

50316  
2003  
225

UNIVERSITÉ DES SCIENCES ET TECHNOLOGIES DE LILLE

Année 2003

N° d'ordre : \* \* \* \*

# THÈSE

pour obtenir le grade de

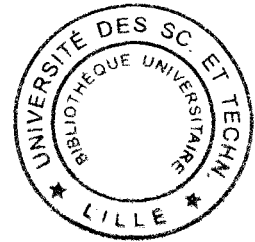
DOCTEUR DE L'UNIVERSITÉ DES SCIENCES ET TECHNOLOGIES DE LILLE

Discipline : AUTOMATIQUE et INFORMATIQUE INDUSTRIELLE

présentée

par

**Marc BAGUELIN**



## MODÉLISATION DE SYSTÈMES COMPLEXES PAR MÉTADYNAMIQUES : APPLICATION À LA MODÉLISATION DE POPULATIONS EN COÉVOLUTION

Soutenance prévue le lundi 1<sup>er</sup> décembre 2003

COMPOSITION DU JURY :

M. Pierre AUGER	Rapporteur	Professeur à l'Université Claude Bernard
M. Alain RICHARD	Rapporteur	Professeur à l'Université Henri Poincaré
M. Jacques LEFÈVRE	Codirecteur de thèse	Professeur PAST à l'École Centrale de Lille
M. Jean-Pierre RICHARD	Codirecteur de thèse	Professeur à l'École Centrale de Lille
M. Jean-Charles CAILLIEZ	Examineur	Institut Pasteur de Lille
M. Jean-Philippe CASSARD	Examineur	Professeur à l'USTL
M <sup>me</sup> Geneviève DAUPHIN-TANGUY	Examineur	Professeur à l'École Centrale de Lille
M. Khashayar PAKDAMAN	Examineur	Chargé de recherche à l'INSERM

Thèse préparée au Laboratoire d'Automatique et d'Informatique industrielle de Lille  
L.A.I.L., UMR 8021 - École Centrale de Lille



בְּזֵיעַת אִפְיֶךָ תֹאכַל לֶחֶם  
עַד שׁוֹבֶכְךָ אֶל-הָאֲדָמָה כִּי מִמֶּנָּה לָקַחְתָּ  
כִּי-עָפָר אַתָּה וְאֶל-עָפָר תָּשׁוּב:

בארשית ג'ט

C'est à la sueur de ton visage que tu mangeras du pain,  
jusqu'à ce que tu retournes dans la terre, d'où tu as été pris ;  
car tu es poussière et tu retourneras dans la poussière.

Genèse 3, 19.

# Table des matières

<b>Introduction</b>	<b>6</b>
<b>1 Le problème</b>	<b>9</b>
1.1 Premier problème : Modélisation et formalisation des systèmes dynamiques adaptatifs complexes . . . . .	10
1.1.1 Définition et objectif du problème . . . . .	11
1.1.2 Réponses déjà existantes . . . . .	15
1.1.3 Une extension dans le cadre de la théorie des systèmes dynamiques : idée générale du travail . . . . .	25
1.2 Deuxième problème : modélisation de systèmes évolutifs . . . . .	28
1.2.1 <i>The evolution in a nutshell</i> . . . . .	29
1.2.2 Les modèles d'évolution existants . . . . .	33
1.2.3 Notre modèle de co-évolution : le modèle Attaque-Défense . . . . .	40
1.2.4 Conclusion . . . . .	41
<b>2 Métadynamiques : aspects formels</b>	<b>43</b>
2.1 Cadre formel général . . . . .	43
2.1.1 Points importants du chapitre précédent . . . . .	44
2.1.2 Définition des systèmes dynamiques de Kalman . . . . .	45
2.1.3 Définition des systèmes à métadynamique . . . . .	47
2.1.4 Un exemple original : le tracé du lapin de Douady . . . . .	53
2.1.5 Extensions . . . . .	57
2.1.6 Conclusion . . . . .	59
2.2 Définitions . . . . .	59
2.2.1 Notions fondamentales . . . . .	59
2.2.2 Caractérisation des méta-états . . . . .	60
2.3 Conclusion . . . . .	62
<b>3 SAM à base continue : aspect stochastique, justification numérique et simulation.</b>	<b>63</b>
3.1 Du modèle multi-agents au modèle particulière . . . . .	64
3.1.1 Modélisation multi-agents . . . . .	66
3.1.2 Modèle particulière . . . . .	66

3.2	Rappels sur les modèles stochastiques . . . . .	67
3.2.1	Fonction de répartition, densité de probabilité et risque instantané . . . . .	67
3.2.2	Loi exponentielle ou loi sans vieillissement . . . . .	69
3.2.3	Application à la simulation stochastique d'une réaction chimique généralisée . . . . .	73
3.2.4	Quelques systèmes classiques revisités . . . . .	77
3.3	Du modèle particulaire au SAM à base continue . . . . .	93
3.3.1	Un modèle non évolutif simple de réacteur bactérien . . . . .	94
3.3.2	Ajout de l'aspect évolutif (Modèle particulaire évolutif) . . . . .	100
3.3.3	Etablissement du SAM équivalent . . . . .	102
3.4	Aspects de simulation . . . . .	107
3.4.1	Algorithmes et structure de données . . . . .	107
3.5	Méthode numérique d'analyse des données . . . . .	108
3.6	Conclusion . . . . .	109
<b>4</b>	<b>Modèle Attaque-Défense</b>	<b>111</b>
4.1	Présentation générale . . . . .	111
4.1.1	Mécanismes physiques et biologiques . . . . .	112
4.1.2	Structure du modèle . . . . .	116
4.2	Description du modèle attaque-défense . . . . .	118
4.2.1	Composition du système . . . . .	119
4.2.2	Hypothèses . . . . .	119
4.2.3	Définition du MAD sous forme de SAM . . . . .	124
4.3	Détails sur l'implémentation . . . . .	127
4.3.1	Le choix de l'environnement . . . . .	128
4.3.2	Les structures de données . . . . .	128
4.3.3	La partie dynamique . . . . .	129
4.3.4	La partie métadynamique . . . . .	130
4.4	Analyse du modèle . . . . .	130
4.4.1	Aspects globaux . . . . .	131
4.4.2	Représentation des populations et définitions utiles . . . . .	137
4.5	Résultats de simulations . . . . .	141
4.5.1	Aspect général de simulation . . . . .	141
4.5.2	Comportement du système sans mutation . . . . .	142
4.5.3	Mobilité des phages dans l'espace des CB . . . . .	143
4.5.4	Fuite des bactéries . . . . .	152
4.5.5	Deux cas particulier du cas général . . . . .	155
4.6	Conclusions . . . . .	163

<b>5</b>	<b>Conclusions et perspectives</b>	<b>165</b>
5.1	Synthèse . . . . .	165
5.2	Perspectives . . . . .	168
5.2.1	Sur le formalisme des SAM . . . . .	168
5.2.2	Sur le modèle attaque-défense . . . . .	168
5.2.3	Sur la modélisation de la théorie de l'évolution . . . . .	169

# Introduction

## Positionnement des travaux

Nos travaux se situent dans le domaine de l'étude des systèmes complexes (dont la dénomination anglo-saxonne est *complexity*) et plus précisément des systèmes dynamiques adaptatifs complexes (SDAC). Les systèmes considérés appartiennent à des domaines où la théorie des systèmes dynamiques classiques à du mal à s'appliquer de manière efficace : système immunitaire, réseaux de neurones, écosystèmes ou encore économie. Nous nous intéressons à deux problèmes spécifiques : la formalisation de notions permettant de modéliser les SDAC et la modélisation des systèmes de populations évoluant au sens de la théorie de l'évolution.

Nos objectifs sont :

- donner une définition permettant de définir de manière rigoureuse, l'équivalent de notions classiques en théorie des systèmes dynamiques (équilibre, attracteur...), non plus pour les grandeurs du système, mais pour la structure elle-même,
- justifier cette définition en montrant comment elle apparaît naturellement dans un processus d'approximations successives d'un modèle,
- proposer un modèle de co-évolution de populations, suffisamment simple pour pouvoir illustrer les notions développées auparavant et suffisamment riche pour présenter un intérêt du point de vue de la modélisation biologique.

## Présentation du mémoire

**Articulation logique de la thèse** Avant de détailler le contenu des différents chapitres, nous allons présenter l'articulation logique de notre travail. Le chapitre 1 définit les deux problèmes qui nous intéressent : la formalisation des systèmes dynamiques adaptatifs complexes et la modélisation de systèmes de populations en évolution. Ce chapitre précise également les liens qui les unissent et décrit les approches déjà existantes. Sur base de ces travaux, nous explicitons au chapitre 2, la définition formelle annoncée, que nous expliquons, illustrons et enrichissons de définitions utiles. Le chapitre 3 établit un lien entre le chapitre 2, théorique et le chapitre 4, application à un exemple biologique précis. Il a pour but de montrer l'utilité et la pertinence de la structure de modélisation décrite au chapitre 2 aux problèmes de populations soumises à l'évolution darwinienne,

auquel le modèle du chapitre 4 est soumis. Le chapitre 5 a pour but de dresser un bilan du travail et d'expliciter les perspectives.

**Premier chapitre** Le chapitre 1 constitue l'introduction proprement dite de ce travail, nous y présentons le problème, ou plutôt le double problème : la formalisation des systèmes adaptatifs complexes (SDAC) et la modélisation des systèmes soumis aux lois de l'évolution (au sens de la théorie de l'évolution en biologie). Nous présentons dans un premier temps un certain nombre de formalismes déjà existants en expliquant quelles sont leur limites dans le cas des SDAC. Cela nous permet de dégager un certain nombre de points auxquels devrait répondre un tel formalisme. Nous abordons ensuite le problème de la modélisation de l'évolution darwinienne et présentons un certain nombre de modèles déjà existants.

**Deuxième chapitre** Nous définirons, au chapitre 2, la notion de système à métadynamique (SAM). Cette notion permet de réunir sous un même formalisme les exemples présentés au chapitre 1. La définition formelle du concept SAM se base sur la définition des systèmes dynamiques formulée par Kalman. Après avoir présenté un certain nombre d'extensions de cette définition, nous développons un exemple original d'algorithme de tracé d'un ensemble de Julia, basé sur un SAM. Ce chapitre se termine par des définitions permettant de caractériser les métatracétoires des SAM nouvellement définis.

**Troisième chapitre** Nous justifierons, au chapitre 3, le passage d'un modèle multi-agents à un modèle basé sur un SAM à base continue dans le cas d'un modèle de population. Nous utiliserons pour cela une modélisation intermédiaire, les réactions chimiques généralisées (RCG), simulées stochastiquement à l'aide de l'algorithme de Gillespie. Nous montrons qu'il est possible de représenter de cette manière un certain nombre de modèles biologiques classiques (modèle de Lotka-Volterra et modèle de Tsuchiya). Nous discutons des vertus respectives des méthodes stochastique et différentielle, en particulier la convergence de l'une vers l'autre, quand le nombre de particules est grand.

**Quatrième chapitre** Le chapitre 4 expose en détail un exemple de modèle évolutif défini par un SAM. Il met en œuvre un petit écosystème bactérien comprenant des populations de bactéries et de bactériophages (les virus des bactéries) en interaction. Ces populations, grâce à une structure de données propre, peuvent évoluer. On s'intéresse aux relations entre la dynamique au sens classique (évolution au cours du temps de certaines grandeurs) d'une part et d'autre part, la métadynamique, c'est-à-dire comment la structure de la population évolue.



# Chapitre 1

## Le problème

“Qui cueille une fleur dérange une étoile.”

THÉODORE MONOD

Cette citation de Théodore Monod évoque de manière particulièrement poétique à quel point toute chose est relié à toute autre dans l'univers. Mais paradoxalement, de l'étoile si lointaine et de la fleur si proche, c'est sûrement l'étoile que nous comprenons le mieux. La fleur si fragile, fait partie, en tant qu'être vivant qui s'est adapté sur la terre, des systèmes adaptatifs complexes, dont l'étude s'est beaucoup développée ces dernières années. Les travaux présentés dans ce mémoire de thèse concernent l'étude de deux problèmes distincts, mais néanmoins très liés comme l'illustre l'exemple de la fleur : le problème de la formalisation des systèmes dynamiques adaptatifs complexes et celui de la modélisation de systèmes soumis aux lois de l'évolution au sens de la théorie de l'évolution en biologie.

Nos travaux se rattachent à la théorie des systèmes dynamiques complexes<sup>1</sup>, mais aussi, par leur utilisation intensive de l'informatique comme outil de calcul et de simulation de phénomènes biologiques, à un domaine d'étude scientifique appelé un peu abusivement « vie artificielle »<sup>2</sup>. Ce domaine tente d'expliquer ou de reproduire au moyen de l'informatique des phénomènes comme l'expression du

---

<sup>1</sup>Dans la littérature anglo-saxonne, la théorie des systèmes complexes est souvent désignée sous le terme de “complexité” (*complexity*) et se réfère aux travaux développés pour beaucoup au sein du Santa Fe Institute [28, 72]. Cet institut, créé en 1984, a pour but de réunir des scientifiques de renom et d'horizons divers (biologistes, économistes, physiciens, informaticiens, mathématiciens, sociologues...), pour réfléchir autour de la notion de système complexe. Regroupant beaucoup de noms prestigieux (dont plusieurs prix Nobel), le Santa Fe Institute (SFI) a beaucoup attiré les regards et suscité beaucoup d'espoir autour d'une approche unifiée de la question de système complexe. Plusieurs scientifiques français se sont distingués également dans ce domaine, parmi ceux-ci, citons G. Weissbuch.

<sup>2</sup>*Artificial life* : ce nom très “publicitaire” a été choisi par C. Langton pour fédérer un congrès autour d'approches très différentes, mais ayant pour caractéristique commune d'allier informatique et systèmes biologiques autour de ce que nous appellerons les systèmes dynamiques adaptatifs complexes.

code génétique [66], l'évolution en biologie [34] ou le développement du système immunitaire [19]. Ce lien étroit entre biologie et informatique théorique qui existe dans la « vie artificielle » peut-être résumé par la question suivante, posée par un intervenant lors d'une réunion du comité de direction du Santa Fe Institute et cité par M. Gell-Mann dans [17] : « *Are we using computation as an aid in understanding biology (e.g., evolution, thinking, etc.) or are we using biology as a metaphor for work on computation ?* ».<sup>3</sup>

Ainsi, pour revenir aux deux problèmes qui nous concernent, la modélisation de la théorie de l'évolution peut être considérée comme un exemple de modélisation de systèmes adaptatifs, mais, comme on le verra dans la suite de ce chapitre, cette théorie de l'évolution<sup>4</sup> a également été le terreau sur lequel se sont nourries les principales approches de formalisation des systèmes dynamiques adaptatifs complexes.

Pour ces deux problèmes (modélisation et formalisation des systèmes adaptatifs complexes et modélisation des systèmes évolutifs), nous exposerons les approches antérieures et leurs limites et nous présenterons les directions que nous avons choisies dans notre étude.

## 1.1 Premier problème : Modélisation et formalisation des systèmes dynamiques adaptatifs complexes

Il est assez frappant de remarquer que, malgré ses succès dans la modélisation et l'analyse de nombreux systèmes physiques (citons par exemple l'automatique, la mécanique, la physiologie, les oscillateurs chaotiques...), la théorie des systèmes dynamiques se heurte à la modélisation des systèmes dynamiques adaptatifs complexes (SDAC). Ces systèmes ont la particularité de présenter un grand nombre d'éléments de types variés en interaction et de fortes non-linéarité. Ces deux points suffisent déjà à poser d'énormes difficultés, mais il s'ajoute à cela la propriété d'adaptation : pour représenter ces systèmes, on doit faire varier au cours du temps, non seulement leurs paramètres, mais également leur structure ajoutant, supprimant et modifiant des variables et équations en fonction de la dynamique, ce qui pose, vu leur grande taille et donc le nombre extrêmement élevé des modifications possibles, des problèmes insolubles pour leur génération, leur stockage et la commutation entre les différentes structures [25]. Nous présentons ci-après notre approche concernant la modélisation de tels systèmes par le mécanisme des

---

<sup>3</sup>Utilisons-nous l'ordinateur comme une aide pour comprendre la biologie (l'évolution, la pensée, etc.) ou sommes-nous en train d'utiliser la biologie comme une métaphore pour un travail d'informatique ?

<sup>4</sup>au sens large car aussi bien d'organismes vivants que d'espèces chimiques (par complexification de molécules, on parle alors plutôt de chimie artificielle)

métadynamiques. Nous montrons que cette approche regroupe un grand nombre d'approches existantes et révèle des propriétés intéressantes pour modéliser et décrire les SDAC.

Nous commencerons par donner une définition de ces systèmes, puis nous analyserons les approches antérieures et finalement nous présenterons les grandes lignes de notre démarche personnelle.

### 1.1.1 Définition et objectif du problème

Nous supposons que le lecteur est familier avec la théorie des systèmes dynamiques. Nous nous contenterons donc ici d'esquisser de façon informelle quelques définitions de base.

**Définition 1** *Un système (du grec συνιστημι, mettre ensemble) est une interconnection d'éléments ou de sous-systèmes, où un élément est un objet indécomposable caractérisé par un certain nombre de propriétés et une interconnection désigne un ensemble de relations entre les propriétés de ces différents éléments.*

**Définition 2** *On appelle système dynamique, un système où intervient la notion du temps au travers de la variation des propriétés des éléments du système.*

Nous nous limiterons dans ce chapitre à cette définition informelle. Nous aurons besoin par la suite (au chapitre 2) d'une définition mathématique formelle de la notion de système dynamique. Nous exposerons alors la définition de Kalman [41] qu'il n'est pas nécessaire de connaître pour l'instant.

Il n'existe pas une seule, mais un grand nombre de définitions de la notion de système complexe. La définition retenue ici, n'est donc qu'une définition particulière parmi beaucoup d'autres possibles.

**Définition 3** *Un système dynamique complexe est un système dynamique constitué d'un grand nombre d'éléments présentant des interactions non-linéaires.*

Dans sa carte du monde des systèmes dynamiques (tableau 1.1), S.H. Strogatz trace une frontière au delà duquel, à l'image des dragons sur les cartes anciennes, se trouvent le monde encore peu connus des systèmes dynamiques complexes, parmi ceux-ci, citons : les tremblements de terre, le système immunitaire, les écosystèmes, la théorie quantique des champs.

L'importance de l'aspect non-linéaire d'un système dynamique complexe est dû au fait que, contrairement à un système dynamique linéaire, celui-ci ne peut être décomposé en sous-parties, rendant impossible la décomposition du problème en sous-problèmes plus simples, permettant d'obtenir la solution en "recollant" les solutions partielles. Pour finir sur l'aspect non-linéaire des systèmes dynamiques complexes, reprenons l'exemple donné par Strogatz dans [71] pour souligner que la vie quotidienne est en règle générale non-linéaire : si vous écoutez en même temps

		Nombre de variables →				
		$n = 1$	$n = 2$	$n \geq 3$	$n \gg 1$	Continu
Non linéarité	Linéaire	<i>Croissance, décroissance ou équilibre</i> Croissance exponentielle Circuit RC Décroissance radioactive	<i>Oscillations</i> Oscillateurs linéaires Masse et ressort Circuit RLC Problème à deux corps (Kepler, Newton)	Génie civil, structures Génie électrique	<i>Phénomène collectif</i> Oscillateurs harmoniques couplés Physique du solide Dynamique moléculaire Mécanique statistique à l'équilibre	<i>Ondes et patterns</i> Elasticité Equations d'ondes Electromagnétisme (Maxwell) Mécanique quantique (Schroedinger, Heisenberg, Dirac) Chaleur et diffusion Acoustique Fluides Visqueux
	Non-linéaire	Points fixes Bifurcations Equation logistique à une seule espèce	Pendule Oscillateurs anharmoniques Cycles limites Oscillateurs biologiques (neurones, cellules cardiaques) Cycles proies-prédateurs Electronique non-linéaire (van der Pol, Josephson)	<i>Chaos</i> Attracteurs étranges (Lorenz) Problème des trois corps (Poincaré) Cinétiques chimiques Fonctions itérées (Feigenbaum) Fractales (Mandelbrot) Oscillateurs non-linéaire forcés (Levinson, Smale)	<b>La frontière</b> Oscillateurs non-linéaires couplés Lasers, optiques non-linéaires Mécanique statistique hors équilibre Synchronisation des cellules cardiaques <b>Réseaux de neurones</b> <b>Système immunitaire</b> <b>Ecosystème</b> <b>Economie</b>	<i>Complexité spatio-temporelle</i> Ondes non-linéaires (solitons) Plasmas Tremblements de terre Relativité générale Théorie quantique des champs Réaction-diffusion, oscillations biologique et chimiques Fibrillation Epilepsy Liquides turbulents <b>Vie</b>

TAB. 1.1 – Classification des systèmes en fonction du nombre de variables et de la non-linéarité (d'après [71]). Au-delà de la frontière se trouvent les SDC et en gras les SDAC.

vos deux morceaux de musique préférés, vous n'en retirerez généralement pas le double de plaisir...

**Définition 4** *Un système dynamique adaptatif (SDA) est un système dynamique dont la structure change au cours du temps de manière constructive en fonction de son état et de son environnement (fig. 1.1).*

Plusieurs points importants ressortent de cette définition, le premier est la variation de la structure, le deuxième est le caractère constructif de cette variation, c'est-à-dire la naissance, la croissance par modifications successives pouvant amener à une complexification présentant une forme d'organisation et enfin, la grande interaction entre le changement de structure, l'état interne du système et son environnement. Cette organisation qui apparaît dans un système est souvent appelée "émergence" lorsqu'elle n'est pas déductible facilement<sup>5</sup> des interactions de base et que l'on doit simuler le système pour la mettre en évidence. Ceci est illustré par la figure 1.1, où l'on voit un système qui, sous l'action du temps et de son environnement, change de structure pour se complexifier de manière organisée. Cette complexification est caractérisée par un accroissement de la taille du système, ce qui permet au système de stocker de l'information (sous forme organisée; l'organisation ayant un rôle de compression de l'information, en cela la complexité s'oppose au simple hasard qu'il n'est pas possible de résumer ainsi par une structure dynamique).

Cette structure, contenant la manière dont le système dynamique réagit avec le monde extérieur, fait que l'on parle souvent des SDAC comme de systèmes englobant en leur sein un schéma du monde extérieur. Un des buts principaux de l'adaptation, est alors de construire ce schéma [27].

Les exemples de tels systèmes sont faciles à trouver. On peut citer, par exemple, les situations suivantes exposées dans le premier chapitre de [36] et que l'on peut retrouver en gras dans le tableau 1.1 :

- l'approvisionnement d'une grande ville comme New York dans un paysage économique changeant ;
- le cerveau où de nouvelles connexions entre les neurones apparaissant et disparaissant continuellement permettent l'apprentissage ;
- la colonisation d'un milieu par un groupe d'espèces qui vont évoluer pour s'adapter à leur nouvel environnement ;
- le système immunitaire qui doit apprendre à reconnaître les corps étrangers qui pénètrent dans l'organisme.

Dans le tableau 1.1, nous avons mis les SDC qui peuvent être considérés comme adaptatifs en gras.

---

<sup>5</sup>Il est à noter la subjectivité de la notion de "déductible facilement", en effet celle-ci dépend de l'expérience de l'observateur qui acquiert une connaissance et une meilleure compréhension du phénomène au fil de son étude.

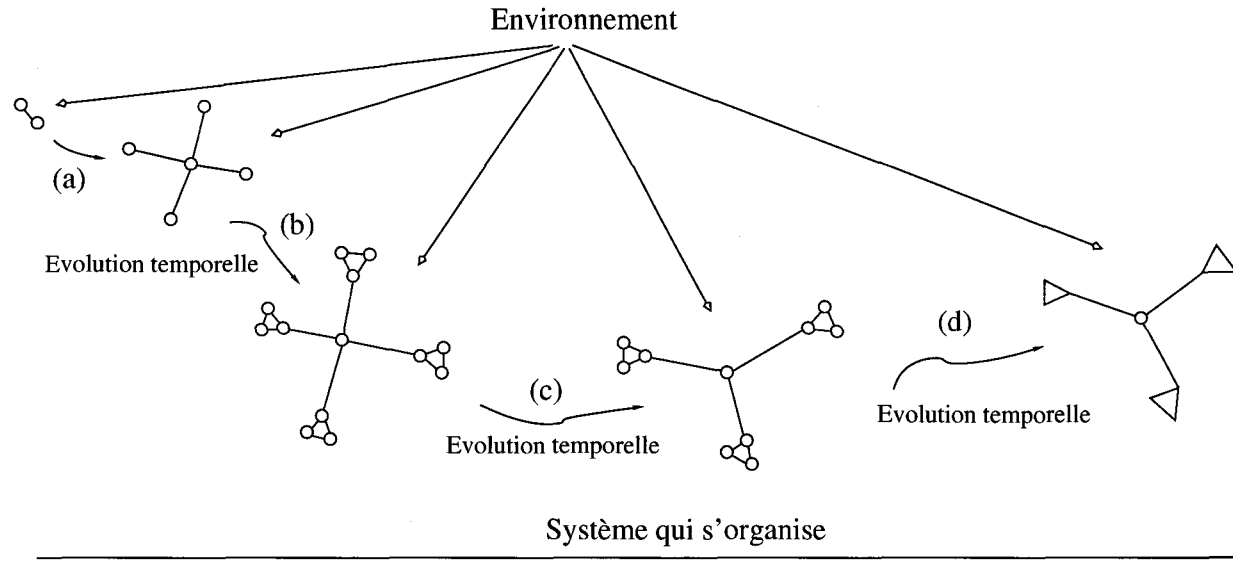


FIG. 1.1 – Schéma d'un système adaptatif en voie de complexification : (a), (b) ajout d'éléments déjà existants, (c) suppression d'éléments, (d) apparition d'un nouvel éléments.

