of a phytoplankton growth model in photobioreactor," in *Proceedings of the IFAC World Congress*, pp. CD-ROM, Barcelona, Spain, 2002.
- [84] O. Bernard and J.-L. Gouzé, "Global qualitative behavior of a class of nonlinear biological systems : application to the qualitative validation of phytoplankton growth models," *Artif. Intel.*, vol. 136, pp. 29–59, 2002.
- [85] O. Bernard and J.-L. Gouzé, "Nonlinear qualitative signal processing for biological systems : application to the algal growth in bioreactors," *Math. Biosciences*, vol. 157, pp. 357–372, 1999.
- [86] O. Bernard and S. Souissi, "Qualitative behavior of stage-structured populations : application to structural validation," *J. Math. Biol.*, vol. 37, pp. 291–308, 1998.
- [87] O. Bernard and J.-L. Gouzé, "Robust validation of uncertain models," in *Proceedings of the Third European Conference on Control* (A. Isidori, ed.), pp. 1261–1266, Rome, Italy, 1995.
- [88] O. Bernard and J.-L. Gouzé, "Transient behavior of biological loop models, with application to the Droop model," *Mathematical Biosciences*, vol. 127, no. 1, pp. 19–43, 1995.

- [89] I. Vatcheva, O. Bernard, H. deJong, and N. Mars, "Discrimination of semi-quantitative models by experiment selection : method and application in population biology," in *Proceedings of the IJCAI conference*, Washington, U.S. 4-10 Aug, 2001, 2001.
- [90] I. Vatcheva, O. Bernard, H. deJong, and N. Mars, "Experiment selection for the discrimination of semi-quantitative models of dynamical systems," *Artif. Intel.*, Soumis.
- [91] J.-L. Gouzé, "Positivity, space scale, and convergence towards the equilibrium," *Journal of Biological Systems*, vol. 3, no. 2, pp. 613–620, 1995.
- [92] L. Chen and G. Bastin, "Structural identifiability of the yield coefficients in bioprocess models when the reaction rates are unknown," *Math. Biosciences*, vol. 132, pp. 35–67, 1996.
- [93] S. R. Hansen and S. P. Hubbell, "Single-nutrient microbial competition," *Science*, vol. 207, no. 28, pp. 1491–1493, 1980.
- [94] R. Arditi and L. R. Ginzburg, "Coupling in predator-prey dynamics : ratio-dependence.," *Journal of Theoretical Biology*, no. 139, pp. 311–326, 1989.
- [95] R. Arditi, N. Perrin, and H. Saïah, "Functional responses and heterogeneity : an experimental test with cladocerans," *Oikos*, no. 60, pp. 69–75, 1991.
- [96] R. Arditi, L. R. Ginzburg, and H. R. Akçakaya, "Variation in plankton densities among lakes : a case for ratio-dependent predation models.," *The American Naturalist*, no. 138, pp. 1287–1296, 1991.
- [97] R. Arditi and H. Saïah, "Empirical evidence of the role of heterogeneity in ratio-dependent consumption," *Ecology*, no. 73, pp. 1544–1551, 1992.
- [98] R. Arditi, L. Ginzburg, and N. Perrin, "Scale invariance is a reasonable approximation in predation models - reply to Ruxton and Gurney," *Oikos*, no. 65, pp. 334–335, 1992.
- [99] H. R. Akçakaya, R. Arditi, and L. Ginzburg, "Ratio-dependent predation : an abstraction that works," *Ecology*, no. 76, pp. 995–1004, 1995.
- [100] C. Jost, *Comparaison qualitative et quantitative de modèles proie-prédateur à des données chronologiques en écologie*. PhD thesis, Institut national Agronomique Paris-Grignon, 1995.
- [101] G. Rhee and I. Gotham, "The effect of environmental factors on phytoplankton growth : light and the interactions of light with nitrate limitation," *Limnol. Oceanogr.*, vol. 26, no. 4, pp. 649–659, 1981.
- [102] P. Falkowski and T. Owens, "Light-shade adaptation : two strategies in marine phytoplankton," *Plant Physiol.*, vol. 66, pp. 592–595, 1980.
- [103] A. Sciandra, J. Gostan, Y. Collos, C. Descolas-Gros, C. Leboulanger, V. Martin-Jézéquel, M. Denis, D. Lefèvre, C. C. Copin, and B. Avril, "Growth compensating phenomena in continuous cultures of *Dunaliella tertiolecta* limited simultaneously by light and nitrate," *Limnol. Oceanogr.*, vol. 46, pp. 1325–1339, 1997.