Le premier objectif de notre travail peut donc se résumer par la question : peut-on trouver une structure formelle générale permettant de regrouper un grand nombre de systèmes répondant à la définition des systèmes dynamiques adaptatifs complexes ?

Cette formalisation n'est pas une théorisation gratuite, puisqu'elle répond à plusieurs enjeux :

- D'un point de vue théorique, elle nous permettra d'étendre les notions de la théorie des systèmes dynamiques au changement temporel, non plus des seules variables comme c'est le cas traditionnellement, mais également de la structure toute entière d'un système. Ainsi, cela nous permettra de parler de structure stable, cyclique, apériodique, ou encore de bassins d'attraction de ces structures ou d'ensembles de structures limites.
- Du point de vue de la modélisation et des méthodes, le formalisme construit capturera une large classe d'exemples. Il sera donc idéalement possible d'appliquer à ces exemples, les différentes notions théoriques issue du formalisme général, de mettre en évidence des conditions de complexification et/ou de simplification, non plus dans un domaine spécifique, mais de façon générique, pour tous les domaines.
- Du point de vue de la simulation, les notions ainsi introduites conduiront à l'élaboration d'algorithmes de simulation et de techniques d'analyse de données spécifiques.

### 1.1.2 Réponses déjà existantes

S'il existe peu de réels formalismes des systèmes dynamiques adaptatifs complexes, il existe en revanche beaucoup d'études de tels systèmes, souvent menées grâce à des simulations. Certains exemples se révèlent fort utiles, car, bien que les auteurs ne recherchent pas une telle formalisation, ils tentent néanmoins de dégager une approche générique commune à plusieurs exemples.

Nous commencerons par parler brièvement de ce que l'on pourrait qualifier de « philosophie orientée objet » et, en particulier, de ses limites, puis nous exposerons en détails ce qui constitue sûrement la tentative la plus aboutie de formalisation des SDAC, issue des travaux de J. Holland sur les algorithmes génétiques et nous exposerons finalement l'approche en niveaux métadynamiques utilisée par Bagley et co-auteurs pour la simulation des modèles de réseaux autocatalytiques de Kaufman et de réseaux immunitaires de Perelson.

#### La philosophie orientée objet et les modèles d'agents

Il ne s'agit pas ici de présenter ici l'approche orientée objet, formalisme très abouti et utile dans beaucoup de domaines (le lecteur intéressé par un texte général sur l'approche objet pourra se référer à [16] et pour la simulation multi-agents à [52]), mais plutôt de voir, en quoi, cette approche, appliquée à la modélisation

des SDAC se révèle insuffisante. L'idée de la philosophie orientée objet et de son pendant en modélisation, appelé modélisation multi-agents (nous utiliserons plus ou moins indifféremment les termes objet ou agent), est la suivante : tout système peut se représenter par un ensemble d'objets et d'interactions. On considère que la base de tout système est l'objet, c'est-à-dire un individu. Cet individu peut aussi bien être une molécule, un animal, une cellule ou une étoile, selon le modèle considéré, mais on le choisit comme un des "atomes" du modèle (au sens premier, c'est-à-dire l'entité élémentaire que l'on ne peut plus diviser). Un autre concept important est celui de "classe" qui permet de regrouper les objets qui ont une structure commune. Les objets d'une même classe partagent une même structure de données, ainsi que les mêmes méthodes de traitement ou d'interaction avec les autres objets. Le dernier avantage de cette partition des objets en classes est de pouvoir créer également des sous-classes qui héritent des caractéristiques de la classe mère.

Mais avant d'aller plus avant, prenons un exemple simple de ce que serait une modélisation multi-agent. Imaginons que nous voulions modéliser le comportement d'une solution comportant un certain nombre de molécules en interaction. La démarche classique consiste à écrire des équations donnant le nombre de molécules des espèces chimiques qui peuvent se trouver dans la solution en fonction du nombre de molécules qui s'y trouvait un peu avant. Il existe bien sûr différentes approches (temps continu ou discret, concentrations considérées comme continues ou nombre entier de molécules, fonctions de transition stochastiques ou déterministes), mais le principe général est celui-là. En revanche, l'approche objet sera très différente. On commencera par définir la classe "molécule" et ses propriétés, c'est-à-dire les structures de données qu'elle contient (par exemple des vecteurs pour la position et la vitesse ou des caractéristiques chimiques générales), ainsi que des méthodes associées (par exemple l'évolution de la position en fonction de la vitesse, la modification de la vitesse s'il y a un choc moléculaire). Ensuite, on définira des sous-classes correspondant aux différentes espèces chimiques : celles-ci hériteront de la structure de leur classe mère, la classe molécule. Il pourra également leur être ajouté des caractéristiques particulières, comme des caractéristiques de réaction par exemple. Il ne reste plus alors qu'à créer les objets ou agents, par exemple `albertine=new enzyme3`, `martine=new enzyme5`, `robert=new sucre`, `philibert=new sucre`. Nous employons volontairement des noms pour les molécules pour souligner que chaque molécule présente dans la solution existe bien individuellement dans le modèle (et lors d'une simulation en mémoire). La suite de la simulation est simple (tout du moins conceptuellement) : les molécules vont évoluer dans la solution en fonction de leur vitesse et interagir selon les méthodes prédéfinies quand elle se rencontrent, pouvant donner naissance à de nouvelles molécules.

Cette approche présente un aspect très séduisant : elle « colle » au réel, en effet, il suffit de décomposer le système en ces pièces élémentaires, les agents, de décrire les interactions du mieux possible et de laisser la simulation agir, ce qui est



d'autant plus facile que, étant issu d'un paradigme informatique très développé, il existe beaucoup de langages (par exemple C++ ou Java) et de programmes (Swarm<sup>6</sup> ou Echo de Holland [34] par exemple) qui lui sont dédiés.

Mais il existe néanmoins deux limitations très importantes. La première est la présence fréquente dans les SDAC d'un nombre très grand d'individus (à titre d'exemple, on estime à  $10^{11}$  environ le nombre de neurones dans le cerveau, mais cette remarque reste valable pour un écosystème, un système immunitaire ou une économie). On est donc amené à n'utiliser que des nombres d'objets de loin inférieurs à la réalité et cela peut introduire des modifications de comportement purement dépendantes du nombre limité d'objets et donc infirmant en tout ou en partie le réalisme « apparent » de la simulation qui n'est alors plus qu'un simple exercice. Ainsi, si nous revenons sur l'exemple de la solution chimique en considérant, une concentration de  $C = 10^{-3} \text{ mol.l}^{-1}$  d'une molécule donnée, on a alors  $\mathcal{A} \times V \times C$  molécules en présence dans la solution, avec  $\mathcal{A} = 6,023 \times 10^{23}$  le nombre d'Avogadro et  $V$  le volume de notre solution. Ainsi, pour simuler l'évolution d'un dé à coudre de solution de quelques millilitres, il faudrait considérer de l'ordre de  $10^{18}$  molécules, c'est à dire que, en supposant que l'on puisse déterminer une molécule par 1 octet (c'est-à-dire 8 bits d'information ce qui est encore extrêmement peu), on atteint des systèmes dont la quantité d'information à mémoriser est de l'ordre de  $10^{12}$  Go. Et que dire de la simulation d'un réacteur chimique...

Mais pourquoi doit-on passer obligatoirement par la simulation ? Parce que, et c'est là que se trouve la deuxième limitation : la compréhension d'un système multi-agents de grande taille est rapidement aussi vaine que la compréhension du système réel lui-même. L'étude nécessite alors de passer par la simulation, avec les problèmes que l'on vient d'évoquer...

Le formalisme objet de la modélisation multi-agents présente donc beaucoup d'intérêt de par son adéquation avec le réel et de sa brique de base très universelle -l'individu. Doublé du partitionnement aristotélicien en classes, il utilise également une base théorique solide et des outils puissants issus de l'informatique. En revanche, la difficulté numérique engendrée par la taille des collections d'objets généralement présente dans les SDAC est telle que cette modélisation peut souvent devenir difficile à utiliser, que ce soit pour l'étude analytique à cause de la trop grande complexité du modèle (quasiment aussi difficile à comprendre que le modèle réel), ou pour la simulation à cause de la trop grande quantité d'information. Et de toute façons, est-il réellement nécessaire de sauvegarder un tel degré d'information, alors qu'on est en présence d'un niveau de précision impossible à atteindre expérimentalement ?

---

<sup>6</sup> développé par le Santa Fe Institute pour fournir un outil de simulation multi-agents pour les systèmes complexes (pour plus d'information, consulter <http://www.swarm.org>). L'utilisation de cet outil dans le cadre de la modélisation en microbiologie a amené le développement de BacSim [47].

### Formalisme de J. Holland basé sur le concept d'algorithmes génétiques

La deuxième approche que nous considérons ici, part du constat suivant : les SDAC étudiés partagent un certain nombre de caractéristiques communes, qu'il convient d'exploiter. Ce que Holland remarque dans [35], c'est que beaucoup des traits des algorithmes génétiques (AG) se retrouvent dans d'autres systèmes adaptatifs complexes<sup>7</sup> (SAC).

Les AG [31] sont une méthode heuristique basé sur l'imitation du processus d'évolution darwinienne permettant en principe de trouver de bonne solution à des « problèmes impliquant de centaines à des centaines de milliers de paramètres interdépendants et présentant des multitudes d'optima locaux » [35]. Le principe standard des AGs est le suivant : partant d'une population de solutions peu efficaces pour un problème donné, l'AG va faire évoluer cette population pour l'améliorer. Il utilise pour cela des fonctions de croisement entre les différentes solutions permettant de générer de nouvelles solutions (les solutions filles), ainsi qu'une fonction de mutation entraînant l'apparition de nouveauté dans cette population. Une fois cette nouvelle génération créée, le principe de sélection darwinienne est appliqué : une compétition entre les différents éléments de la population est effectuée (il s'agit souvent d'un tirage aléatoire) où la survie est proportionnelle à leur valeur sélective<sup>8</sup>. Ainsi, petit à petit, la population évolue -s'adapte- jusqu'à être en moyenne très adaptée au problème. Cette méthode, même si elle est bien sûr loin d'être évidente à mettre en œuvre (en particulier la détermination du codage des solutions sous formes de données que l'on peut faire évoluer ou le croisement de deux bonnes solutions en conservant des propriétés des parents comme cela se passe dans la nature), donne néanmoins de très bons résultats dans beaucoup de domaines (application à la productique par exemple).

On peut dégager des AG une structure plus générale, si l'on remplace les "chromosomes" de données qui codent une solution du problème par la notion de structure au sens large (voir les équivalence données par Holland dans les différents domaines, tableau 1.2), les mécanismes de mutation, recombinaison, sélection par le concept d'opérateurs modifiant la structure et enfin l'adaptation par une mesure de performance. Cette notion d'une structure qui évolue sous l'action d'opérateurs, déterminés par un plan adaptatif, qui agit en fonction de l'environnement et d'une mesure de performance de la structure présente, se retrouve dans beaucoup de domaines (voir également tableau 1.2).

---

<sup>7</sup>Nous omettons volontairement le D de "dynamique" pour conserver la formulation originale de Holland de *complex adaptive system*, mais il s'agit fondamentalement du même concept que celui défini auparavant. Il nous a semblé important en ce qui nous concerne d'insister sur le caractère dynamique de tels systèmes, même si cela doit donner raison à M. Gell-Mann qui cite dans [17] la phrase d'un "professeur distingué" qui déclare : "*a scientist would rather use someone else's toothbrush than another scientist's terminology*" (Un scientifique préférerait utiliser la brosse à dents de quelqu'un d'autre plutôt que la terminologie d'un de ses confrères)...

<sup>8</sup>ou en anglais *fitness*, fonction qui donne l'adaptation d'un objet à son environnement. Dans un algorithme génétique, elle est décidée par son concepteur pour orienter la forme des solutions.

1.1. PREMIER PROBLÈME : MODÉLISATION ET FORMALISATION DES SYSTÈMES DYNAMIQUES

Domaine	Structures	Opérateurs	Mesure de performance
Génétique	chromosomes	mutation, recombinaison, etc.	adaptation
Planification économique	échanges de biens	activités de production	utilité
Contrôle	stratégies	règle de Bayes, approximations, etc.	fonctions d'erreur
Psychologie physiologique	groupe de cellules	modifications synaptiques	taux de performance (dans certains contextes, mais souvent non spécifiés)
Théorie des jeux	stratégies	règles itératives d'approximation d'une stratégie optimale	gain
Intelligence artificielle	programmes	règles d'apprentissage	efficacité comparative (si spécifié)

TAB. 1.2 – Equivalence des notions pour les différents systèmes adaptifs complexes, selon Holland [35].

Ainsi, une fois établies ces connexions inter-disciplinaires, Holland considère-t-il qu'un SAC peut se décrire par le schéma suivant : l'objet cherche à s'adapter à son environnement, bien qu'il en ait une connaissance incomplète. Il doit donc tester différentes configurations ou structures et dispose pour cela de deux « outils » : le plan adaptatif qui permet d'explorer l'espace des structures possibles et la *fitness* qui permet de mesurer la performance de la nouvelle configuration, bien que ne connaissant pas en détail l'environnement (voir figure 1.2).

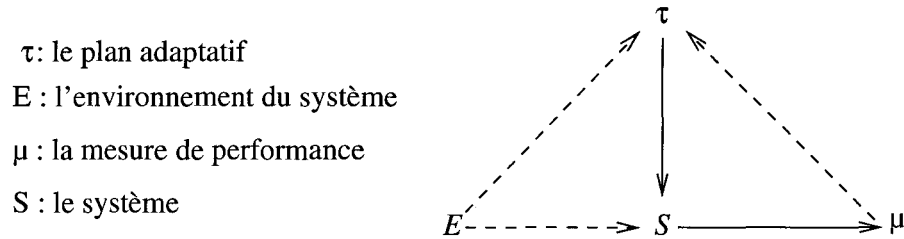


FIG. 1.2 – Le formalisme de Holland : un plan adaptatif  $\tau$  vient, au travers d'une mesure de performance  $\mu$ , analyser l'adéquation du système  $S$  à son environnement  $E$  et éventuellement modifier sa structure.

A partir de ces considérations, Holland a donné une définition formelle [35] de la notion de système adaptatif :

**Définition 5** *Un système adaptatif au sens de Holland est défini par l'ensemble des objets  $(\mathcal{A}, \Omega, I, \tau)$  où :*

$\mathcal{A} = \{A_1, A_2, \dots\}$  est l'ensemble des structures atteignables, le domaine d'action du plan adaptatif,

$\Omega = \{\omega_1, \omega_2, \dots\}$  est l'ensemble des opérateurs pour modifier les structures avec  $\omega \in \Omega$  une fonction  $\omega : \mathcal{A} \rightarrow \mathcal{P}$ , où  $\mathcal{P}$  est un ensemble de distributions de probabilités sur  $\mathcal{A}$ ,

$I$  est l'ensemble des entrées possibles du système venant de l'environnement, et

$\tau : I \times \mathcal{A} \rightarrow \Omega$  est le plan adaptatif qui, sur la base de l'entrée et de la structure courante, détermine quel opérateur doit être appliqué au temps  $t$ .

On a alors :

$$\tau(I(t), \mathcal{A}(t)) = \omega_t \in \Omega \text{ et } \omega_t(\mathcal{A}(t)) = \mathcal{P}(t+1),$$

avec  $\mathcal{P}(t+1)$  une distribution particulière sur  $\mathcal{A}$ . On peut alors faire évoluer le système selon  $t$  en itérant ce processus.

On a donc une suite de modifications de structures, modifications effectuées à partir d'opérateurs renvoyant des distributions de probabilités. Ces opérateurs sont choisis en fonction de la structure courante et des entrées. Ce terme de structure reste très vague, il n'y a pas ici de distinction explicite entre structure et variables

du système et donc pas de distinction explicite entre variation temporelle de la structure et variation temporelle des variables.

Ce formalisme se développe sur une base temporelle exclusivement discrète. Holland évoque d'ailleurs la difficulté de l'étendre au continu en ces termes : *“Where continuity is more appropriate, it is often straightforward to obtain continuous counterparts of definitions and theorems, though in some cases appropriate restatements are full-fledged research problems with the discrete results serving only as guidelines.”*<sup>9</sup>.

Le deuxième point est l'importance de la notion de *fitness* comme moyen de connaissance par le système de son environnement. Il est souvent hautement spéculatif de considérer cette mesure comme quelque chose d'existant et d'appartenant au système. Il nous semble que cette valeur sélective (sauf dans les cas où elle directement codée, comme dans certains systèmes artificiels) est souvent plutôt le résultat d'un processus dynamique : dès que la structure et ses interactions deviennent un peu compliquées, ce n'est généralement qu'a posteriori que l'on peut dire si telle ou telle caractéristique procure un avantage sélectif ou non. Cette remarque est encore plus vraie dans un environnement changeant, pouvant survenir par exemple dans des problèmes de coévolution.

### Un début de généralisation : Bagley et co-auteurs

La dernière approche présentée ne consiste pas en un formalisme, au sens de l'approche objet ou de celle de Holland. Il s'agit en fait d'une structure de simulation présentée dans [2], servant à modéliser et à simuler deux exemples fondamentaux de SDAC (les réseaux autocatalytiques de Kauffman [42] et le système immunitaire [19]).

Etant donné l'importance que cette structure a eu sur le développement de notre approche, ainsi que sur l'exemple que nous développons au chapitre 4, nous avons choisi de présenter ces travaux plus en détails en annexe (voir annexe ??).

L'idée de ces travaux est la suivante : on a affaire à des populations d'un certain type d'objets (des polymères dans le cas des réseaux autocatalytiques, des anticorps dans le cas du système immunitaire), avec un nombre de variants fini non fixé et relativement petit, mais appartenant à un espace de possibilités énorme.

Par exemple, si l'on considère les chaînes de moins de cent monomères symbolisés par un alphabet {a,b} (par exemple : aab, a, abbbbababaaaaa, baabbbaab), sauvegarder l'information d'un système de 20 polymères sera relativement aisé (taille inférieure à 2000 octets) ; en revanche la taille de l'espace des possibilités est considérable : les différentes chaînes de moins de 100 monomères sont au nombre de  $2(2^{100} - 1)$  (sans compter la chaîne vide). Prendre 20 polymères parmi ces chaînes

---

<sup>9</sup>Là où le continu est plus approprié, il est souvent simple d'obtenir des équivalents continus aux définitions et théorèmes, bien que dans certains cas, les reformulations appropriées soient des problèmes de recherche à part entière où les résultats du discret servent seulement d'indications.

donne donc  $C_{2(2^{100}-1)}^{20} = \frac{[2(2^{100}-1)]!}{[2(2^{100}-1)]!20!}$  possibilités différentes<sup>10</sup>.

Il est donc évidemment hors de question de décrire en extension le système (c'est-à-dire en décrivant les équations différentielles décrivant tous les cas possibles d'existence simultanée de différents types de molécules et en commutant entre ces différents cas lorsque différentes espèces apparaissent ou disparaissent). La solution retenue est la suivante : le système biologique est modélisé à chaque instant par un système d'équations différentielles qui représente l'ajustement dans le temps des différentes concentrations résultant des interactions entre les éléments du système. D'autre part, un processus innovatif "regarde" l'état du système dynamique et en fonction de cet état et d'entrées extérieures génère de nouveaux composants et couplages, modifiant ainsi le système d'équations différentielles pré-existant. L'implémentation de ce système a été faite en C en utilisant des structures de listes [2]. Ce type de structure de données semble particulièrement adapté car il autorise le système à changer de dimension, lors de la simulation, par l'ajout ou la suppression de composants.

L'approche de Bagley et co-auteurs est particulièrement intéressante, car elle propose une méthode générique pour la simulation de systèmes adaptatifs complexes en biologie. En revanche, la formalisation -traduction en langage mathématique de la structure de simulation- n'a pas été réalisée, ce qui prive de la possibilité de définir rigoureusement d'autres notions. Enfin la dernière critique que nous formulerons, concerne l'aspect très spécifique de ce schéma : pourquoi se limiter à une dynamique de base continue et différentielle ?

En résumé, les travaux de Bagley et co-auteurs apportent une idée fondamentale : dans beaucoup de SDAC, il est possible de découpler l'aspect dynamique et l'aspect adaptatif en dégageant des règles de construction de la structure du système durant la simulation, permettant ainsi d'éviter la description de l'espace des structures possibles qui est énorme<sup>11</sup>.

En revanche, les auteurs se bornent à une seule type de dynamique (un système d'ODE) et n'utilise pas la portée de la généralisation pour étudier des propriétés communes aux différents systèmes représentés. C'est donc vers ces pistes que nous avons orientés nos recherches.

---

<sup>10</sup>pour information, on peut évaluer ce nombre conséquent :

$$\begin{aligned} \frac{[2(2^{100}-1)]!}{[2(2^{100}-1)]!20!} &= \frac{(2^{101}-2)(2^{101}-3)\dots(2^{101}-21)}{20!} \\ &= \frac{2^{101 \times 20}}{2^{20 \times 20}} \left(1 - \frac{230}{2^{101}} + \frac{24795}{2^{202}} \dots\right) \\ &\approx \frac{2^{2020}}{20!} \\ &\approx 4,95 \times 10^{589} \end{aligned}$$

<sup>11</sup>Cette même idée est d'ailleurs aussi implicite dans la notion de plan adaptatif de Holland.

### Approches complémentaires et émergentes

Il existe beaucoup d'autres travaux possédant des liens plus ou moins forts avec nos travaux. Nous ne les développerons pas tous. Le lecteur pourra en trouver un panorama assez exhaustif dans les publications du SFI<sup>12</sup> et dans les livres suivants [10, 35, 37, 43, 68]). Il faut ajouter à cela les travaux émergents actuellement sur les réseaux à connectivité élevée ("*small worlds*") de Watts [73].

Nous donnons ici toutefois quelques détails sur certains travaux dont les idées nous ont servi d'inspiration, mais qui ne s'intègrent pas naturellement dans l'exposé qui vient d'être donné.

**Les hypercycles de Eigen** Eigen [23] a établi un modèle différentiel (appelé "hypercycle") de la diversification évolutive d'une espèce chimique autocatalysante, autorépliquante avec erreurs de copie. Il présuppose une organisation en cycles emboîtés (voir figure 1.3) et des termes d'apparition d'espèce sous pression évolutive due à une contrainte de masse totale constante. Ainsi, il existe un niveau premier de cycles d'espèces chimiques qui forment une chaîne autocatalytique fermée, c'est-à-dire où chaque espèce catalyse la formation d'une autre espèce (figure 1.3, premier panneau). Ces cycles peuvent eux-mêmes former une chaîne catalytique, chaque cycle favorisant la formation d'un autre cycle de la chaîne (deuxième panneau). On peut répéter ces « emboîtements » formant ainsi une structure d'hypercycle (troisième panneau).

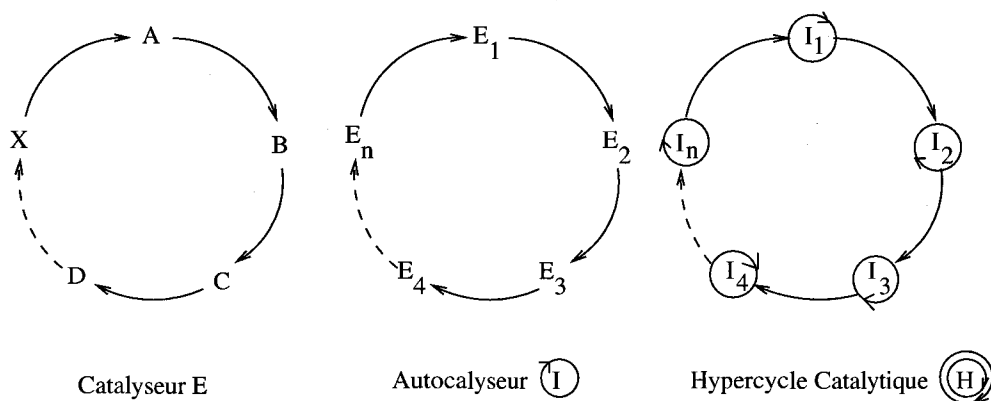


FIG. 1.3 – La hiérarchie des réseaux de réactions cyclique et leur organisation en cycles emboîtés (d'après [22]).

Toutefois, il ne détaille pas l'aspect "mutation" et donc ne dégage pas vraiment ce que nous appelons une "métadynamique" (une ou des règles qui viennent changer la structure du système). De plus son organisation est trop stricte et U. Niesert a démontré qu'elle était instable lorsque le taux de mutation n'était pas très faible

<sup>12</sup><http://www.santafe.edu/sfi/publications/working-papers.html>

grand [58]. Nous verrons au chapitre 4 un modèle non pas d'équations chimiques, mais d'un système bactéries-phages, s'inspirant de façon éloignée de Eigen, mais essayant de lever les critiques ci-dessus.

**Les modèles hiérarchiques de Auger** P. Auger [1] a établi une théorie des systèmes de population organisées de façon hiérarchique. Il l'a alors appliquée à la mécanique des particules, aux réseaux biochimiques, aux réseaux de neurones, aux écosystèmes et aux systèmes économiques.

Pour couvrir tous ces différents cas d'étude, il a développé une théorie générale de l'émergence de niveaux hiérarchiques basés sur la différence des constantes de temps et il a aussi étudié une thermodynamique généralisée.

Cependant, ses systèmes sont déterministes et obtenus à partir de développements en série, donc locaux.

Son approche, tout comme la nôtre, fait apparaître une hiérarchie, mais ses systèmes existent a priori et ne sont pas constructifs, ils n'ont pas de métadynamique. Il s'agit donc d'une approche différente.

**Autres approches** D'autres auteurs sont reliés de plus loin à nos travaux :

- Fontana [25] : a développé une approche évolutive et constructive d'un ensemble d'objets similaire à ceux d'une chimie artificielle. Ses objets se complexifient et s'autoreproduisent. Il a appelé cette approche le "calcul de objets" et nous avons vu que l'évolution de tels systèmes pose un problème de compréhension causale que nous abordons en proposant de coupler à un système objet un niveau supérieur de modélisation basé sur les systèmes à métadynamique.

De plus, il utilise des objets abstraits basés sur le  $\lambda$ -calcul de Church [14, 15] et a peut-être biaisé les évolutions qu'il observe suite à ce choix.

- Solé [68] a publié un grand nombre de travaux sur les SAC en abordant à peu près tous les aspects. En ce qui nous concerne, nous exposerons un de ces modèles d'ARN-virus à la section 1.2.2 au moment où nous aborderons les modèles d'évolution existants.
- Bersini [11] a abordé récemment l'étude des systèmes à métadynamique en se basant surtout sur des exemples de chimie artificielle. Nous n'avons vu de lui qu'une présentation powerpoint résumant ses projets de travaux. Ceux-ci nous semblent aller dans une ligne parfaitement compatible avec la nôtre. Toutefois, nous n'avons pas assez de détail sur ses travaux pour en dire plus ici.
- Banzhaf et Dittrich [76] ont étudié en détail la notion et les propriétés des "chimies artificielles". Ils travaillent donc essentiellement au niveau objet et ne semblent pas avoir étudié en détail des représentations plus globales à base de systèmes différentiels lorsqu'ils parlent de ce type de représentation, ils se bornent essentiellement à discuter des travaux de Kauffman et de Bagley



mentionnés auparavant 1.1.2.

- Au cours des dernières années s’est développée à l’université du Sussex à Brighton, un pôle de recherche autour des aspects de la vie artificielle et de la modélisation de l’évolution, au sein du groupe EASY (Evolutionary and Adaptive System) du COGS (Computing and Cognitive Science). Bien que leurs recherches concernent plus l’aspect robotique évolutive et neuroscience, il existe des liens très étroits entre leur approche et la nôtre. Nous nous sommes d’ailleurs directement inspirés de certaines de leurs visualisations.

### 1.1.3 Une extension dans le cadre de la théorie des systèmes dynamiques : idée générale du travail

Il est intéressant de constater qu’une modélisation efficace des SDAC (du point de vue de la réduction d’information et de la capacité d’explication) passe par une structure à deux niveaux, avec une structure dynamique de base constituée d’interactions dynamiques et une structure supérieure qui vient modifier la structure de base lorsque certains événements se produisent.

Nous allons, dans ce travail, formaliser cette notion en la rapprochant des systèmes dynamiques de Kalman [41]. Nous verrons notre SDAC comme un automate opérant de façon constructive (ou destructive) sur un ensemble de systèmes dynamiques (des grandeurs qui évoluent au cours du temps selon des fonctions de transition données en fonction du temps et de l’environnement). Cet automate pourra être déterministe ou stochastique. Nous appellerons ce niveau supérieur “métadynamique” en référence aux travaux de Bagley (voir section 1.1.2). Quant au niveau dynamique de base, il pourra aussi bien être continu que discret.

Ainsi, grâce à cette approche, l’on voit que l’on peut regrouper et classer un certain nombre d’exemples en tenant compte de la nature du niveau supérieur (métadynamique) et du niveau inférieur (dynamique). Par exemple, si l’on divise le niveau métadynamique en stochastique et déterministe et le niveau dynamique en continu et en discret, on obtient quatre classes de modèles, avec dans chaque classe des représentants bien connus (voir tableau 1.3) :

- les systèmes hybrides (un automate déterministe sur un système continu [75]) appartient à la classe “Métadynamique déterministe et dynamique continue”,
- l’approche de Bagley vue à la section 1.1.2) appartient à la classe “Métadynamique stochastique et dynamique continue”,
- le formalisme de Holland, où le plan adaptatif est constitué de règles probabilistes et le niveau inférieur est discret appartient à la classe “Métadynamique stochastique et dynamique discrète”,
- finalement, on peut classer certains réseaux de neurones évolués, les *recurrent neural nets* d’Elmann dans la classe des “Métadynamique déterministe et dynamique discrète”.

Dynamique \ Métadynamique	déterministe	stochastique
continue	systèmes hybrides	Bagley
discrète	<i>recurrent neural nets</i>	Holland

TAB. 1.3 – Situation d'exemples antérieures dans le nouveau formalisme

### Concepts et exploitation

Notre formalisation nous permettra de généraliser aux structures du système certains concepts de la théorie des systèmes dynamiques, comme les concepts d'équilibre, de stabilité, d'attracteurs et de bassins d'attraction. On pourra ainsi parler d'un ensemble de structures stable, ou de bassins d'attraction vers des structures différentes, permettant de délimiter les zones de conditions initiales qui évoluent vers une même structure ou ensemble de structures. Ainsi, la figure 1.4 illustre le cas, où le système présente deux bassins d'attractions, l'un menant vers des populations ayant certaines caractéristiques (symbolisées par des ellipses), l'autre vers des populations ayant d'autres caractéristiques (symbolisées par des rectangles). Ces concepts, une fois définis, s'appliqueront alors à tous les exemples couverts par le formalisme. L'autre intérêt sera de suggérer des méthodes de simulation et d'extraction de l'information communes.

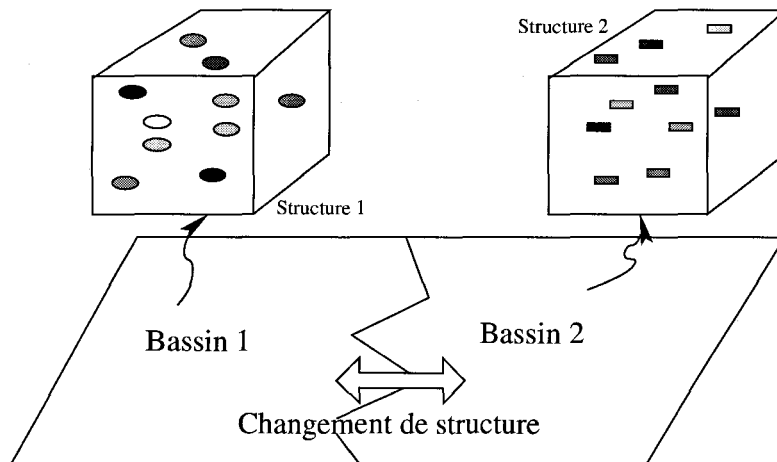


FIG. 1.4 – Exemple de bassin menant à deux classes de structures différentes

### Un besoin d'interprétation

Il est certain que la distinction entre niveaux dynamique et métadynamique n'existe pas telle quelle dans la majorité des cas réels; une modélisation du type multi-agents -avec les limitations évoquées auparavant à la section 1.1.2- serait sans aucun doute plus proche du réel. Pourtant, le fait que cette hiérarchisation

## 1.1. PREMIER PROBLÈME : MODÉLISATION ET FORMALISATION DES SYSTÈMES DYNAMIQUES

soit récurrente en modélisation des SDAC prouve qu'il existe des raisons profondes à cela.

Le problème est donc de trouver une connexion entre des systèmes de types agents et nos systèmes à métadynamique. Dans le chapitre 3, nous montrons sur un exemple simple de réacteur bactérien comment, quand les probabilités de transition d'un système particulière sont hétérogènes, deux échelles de temps peuvent apparaître donnant naissance à un système à métadynamique à base continue (voir figure 1.5).

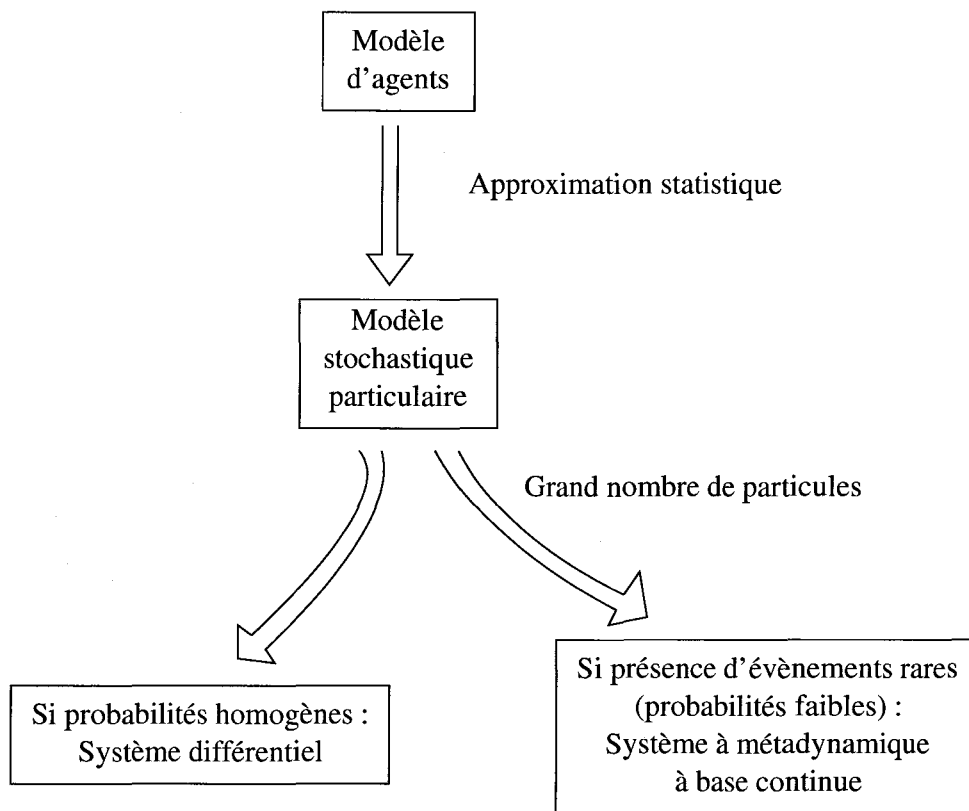


FIG. 1.5 – Passage de niveaux de complexité : du modèle d'agents au système à métadynamique à base continue

La métadynamique peut alors s'interpréter comme l'approximation successive d'un modèle sous-jacent très précis et de grande taille -le modèle d'agents- par des modèles plus "manipulables" constitués d'ODE. Bien qu'il y ait évidemment perte d'information au cours de ce processus, un tel modèle peut reproduire des comportements et/ou des propriétés du modèle par agents sous-jacent et à l'avantage d'être beaucoup plus rapide.

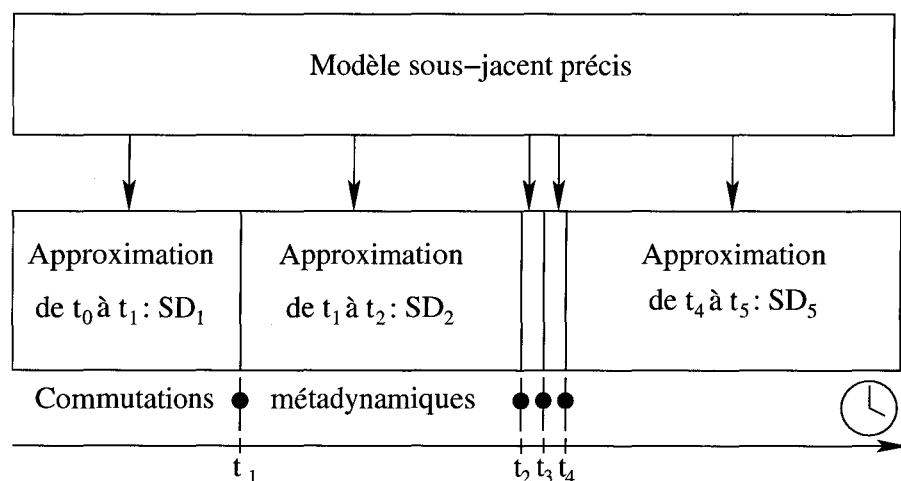


FIG. 1.6 – Une métadynamique comme approximation successive d'un modèle complexe

## 1.2 Deuxième problème : modélisation de systèmes évolutifs

Les systèmes à métadynamique introduits à la section précédente ont évidemment une portée très générale. Cependant, nous les appliquerons surtout par la suite au problème de la coévolution darwinienne de deux espèces. Aussi dans cette section, nous allons présenter brièvement certaines des idées générales de la théorie de l'évolution qui sous-tendent notre application.

La question de la modélisation des systèmes évolutifs est particulièrement importante, car on assiste à l'heure actuelle au creusement d'un écart de plus en plus grand entre la somme des données biologiques accumulées lors d'expériences diverses, d'une part, et le faible nombre de ces observations qu'il est possible d'expliquer par des modèles ayant une base scientifique solide [21].

Ces données collectées concernent principalement deux aspects :

- la biologie moléculaire (avec le déchiffrement systématique du code génétique, le séquençage d'un grand nombre de gènes et la détermination en protéomique de la structure des protéines qu'ils codent) et le comportement évolutifs d'organismes « simples » comme les bactéries ou les mouches drosophiles (réponses aux phénomènes de sélection d'un point de vue génotypique et phénotypique et connexion entre les deux),
- le comportement des écosystèmes sur lesquels on dispose de plus en plus de données concernant les interactions entre compétition et coopération.

Toutes ces découvertes obligent à raffiner ou à modifier la théorie. La question de savoir quels mécanismes sont importants pour l'évolution sur une grande échelle de temps suscite un débat vivant et génère des opinions très diverses [21].

Les approches théoriques restent très en recul de ces résultats expérimentaux :

la plupart du temps, les phénomènes décrits ne sont expliqués que par des *verbal arguments* ou des « histoires possibles ». En fait, beaucoup des principales avancées théoriques ne viennent pas de la biologie directement, mais d'autres sciences, en particulier la physique statistique ou l'informatique. Mais avant de présenter quelques modèles parmi ceux qui permettront de comprendre notre démarche personnelle sur le sujet, commençons par rappeler quelques notions de la théorie de l'évolution.

### 1.2.1 *The evolution in a nutshell*

Parmi les nombreux modèles théoriques mis au point par les spécialistes de ce domaine, nous allons présenter les idées qui vont nous servir par la suite. Nous ne cherchons en aucun cas à être exhaustifs, le lecteur intéressé par plus de précision sur la théorie de l'évolution pourra se référer à [18], [32] ou [48].

#### Origine et but de la théorie de l'évolution

La théorie de l'évolution a pour objectif de rendre compte et d'expliquer l'origine de la diversité des êtres vivants, qui se traduit par des formes de vie très différentes, que l'on peut regrouper en grandes classes, présentant elles mêmes de grandes variations. On considère C. Darwin comme le père de cette théorie après qu'il ait publié en 1859 "L'origine des espèces".

Darwin expose dans ce livre des preuves de la non fixité des espèces au cours du temps (il s'oppose en cela au fixisme, théorie très répandue à son époque, qui soutient que le monde a été créé tel quel, d'un seul tenant). Bien qu'il ne soit pas le premier à défendre cette thèse (on pense à Lamarck un peu avant ou même à Anaximandre de Milet au VI<sup>e</sup> s. av. JC), il est en revanche le premier qui ait donné un mécanisme pour cette évolution : à partir de la diversité qui apparaît, la sélection naturelle assure la survie du plus apte (celui dont la descendance a le plus de chances d'arriver à l'âge adulte) garantissant ainsi la progression de l'espèce vers plus d'efficacité face à son milieu. La théorie de l'évolution s'enrichira par la suite des travaux de Mendel sur la génétique et l'hérédité, de la génétique des populations (Fischer), ainsi que de l'avancée de la biologie moléculaire (découverte de la structure de l'ADN en 1953 par Crick, Watson et Wilkins récompensée par le prix Nobel en 1962). Récemment, elle a connue de nombreux développements théoriques comme le *neutral drift* de Kimura (1968) [45], les équilibres ponctués de Gould et Eldredge [33] ou la criticalité auto-organisée évolutionnaire<sup>13</sup> [10].

Des deux versants de la théorie de l'évolution, l'historique (systématique et histoire des espèces, par des descriptions et des arbres phylogéniques) et le mécanistique (explication des mécanismes générant la diversité), nous ne nous intéressons qu'au second.

---

<sup>13</sup> *evolutionary self-organized criticality*

## Mécanismes généraux

Nous allons maintenant présenter les mécanismes généraux de la génétique qui sont une des bases de la théorie moderne de l'évolution.

Chaque cellule vivante porte une longue molécule en forme de double hélice appelée ADN (acide désoxyribonucléique). Cette molécule est constituée d'une suite de paire de "bases" au nombre de quatre (Adénosine, Thymine, Guanine, Cytosine). Cette suite de bases définit une carte d'identité propre à chaque être vivant. Ce code reste le même pour toutes les cellules de cet être vivant tout au long de sa vie.

Ce code permet, après un mécanisme complexe de traduction, de produire les protéines nécessaires au métabolisme de l'être vivant. C'est donc après un processus complexe (par le biais de mécanismes physico-chimiques sensibles à l'environnement) que l'on passe du code génétique -le génotype- aux caractères exprimés (taille, forme...) -le phénotype.

Si le code utilisé est le même pour tous les êtres vivants ("de la bactérie à l'éléphant"), le mécanisme créateur de la diversité génétique est différent selon que la reproduction de l'individu soit sexué ou non. Pour des individus à reproduction sexuée, la diversité est créée par le mélange des gènes de la mère et du père. En revanche, pour les individus asexués (comme les bactéries) les descendants possèdent les mêmes gènes que le parent, la diversité est donc créée par les erreurs de copie du génôme de ce parent : les mutations.

Dans le cas des bactéries, il existe au sein de celles-ci des mécanismes de réparation très efficaces, qui permettent de limiter les mutations. Dans certaines situations (de stress par exemple ou de changement dans sa population que la bactérie détecte par des mécanismes comme le "*quorum sensing*" qui permet aux bactéries par l'émission de molécules spécifiques d'avoir une connaissance de la concentration de son espèce<sup>14</sup>), la bactérie inhibe ces mécanismes de réparation donnant lieu à des poussées mutatoires, on parle alors de mutations dirigées. En plus de leur chromosome propre (les bactéries ont un seul chromosome ou filament d'ADN, les cellules plus complexes en ont plusieurs), les bactéries possèdent des petites molécules circulaires d'ADN, les plasmides, dont la reproduction se fait indépendamment de celle de la bactérie. Ces molécules peuvent coder par exemple la résistance à un antibiotique. La particularité de ces plasmides est que les bactéries peuvent facilement se les échanger, même entre bactéries d'espèces différentes, créant ainsi un bassin de gènes en commun. Il existe également une pseudo-sexualité chez les bactéries : indépendamment de la reproduction, elles peuvent s'échanger des bouts d'ADN en créant un pont entre leurs deux cellules.

---

<sup>14</sup>Vu les échanges horizontaux (sans reproduction) génétiques fréquents par échange de morceaux de gènes et de plasmides entre bactéries éventuellement de classifications différentes, cette notion d'espèce devient fort floue quand on l'applique au monde bactérien. Nous l'utilisons ici dans un sens approximatif.

### Principe de Hardy-Weinberg

Le principe de Hardy-Weinberg, ou hypothèse nulle de l'évolution, peut être comparé au principe d'inertie de la mécanique newtonienne (en l'absence de force appliquée à un système, celui-ci suit un mouvement rectiligne et uniforme). Il a été énoncé au début du XX<sup>ème</sup> siècle (1906) de manière séparée par T. Hardy et H. Weinberg.

Supposons une population fermée dans laquelle deux variants génétiques coexistent, notés  $A$  et  $B$ . On suppose la sélection naturelle absente du système, c'est-à-dire qu'aucun des deux n'est meilleur que l'autre pour se perpétuer. Un variant pris au hasard dans la population doit être soit de type  $A$  ou soit de type  $B$ . L'état de la population peut être décrit par les fréquences de  $A$  et de  $B$  (notées  $p$  et  $q$ ). Le principe s'énonce ainsi : Si la population a un effectif infini, et en l'absence de mutations, les fréquences  $p$  et  $q$  restent indéfiniment constantes de génération en génération.

Aucune population réelle ne répond bien sûr à ces hypothèses (toute population réelle est finie). Il faut comprendre ce principe comme donnant un état de référence : toute variation des fréquences des caractères s'explique par la violation des hypothèses de départ. Ce changement est alors dû, soit à la sélection naturelle, soit à des mutations qui apparaissent au cours du temps, soit à l'effectif de la population. Ainsi peut-on distinguer une différence entre les populations de très grande taille (que l'on peut considérer comme infinie) et les populations de petite taille, où apparaissent des phénomènes de dérive génétique (voir un modèle de dérive génétique dans les modèles présentés ci-après en section 1.2.2).

Ce principe intervient dans notre modèle du chapitre 4 de la façon suivante : nous nous intéressons à l'évolution (au sens des caractères de la population et non seulement des changements de concentrations) de populations de phages et de bactéries en interaction. Comme les populations sont considérées de grande taille (approximation continue), entre deux mutations, le changement de fréquence ne peut se faire d'après le principe de Hardy-Weinberg que par une différence de sélection naturelle, différence que nous choisissons de représenter sous forme d'une chaîne binaire codant la capacité d'attaque (pour les phages) ou de défense (pour les bactéries) à l'infection.

### Coévolution et effet reine rouge

"Alice looked round her in great surprise. 'Why, I do believe we've been under this tree the whole time! Everything's just as it was!'

'Of course it is,' said the Queen. 'What would you have it?'

'Well, in *our* country,' said Alice, still panting a little, 'you'd generally get to somewhere else-if you ran very fast for a long time, as we've been doing.'

'A slow sort of country!' said the Queen. 'Now, *here*, you see, it

takes all the running *you* can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!"

Lewis CARROLL 'Through the looking-glass'.



FIG. 1.7 – Alice courant avec la Reine Rouge

L'un des concepts très importants de la théorie de l'évolution et dont nous allons nous servir par la suite, est le concept de coévolution. C'est précisément cette propriété là que nous cherchons à mettre en évidence dans le modèle évolutifs d'interactions bactéries-phages du chapitre 4.