- [104] R. Geider, H. Macintyre, and T. Kana, "A dynamic regulatory model of phytoplankton acclimation to light, nutrients and temperature," *Limnol. Oceanogr.*, vol. 43, no. 4, pp. 679–694, 1998.
- [105] E. Laws and T. Bannister, "Nutrient- and light-limited growth of *Thalassiosira fluviatilis* in continuous culture with implications for phytoplankton growth in the ocean," *Limnol. Oceanogr.*, vol. 25, no. 3, pp. 457–473, 1980.



- [106] M. Droop, "Vitamin B12 and marine ecology. IV. the kinetics of uptake growth and inhibition in *Monochrysis lutheri*," *J. Mar. Biol. Assoc.*, vol. 48, no. 3, pp. 689–733, 1968.
- [107] M. Droop, "25 years of algal growth kinetics, a personal view," *Botanica marina*, vol. 16, pp. 99–112, 1983.
- [108] M. Chalup and E. Laws, "A test of the assumptions and predictions of recent microalgal growth models with the marine phytoplankter *Pavlova lutheri*," *Limnol. Oceanogr.*, vol. 35, no. 3, pp. 583–596, 1990.
- [109] L. Perko, *Differential equations and dynamical systems*. Springer, 1991.
- [110] B. Kuipers, *Qualitative reasoning*. Cambridge : MIT Press, 1994.
- [111] H. de Jong, "Modeling and simulation of genetic regulatory systems : A literature review," *Journal of Computational Biology*, vol. 9, no. 1, pp. 69–105, 2002.
- [112] D. Berleant and B. Kuipers, "Qualitative and quantitative simulation : bridging the gap," *Artif. Intell.*, vol. 95, pp. 215–255, 1997.
- [113] E. Sacks, "Automatic qualitative analysis of dynamic systems using piecewise linear approximations," *Artif. Intell.*, vol. 41, pp. 313–364, 1990.
- [114] E. Sacks, "A dynamic systems perspective on qualitative simulation," *Artif. Intell.*, vol. 42, pp. 349–362, 1990.
- [115] O. Dordan, "Mathematical problems arising in qualitative simulation of a differential equation," *Artif. Intell.*, vol. 55, pp. 61–86, 1992.
- [116] B. C. Goodwin, "Oscillatory behaviour in enzymatic control processes," in *Advances in Enzymatic Regulation* (G. Weber, ed.), Oxford : Pergamon, 1965.
- [117] D. S. Levine, "Qualitative theory of a third order nonlinear system with examples in population dynamics and chemical kinetics," *Mathematical Biosciences*, vol. 77, pp. 17–33, 1985.
- [118] D. Burmaster, "The unsteady continuous culture of phosphate-limited *Monochrysis lutheri* Droop : experimental and theoretical analysis," *J. Exp. Mar. Biol. Ecol.*, vol. 39, no. 2, pp. 167–186, 1979.
- [119] H. Caswell, *Matrix Population Models*. Sunderland, Massachusetts : Sinauer Associates, Inc. Publishers, 1989.
- [120] F. J. Oyarzun and K. Lange, "The attractiveness of the Droop equations. II : Generic uptake and growth functions," *Mathematical Biosciences*, vol. 121, pp. 127–139, 1994.
- [121] D. Contois, "Kinetics of bacterial growth : Relationship between population density and specific growth rate of continuous cultures.," *Journal of General Microbiology*, vol. 21, pp. 40–50, 1959.
- [122] J. Caperon and J. Meyer, "Nitrogen-limited growth of marine phytoplankton. ii. uptake kinetics and their role in nutrient limited growth of phytoplankton," *Deep-Sea Res.*, vol. 19, pp. 619–632, 1972.
- [123] J.-P. Steyer and O. Bernard, "An example of the benefits obtained from the long term use of mathematical models in wastewater biological treatment," in *Proceedings of the MathMod conference*, pp. 245–251 , February,5-7, Vienna, Austria, 2003.
- [124] J. H. Lawton, "Ecological experiments with model systems," *Science*, no. 269, pp. 328–331, 1995.
- [125] J. H. Lawton, "The Ecotron facility at Silwood Park : the value of "big bottle" experiments," *Ecology*, no. 77, pp. 665–669, 1996.