L'idée sous jacente de la coévolution et de l'effet reine rouge est la suivante : les changements évolutifs ne sont pas seulement initiés par des changements externes au monde vivant (si le climat change par exemple) mais ceux-ci peuvent aussi être générés par une dynamique évolutive créée par les êtres vivants eux-mêmes. Par exemple, il existe à Madagascar une orchidée, appelée *Angraecum sesquipedale*, dont le nectaire a une longueur extraordinaire d'une trentaine de centimètres (voir figure 1.8). Le nectaire est un tube cylindrique fixé à la base de la fleur, qui contient un liquide sucré pour attirer les insectes. Lorsque les insectes viennent boire le liquide, ils heurtent de la tête un point de la fleur où se trouvent les pollinies, nom que l'on donne aux masses de pollen collantes. L'insecte, sans le vouloir, emporte alors les pollinies et les transporte sur les organes femelles d'autres orchidées. Darwin, frappé par la taille incroyable de ce nectaire, eut -on était en 1865- l'intuition géniale qu'à Madagascar devait exister un papillon dont la trompe avait une longueur comparable. Ce papillon, *Xanthopan morgani*, fut découvert en effet, mais en 1920 seulement. L'explication la plus souvent donnée est la suivante : pour que les pollinies se collent sur la tête du papillon, il faut que la tête de celui-ci les heurte avec une certaine force. Si l'accès au nectar est trop facile, le papillon ingurgite du nectar mais repart sans pollinies. Par conséquent, seules les plantes à nectaires longs, qui contraignent l'insecte à heurter la base des pollinies pour atteindre le nectar, se reproduisent : le caractère "nectaire long" est donc favorisé par la sélection naturelle. Parallèlement, la sélection naturelle favorise chez le papillon



le caractère “trompe longue”, puisque les papillons à trompe courte n’atteignent pas le précieux nectar et, mal nourris, se reproduisent mal. Bref, ce processus co-évolutif a abouti à des orchidées aux nectaires interminables et à des papillons à la trompe démesurée. On arrive donc à deux populations interdépendantes : la survie de l’une dépend de la survie de l’autre ; la course aux armements s’est transformé en une forme de coopération.

Pour L. Van Valen, le moteur principal de l’évolution de toute espèce vivante est représenté par les autres espèces avec lesquelles cette espèce partage des ressources. Tout progrès dans la valeur adaptative d’une espèce modifie l’environnement des autres espèces qui l’entourent et les oblige à s’adapter. Cette adaptation provoque à son tour un changement dans l’environnement de la première espèce, ce qui la pousse à un nouvel épisode de sélection, et ainsi de suite. Cela se produit parce que les ressources sont limitées. Van Valen a baptisé cette proposition du nom d’hypothèse de la Reine Rouge.

L’expression “Reine Rouge” est empruntée au roman de Lewis Carroll “A travers le miroir” (*Through the looking-glass*), dans laquelle Alice tient la Reine Rouge par la main et court avec elle au pays des Merveilles (voir figure 1.7). Alice, surprise que le paysage autour d’elles ne change pas, interroge la Reine, qui lui répond qu’elles courent pour rester sur place et que c’est pourquoi le paysage leur paraît immobile. Il en va de même dans les co-évolutions : les espèces en conflit courent, c’est à dire “inventent” sans cesse de nouvelles adaptations, mais la valeur relative de chacune de ces adaptations ne change pas : nous, les humains, par exemple, ne sommes pas mieux adaptés pour exploiter notre milieu que les bactéries pour exploiter le leur -les bactéries font même beaucoup mieux que nous, puisqu’elles sont capables de vivre dans des milieux qui nous sont interdits, comme le fond des océans, le cœur des glaciers ou les nappes pétrolifères. Mais si, en termes d’adaptation, nous ne sommes pas supérieurs aux bactéries<sup>15</sup>, nous leur sommes en revanche très supérieurs en termes de complexité (pour plus de détails, lire la partie consacrée à l’effet Reine Rouge par C. Combes dans [48]).

### 1.2.2 Les modèles d’évolution existants

Il serait vain de prétendre dresser un panorama exhaustif du nombre considérable de modèles traitant d’aspects évolutifs dans ce court chapitre de présentation du problème. Nous nous contenterons donc de dégager ceux qui nous semblent les plus intéressants pour comprendre ce qu’est et ce qu’apporte notre modèle. Nous nous arrêterons seulement sur un certain type de modèles que nous pourrions qualifier de systèmes dynamiques aspaciaux. Nous ne considérons pas, par exemple, de modèles de type théorie des jeux, bien qu’ils soient par ailleurs très utiles en

---

<sup>15</sup>Comme nous le verrons à la section 4.1.1 du chapitre 4, leurs capacités d’adaptation sont assez incroyables : elles peuvent proliférer dans des endroits aussi divers que les nappes pétrolifères ou les réacteurs bactériens...



FIG. 1.8 – le *Xanthopa morgani* et sa trompe peu ordinaire

théorie de l'évolution (explication des stratégies évolutivement stables (ESS<sup>16</sup>) par J. Maynard Smith [39, 53] par exemple).

Il est intéressant de noter qu'une bonne proportion de ces modèles vient de la physique statistique. Pour plus d'informations sur les modèles évolutifs issus de la physique statistique, on pourra consulter [21]. Pour clarifier et synthétiser la présentation, nous avons regroupé ces exemples sous trois grandes classes de modèles que nous appelons les modèles de type urne, géométrique et dynamique.

### Modèles de type urne

Ces modèles sont très faciles à mettre en œuvre et sont de ce fait très courants. Nous avons choisi deux modèles en particulier à causes des liens qu'ils entretiennent avec la suite de notre travail. L'exemple 1 donne un modèle de dérive génétique, c'est-à-dire de comment un gène arrive à supplanter un autre gène dans une population. Nous donnerons un scénario particulière, puis métadynamique à ce mécanisme au chapitre 3. Il nous a semblé également pertinent de présenter l'exemple des populations d'ARN-virus de Solé parce qu'il met en avant un modèle utilisant l'ajout de chaînes binaires à un mécanisme dynamique comme moyen de stocker de l'information (rôle que tient l'ADN dans l'évolution), c'est ce que nous ferons au chapitre 4, mais cette fois-ci couplé à une métadynamique.

**Définition 6** Nous appelons *modèle de type urne* un modèle présentant les caractéristiques suivantes :

- à chaque génération, on a un nombre fini constant d'individus ;
- on passe d'une génération à l'autre par une ou des lois stochastiques.

Ce type de modèle se traduit souvent sur le plan informatique par un tirage, à chaque pas de simulation, d'un ou plusieurs élément et remplacement après une action, d'où le nom que nous avons donné de modèle de type urne.

<sup>16</sup>de l'anglais *Evolutionary Stable Strategy*

Il est à noter que, dans le cas où, en notant  $X_n$  l'état du système à la génération  $n$  et  $P$  une probabilité sur l'espace d'état de ce système, on a  $P(X_n|X_{n-1}) = P(X_n|X_{n-1}|X_{n-2}|X_{n-3}|\dots|X_1|X_0)$ <sup>17</sup> ce qui est en pratique quasiment toujours le cas, on parle alors de processus de Markov à temps discret et d'espace d'état fini.

Nous présentons deux exemples de tels modèles : un modèle de dérive génétique [18, 32, 39] et le modèle de population d'ARN-virus de Solé et co-auteurs [69].

**Dérive génétique** Nous commençons par exposer ce modèle très simple, mais possédant néanmoins un intérêt dont témoigne son omniprésence dans les manuels de théorie de l'évolution [18, 32] ou de modélisation biologique [39].

Il s'agit d'un modèle haploïde<sup>18</sup> de reproduction aléatoire, sans mutation ni sélection. L'effectif constant est égal à  $2N$ , nous nous intéressons à deux allèles  $a$  et  $A$  d'un gène donné. On considère qu'à l'instant initial, il y a  $X_0$  gènes  $a$  ( $0 < X_0 < 2N$ ) et  $2N - X_0$  gènes  $A$ . Le tirage des gènes à la génération suivante suit une loi binomiale. On note  $X_n$  le nombre de gène  $a$  à la génération  $n$ . On obtient donc une chaîne de Markov définie par les probabilités de transition suivantes :

$$\begin{aligned} P(X_{n+1} = k | X_n = j) &= P(j, k) \\ &= C_{2N}^k \left(\frac{j}{2N}\right)^k \left(\frac{2N-j}{2N}\right)^{2N-k} \\ & \quad j, k = 0, \dots, 2N \end{aligned}$$

avec les calculs d'espérance classiques :

$$\begin{aligned} E(X_{n+1} | X_n = j) &= \sum_{k=0}^{2N} k P(j, k) \\ &= \sum_{k=0}^{2N} k C_{2N}^k \left(\frac{j}{2N}\right)^k \left(\frac{2N-j}{2N}\right)^{2N-k} \\ &= j. \end{aligned}$$

D'où l'on déduit :

---

<sup>17</sup>qui se lit : la probabilité de passer de l'état  $X_{n-1}$  à l'état  $X_n$  ne dépend pas des états antérieurs du système : le système n'a pas de "mémoire"

<sup>18</sup>Cela veut dire que les individus ne portent qu'un seul exemplaire des gènes présents dans leur ADN, par opposition à diploïde quand le génôme contient chaque gène en double exemplaire, comme chez les humains par exemple.

$$\begin{aligned}
EX_n &= E(E(X_n|X_{n-1})) \\
&= \sum_{j=0}^{2N} P(X_{n-1} = j)E(X_n|X_{n-1}) \\
&= \sum_{j=0}^{2N} P(X_{n-1} = j)j \\
&= EX_{n-1} \\
&= X_0.
\end{aligned}$$

Si l'hypothèse de Hardy-Weinberg stipule que pour une population infinie, les fréquences des allèles restent constantes, l'étude de ce petit modèle montre que pour des populations finies, cela est différent, la population finit par perdre sa variabilité génétique et se fixe dans un des états absorbants  $X_n = 0$  (que des  $A$ ) avec une probabilité  $\frac{2N-X_0}{2N}$  ou  $X_n = 2n$  (que des  $a$ ) avec une probabilité  $\frac{X_0}{2N}$ .

**Populations d'ARN-virus (modèle de Solé et co-auteurs)** Le modèle que nous allons présenter maintenant est un modèle évolutif d'une population d'ARN-virus. Il a été présenté et analysé par Solé et co-auteurs dans [69]. Chaque ARN-virus est représenté par une chaîne binaire (CB) de  $\nu$  bits. La population est formée par  $N$  CB notées  $S_i$  et dont la séquence peut s'écrire :

$$S_i = S_i^1 S_i^2 \dots S_i^\nu, i = 1, 2, \dots, N$$

avec  $S_i^j \in \{0, 1\}$ . A chaque génération, on exécute  $N$  fois l'ensemble de règles suivantes :

- On prend une CB au hasard dans la population, par exemple  $S_i$  et on la réplique. La probabilité de réplique  $r_i$  est dépendante de la CB, c'est-à-dire :

$$r_i = \phi(S_i^1 S_i^2 \dots S_i^\nu),$$

avec  $\phi = \sum_k S_i^k$ , soit la fonction la plus simple possible.

- La nouvelle CB formée remplace au hasard une des CB de la population, par exemple  $S_i$  remplace  $S_j$ . Ce mécanisme présente des erreurs, à un taux  $\mu$  (par bit et par cycle de réplique). On a donc :  $S_j^k = S_i^k$  avec une probabilité  $1 - \mu$  et  $S_j^k = 1 - S_i^k$  avec une probabilité  $\mu$ .

Ce modèle bien que très simple, permet de rendre compte de bifurcations et de phénomènes de compétition menant à l'effet Reine Rouge.

## Modèles géométriques

Ce type de modèle (Kauffman [44]) utilise les notions d'adaptation et de paysage adaptatif. L'adaptation (auquel nous préférons les termes de valeur selective

ou *fitness* quand ce terme concerne le cadre de la théorie de l'évolution), est la capacité d'un individu de transmettre son patrimoine génétique en assurant à sa progéniture d'arriver à l'âge adulte. La notion de paysage d'adaptation stipule que l'on peut définir une fonction qui associe à chaque génotype existant une valeur numérique réelle donnant l'adaptation de cet individu. Cette assertion est hautement critiquable, car l'adaptation d'un individu ne dépend pas seulement de son génôme, mais aussi par exemple des autres individus en présence ou de son histoire, de plus cette fonction peut varier au cours du temps. Néanmoins, ce concept, parce que fort pratique, est très utilisé. Les modèles géométriques partent de l'hypothèse que la fixation d'un allèle mutant se fait quasi immédiatement s'il est plus apte, c'est-à-dire que, dès qu'il apparaît il se répand dans la population et remplace l'ancien. Les mutations défavorables, elles, sont rapidement éliminées.

Le concept de "marche adaptative" (*adaptive walk*) a été introduit par Kauffman [44]. Il consiste, considérant les hypothèses précédentes, en un déplacement de locus à locus (le locus est le nom que l'on donne à un emplacement de gène), de telle manière que l'adaptation soit croissante. On peut alors définir les propriétés statistiques du système selon la forme du paysage adaptatif. Parmi les nombreux paysages possibles, on peut citer le plus connu et le plus simple, appelé "paysage Fujiyama" : c'est un paysage contenant un seul pic, qui est atteint par une simple ascension quelque soit l'endroit d'où l'on part.

L'un des principaux raffinements de ce modèle est le paysage  $NK$  de Kauffman (inspiré des modèles à verres de spin).  $N$  est le nombre de gènes dans un chromosome haploïde et  $K$  le nombre de liens qui existent entre chaque gène et les autres gènes du chromosome. La question est alors de connaître l'allure du paysage d'adaptation en fonction de la valeur de  $K$ . Le modèle prédit que, plus  $K$  est grand (au maximum  $N - 1$ ), plus le paysage d'adaptation résultant est accidenté, présentant de nombreux pics. En revanche, quand  $K = 0$ , on retrouve un paysage avec un seul pic, comme le Fujiyama.

L'intérêt de se type de modèle est de voir l'évolution d'une manière très intuitive et métaphorique, comme le déplacement dans un paysage en relief. L'image peut être trompeuse car il n'est pratiquement jamais possible de représenter ces paysages, ni donc d'avoir une visualisation comme support. Néanmoins, cette idée de déplacement dans un espace de chromosomes est intéressante et très parlante, elle permet de donner une vision de l'évolution sur de grands intervalles de temps en terme d'optimisation de plusieurs critères (les pics du paysage adaptatif). Ainsi, l'utilisation de cette métaphore pour parler de la coévolution amène à concevoir des paysages qui se déforment avec les changements intervenants au sein des autres espèces. L'espèce essaie "d'escalader" une hauteur (augmenter sa *fitness*), mais au même moment cette montagne se déforme.

Nous préférons utiliser comme image de l'évolution d'une population, l'image d'un nuage en mouvement dans un espace de caractères (un nuage est alors un ensemble de points représentant un ensemble de caractères pondérés par l'effectif

des individus qui présentent cet ensemble de caractère). Nous en viendrons donc à tracer des grandeurs caractéristiques de ce nuage : son centroïde (le "centre de gravité") et sa dispersion.

## Modèles dynamiques

Nous considérons, comme appartenant à cette catégorie, des modèles issus de la théorie des systèmes dynamiques où les changements dus aux mutations apparaissent dans les équations comme des termes normaux, au même titre que les autres éléments du modèle (termes de croissance des populations, de pression de sélection etc.). Le modèle le plus célèbre de ce type est sans aucun doute, le modèle de quasi-espèce proposé par Manfred Eigen en 1971 et développé plus tard en collaboration avec Peter Schuster. Bien que déjà relativement ancien, il est encore très souvent cité et est à l'origine de beaucoup de modèles simplement obtenus par rajout ou modification des interactions [64, 69].

Il est intéressant de constater que ce modèle n'est à l'origine pas un modèle biologique mais un modèle physique portant sur des molécules capables de se répliquer, bien que le raisonnement s'applique également à la réplication des molécules du vivant, c'est-à-dire l'ADN.

Le modèle considère qu'il y a  $n$  espèces moléculaires présentes dans le système. Chaque espèce a la faculté de se reproduire mais, la reproduction n'étant pas parfaite, il existe des espèces mutantes. Une autre hypothèse du modèle est que le système est fermé par mutation, c'est-à-dire qu'une molécule du système ne peut muter qu'en une autre molécule du système : il n'y a pas apparition de nouvelle espèce.

Soit  $x_i (i = 1, \dots, n)$  le nombre de molécules du type  $i$  présentes dans le système.

On suppose que les molécules de type  $i$  se reproduisent au taux constant de  $A_i$  « descendants » par parent par unité de temps. Le taux de production (de quelques espèces que ce soit) de descendants de membres de l'espèce  $i$  serait donc  $A_i x_i$ . Mais toutes les copies n'étant pas fidèles, il existe un paramètre  $Q_i$ , appelé facteur de qualité, qui spécifie le nombre des copies de  $i$  qui appartiennent bien à l'espèce  $i$ . Le taux de production de  $x_i$  par lui-même est donc de  $A_i Q_i x_i$ , les  $(1 - A_i Q_i)x_i$  restant se répartissant entre les autres espèces.

Ces molécules, à cause des chocs moléculaires, ont une durée de vie limitée (cela s'applique évidemment aussi aux êtres vivants) : on suppose que cette dégradation est proportionnelle au nombre de molécules et est donc de la forme  $-D_i x_i$ .

En plus de leur propre réplication, les différentes molécules bénéficient de l'apport de certaines copies erronées venant des autres espèces du système. Nous définissons pour cela  $w_{ik}$ , le taux à laquelle une molécule de l'espèce  $k$  produit une molécule de l'espèce  $i$ .

En combinant tous ces termes on aboutit à l'équation suivante :

$$\dot{x}_i = (A_i Q_i - D_i)x_i + \sum_{k \neq i} w_{ik} x_k.$$

Telle quelle, cette équation autorise des populations infinies (si  $A_i Q_i > D_i$ ). Or dans la réalité ce n'est jamais le cas, à cause de la compétition pour les ressources. Il faut donc rajouter un terme pour traduire cela. M. Eigen a choisi d'ajouter un flot de dilution hors du système qui maintient la population constante (l'autre alternative aurait été de rajouter une équation pour une ou des ressources renouvelables, mais finies).

Considérant que l'on veut garder la population constante, le flot total (sur tout les  $x_i$ ) doit être égal à :

$$\phi_t = - \sum_i (A_i - D_i)x_i \equiv - \sum_i E_i x_i$$

et si l'on pose

$$W_{ii} \equiv A_i Q_i - D_i$$

$$\bar{E} \equiv \frac{\sum_k E_k x_k}{\sum_k x_k}$$

on obtient l'équation finale

$$\dot{x}_m = \sum_k W_{mk} x_k - \bar{E} x_m$$

où la matrice  $W$  a les éléments suivants :

$$W_{ij} = \begin{cases} W_{ii} = A_i Q_i - D_i & \text{si } i = j \\ w_{ij} & \text{si } i \neq j \end{cases}$$

(pour plus de détails dans les calculs, consulter [64]).

La résolution analytique de ce systèmes montre que les différentes espèces se stabilisent (les mutations/dégradations équilibrant les réplifications réussies) selon différentes valeurs dépendant des valeurs propres de la matrice  $W$ . Cette combinaison d'espèces est appelé par Eigen et Schuster quasi-espèce, puisqu'il s'agit d'une entité stable composée de plus d'une espèce "pure".

La critique la plus évidente que l'on peut faire est évidemment qu'il n'y a pas d'apparition de nouveauté dans ce modèle et, donc d'une certaine manière, pas d'évolution. Ce modèle semble donc seulement valide dans deux cas :

- quand on est présence d'une population tellement grande que chaque variant possible est présent en grand nombre ;
- si l'on considère qu'il n'y a que très peu (un ou deux) de gènes différents qui interviennent dans la *fitness*.

Nous essayerons avec le formalisme des systèmes à métadynamiques de généraliser ce type de modèle en couplant à un modèle dynamique la possibilité d'évolution adaptative, processus discret venant modifier la structure du système.

### 1.2.3 Notre modèle de co-évolution : le modèle Attaque-Défense

On remarque que chaque grande classe de modèles s'inscrit dans un formalisme donné, ainsi les modèles que nous avons appelé de type urne s'apparentent aux processus markoviens, ou les modèles de type quasi-espèces à la théorie des systèmes dynamiques. Il est donc logique au vu de cela et de la grande interaction entre les systèmes adaptatifs complexes et les modèles évolutifs que notre formalisme nous permettent de créer un modèle évolutif. Notre modèle est un modèle de coévolution bactéries-phages, les phages étant les virus des bactéries. Les bactéries peuvent faire évoluer leurs moyens de défense, en se rendant, moins reconnaissables par les phages. De la même façon, les virus peuvent mieux reconnaître les bactéries en mutant.

Nous allons esquisser dans ce premier chapitre les originalités et les apports qui seront développés au chapitre IV.

#### Deux niveaux de dynamique

Dans les modèles pré-cités, on observe deux cas de figures :

- les événements fréquents tels que la croissance ou la régulation et les événements rares tels que les mutations ou l'extinction d'une classe de population sont vu sur le même niveau dynamique (modèle de type urne, quasi espèce),
- on suppose la fixation d'un allèle immédiate et le seul changement dans le temps représenté est alors celui du ou des allèles dominant, sans tenir compte de l'aspect dynamique de ce changement puisque celui-ci se fait de manière instantané ("marche" dans un paysage d'adaptation).

Comme dans les travaux de Bagley, notre modèle dynamique est constitué de deux niveaux, l'un, appelé niveau dynamique, est constitué d'un système d'équations différentielles ordinaires du type dynamique des populations (modélisant les interactions entre des populations de bactéries et des populations de virus appelés phages), l'autre, le niveau métadynamique, est composé de règles stochastiques qui, au cours de la simulation, dépendant de l'état du système (des effectifs des différentes populations), viennent modifier la structure du système :

- une mutation vient rajouter des équations liées à la nouvelle variable créée et modifier (ajout de nouveaux termes) les équations déjà présentes.
- la disparition d'une population (dépendant d'un seuil) provoque la disparition des équations et termes associés.

Les deux niveaux sont reliés par l'emploi d'une structure de données (une chaîne binaire) associée à chaque population codant pour les caractères d'attaque (pour les phages) ou de défense (pour les bactéries). Cette chaîne est prise en compte dans la forme des équations différentielle qui définissent la dynamique du système, mais c'est également de cette chaîne que dépend la forme des populations mutantes : elles ne diffèrent que d'un caractère (un bit de la chaîne binaire) de la population



mère.

### Une dynamique dans l'espace des objets

Notre formalisme nous permet de différencier une dynamique et une métadynamique et donc de dégager des propriétés plus globales du système : sans regarder la dynamique précise de chaque classe, on peut regarder par exemple le déplacement du nuage dans l'espace des gènes (ce déplacement n'existe pas en soit, mais est une résultante de la métadynamique et des dynamiques sous jacentes) et en déduire des propriétés du type classique en théorie des systèmes dynamique, du type :

- Est-ce que la structure se stabilise ?
- Y-a-t-il une structure qui apparaît naturellement ?
- Dans un système de grande taille, peut-on réduire l'information sur la structure du système tout en conservant des propriétés intéressantes ?

Nous utiliserons pour répondre à ces questions et en particulier pour visualiser les évolutions du système, des méthodes issues de l'analyse factorielle (une branche des statistiques), comme l'analyse en composantes principales. Celle-ci nous permettra de visualiser les trajectoires des centres de gravité (ou centroïdes) du nuage d'une population dans l'espace des caractères.

### Un modèle de co-évolution

Le dernier apport de notre modèle est la prise en compte de phénomènes de co-évolution. En effet, dans les modèles évolutifs traditionnels, l'évolution se fait par rapport à un environnement constant. La valeur selective de la population (souvent codée directement dans le modèle au lieu de se déduire de l'interaction entre la population et son environnement) progresse sous l'influence de la selection naturelle. Dans notre modèle, on a une réelle coévolution entre les populations de bactéries et de phages, chacune modifiant l'environnement de l'autre en changeant. Les questions sont alors les suivantes :

- voit-on apparaître des phénomènes de type effet Reine Rouge ?
- jusqu'à quel point peut-on corrélérer l'évolution de l'une par rapport à celle de l'autre ?

#### 1.2.4 Conclusion

Nous avons présenté dans ce premier chapitre la filiation de notre travail, les réponses qu'il va tenter d'apporter et les questions qu'il soulève.

Deux problèmes fortement connectés ont été étudiés :

- la formalisation des modèles de systèmes dynamiques adaptatifs complexes,
- la modélisation de systèmes évolutifs.

Nous avons monté l'intérêt et la pertinence d'un formalisme plus général basé sur deux niveaux (dynamique et méta-dynamique). Le chapitre suivant va décrire plus

précisément la formalisation annoncée.

# Chapitre 2

## Métadynamiques : aspects formels

“Mathematics may be defined as the subject where we never know what we are talking about, nor whether what we are saying is true.<sup>1</sup>”

“Recent Work on the Principles of Mathematics”, a.k.a. “Mathematics and the Metaphysicians” BERTRAND RUSSEL

Nous avons vu, dans le chapitre précédent, les différents problèmes posés par la modélisation des systèmes dynamiques adaptatifs complexes (SDAC) et les principales approches existantes. Les différentes remarques formulées nous ont amené à concevoir une structure appelée “système adaptatif à métadynamique” (SAM) permettant de modéliser un grand nombre de SDAC. Le but de ce chapitre est de proposer un cadre formel, constitué de différentes définitions, autour de cette notion de SAM.

Nous commençons par rappeler les points importants concernant les SDAC évoqués dans le chapitre précédent, c’est-à-dire, en quoi les différentes approches entreprises pour modéliser les SDAC comportent des points communs que nous allons utiliser pour dégager une définition formelle en 2.1.3 dans le cadre du formalisme des systèmes dynamiques de Kalman, que nous exposons préalablement en 2.1.2. Une fois développé notre formalisme de base, nous envisageons des extensions possibles en 2.1.5, puis nous construisons des définitions associées pour étendre les notions classiques de la théorie des systèmes dynamiques à ce nouveau formalisme (section 2.2).

### 2.1 Cadre formel général

Cette première section est consacrée à l’élaboration de la définition formelle des SAM. Après une courte synthèse des problèmes concernant la modélisation des

---

<sup>1</sup>Les mathématiques sont une étude où l’on ignore de quoi l’on parle et où l’on ne sait pas si ce qu’on dit est vrai.

SDAC, nous définissons sur la base de la formalisation des systèmes dynamiques, la notion de SAM.

### 2.1.1 Points importants du chapitre précédent

Les SDAC sont un type de système où de nombreux éléments différents doivent être pris en compte avec leurs interactions. Certains formalismes, bien que disposant d'une base théorique et pratique solide en modélisation se révèlent mal adaptés. Ainsi, l'approche classique de la théorie des systèmes dynamiques et celle de la modélisation multi-agents (autour de l'approche objet) n'offrent pas de solution réellement satisfaisante en raison de la très grande taille des SDAC et de la complexité engendrée. Nous avons donc vu qu'il existe de nombreux travaux visant à proposer, soit un formalisme original pour ce type de système (voir concernant le formalisme de Holland la section 1.1.2 ou [35]), soit une approche méthodologique où est définie une structure de simulation permettant de prendre en compte un grand nombre d'exemples de SDAC tirés de la biologie (voir pour les travaux de Bagley et co-auteurs la section 1.1.2 ou [2]).

Ces approches présentent des différences importantes. Holland, bien que gardant un aspect très général à son formalisme, ne le développe que sur une base temporelle discrète. Il précise d'ailleurs que l'adaptation de son formalisme à une base temporelle continue n'aurait rien de trivial, fournissant pour certains points "des problèmes de recherche à part entière" (1.1.2). Quant à l'approche de Bagley, elle se limite à l'implémentation et à la simulation de systèmes d'équations différentielles ordinaires modifiées durant la simulation par des règles stochastiques.

Au delà de leurs différences, le point commun de ces travaux est de proposer une structure à deux niveaux : un premier niveau qui représente le système en interaction avec son environnement et un deuxième niveau qui est en charge de modifier ("d'adapter") cette structure, selon un schéma donné. Pour Holland, ce schéma (appelé plan adaptatif), est considéré comme une fonction renvoyant un opérateur qui modifie la structure du système (de façon probabiliste). Pour Bagley, ce niveau supérieur se matérialise par un certain nombre de règles, qui (également de manière probabiliste) modifient la structure pour construire le nouveau système en ajoutant ou en supprimant des équations.

En terme de théorie de l'information, ce niveau supérieur peut souvent être interprété comme la nécessité de réintroduire de l'information nouvelle, information qui n'apparaît plus dans le système parce qu'elle a été perdue lors d'un processus d'approximation. Cette information perdue peut ne pas être intéressante pendant un temps important et le devenir par la suite, d'où la nécessité d'une fonction supérieure qui vient modifier la structure courante et recréer ainsi les changements dans le système.

Notre idée pour exploiter les similitudes des approches antérieures est donc la suivante : nous définirons notre SAM comme une structure à deux niveaux, un premier niveau composé d'un ensemble de différents systèmes dynamiques et un

deuxième niveau définissant une fonction qui vient changer au cours du temps le système dynamique (SD) courant, la modification de structure est donc vue comme le passage d'un  $SD_1$  à un autre  $SD_2$ . L'avantage étant que, si l'on dispose d'une définition suffisamment générale de ce qu'est un SD alors notre définition héritera de cette généralité.

Dans la section suivante, nous développons l'aspect intuitif de la notion de SD, de telle façon que la définition générale de Kalman apparaisse ensuite comme naturelle.

## 2.1.2 Définition des systèmes dynamiques de Kalman

Un **système dynamique** est, en termes non mathématiques (définition du Petit Robert), un système "où intervient la notion de temps".

Si l'on veut chercher une définition mathématique et formelle, la question est alors de savoir quels objets ou assemblages mathématiques permettent de donner une bonne représentation de ce que nous nommons temps et système.

Commençons par le **temps**. Que cela soit vrai ou non (si cela a même un sens de répondre à cette question), la perception que nous avons du temps nous le fait considérer comme continu et ayant un sens ("la flèche du temps"). Ce sont deux propriétés qu'il partage avec l'ensemble des réels orientés (que l'on y raccroche l'idée d'infini comme l'on le fait classiquement ou celle d'indiscernabilité que l'on utiliserait en analyse non standard). Il est donc naturel de modéliser le temps par la droite réelle<sup>2</sup>.

Pour parler de la description d'un système, nous resterons volontairement dans un cadre très général. Si l'on veut parler d'un objet de manière précise, cela suppose que nous disposons d'une façon de le décrire. Si nous regroupons maintenant les différentes descriptions possibles, nous pouvons identifier l'**état** du système à un élément de cet ensemble appelé **ensemble d'état**. L'objet étudié n'existe pas en lui-même, mais au travers de son **environnement** avec lequel il interagit. Comme nous regardons le monde dans l'optique de notre système, nous nous intéressons aux informations que l'extérieur peut lui communiquer et nous les appelons **entrées**. De même, étant nous même extérieurs à ce système, il se peut que nous n'ayons pas totalement accès à l'état du système, mais seulement à des informations que nous appelons **sorties**.

En connectant ces deux notions, la notion de système donne un sens à l'évolution au cours du temps des composants d'un système. Précisons bien que nous ne sommes aucunement forcé de nous intéresser tout le temps à notre système, on considère donc seulement un sous-ensemble de notre "temps absolu" représenté

---

<sup>2</sup>Pour une approche scientifique de la notion de temps et de l'évolution de sa conception dans l'histoire de la science, on pourra lire "Entre le temps et l'éternité" de Isabelle Stengers et Ilya Prigogine [70].

par la droite réelle. Qu'entend-t-on maintenant par évolution<sup>3</sup>? Que le système change avec le temps, ce qui s'exprime par le fait que, lorsque le temps s'écoule, on passe d'une description à une autre dépendant de certaines données comme le temps écoulé, l'état de l'environnement extérieur durant celui-ci ou encore l'état duquel le système est parti. On définit une **fonction de transition** à laquelle on assigne cette tâche : renvoyer l'état final sous les conditions d'évolution.

Une traduction de cette perception relativement intuitive de la notion de système dynamique a été développée sous forme axiomatique dans la première partie d'un livre co-écrit par Kalman, Falb et Arbib<sup>4</sup> [41]. Nous rappelons cette définition fondamentale en nous basant sur la traduction de [41] :

**Définition.** Un **système dynamique**  $\Sigma$  est un concept mathématique composite défini par les axiomes suivants :

1. Il existe un **ensemble temps**  $T$  donné, un **ensemble d'état**  $X$ , un ensemble de **valeurs d'entrée**  $U$ , un ensemble de **fonctions d'entrée** admissibles  $\Omega = \{\omega : T \rightarrow U\}$ , un ensemble de **valeurs de sortie**  $Y$ , et un ensemble de **fonctions de sortie**

$$\Gamma = \{\gamma : T \rightarrow Y\}.$$

2. (Direction du temps).  $T$  est un sous-ensemble ordonné des réels.
3. L'espace des entrées  $\Omega$  satisfait les conditions suivantes :
  - (a) (Non trivialité).  $\Omega$  est non vide.
  - (b) (Concaténation des entrées). Un **segment d'entrée**  $\omega_{(t_1, t_2]}$  est  $\omega \in \Omega$  restreint à  $(t_1, t_2]T$ . Si  $\omega, \omega' \in \Omega$  et  $t_1 < t_2 < t_3$ , il existe un  $\omega'' \in \Omega$  tel que  $\omega''_{(t_1, t_2]} = \omega_{(t_1, t_2]}$  et  $\omega''_{(t_2, t_3]} = \omega'_{(t_2, t_3]}$ .
4. Il existe une **fonction de transition d'état**

$$\varphi : T \times T \times X \times \Omega \rightarrow X$$

dont la valeur est l'état  $x(t) = \varphi(t; \tau, x, \omega) \in X$  résultant à l'instant  $t \in T$  à partir de l'état **initial**  $x = x(\tau) \in X$  à l'**instant initial**  $\tau \in T$  sous l'**action de l'entrée**  $\omega \in \Omega$ .  $\varphi$  possède les propriétés suivantes :

- (a) (Direction du temps).  $\varphi$  est définie pour tout  $t \geq \tau$ , mais pas nécessairement pour tout  $t < \tau$ .<sup>5</sup>
- (b) (Consistance).  $\varphi(t; t, x, \omega) = x$  pour tout  $t \in T$ , tout  $x \in X$  et tout  $\omega \in \Omega$ .

<sup>3</sup>On parle ici bien sûr de l'évolution temporelle d'un système et non de l'évolution darwinienne dont nous avons développé les concepts de base à la section 1.2.1.

<sup>4</sup>Il n'est pas étonnant que ce livre ait pris une place importante dans la littérature scientifique. En effet, il a été placé dès le départ sous les meilleurs auspices, étant dédié "To CONSTANTINA, FORTUNA, and PRUDENCE" les prénoms respectifs des femmes des trois co-auteurs...

<sup>5</sup>léger abus de notation  $\varphi : T \times T \times X \times \Omega \rightarrow X$

(c) (Propriété de composition). Quels que soient  $t_1 < t_2 < t_3$ , on a

$$\varphi(t_3; t_1, x, \omega) = \varphi(t_3; t_2, \varphi(t_2; t_1, x, \omega), \omega)$$

pour tout  $x \in X$  et tout  $\omega \in \Omega$ .

(d) (Causalité). Si  $\omega, \omega' \in \Omega$  et  $\omega_{(\tau, t]} = \omega'_{(\tau, t]}$ , alors

$$\varphi(t; \tau, x, \omega) = \varphi(t; \tau, x, \omega').$$

5. Il existe une **application lecture**<sup>6</sup>  $\eta : T \times X \rightarrow Y$  donnée qui définit la sortie  $y(t) = \eta(t, x(t))$ . L'application  $(\tau, t) \rightarrow Y$  donnée par  $\sigma \mapsto \eta(\sigma, \varphi(\sigma; \tau, x, \omega))$ ,  $\sigma \in (\tau, t]$ , est un **segment de sortie**, c'est-à-dire, la restriction  $\gamma_{(\tau, t]}$  d'un  $\gamma \in \Gamma$  à  $(\tau, t]$ .

Le grand intérêt de cette définition est qu'elle conserve un caractère très général à la définition de système dynamique, tout en employant un formalisme précis et rigoureux. Nous nous appuyons sur cette définition pour servir de base à notre définition des SAM : à un ensemble de SD tels que définis dans cette section, nous ajoutons "au-dessus" une fonction qui vient commuter dynamiquement et passer d'un SD à un autre en fonction de l'état du système. Nous avons essayé de nous inspirer de la précision et de la rigueur de la définition précédente pour établir dans la section suivante une définition des SAM.

### 2.1.3 Définition des systèmes à métadynamique

Nous commençons par donner quelques explications sur les axiomes de la définition qui est le cœur de cette section. Ces explications ont pour but de clarifier et de faciliter la lecture de la définition qui suit sans pour autant venir l'alourdir en s'insérant à l'intérieur de celle-ci. Nous finissons par des remarques et un exemple que nous avons mis au point pour mettre en avant certains aspects de la définition présentée et montrer sa généralité.

#### Détails sur les axiomes choisis

La définition se compose de trois parties :

- la première partie présente le niveau dynamique, comme un ensemble de systèmes dynamiques,
- la deuxième partie expose le domaine d'application de la règle métadynamique et finalement,
- la troisième partie décrit comment évolue le système une fois définis ces deux niveaux, elle donne ainsi les critères pour qu'un SAM soit défini.

---

<sup>6</sup> *readout*

1. La première partie de la définition intitulée “axiome dynamique” concerne la définition des objets constituant le premier niveau de notre système, le niveau dynamique. L’axiome dynamique stipule que ces objets sont agencés de telle façon qu’ils constituent un ensemble de SD au sens de Kalman. C’est dans cet ensemble que le niveau supérieur viendra commuter. Il est possible de distinguer ces objets en deux classes :
  - les objets communs à tous les SD : le temps  $T_0$ , l’ensemble des valeurs d’entrée  $U$ , l’ensemble des fonctions d’entrée admissibles  $\Omega$ , l’ensemble des valeurs de sortie  $Y$  et l’ensemble des fonctions de sortie  $\Gamma$ ,
  - les objets propres à chaque SD et indexés par deux familles d’indices,  $I$  et  $J$  : des ensembles d’état  $X_i$  et des fonctions de transition  $\varphi_{i,j}$  regroupés en  $\{X_i\}_{i \in I}$  et  $\{\varphi_{i,j}\}_{i \in I, j \in J}$ .

Dans la notation  $\varphi_{i,j}$ , le premier indice ( $i \in I$ ) représente l’ensemble d’état sur lequel  $\varphi_{i,j}$  opère, le second ( $j \in J$ ) représente le numéro d’ordre de  $\varphi_{i,j}$ ,  $j^{\text{ième}}$  fonction de transition sur  $X_i$ .

Un premier axiome qui sera numéroté 1.(a) en page 1 précisera que les  $\Sigma_{i,j} = \{T_0, X_i, \varphi_{i,j}, U, \Omega, Y, \Gamma\}$  sont des SD qui ont leur ensemble d’état et leur fonction de transition propre, mais partagent le temps et “l’interfaçage”, c’est-à-dire, les “langages” d’entrée et de sortie (figure 2.1).

Il est à remarquer que cet ensemble de systèmes dynamiques n’a pas de propriété particulière, il ne doit par exemple pas forcément être de cardinal dénombrable ; nous présenterons d’ailleurs après la définition un exemple où le premier niveau est constitué d’un ensemble non dénombrable de SD (voir section 2.1.4).

2. La deuxième partie de la définition concernera le niveau métadynamique en tant que fonction  $r$ , appelée règle métadynamique, agissant sur le niveau dynamique inférieur que l’on vient de définir. Nous désignons donc cette deuxième partie sous le nom “d’axiomes métadynamiques”.

Nous commençons par définir les ensembles sur lesquels agit la règle métadynamique. Comme les commutations ne dépendent pas des ensembles d’états en eux-mêmes, mais éventuellement de certains états critiques, la métadynamique doit être définie sur l’ensemble de tous les états possibles des SD, c’est-à-dire la réunion  $X$  de tous les ensembles d’états  $X_i$ . La commutation peut dépendre également de la dynamique en présence représentée par la fonction de transition courante. La règle métadynamique  $r$  dépend donc également de  $D$ , ensemble de toutes les fonctions de transition.

On définit finalement l’ensemble de temps  $T_1$  propre à  $r$ .

Un fois les objets donnés, on pourra exposer les axiomes auxquels ils obéissent :

- Axiome 2.(a) Hiérarchie dynamique :  $T_1$  est inclu dans  $T_0$ , c’est-à-dire que l’exécution de  $r$  est moins “fréquente” que celle des dynamiques sous-jacentes. Deux arguments justifient cette hypothèse :



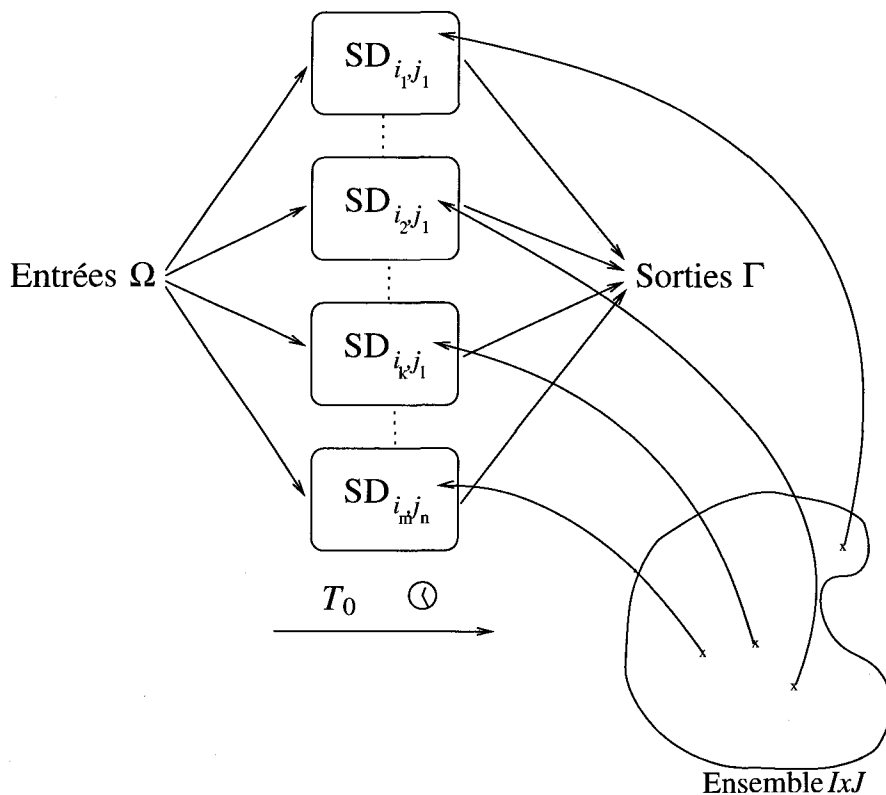


FIG. 2.1 – Le niveau dynamique : un ensemble de SD partageant les mêmes ensembles d'entrée et de sortie.

- le premier est le principe de Weiss [74] qui stipule que dans un système hiérarchiques, les niveaux inférieurs sont plus rapides que les niveaux supérieurs. Ainsi, les échelles de temps au niveau atomique sont plus rapides qu'au niveau moléculaire lui même plus rapide que le niveau cellulaire, lui même plus rapide, etc.
- Le deuxième argument est plus élémentaire, étant donnée l'architecture de notre définition, on connaît le système au travers de la fonction  $\Gamma$  de sortie des SD. Imaginons maintenant que  $T_1$  (temps métadynamique) soit plus fin que  $T_0$  (temps dynamique), cela voudrait dire qu'entre deux instants  $t_1$  et  $t_3$  de  $T_0$ , la structure du SD changerait, mettons en  $t_2 \in [t_1, t_3]$ ,  $t_2 \notin T_0$  et  $t_2 \in T_1$ . De l'extérieur, on ne verrait que  $SD_{t_1}$  et  $SD_{t_3}$  et tout se passerait comme si la structure passait directement de  $SD_{t_1}$  à  $SD_{t_3}$  sans passer par  $SD_{t_2}$ . Il est donc seulement nécessaire de définir  $T_1$  sur un sous-ensemble de  $T_0$ .
- Axiome 2.(b) l'axiome suivant est un axiome de cohérence, il doit y avoir cohérence entre l'état et la fonction de transition d'arrivée : cette fonction doit définir sur cet ensemble d'état un SD. Pour prendre un exemple imagé, si l'ensemble d'état est {herbe, lapins, renards}, la fonction de transition

pourrait être : l’herbe pousse, les lapins mangent l’herbe et se reproduisent, les renards mangent les lapins et se multiplient, les trois meurent. Mais la fonction de transition associée ne pourrait être en aucun cas : les ours mangent du miel et les chasseurs tuent les ours.

3. La dernière partie de la définition (axiome 3) explicitera de quelle manière on déduit l’évolution du SAM au cours du temps, en faisant interagir la règle métadynamique et la partie dynamique. Quand la règle est “muette”, c’est-à-dire quand elle ne provoque aucune commutation, l’évolution du système suit la dynamique courante. Quand elle arrive sur un point de commutation, elle commute : l’espace d’état (et donc l’état) et la fonction de transition changent, on a apparition d’un nouveau système dynamique. Si la métadynamique est définie sur deux intervalles consécutifs, l’évolution s’obtient par la concaténation des évolutions sur les deux intervalles considérés. Dernier point, et non des moindres, la règle d’arrêt : on ne peut définir un SAM sur un intervalle de temps que si on se limite à un nombre fini de commutations. Le but de cet axiome est d’éviter les systèmes de type Zénon<sup>7</sup> qui contiennent un nombre infini de commutations en un temps fini. On pourrait, par exemple, ainsi imaginer un système contenant une variable  $x(t) = \sin\left(2\pi\frac{1}{1-t}\right)$ . Si l’on force le système à commuter chaque fois que  $x(t) = 0$ , on obtient sur  $T_0 = [a, b] \ni 1$  un système de type Zénon que notre règle d’arrêt exclut.

### Définition d’un système adaptatif à métadynamique (SAM)

Nous donnons maintenant, après avoir défini un SD au sens de Kalman et avoir développé notre intuition sur les axiomes qui suivent, la définition formelle d’un SAM.

**Définition.** Un système adaptatif à métadynamique (SAM) est un concept mathématique composite défini par les axiomes suivants :

#### 1. Axiome dynamique

Il existe un ensemble temps  $T_0$ , un ensemble  $\{X_i\}_{i \in I}$  de tous les ensembles d’état possibles  $X_i$ , un ensemble de valeurs d’entrée  $U$ , un ensemble de fonctions d’entrée admissibles  $\Omega = \{\omega : T_0 \rightarrow U\}$ , un ensemble de valeurs de

---

<sup>7</sup>du nom du fameux paradoxe de Zénon d’Elée, où Achille essaie de rattraper la tortue. La question est la suivante : sachant qu’Achille avance deux fois plus vite que la tortue, mais qu’elle possède de l’avance au départ, arrivera-t-il à la rattraper ? Le raisonnement suivant nous prouve que non : pendant qu’Achille parcourt la distance  $d$  qui le sépare de la tortue, celle-ci parcourt une distance  $\frac{d}{2}$ . En itérant, on s’aperçoit qu’Achille se rapproche de la tortue (il divise la distance qui le sépare par deux à chaque étape), mais ne la rejoint jamais. Bien sûr, les iconoclastes briseurs de paradoxe diront impitoyablement que la série  $\left\{\frac{d}{2^n}\right\}$  converge ( $\sum_{n=0}^{\infty} \frac{d}{2^n} = 2d$ ) et que donc Achille rejoint la tortue au bout de deux fois la distance initiale. On retrouve néanmoins également ce paradoxe (appelé aussi paradoxe de la dichotomie) dans des domaines très divers comme par exemple la mécanique quantique [70].

sortie  $Y$ , un ensemble de fonctions de sortie  $\Gamma = \{\gamma : T_0 \rightarrow Y\}$  et un ensemble de fonctions de transition  $\{\varphi_{i,j}\}_{i \in I, j \in J}$  sur les ensembles d'états  $X_i$ ,

$$\varphi_{i,j} : T_0 \times T_0 \times X_i \times \Omega \rightarrow X_i.$$

(a)  $\forall i \in I, j \in J, \{T_0, X_i, \varphi_{i,j}, U, \Omega, Y, \Gamma\}$  est un système dynamique respectant les axiomes de Kalman [41].

## 2. Axiomes métadynamiques

Soit  $X = \bigcup_{i \in I} X_i$  l'ensemble de tous les états possibles du système.

Soit  $D = \{\varphi_{i,j}\}_{i \in I, j \in J}$  l'ensemble de toutes les fonctions de transitions possibles.

Il existe un ensemble temps métadynamique donné  $T_1$  et une règle métadynamique<sup>8</sup>

$$r : T_1 \times X \times D \rightarrow X \times D.$$

(a) (Hiérarchie temporelle)  $T_1 \subseteq T_0$

(b) (Cohérence états/transitions) Si  $r(t, x_1, \varphi_{i_1, j_1}) \mapsto (x_2, \varphi_{i_2, j_2})$ ,  
 $\forall t \in T_1, x_1 \in X_{i_1} \implies x_2 \in X_{i_2}$ .

Si  $t \in T_1$  et  $r(t, x, \varphi) = (x, \varphi)$ ,  $r$  est dite *muette* en  $(t, x, \varphi)$  sinon  $(t, x, \varphi)$  est un *point de commutation*.

## 3. Axiomes de définition de la fonction de méta-transition

Soit  $x_t = \varphi(t, t_1, x_1, \omega)$  avec  $\omega$  une fonction d'entrée.

Il existe une fonction de méta-transition  $\Phi : T \times T \times X \times D \times \Omega \rightarrow X \times D$

(a) (Phase dynamique pure) Si  $r$  est muette sur  $(t, x_t, \varphi)$ ,  $\forall t \in [t_1, t_2[T_1$   
 Alors  $\Phi$  est définie entre  $t_1$  et  $t_2$  et en ce qui concerne l'instant final  $t_2$  :  
 – si  $t_2 \notin T_1$ , alors

$$\Phi(t_2, t_1, x_1, \varphi, \omega) = (x_{t_2}, \varphi),$$

– sinon  $t_2 \in T_1$  et

$$\Phi(t_2, t_1, x_1, \varphi, \omega) = r(t_2, x_{t_2}, \varphi).$$

(b) (Règle de concaténation) Si  $\exists t_2 \in ]t_1, t_3[$  tel que  $\Phi$  est définie entre  $t_1$  et  $t_2$  et entre  $t_2$  et  $t_3$  alors  $\Phi$  est définie entre  $t_1$  et  $t_3$  et

$$\Phi(t_3, t_1, x_1, \varphi, \omega) = \Phi(t_3, t_2, \Phi(t_2, t_1, x_1, \varphi, \omega), \omega)$$

---

<sup>8</sup>Nous ne souhaitons pas considérer ici les règles métadynamiques probabilistes que nous explicitons (en 2.1.5) comme une extension. Dans ce cas, le résultat n'est plus un couple  $X \times D$ , mais une distribution  $Pr(t, x, \varphi)$  sur  $X \times D$ . Cette extension bien que simple est très importante dans l'étude des SDAC, car souvent, l'information perdue que réintroduit la règle métadynamique se modélise par un changement stochastique de structure.

- (c) (Règle d'arrêt)  $\Phi$  est définie entre  $t_1$  et  $t_2$  si, en respectant les axiomes précédents, il existe seulement un nombre fini de points de commutations sur  $[t_1, t_2]$ .

**Remarque 7** En 3.(a) dans le second cas, quand  $t_2 \in T_1$ , nous supposons pour ne pas alourdir la définition que  $r$  est muette en  $(t_2, r(x_{t_2}, \varphi))$ . Nous "oublions" le cas où la métadynamique "rebondirait" (c'est-à-dire que  $r$  aurait comme résultat un système dynamique et une fonction de transition qui définirait un point de commutation à l'instant donné). En fait, si le système est soumis à plusieurs commutations successives au même instant, il est toujours possible de considérer l'ensemble des commutations comme une seule transition globale qui a pour état final l'état résultant de la dernière commutation (sachant que, par 3.(c), le nombre de commutations est fini).

**Remarque 8** La définition des SAM acquiert toute sa puissance lorsque  $r$  est définie de manière constructive. On entend par constructif qu'elle ne consiste pas seulement en un simple switch (on passe d'un système prédéfini 1 à un système prédéfini 2) (définition en extension), mais qu'elle donne un algorithme pour définir le nouveau système (définition en intention). Un simple exemple numérique permettra de clarifier cette notion : voici deux manières de définir la fonction

$$\begin{aligned} \{0, 1\}^n &\rightarrow \{0, 1\}^n \\ s &\mapsto 2 \times s + 1 \end{aligned}$$

1. définition en extension

On associe à chaque élément du premier ensemble un résultat appartenant à un ensemble connu (définition en extension) :

argument	résultat
000...000	→ 000...000
000...001	→ 000...011
000...010	→ 000...101
⋮	⋮
011...111	→ 111...111
100...000	→ non défini
	etc.

2. définition en intention (et constructive<sup>9</sup>)

On décrit comment obtenir le résultat grâce à un algorithme :

<sup>9</sup> Une définition en intention n'est pas forcément constructive. Par exemple, si l'on pense à la définition des nombres premiers, comme un nombre divisible seulement par 1 et par lui-même, on a bien affaire à une définition qui n'est pas en extension (elle ne donne pas tous les nombres premiers), mais elle n'est pas pour autant constructive. En effet, elle ne donne pas d'algorithme pour construire cet ensemble.

*Si le premier bit est 0, on décale à gauche et on met le dernier bit à 1.  
Sinon on renvoie non défini.*

**Remarque 9** *On peut se demander si, un SAM est un SD d'un type un peu particulier, où l'on définirait la fonction de méta-transition modifiée comme une fonction de transition sur un certain espace. En fait, lors d'une commutation, on considère au même instant deux représentations du système comme "équivalente" (l'adaptation est vue comme ce changement), on viole donc l'axiome 4.(b) de consistance des SD. On peut en fait considérer les SAM comme des SD, si on admet que l'on puisse identifier les états "équivalents" (les points d'entrées et de sortie des commutations).*

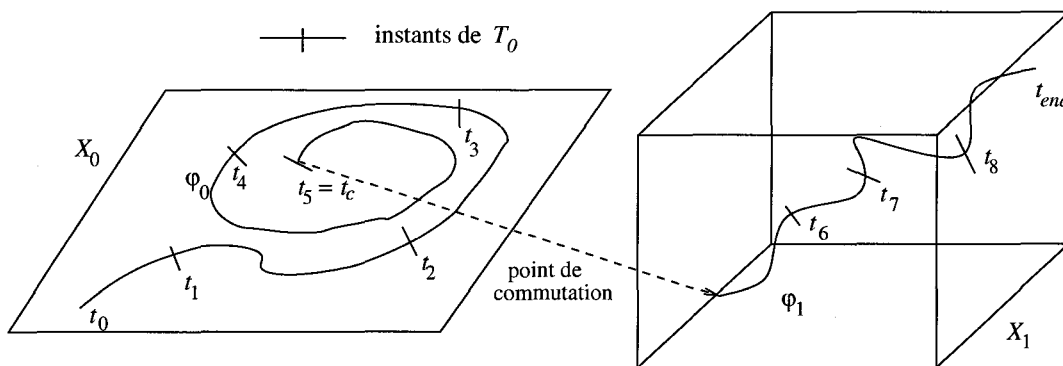


FIG. 2.2 – Illustration de certains axiomes d'un SAM

Pour illustrer ces notions, considérons un SAM en action (figure 2.2). A  $t_5 = t_c$ , on passe d'un ensemble d'état à deux dimensions à un ensemble d'état à trois dimensions. Les dynamiques de base sont continues, donc  $T_0 = [t_0, t_{end}]$ . En revanche, la métadynamique est à base discrète :  $T_1 = \{t_1, t_2, t_3, \dots, t_8\}$ . Etant donné que la métadynamique est muette en  $t_1, t_2, t_3$  et  $t_4$ , l'évolution de  $t_0$  à  $t_5$  est une phase purement dynamique et se termine par une commutation (axiome 3.(a)). L'évolution de  $t_5$  à  $t_{end}$  est également purement dynamique (la métadynamique est muette en  $t_6, t_7$  et  $t_8$ ). La jonction entre les deux intervalles temporels est faite en utilisant la règle de concaténation. Comme il y a une seule commutation, le système respecte l'axiome 3.(c) et est donc défini.

#### 2.1.4 Un exemple original : le tracé du lapin de Douady

Nous avons mentionné que la famille  $\{X_i\}_{i \in I}$  ne devait pas nécessairement être finie, ni même dénombrable. Bien entendu, les exemples "classiques" présentés au chapitre 1 et sur base desquels nous avons établi notre axiomatique générale sont restreints à des cas finis ou au moins dénombrables (par exemple, un système dont les systèmes dynamiques de base sont décrits par des équations différentielles ordinaires à un indice  $i$  fini ou entier).

Sur base de ces exemples, le lecteur peut se demander si admettre que  $I$  est non dénombrable a un sens. Pour analyser cette question, nous donnons ci-après un exemple de SAM à  $I$  non dénombrable et présentant un intérêt pratique réel. Cet exemple, et de nombreux autres que nous pourrions formuler, montre que notre axiomatique unifie un grand nombre de problèmes. Nous montrerons à la fin de ce chapitre qu'elle permet aussi de préciser des concepts et questions génériques pour tous ces systèmes qui sans elle, sont entièrement déconnectés conceptuellement. Nous allons maintenant décrire notre exemple, un algorithme original, basé sur un SAM pour tracer un ensemble de Julia de manière efficace.

### Ce qu'il faut savoir sur les ensembles de Julia

Les ensembles de Julia (EJ) sont définis comme étant des frontières fractales entre des bassins d'attraction de dynamiques quadratiques complexes du type

$$z_{n+1} = z_n^2 + c \quad (2.1)$$

avec  $c \in \mathbb{C}$ . Ces ensembles sont célèbres pour les images très spectaculaires et colorées auxquels ils donnent naissance (voir [59]).

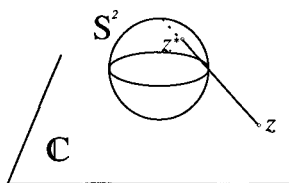
Il existe deux sortes d'EJ :

- les EJ à intérieur vide de types poussières de Fatou qui sont constitués d'un amas de points du plan complexe,
- les EJ à intérieur non vide, séparant le plan complexe en deux parties par une frontière fractale<sup>10</sup>.

Ce que nous développons maintenant concerne exclusivement ce dernier type d'EJ. Ces EJ séparent donc le plan complexe en deux bassins d'attraction : l'itération de la relation 2.1 en partant de conditions initiales à l'intérieur de l'EJ converge vers un cycle limite, alors que l'itération de cette même relation en partant de l'extérieur converge vers l'infini<sup>11</sup>. Au fur et à mesure que l'on se rapproche de cette frontière, la vitesse de convergence des points ralentit. Les points de la frontière, quant à eux, errent de manière chaotique sur celle-ci. L'idée est donc de faire évoluer un nuage de points en utilisant la vitesse de convergence comme valeur d'adaptation de ces points.

<sup>10</sup>On peut noter que les paramètres du plan complexe donnant naissance au deuxième type d'EJ définissent l'ensemble de Mandelbrot, autre fractale très célèbre.

<sup>11</sup>Nous nous autorisons à parler de "convergence" vers l'infini car on considère le plan complexe compactifié, i.e.  $\mathbb{C} \cup \{\infty\}$ , identifié par la projection stéréographique  $z \rightarrow z^*$  à la sphère unité euclidienne  $S^2$ .



## Principe de l'algorithme

Le but de cet algorithme basé sur une métadynamique est donc de faire évoluer un ensemble de point du plan complexe vers la frontière fractale. Pour cela, nous commençons par donner un poids à chacun des points de cet ensemble. Ces poids varient grâce à une dynamique qui récompense les points les plus adaptés (les points qui sont les moins attirés par les valeurs limites) et pénalise les moins intéressants. Quand les points les plus performants ont un certain capital, ils sont autorisés à se dupliquer dans leur voisinage et peuvent ainsi se propager. Les points les moins efficaces sont eux ammenés à disparaître une fois atteint un seuil minimal. Ce système de seuil qui change la structure du niveau dynamique (c'est-à-dire notre ensemble pondéré de points de  $\mathbb{C}$ ) constitue notre métadynamique.

## Implémentation pour le lapin de Douady

Le lapin de Douady correspond au cas  $c = -\frac{3}{2} + \frac{1}{2}(a + b) + i\frac{\sqrt{3}}{2}(a - b) \approx -0.12256 + 0.744862i$ , avec  $a = \sqrt[3]{\frac{25+\sqrt{621}}{54}}$  et  $b = \sqrt[3]{\frac{25-\sqrt{621}}{54}}$ . Le cycle attractif correspondant est  $\{0, c, c^2 + c\}$ . Nous avons représenté le lapin à la figure 2.3.

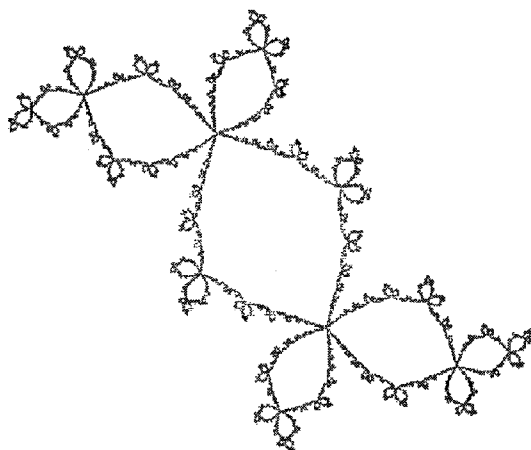


FIG. 2.3 – L'ensemble de Julia appelé "lapin de Douady" du nom du mathématicien français Adrien Douady.

**Niveau dynamique** Nos dynamiques de base sont discrètes donc  $T_0 = \mathbb{N}^+$ .  $J$ , l'ensemble des indices<sup>12</sup> est l'ensemble des ensembles finis de points de  $\mathbb{C}$  (et est donc par conséquent non dénombrable).

<sup>12</sup>Nous avons choisi d'utiliser  $J$  et  $j$ , et non  $I$  et  $i$ , pour éviter la confusion avec le nombre complexe  $i = \sqrt{-1}$ .

$X_{j \in J} = \mathbb{N}^{\text{card}(j)}$  est l'ensemble des poids des points de  $j$  (en effet  $j$  est par définition un ensemble de points de  $\mathbb{C}$ ).

On définit :

$$P_c : \mathbb{C} \rightarrow \mathbb{C} \\ z \mapsto z^2 + c \quad \text{et} \quad P_c^n(z) = \underbrace{P_c \circ \dots \circ P_c}_n(z)$$

avec  $c = -\frac{3}{2} + \frac{1}{2}(a+b) + i\frac{\sqrt{3}}{2}(a-b) \approx -0.12256 + 0.744862i$ ,

avec  $a = \sqrt[3]{\frac{25+\sqrt{621}}{54}}$  et  $b = \sqrt[3]{\frac{25-\sqrt{621}}{54}}$ .

A chaque point  $z$  from  $\mathbb{C}$  est donné une valeur sélective :

$$\mu_n : \mathbb{C} \rightarrow \mathbb{R}^+ \\ z \mapsto \min(|P_c^n(z)|, |P_c^n(z) - c|, |P_c^n(z) - c^2 - c|, |\frac{1}{P_c^n(z)}|)$$

$\{0, c, c^2 + c\}$  sont les points du cycle super attractif de période 3

$\frac{1}{P_c^n(z)}$  se réfère à l'attraction vers l'infini (sachant que  $P_c^n(z) \rightarrow \infty$  équivaut à  $\frac{1}{P_c^n(z)} \rightarrow 0$ ).

On augmente le poids des points près de la frontière, au pro rata de leur lenteur de convergence. Le poids du point  $j$  au cours du temps est donc donné par :

$$w_{z \in j}(t+1) = w_z(t) + \sum_{q \in j, q \neq z} \delta(z, q) \quad \text{avec} \quad \delta(z, q) = 1 \text{ si } \mu_t(z) > \mu_t(q) \text{ et } -1 \text{ sinon}$$

**Niveau métadynamique** Quand le poids d'un point atteint :

- un seuil supérieur, le point est récompensé en étant autorisé à se "reproduire" aléatoirement dans son voisinage,
- un seuil inférieur, le point est supprimé.

**Remarque 10** *Au fur et à mesure que le temps croît, la valeur sélective devient de plus en plus précise, ce qui est un des points intéressants puisque nous concentrons ainsi le calcul, sur les points délicats qui sont à la frontière fractale, par définition très découpée. Dans beaucoup d'algorithmes classiques de calculs d'ensemble de Julia, le temps de calcul est gaspillé dans le calcul de points à l'intérieur de l'ensemble de Julia<sup>13</sup>. En effet, ces algorithmes calculent un grand nombre de points sans distinguer si le point est intéressant ou non. En faisant évoluer le nuage et en améliorant le calcul au fur et à mesure que le nuage est se rapproche de la frontière, nous introduisons une telle distinction.*

Ce SAM produit un nuage de points qui vient s'agglutiner sur l'ensemble de Julia représenté par la figure 2.3 (appelé "lapin de Douady" d'après le mathématicien français Adrien Douady). Dans ce cas précis, il est probable que l'algorithme

<sup>13</sup>En parlant d'intérieur, on sous entend le bassin d'attraction du cycle limite, car d'un point de vue strictement mathématique, l'intérieur de l'ensemble de Julia est vide.



proposé n'est pas plus efficace que certains algorithmes déjà existants (comme par exemple la méthode d'itération inverse modifiée<sup>14</sup> [59, 60]). Cela est dû au fait, que la structure mathématique des ensembles de Julia a été étudiée de manière intensive ([20, 24, 40]), et que, par conséquent, certaines propriétés spécifiques peuvent être exploitées. Dans d'autres cas, avec des frontières dont les propriétés sont moins connues, ce type d'algorithme est extrêmement intéressant, parce qu'il permet de concentrer le calcul dans des régions spécifiques. Ce type d'algorithme inspiré par notre concept de métadynamique n'a pas, à notre connaissance, été étudié jusqu'ici dans le calcul de structures fractales. Dans un cadre plus vaste, il se rapproche des *particle algorithms* [?] domaine qui se développe actuellement dans le cadre du calcul numérique et qui est donc couvert conceptuellement par notre axiomatique des SAM.

### 2.1.5 Extensions

En suivant la même approche métadynamique, on peut étendre notre formalisme de base dans plusieurs directions. Nous présentons ici trois extensions possibles : les métadynamiques récursives multi-niveaux, les métadynamiques stochastiques et l'ajout d'entrées à la règle métadynamique.

#### Métadynamiques récursives multi-niveaux

Nous avons présenté plus haut, une structure possédant un seul niveau supérieur contrôlant un niveau dynamique inférieur. Il est en fait tout à fait possible, au prix d'une complexification de la définition, d'établir un nombre  $m$  de niveaux hiérarchiques définis par des règles  $r_{n,k}$ ,  $1 < n < m$  :

$$r_{n,k} : T_n \times X \times D \times \prod_{l=1}^{n-1} R_l \rightarrow X \times D \times \prod_{l=1}^{n-1} R_l$$

avec  $R_{n-1} = \{r_{n-1,k}\}$  et  $T_n \subseteq T_{n-1}$ . C'est-à-dire que les règles d'un niveau  $n$  donné, n'agissent plus seulement sur  $X$  et  $D$ , mais également sur les règles des niveaux inférieurs. Comme on le voit, l'axiome de hiérarchie temporelle se généralise également : plus l'ensemble temporel correspond à un niveau élevé, plus il est "lent" (en accord avec le principe de Weiss).

Ce remaniement est très simple d'un point de vue théorique, même si s'opère alors une multiplication des indices. On peut donner un exemple d'un tel système : on connaît assez bien les réseaux de neurones qui associent à un certains nombre d'entrées, des sorties, après le passage de ces entrées dans un réseau pondéré de petits sommateurs à seuil considérés comme des modèles très simplifiés de neurone. Ce système est statique, mais peut-être rendu dynamique si l'on reboucle

---

<sup>14</sup> *Modified Inverse Iteration Method*

certaines des sorties sur des entrées. On a alors un signal de sortie qui change au cours du temps. Ce type de réseaux de neurones est appelé réseaux neuronaux d'Elman [61]. Les algorithmes d'apprentissage de ce type de réseaux (comme des autres réseaux de neurones d'ailleurs), modifient les poids des liens du réseau (les "synapses") pour permettre à ceux-ci d'effectuer des tâches (reconnaissance de caractères par exemple). Il s'agit là, d'une métadynamique typique. Si l'on imagine, maintenant que cet apprentissage est dirigé par un algorithme génétique qui modifie les paramètres de l'apprentissage et la topologie du réseau en fonction des performances enregistrées dans un horizon temporel passé, on a affaire à un système à méta-métadynamique. Sur le même principe, il est relativement facile d'imaginer plus de niveaux hiérarchiques.

### Métadynamiques stochastiques

Une autre extension possible est de remplacer la règle  $r$  de transition déterministe par une règle de transition probabiliste (dans le style markovien ou non). Dans ce cas, la règle métadynamique devient :

$$\mathcal{P}r : T_0 \times X \times D \rightarrow (X \times D \rightarrow [0, 1])$$

avec  $\mathcal{P}r(t, x, \varphi)$  une distribution de probabilité sur  $X \times D$  qui représente les probabilités  $\mathcal{P}r(t, x, \varphi) \cdot (x_f, \varphi_f)$ , que  $(x, \varphi)$  passe à l'instant  $t$  en  $(x_f, \varphi_f)$ .

Les chapitres 3 et 4 seront consacrés à l'étude approfondie d'une classe de systèmes de ce type, suffisamment complexe pour illustrer tous les aspects donnés ici : les systèmes bactériens évolutifs. Dans ces systèmes, la règle métadynamique exprimera l'apparition de nouvelles espèces microbiologiques par mutation et leur disparition par extinction. Les mutations, qu'elles soient dirigées [67] ou non, ont une base aléatoire et l'extension probabiliste sera donc nécessaire.

### Métadynamiques avec entrées

Dans notre définition, nous ne considérons que des entrées au niveau dynamique. Néanmoins, il est tout à fait envisageable d'imaginer d'inclure des entrées au niveau de la règle métadynamique  $r$ . La règle prend alors la forme :

$$r : T_1 \times X \times D \times \Omega_1 \rightarrow X \times D$$

avec  $\Omega_1 = \{\omega : T_1 \rightarrow U_1\}$  un ensemble de fonction de méta-entrées admissibles et  $U_1$  un ensemble de valeurs d'entrées pour la métadynamique.

Dans ce cas, on peut s'attendre à un accroissement de la complexité aussi important que celui qui apparaît pour les automates quand on leur ajoute des entrées [13].

### 2.1.6 Conclusion

Nous disposons donc maintenant, d'une définition formelle d'une structure mathématique appelée système adaptatif à métadynamique (SAM) qui nous permet de représenter un grand nombre de SDAC. Cette définition englobe en particulier la définition formelle de Holland et la méthodologie de simulation de Bagley. Mais le principal intérêt de cette définition est de permettre maintenant l'élaboration précise d'un certain nombre de notions liées à l'évolution à long terme des SDAC. C'est l'objet de la section suivante.

## 2.2 Définitions

Nous avons proposé, dans la première partie de ce chapitre, une définition formelle des SAM. Nous disposons donc maintenant, d'un langage commun pour parler des différents SDAC qui peuvent être représenté par cette définition. Il nous est donc possible, sur base de ce formalisme, de développer un ensemble de définitions permettant d'étendre des notions classiques en théorie des systèmes dynamiques, telles que les notions de trajectoires ou d'attractivité...

### 2.2.1 Notions fondamentales

Un des aspects intéressant de notre formalisme, est qu'il fournit les outils théoriques pour suivre ce que W. Fontana et L.W. Buss appellent "les dynamiques dans l'espace des objets" [25]. La section suivante est consacrée à l'étude des "trajectoires métadynamiques", c'est-à-dire non pas le suivi des trajectoires dynamiques individuelles, mais celui du changement des structures au cours du temps dans un "espace des structures". Nous suivons pour cela les séquences du couple espace d'état/fonction de transition.

**Définition 11** Soit  $\mathcal{M}$  une métadynamique de fonction de méta-transition  $\Phi$ .

Si en  $t$ , on a  $\Phi(t, t_0, x_0, \varphi_0, \omega) = (x_t, \varphi_t)$  avec  $x_t \in X_t$  alors  $(X_t, \varphi_t)$  est appelé **méta-état** de  $\mathcal{M}$  en  $t$ .

La suite  $(X_{t_i}, \varphi_{t_i})$ ,  $i \in \llbracket 0, n \rrbracket$ ,  $0 \leq n \leq \infty$ , des méta-états de  $\mathcal{M}$  telle que  $\forall t \in [t_i, t_{i+1}[$ ,  $(X_t, \varphi_t) = (X_{t_i}, \varphi_{t_i})$  et  $(X_{t_i}, \varphi_{t_i}) \neq (X_{t_{i+1}}, \varphi_{t_{i+1}})$ <sup>15</sup> est appelé **méta-orbite** de  $\mathcal{M}$ .

Si  $n \leq \infty$ ,  $(X_{t_n}, \varphi_{t_n})$  est dit **méta-état absorbant**. La suite  $(t_i, X_{t_i}, \varphi_{t_i})$  répondant aux mêmes conditions est appelée **méta-trajectoire** de  $\mathcal{M}$ .

**Exemple 12** Supposons que l'on puisse définir chaque SD d'un SAM par un graphe dont les nœuds représenteraient les variables et la topologie la fonction de transition (les interactions entre les variables). Pour avoir une représentation de la

<sup>15</sup> c'est-à-dire  $X_{t_i} = X_{t_{i+1}}$  ou  $\varphi_{t_i} = \varphi_{t_{i+1}}$

métra-trajectoire, il faudrait dessiner la suite des graphes. On ne dispose alors plus de l'information précise sur les valeurs au cours du temps des différentes variables, mais on peut suivre l'évolution de la structure.

### 2.2.2 Caractérisation des méta-états

Nous présentons ici quelques définitions permettant de caractériser les méta-états. Ces définitions s'inspirent des définitions classiques de la théorie des systèmes dynamiques, elles sont illustrées figure 2.4.

Soit un SAM  $\mathcal{M}$ .

**Définition 13**  $x \in X$  est un **point de départ** du méta-état  $(X, \varphi)$  de  $\mathcal{M}$  si  $\exists t$  tel que  $(t, x, \varphi)$  soit un point de commutation.

**Définition 14**  $x \in X$  est **point d'arrivée** du méta-état  $(X, \varphi)$  si  $\exists t, x_2, X_2, \varphi_2$  tels que  $x_2 \in X_2$  et  $(X_2, \varphi_2)$  méta-état avec  $r(t, x_2, \varphi_2) = (x, \varphi_t)$

**Définition 15** La **sous-partie absorbante** d'un méta-état  $M = (X, \varphi)$  est

$$A(M) = \{x \in X_t \mid \forall (t_1, t_2) \in T_0^2, t_2 > t_1, r \text{ est muette en } \varphi(t_2, t_1, x, \omega)\}$$

**Définition 16** La **sous-partie émigrante** d'un méta-état  $M = (X, \varphi)$  est

$$E(M) = X - A(M).$$

**Remarque 17**  $E(M)$  est le bassin d'attraction des points de départ de  $X$ .

**Définition 18** Soit  $M = (X, \varphi)$  un méta-état d'un SAM  $\mathcal{M}$ .  $M$  est **répulsif** si :

1.  $X$  ne contient pas de points d'arrivée de  $M$ ,
2.  $A(M) = \emptyset$ ,
3.  $E(M) = X$ .

**Définition 19** Soit  $M = (X, \varphi)$  un méta-état d'un SAM  $\mathcal{M}$ .  $M$  est **attracteur** si :

1.  $M$  contient au moins un point d'arrivée,
2.  $A(M) = X$ .

Ces définitions peuvent être étendues à des ensembles de méta-états avec quelques modifications. Dans le cas des métadynamiques stochastiques, on considèrera en pratique qu'un méta-état est répulsif ou attracteur lorsque la probabilité qu'il se comporte de la même manière que dans le cas déterministe est très proche de un.

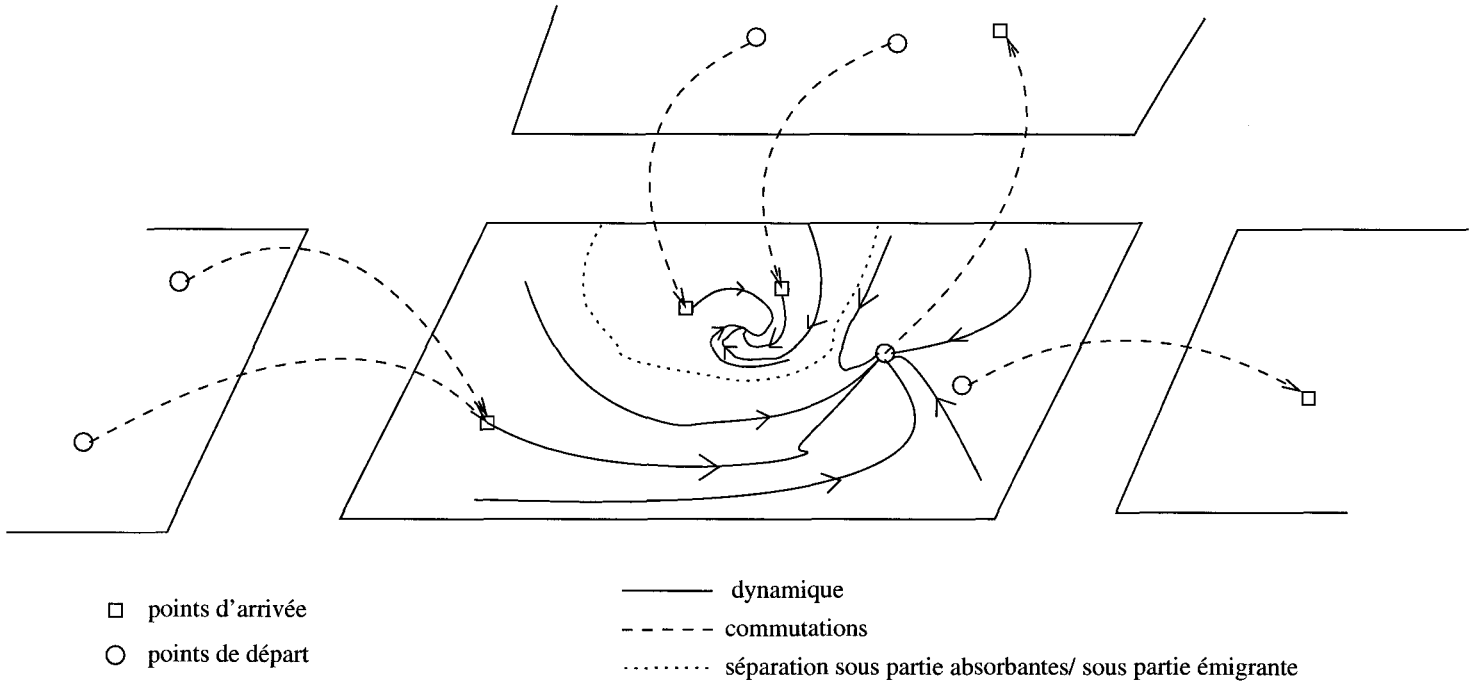


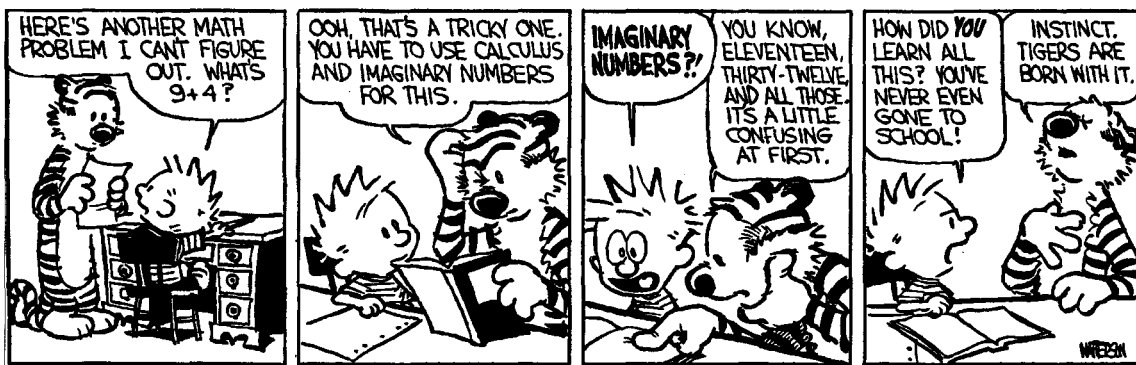
FIG. 2.4 – Propriétés des méta-états, sous-parties absorbantes et émigrantes, point de départ et d'arrivée.

## 2.3 Conclusion

Nous avons présenté dans ce chapitre un cadre formel pour les SDAC, qui se concentre autour de la définition de SAM, ainsi que des notions de trajectoires métadynamiques. Nous disposons donc, maintenant d'un certain nombre d'outils théoriques pour la description et l'étude des SDAC qui rentrent dans ce cadre. Le chapitre suivant se consacre aux aspects numériques et de simulation de ce cadre formel appliqué au SAM à base continue. Nous développons également un exemple de passage de niveau de complexité (d'un système particulière à un système différentiel) qui justifie l'utilisation des SAM.

## Chapitre 3

# SAM à base continue : aspect stochastique, justification numérique et simulation.



- Voici un autre problème de maths que j'arrive pas à résoudre. Ça fait combien  $9+4$  ?
- Oh! C'est un compliqué. Tu dois utiliser les dérivées et les nombres imaginaires pour ça.
- **LES NOMBRES IMAGINAIRES ?!**
- Tu sais, dix-onze, trente-douze et tout ça. C'est un peu déroutant au départ.
- Comment ça se fait que *toi*, tu a appris tout ça ? Tu n'as jamais même été à l'école.
- L'instinct. Les tigres naissent avec.

Calvin & Hobbes, BILL WATERSON

Nous avons vu au chapitre 1 les problèmes inhérents à la modélisation des SDAC ainsi qu'un certain nombre de modèles d'évolution. Le deuxième chapitre a été consacré à la définition d'une structure formelle appelée système à métadynamique (SAM).

Nous avons montré que la classe de systèmes couverte par ce formalisme était très large, couvrant en particulier les exemples du chapitre 1, mais présentant également des idées pour des algorithmes originaux comme l'exemple de l'ensemble de Julia de la section 2.1.4. Néanmoins, beaucoup des exemples de la littérature

que nous regroupons sous le vocable de SAM concerne des systèmes dont la bas dynamique est un ensemble variable d'ODE sur lequel agit une métadynamique, stochastique ou non, discrète souvent tiré d'une simulation d'un mécanisme évolutif. Ces exemples, de par leur grande taille posent des problèmes quant à la nature des phénomènes métadynamiques qu'ils présentent et à la façon de les illustrer ou de les détecter. Vu la généralité de cette classe, il est nécessaire d'aborder ces questions en étudiant de façon approfondie des exemples suffisamment complexes.

Dans ce but, nous développerons au chapitre 4, un modèle de coévolution bactéries-phages. Nous utiliserons pour cela une structure du type de celle établit au chapitre 2. Ce chapitre a donc pour but de justifier l'utilisation d'une certaine classe de SAM, les SAM à base continue, dans la modélisation de certains phénomènes, en particulier l'évolution de populations. Ainsi, les SAM à base continue se présentent comme une alternative aux ODE quand on veut modéliser une population de très grande taille qui met en jeu des événements dynamiques de probabilités hétérogènes (c'est-à-dire que certains événements sont beaucoup plus rares).

Le cœur de ce chapitre concernera le passage entre modèles de niveaux de complexité différent. Nous montrerons que les SAM à base temporelle continue constituent une approximation du même ordre que celle des ODE : passage à la limite de variables discrètes en grand nombre à des variables continues. Pour cela nous étudierons le comportement (en particulier les comportements limites) d'une classe de modèles intermédiaires entre les modèles agents et les modèles à espace d'état continu : les modèles stochastiques particuliers de réactions chimiques généralisés (RCG).

Nous commencerons par montrer que ce type de modèles peut servir pour modéliser le comportement de populations de grande taille, présentant des interactions biologiques complexes : ce chapitre a donc pour but de justifier l'utilisation des SAM dans la modélisation de systèmes biologiques complexes, comme celui étudié au chapitre 4.

Nous finirons en présentant quelques idées sur la simulation et l'analyse de données issues de tels systèmes, idées qui seront largement reprises et développées lors de l'analyse du modèle au chapitre 4.

### 3.1 Du modèle multi-agents au modèle particulière

Nous avons souligné dans le chapitre 1, la rigueur de l'approche orientée objet, sa pertinence concernant la modélisation du réel, ainsi que la souplesse et l'efficacité qu'offrent au niveau de la simulation les liens privilégiés qu'elle entretient avec l'informatique. Nous avons néanmoins dégagé deux limites importantes dans la modélisation des SDAC :

- le fait de sauvegarder individuellement chaque composant (ou objet) élémentaire du système est impossible quand le système est de très grande



taille (même en prenant en compte l'amélioration continue des capacités de calcul),

- une fois le phénomène naturel “mimé” de manière exacte, l'interprétation et l'analyse du modèle deviennent aussi difficiles que celles du système original.

Une manière de répondre à ces difficultés est de ne plus sauvegarder chaque objet individuellement, mais de regrouper les objets possédant certaines caractéristiques communes. Les objets ainsi regroupés sont considérés comme des particules indistinguables : on parle alors de modèle particulaire. Les propriétés et les comportements individuels des objets  $y$  sont généralement modélisés par des lois stochastiques, que l'on obtient en faisant une analyse statistique du système.

Pour mieux comprendre le principe d'une telle approche, supposons un ensemble d'êtres dont les propriétés sont la couleur (blanche ou noire) et la position, cette position étant caractérisée par le niveau sur lequel ils se situent (niveau 1, 2 ou 3) (figure 3.1).

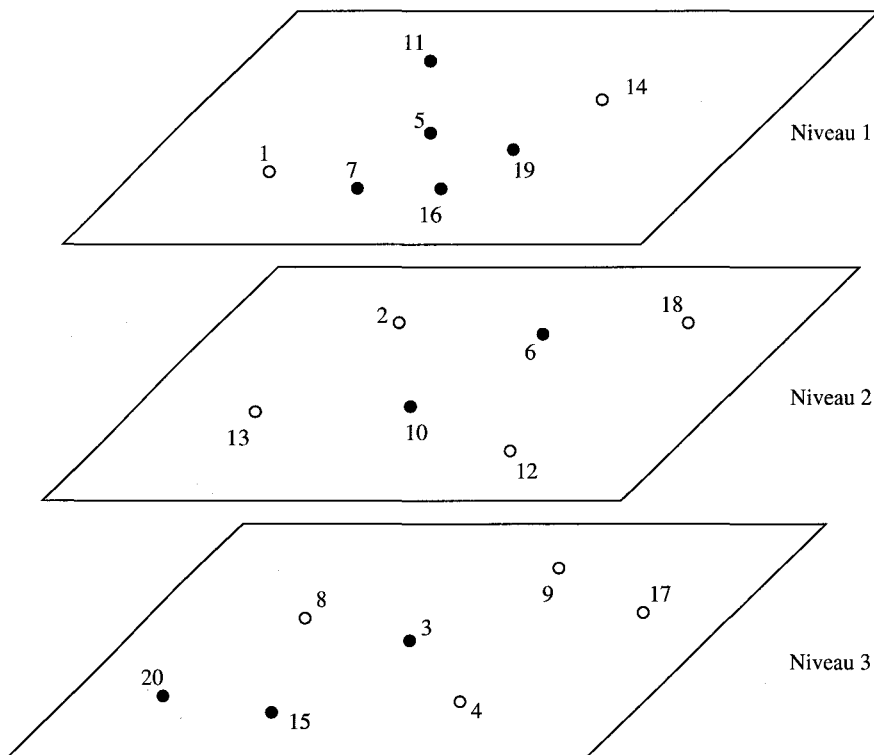


FIG. 3.1 – Exemple d'une population d'individus

Nous allons maintenant voir deux façons de représenter cette réalité, tout d'abord une modélisation de type multi-agent, où toutes les informations concernant chaque individu sont stockées et une modélisation particulaire, où l'information est condensée par l'utilisation de classes d'équivalence d'individus, l'information perdue étant résumée par l'ajout de distributions de probabilités.

Numéro	Couleur	Niveau	Numéro	Couleurs	Niveau
1	blanche	1	11	noire	1
2	blanche	2	12	blanche	2
3	noire	3	13	noire	1
4	blanche	3	14	blanche	1
5	noire	1	15	noire	3
6	noire	2	16	noire	1
7	noire	1	17	blanche	3
8	blanche	3	18	blanche	2
9	blanche	3	19	noire	1
10	noire	2	20	noire	3

TAB. 3.1 – Sauvegarde des informations relatives à tous les individus dans un modèle de type agents

### 3.1.1 Modélisation multi-agents

Chaque élément de la population est décrit individuellement par les caractères intéressant le modélisateur (tableau 3.1).

L'étape suivante de la modélisation serait de donner les lois d'évolution des éléments du système en fonction de leurs caractéristiques (les méthodes des objets). Celles-ci pourraient décrire par exemple décrire comment un individu passe d'un niveau à l'autre en fonction du nombre d'individus de chaque couleur dans son niveau.

### 3.1.2 Modèle particulière

Pour passer à une modélisation particulière, on doit regrouper les particules qui ont des caractéristiques communes. On a ici le choix entre la couleur et la position. On peut supposer par exemple que l'on ne conserve que l'information relative à la position. On obtient alors l'état suivant pour le système :

- 7 particules au niveau 1, indistingables,
- 6 particules au niveau 2, indistingables,
- 7 particules au niveau 3, indistingables.

L'information "oubliée" va se retrouver sous forme d'éléments stochastiques dans la fonction d'évolution. Par exemple, si l'on considère que les individus conservent leur couleur au cours du temps et que cette couleur a une influence sur les mouvements possibles de l'individu, la modélisation pourra tenir compte de cela en ne considérant plus seulement une évolution possible, mais deux évolutions possibles (une pour chaque couleur) avec chacune une probabilité  $\frac{1}{2}$  de réalisation (puisque qu'il y a autant de particules blanches que de particules noires dans le système).

Le temps ne fait plus alors seulement évoluer un état, mais un ensemble d'états possibles, pondérés par des lois de probabilité. Une trajectoire ne représentant qu'une trajectoire parmi un grand nombre d'autres possibles, on doit, pour se faire une idée du comportement du système, considérer un certain nombre de trajectoires correspondant à diverses réalisations possibles (au cours du temps) d'un même état initial (approche statistique).

Pour prendre un exemple qui sera repris dans la section suivante, on peut appliquer ce qui vient d'être dit aux réactions chimiques (éventuellement généralisées). On parle dans ce cas de modèle particulaire parce que l'on considère comme indistinguable toute molécule d'une même espèce chimique. La position dans la solution (qui serait présente dans une modélisation multi-agent) n'est pas conservée, même si elle joue un rôle fondamental dans la production des événements cinétiques élémentaires (la rencontre de deux molécules qui réagissent, provoquée par un choc moléculaire). On considère que, sous certaines hypothèses (grand nombre de particules convenablement réparties dans le mélange), cette information sur la position peut se résumer en une loi de probabilité de rencontre (et donc de manière globale, de réaction) ne dépendant que du nombre de particules des différentes espèces présentes.

## 3.2 Rappels sur les modèles stochastiques

Nous ne faisons ici que présenter des rappels sur la théorie des probabilités et des modèles de réactions stochastiques. Le lecteur qui désire plus de détails et de développements pourra se référer à [65], quand à la partie plus spécialement dédiée aux réactions chimiques stochastiques, elle est largement inspirée de notes prises lors d'une session du cours de modélisation en génomique fonctionnelle donnée par Pierre-Yves Boelle à l'Université de Paris VII.

Il est important de bien comprendre le fonctionnement de ce type de modèle car nous allons voir à la section suivante, comment il peut être approché, soit par un système d'équations différentielles, soit par un SAM à base continue.

### 3.2.1 Fonction de répartition, densité de probabilité et risque instantané

La plupart des définitions qui suivent sont extrêmement classiques et peuvent se retrouver dans n'importe quel manuel de théorie des probabilités. Nous ne faisons ici que donner ces définitions sans développements ni démonstrations (le lecteur intéressé pourra consulter [12] ou [65]), dans le but d'introduire une notion moins courante : le risque instantané. La compréhension de cette notion est indispensable à l'analyse et à la simulation des modèles stochastiques de réactions.

Nous nous limitons à un niveau d'abstraction accessible aux lecteurs disposant seulement des bases de la théorie des probabilités tout en essayant de conserver la

plus grande rigueur possible<sup>1</sup>.

Nous commençons par définir la notion de variable aléatoire réelle :

**Définition 20** Une *variable aléatoire réelle (v.a.r.)* est une application mesurable de l'espace probabilisé  $(\Omega, \mathcal{C}, P)$  dans  $\mathbb{R}$  muni de sa tribu borélienne  $(\mathbb{R}, \mathcal{B})$ .

**Définition 21** Soit  $T$  une v.a.r.

On appelle *loi de probabilité de  $T$*  la mesure image de  $P$  par  $T$  et on la note  $P_T$ .

Ce qui se traduit par :  $\forall B \in \mathcal{B}, P_T(B) = P(\{\omega | T(\omega) \in B\}) = P(\{T^{-1}(B)\})$ .

**Définition 22** La *fonction de répartition* d'une v.a.r.  $T$  est l'application  $F$  de  $\mathbb{R}$  dans  $[0, 1]$  définie par :

$$F(t) = P(T < t).$$

**Propriété 23**  $F$  est une fonction monotone croissante continue à gauche qui admet un nombre de points de discontinuité au plus dénombrable et telle que  $F(-\infty) = 0$  et  $F(+\infty) = 1$ .

**Définition 24** Une loi de probabilité  $P_T$  admet une *densité  $f$*  si pour tout intervalle  $I$  de  $\mathbb{R}$  on a :

$$P_T(I) = \int_I f(t) dt = \int_{\mathbb{R}} \mathbb{1}_I(t) f(t) dt$$

avec  $\mathbb{1}_I$  la fonction indicatrice de  $I$  (égale à 1 si  $t \in I$ , à 0 sinon).

$F$  est alors dérivable et admet  $f$  pour dérivée. On a donc :

$$P(a < T < b) = \int_a^b f(t) dt = F(b) - F(a).$$

Nous définissons maintenant le risque instantané qui nous sert à établir le temps entre deux réactions chimiques à partir des réactions élémentaires grâce à ses propriétés particulières. Le risque instantané d'une v.a.r.  $T$  correspond à la densité de probabilité locale de réalisation de  $T$  sachant que  $T$  ne s'est pas réalisée avant. On le définit donc en utilisant une probabilité conditionnelle.

---

<sup>1</sup>Nous donnons ici la définition précise d'une v.a.r. dont nous étudions certaines propriétés dans la suite du chapitre. Néanmoins, le lecteur plus soucieux d'une compréhension générale du mécanisme que de stricte rigueur théorique (nécessitant de comprendre les axiomes fondamentaux de la théorie des probabilités) pourra considérer qu'une v.a.r. est l'association de probabilités à des ensembles de réels correspondant aux valeurs résultat d'une expérience, sans que cela ne gêne la compréhension de ce qui suit. Ainsi  $P(T \in [a, b])$  avec  $T$  une v.a.r. se lira : la probabilité que la réalisation de  $T$  ait une valeur comprise entre  $a$  et  $b$ .

**Définition 25** *Le risque instantané  $\lambda(t)$  de  $T$  est défini par*

$$\lambda(t)dt = P(T \in [t, t + dt] | T > t).$$

En utilisant la formule  $P(A|B) = \frac{P(AB)}{P(B)}$ , on peut exprimer  $\lambda(t)$  en fonction de  $f(t)$  et  $F(t)$  :

$$\begin{aligned} \lambda(t)dt &= P(T \in [t, t + dt] | T > t) \\ &= \frac{f(t)}{1 - F(t)}dt, \end{aligned}$$

d'où :

$$\lambda(t) = \frac{f(t)}{1 - F(t)}.$$

**Remarque 26** *Il est également possible de donner une expression de la fonction de répartition et de la densité en fonction du risque instantané. Pour cela, on remarque que  $\int \frac{f(t)}{1 - F(t)}dt = -\ln(1 - F(t))$  et donc,  $F(t) = 1 - e^{-\int_0^t \lambda(u)du}$  et  $f(t) = \lambda(t)e^{-\int_0^t \lambda(u)du}$ .*

### 3.2.2 Loi exponentielle ou loi sans vieillissement

Nous allons maintenant étudier une loi de probabilité particulière : la loi exponentielle. Nous présentons cette loi car elle permet de modéliser de manière naturelle la rencontre entre deux molécules chimiques, base de la modélisation de réactions chimiques. De plus, sa relative simplicité permet de déduire des propriétés intéressantes quand on met, par exemple, plusieurs de ces lois en concurrence. Nous commencerons par en rappeler la définition, puis nous étudierons le cas de lois en concurrence et, finalement nous terminerons en exposant les méthodes de simulation.

#### Définition

La loi exponentielle correspond à l'un des cas les plus simples de risque instantané : le cas où  $\lambda(t) = \lambda$ , c'est-à-dire que le risque instantané est constant. Sachant que la réalisation n'a pas encore eu lieu, les chances de voir l'évènement se produire restent les mêmes au cours du temps ; pour cela, la loi exponentielle est aussi appelée "loi sans vieillissement". Attention, cela ne veut pas dire que le risque de voir l'évènement se produire n'augmente pas avec le temps : dans un cas, on compare  $F(a) = P(T < a)$  et  $F(b) = P(T < b)$  avec  $a < b$ , dans l'autre cas, on compare  $\lambda(a)$  et  $\lambda(b)$ .

### Fonction de répartition, distribution et espérance

En utilisant la remarque 7, on obtient la fonction de répartition de la loi sans vieillissement en fonction de  $\lambda$  :  $F(t) = 1 - e^{-\lambda t}$  et sa densité  $f(t) = \lambda e^{-\lambda t}$  (voir figure 3.2). On obtient donc, une distribution exponentielle, distribution qui a donné son nom à la loi. Ainsi, quand un évènement a un risque instantané constant sur une période de temps donné, l'attente avant la réalisation possède une distribution exponentielle, asymétrique : la majorité des évènements se concentre dans les temps courts.

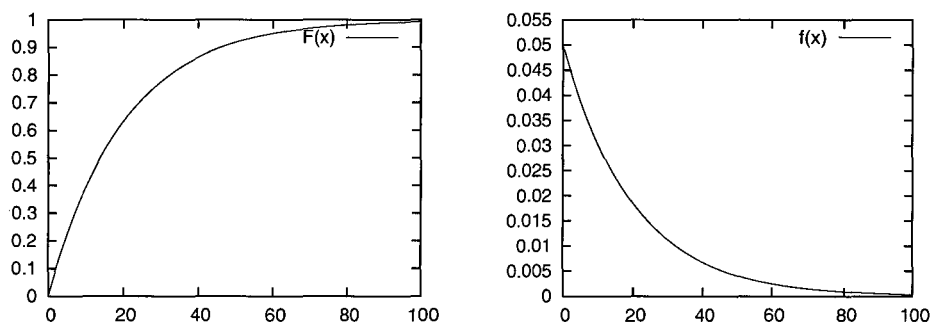


FIG. 3.2 – Fonction de répartition et densité de la loi exponentielle avec  $\lambda = 0,05$ .

On peut se demander combien de temps, en moyenne, on doit attendre avant que la réalisation ait lieu. Pour cela, on calcule l'espérance de  $T$  :

$$E(T) = \int_{\mathbb{R}} t dP_T(T),$$

or on a :  $dP_T(t) = f(t)dt$ , d'où :

$$\begin{aligned} E(T) &= \int_0^{+\infty} \lambda t e^{-\lambda t} dt \\ &= \left[ \frac{(1 + \lambda t)e^{-\lambda t}}{\lambda} \right]_0^{+\infty} \\ &= \frac{1}{\lambda}. \end{aligned}$$

On doit donc attendre en moyenne un temps  $\frac{1}{\lambda}$ , quand la réalisation d'un évènement suit une loi exponentielle de paramètre  $\lambda$ .

### Lois exponentielles en concurrence

On se place maintenant dans la situation où l'on a plusieurs v.a.r. indépendantes  $T_i$ ,  $0 < i \leq n$ , de lois exponentielles de paramètres  $\lambda_i$ . On désire savoir quelle sera

la prochaine réalisation d'une de ces lois et quand elle aura lieu, c'est-à-dire que l'on cherche la loi de distribution de  $T = \min(T_i)$  et les probabilités respectives de chaque réalisation.

**Propriété 27** Soient  $T_i$ ,  $0 < i \leq n$ ,  $n$  v.a.r. indépendantes de lois exponentielles de paramètres respectifs  $\lambda_i$ .

La loi  $T = \min(T_i)$  suit une loi exponentielle de paramètre  $\lambda = \sum_{i=1}^n \lambda_i$  et la distribution une loi donnant à chaque réalisation  $i$  un poids de  $\frac{\lambda_i}{\lambda}$ .

**Preuve.** Soit  $T_1$  et  $T_2$  deux v.a.r. indépendantes de loi exponentielles de paramètre  $\lambda_1$  et  $\lambda_2$ . Soit  $T = \min(T_1, T_2)$ . Soit  $\lambda(t)$  le risque instantané de  $T$ . On a :

$$\begin{aligned} \lambda(t)dt &= P(T_1 \text{ ou } T_2 \in [t, t + dt] | T_1 \text{ et } T_2 > t) \\ &= P(T_1 \in [t, t + dt] | T_1 \text{ et } T_2 > t) + P(T_2 \in [t, t + dt] | T_1 \text{ et } T_2 > t) \\ &\quad - P(T_1 \in [t, t + dt] | T_2 \in [t, t + dt] | T_1 \text{ et } T_2 > t), \end{aligned}$$

or  $T_1$  et  $T_2$  sont indépendants, d'où :

$$\begin{aligned} \lambda(t)dt &= P(T_1 \in [t, t + dt] | T_1 > t) + P(T_2 \in [t, t + dt] | T_2 > t) \\ &\quad - P(T_1 \in [t, t + dt] | T_1 > t)P(T_2 \in [t, t + dt] | T_2 > t) \\ &= (\lambda_1 + \lambda_2 - \lambda_1\lambda_2 dt)dt \end{aligned}$$

d'où  $\lambda(t) = \lambda_1 + \lambda_2$ .

Donc  $T$  suit une loi exponentielle de paramètre  $\lambda_1 + \lambda_2$  et par récurrence (si  $T_1$ ,  $T_2$  et  $T_3$  sont indépendantes  $\min(T_1, T_2)$  et  $T_3$  le sont également), dans le cas de  $n$  variables indépendantes,  $T$  suit une loi exponentielle de paramètre  $\lambda = \sum_{i=1}^n \lambda_i$ . On suppose que  $T$  se réalise sur  $[t, t + dt]$ , la probabilité que ce soit la variable  $T_i$  qui se réalise est alors :

$$\begin{aligned} P(T_i \in [t, t + dt]) &= P(T_i \in [t, t + dt] | T_i > t) \\ &= \lambda_i. \end{aligned}$$

D'où la distribution des variables réalisées suit une loi de poids  $\frac{\lambda_i}{\lambda}$ .

■

### Simulation d'une loi exponentielle

“Any one who considers arithmetical methods of producing random digits is, of course, in a state of sin.”<sup>2</sup>

<sup>2</sup>“Quiconque envisage des méthodes arithmétiques de génération de nombres aléatoires est, de toute évidence, en état de péché.” Bien qu'il y ait sûrement des péchés plus graves que de considérer de telles méthodes. On peut citer par exemple, sur l'activité débordante de von

MARGITTAI NEUMANN JÁNOS dit JOHN VON NEUMANN (1951) cité  
par D. KNUTH dans [46].

Dans un logiciel de simulation ou dans un langage de programmation, l'accès au hasard s'établit au travers de la génération de nombres aléatoires, qui fournit (plus ou moins bien selon l'algorithme utilisé) une loi uniforme sur une plage donnée. On obtient la plupart du temps un entier  $r$  satisfaisant  $-N_{\max} \leq r \leq N_{\max} - 1$ , avec  $N_{\max}$  le plus grand entier codé par le système (dépendant de la taille alloué à un entier). Pour simuler n'importe quelle loi de fonction de répartition connue, à partir de ce générateur de nombre aléatoire, on utilise la propriété classique qui stipule que la v.a.r.  $F(T)$  a une fonction de répartition de loi uniforme sur  $[0,1]$ . Ainsi, en utilisant le générateur aléatoire, on ramène le nombre  $r$  au nombre  $r'$  entre 0 et 1 par  $r' = \frac{r+N_{\max}}{2N_{\max}-1}$ , et on construit  $T$  par  $F^{-1}(r')$  (voir figure 3.3).

Ainsi, pour simuler une observation  $t$  issue d'une loi exponentielle de paramètre  $\lambda$ , on calcule :

$$t = -\frac{1}{\lambda \ln(1 - r')}.$$

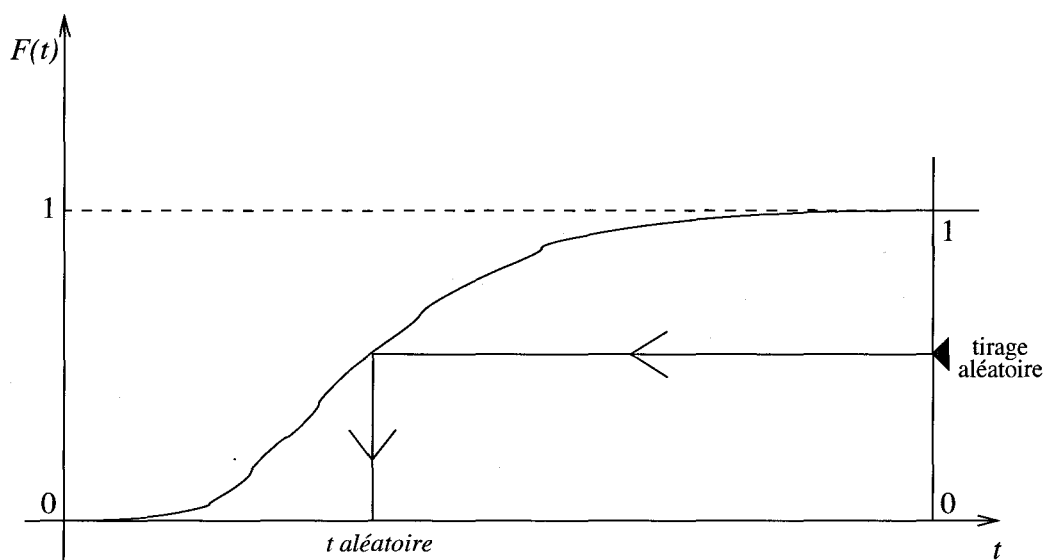


FIG. 3.3 – Tirage d'une loi à partir de sa fonction de répartition et d'un tirage uniforme sur  $[0,1]$ . On a choisi ici une fonction de répartition quelconque, mais cette même méthode s'applique dans le cas d'une loi exponentielle.

Neumann, une phrase de Roger Godement tiré de [30] (formulée à propos du côté pratique d'une définition des entiers naturels issu de la théorie des ensembles de von Neumann et reprise par les Bourbakistes) : "Une liste, même incomplète, des activités de von Neumann à partir de 1937 -explosifs classiques, recherche opérationnelle et théorie des jeux de stratégies, bombe atomique, ordinateurs programmables, bombe H, missiles intercontinentaux- devrait rassurer les philistins quant à son sens du concret".



Il subsiste malgré tout un problème -inévitabile- inhérent au caractère discret des ordinateurs. La loi "uniforme" simulée n'est pas continue mais discrète, c'est-à-dire que l'on ne donne pas, par exemple, un poids  $\frac{1}{2^{N_{\max}-1}}$  à l'intervalle  $\left[0, \frac{1}{2^{N_{\max}-1}}\right]$ , mais un poids  $\frac{1}{2^{N_{\max}}}$  au point 0. Or, dans le cas d'une loi continue, chaque point a une probabilité nulle d'apparaître. Ce n'est pas gênant dans la plupart des cas (surtout si  $N_{\max}$  est grand), sauf quand un point a un rôle particulier, comme c'est le cas dans la loi exponentielle pour  $r' = 1$  (parce qu'alors  $t = +\infty$ ). Dans le cas où la simulation fait un grand nombre de fois appel au générateur de nombre aléatoire, cette situation apparaît inévitablement. Une solution pour résoudre ce problème peut-être de mettre un test pour détecter l'apparition d'un 1.

### 3.2.3 Application à la simulation stochastique d'une réaction chimique généralisée

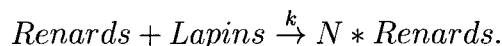
Nous allons maintenant montrer comment appliquer ces notions de théorie des probabilités à la modélisation et à la simulation des réactions stochastiques.

#### Petit rappel de terminologie

Nous allons utiliser pour l'étude de nos modèles particuliers des notations et une terminologie provenant de la chimie. On parle de la **réaction chimique**



pour désigner le mécanisme macroscopique qui en consommant du  $R_1$  et du  $R_2$  (les réactifs) produit des  $P_i$  (les produits) avec un taux caractérisé par  $k$ . En chimie  $R_1$ ,  $R_2$  et  $P_i$  désignent des molécules, mais on peut utiliser par analogie la même notation pour désigner une "réaction" mettant en jeu d'autres acteurs. Par exemple, si dans un modèle écologique, une population de renards consomme des lapins pour croître, on pourra écrire :



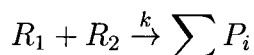
On ne parle plus alors de réaction chimique, mais de **réaction chimique généralisée** (RCG).

Pour désigner, la rencontre physique entre la molécule  $R_1$  et la molécule  $R_2$ , constituant la base de la réaction 3.1, on parle d'**événement cinétique élémentaire de la réaction**.

La loi qui donne la vitesse de production des  $P_i$  (et, de par la stoechiométrie, la vitesse de consommation des réactifs) est appelée **loi cinétique de la réaction**. Pour la réaction 3.1, on peut considérer par exemple qu'elle est égale à  $kR_1R_2$  dans le cas où  $R_1 \neq R_2$ .

### Interprétation stochastique des réactions chimiques généralisées

On interprète un évènement cinétique élémentaire de la réaction

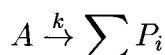


comme le risque instantané  $k$  de voir se rencontrer une particule du réactif<sup>3</sup>  $R_1$  et une particule du réactif  $R_2$  et de former les produits  $\sum P_i$ . Comme  $k$  est la plupart du temps constant, la loi est alors sans vieillissement.

Considérant un ensemble de particules de réactifs en interaction, on a donc un ensemble d'évènements cinétiques élémentaires possibles en concurrence. Pour établir la loi cinétique globale d'une réaction, on doit donc dénombrer le nombre  $N$  d'évènements cinétiques élémentaires, la loi cinétique globale étant alors une loi exponentielle de paramètre  $\lambda_{globale} = Nk$ .

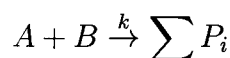
Voici les différents cas qui peuvent se présenter :

– **Transformation :**



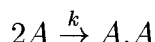
Chaque  $A$  peut se transformer avec un risque instantané  $k$ , le paramètre de la loi cinétique globale est donc<sup>4</sup>  $kA$ .

– **Réaction à deux réactifs :**



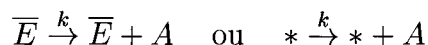
On dénombre les réactions possibles, il y a  $AB$  combinaisons possibles, la loi globale a donc un paramètre  $kAB$ .

– **Dimérisation :**



Les réactions possibles sont au nombre de  $A(A - 1)$ , puisque chaque  $A$  peut réagir avec les  $(A - 1)$  autres  $A$ . Le paramètre de la loi cinétique est donc  $kA(A - 1)$ . On peut noter que, pour un nombre de  $A$  élevé,  $A(A - 1) \approx A^2$  et on retrouve bien la loi cinétique  $kA^2$  classique.

– **Création à taux constant :**



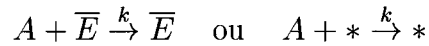
On suppose qu'un composé en nombre constant  $\bar{E}$  produit du  $A$ . La loi cinétique est donc  $K = k\bar{E}$  constante. Une autre convention remplace les

<sup>3</sup>On considère qu'une réaction où trois réactifs ou plus interviennent est improbable, en effet, cela supposerait que les trois réactifs se trouvent très près les uns des autres au même moment. Les réactions à trois éléments ou plus qu'on utilise en chimie, sont en fait des réactions composées.

<sup>4</sup>On confond le composé  $A$  et le nombre de particules de composé  $A$ , comme cela se fait couramment, parce qu'il n'y a pas ici d'ambiguïté.

quantités constantes considérées par un signe (par exemple  $*$ ) et le  $k$ , au dessus de la flèche est alors le  $K$  global de la réaction.

– **Dégradation simple :**



La loi cinétique est  $kA\bar{E}$  ou  $kA$  selon la convention.

### Algorithme de Gillespie

Cette méthode de simulation a été proposée en 1977 par D.T. Gillespie dans [29]. Elle a pour but de simuler de manière stochastique un ensemble de réactions chimiques et a été utilisée récemment par McAdams et co-auteurs [54] pour modéliser les mécanismes chimiques de régulation des gènes dans les cellules. On veut simuler un ensemble de  $n$  réactions données, sachant qu'entre  $t_i$ , l'instant de réalisation de la  $i^{\text{ième}}$  réaction, et  $t_{i+1}$ , l'instant de réalisation de la  $i+1^{\text{ième}}$  réaction, il ne se passe rien. Les quantités sont donc constantes de  $t_i$  à  $t_{i+1}$ . Les concentrations sont donc en escalier.

**Etape 1.** On initialise<sup>5</sup>  $t \leftarrow t_0$ , ainsi que le nombre  $i$  d'évènements cinétiques élémentaires réalisés depuis  $t_0$ ,  $i \leftarrow 0$ .

**Etape 2.** On détermine l'instant  $t_{i+1} \leftarrow t_i + t_r$  en simulant la loi exponentielle comme expliqué en 3.2.2 avec comme paramètre  $\lambda = \sum \lambda_i$  et  $\lambda_i$  les paramètres des différentes réactions. On obtient alors le temps  $t_r$  séparant la réaction  $i$  de la réaction  $i+1$ .

**Exemple 28** Soient les réactions chimiques suivantes :

dimérisation	$2A \xrightarrow{k_1} A.A$	$\lambda_1 = k_1 A(A-1)$
création à taux constant	$* \xrightarrow{k_2} A + *$	$\lambda_2 = k_2$
dégradation	$A + * \xrightarrow{k_3} *$	$\lambda_3 = k_3 A$
dégradation	$A.A + * \xrightarrow{k_4} *$	$\lambda_3 = k_4 A.A$

La loi globale est donc de paramètre :

$$\begin{aligned} \lambda &= \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 \\ &= k_1 A(A-1) + k_2 + k_3 A + k_4 A.A \end{aligned}$$

Pour obtenir l'intervalle de temps entre  $t_i$  et  $t_{i+1}$ , on tire un nombre  $r_1$  au hasard entre 0 et 1 et on a alors :

$$t_r = -\frac{1}{[k_1 A(A-1) + k_2 + k_3 A + k_4 A.A] \ln(1-r_1)}$$

<sup>5</sup>On rappelle pour les non-informaticiens que la notation " $a \leftarrow b$ " désigne l'affectation de la valeur  $b$  à la variable  $a$  (noté  $:=$  en Pascal et  $=$  en C).

et  $t_{i+1} \leftarrow t_i + t_r$ .

On connaît donc maintenant le temps d'apparition de la prochaine réaction, on doit maintenant "choisir" la réaction qui a lieu après avoir vérifié que  $t_{final}$  n'est pas dépassé.

**Etape 3** Si  $t > t_{final}$ , on arrête.

**Etape 4.** La probabilité que la réaction  $j$ ,  $0 < j \leq n$ , ait lieu est  $\frac{\lambda_j}{\lambda}$ . On tire un nombre  $r_2$  au hasard entre 0 et 1. Si, en posant  $\lambda_0 = 0$ , on a

$$\sum_{k=0}^{j-1} \lambda_k < \lambda r_2 < \sum_{k=1}^j \lambda_k,$$

alors la réaction  $j$  a lieu (voir figure 3.4). On actualise alors les variables en fonction de la réaction tirée et on incrémente  $i$ .

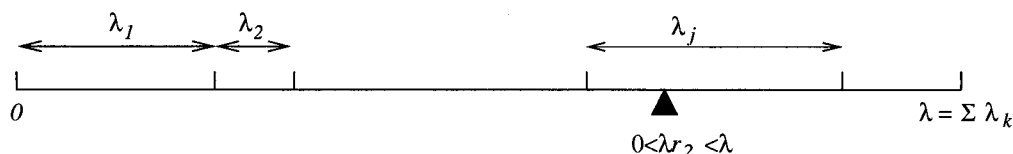


FIG. 3.4 – Tirage de la réaction à effectuer, chaque réaction a un poids  $\frac{\lambda_i}{\lambda}$ .

**Exemple 29** Considérons les réactions de l'exemple 16, avec  $k_1 = 2$ ,  $A = 503$ ,  $\lambda = 100000$ . On tire un nombre entre 0 et 1, par exemple  $r_2 = 0,3$ .

$$\lambda_0 = 0 \leq \lambda r_2 = 300000 < \lambda_1 = k_1 \times A \times (A - 1) = 505012$$

C'est donc la réaction 1 qui a lieu, soit la dimérisation. Le nombre de  $A$  diminue de 2 et le nombre de  $A.A$  augmente de 1.

**Etape 5.** On repasse à l'étape 2.

L'algorithme de Gillespie est extrêmement simple à mettre en œuvre, il suffit d'effectuer deux tirages pour chaque pas de simulation, ainsi qu'une série de tests. Sa rapidité dépend donc exclusivement du nombre d'itérations pour arriver à  $t_{final}$ . Ce nombre d'itérations dépend de beaucoup de paramètres, en particulier du nombre de particules. En effet, on peut remarquer que, l'espérance de la loi exponentielle étant en  $\frac{1}{\lambda}$ , si durant la simulation,  $\lambda$  est de l'ordre de  $K$ , le nombre d'itérations sera de l'ordre de  $K \times (t_{final} - t_0)$ . Donc, plus  $K$  est grand (par exemple, si le nombre de particules est grand, ou si les réactions sont "rapides"), plus l'intervalle de temps est petit entre deux réactions et plus le temps de simulation est long.

### 3.2.4 Quelques systèmes classiques revisités

Etant donné que la base dynamique des SAM que nous considérerons est un modèle cinétique, nous devons étudier, afin de justifier ce choix, soit les modèles d'agents, soit les modèles particuliers, qui constituent des modèles de niveaux de détails plus fin dont les SAM sont issus. Comme nous l'avons vu à la section 3.1, la modélisation stochastique particulière représente déjà un niveau moins "détaillée" que la modélisation "orientée-agent". Nous devons donc maintenant étudier le passage d'un modèle particulier à un SAM. L'idée générale de cette étude est de considérer des systèmes particuliers, tout d'abord simples et ensuite se rapprochant du type de phénomènes que nous considérons au chapitre 4 (interaction bactéries-phages). On doit ensuite montrer dans chaque cas comment sous certaines conditions, on obtient un modèle continu différentiel (cas "classique") ou de type SAM (dans le cas évolutif).

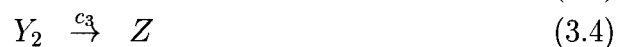
Dans ce but, nous allons maintenant présenter plusieurs systèmes cinétiques classiques auxquels nous appliquerons la modélisation par RCG (réactions chimiques généralisées) et que nous simulerons en utilisant l'algorithme de Gillespie.

Les deux premiers systèmes (Lotka et l'orégonateur<sup>6</sup>), dont nous avons programmé les simulations sur base des informations contenues dans [29], nous amèneront à nous intéresser aux liens entre les aspects stochastiques et différentiels de ces systèmes.

Dans l'optique du modèle bactérien du chapitre 4, nous établirons ensuite le pendant stochastique du modèle de Tsuchiya, modèle différentiel classique en microbiologie. Ce modèle est particulièrement important car il met en œuvre des mécanismes plus complexes que ceux du modèle de Lotka et de l'*Oregonator* (des termes de saturation). Nous montrerons ainsi qu'il est possible d'utiliser la simulation stochastique des RCG pour l'étude de modèles biologiques de population, et qu'ainsi, certaines propriétés mises en avant pour les systèmes chimiques peuvent également se révéler utiles à la modélisation des systèmes biologiques et plus spécialement "microbiologiques".

#### Les réactions de Lotka

Lotka, chimiste américano-austro-ukrainien<sup>7</sup>, observa en 1920 que l'ensemble de réactions autocatalytiques couplées




---

<sup>6</sup>ou *Oregonator*

<sup>7</sup>Alfred Lotka est né en 1880 à Lviv (Lemberg) à l'époque en Autriche et maintenant en Ukraine. Il a émigré en 1902 aux États-Unis.

possède des propriétés dynamiques remarquables. En effet, les concentrations des espèces chimiques soumises à de telles réactions oscillent au cours du temps. Peu de temps après, le mathématicien italien V. Volterra découvrit indépendamment le système d'ODE correspondant aux équations des taux de réaction :

$$\begin{cases} \frac{dY_1}{dt} = c_1 \bar{X} Y_1 - c_2 Y_1 Y_2 \\ \frac{dY_2}{dt} = c_2 Y_1 Y_2 - c_3 Y_2 \end{cases}$$

en cherchant à modéliser un écosystème proie-prédateur simple.

Dans ce modèle, extrêmement simpliste, les  $Y_1$  sont des proies (disons des lapins) et se nourrissent de  $\bar{X}$  (l'herbe), supposée en quantité toujours suffisante ; les  $Y_2$ , quant à eux, sont les prédateurs (disons des renards) et se reproduisent quand ils arrivent à capturer des lapins. La dynamique du modèle différentiel est constituée de cycles concentriques autour du point d'équilibre  $\left(\frac{c_3}{c_2}, \frac{c_1 \bar{X}}{c_2}\right)$ . Ce type de dynamique appelée "centre" est très peu réaliste, car elle correspond au cas où les valeurs propres de la jacobienne<sup>8</sup> au point d'équilibre sont complexes pures, évènement de probabilité nulle et non robuste (une petite perturbation introduit un élément réel, rendant le système stable ou instable selon le signe de cette partie réelle). Les questions que l'on peut se poser sont :

- comment va réagir le modèle stochastique correspondant aux réactions 3.2 à 3.4 ?
- va-t-on voir les mêmes oscillations que pour le système différentiel ?
- Le système a-t-il un ou des états asymptotiques privilégiés ?

Nos simulations (réalisées suivant l'algorithme de Gillespie, donc stochastiques) montrent que le comportement des réactions de Lotka est le suivant : les nombres d'individus oscillent autour des valeurs d'équilibres, comme pour le système différentiel. En revanche l'amplitude de ces oscillations est aléatoire. Comme il existe une zone de contraction (près de 0), le moindre petit écart dans cette zone est amplifié dans la partie "dilatée" du cycle, engendrant ainsi de grandes variations.

Dans la zone de contraction, les variables sont "fragiles" et sont susceptibles d'atteindre 0 et ainsi de disparaître. Ainsi dans les exemples des figures 3.5, 3.6 et 3.7, une des populations finit toujours par s'éteindre. On est alors face à deux possibilités :

- les lapins disparaissent en premier. Les renards s'éteignent alors faute de nourriture. L'état asymptotique final est lapins = 0 et renards = 0 (cas (d) et (c)).
- les renards disparaissent en premier. Les lapins croissent alors exponentiellement (absence de prédateur). Comme le niveau de nourriture est en quantité suffisante (hypothèse de modélisation), l'effectif de la population tend vers l'infini. L'état asymptotique final est alors lapins =  $\infty$  et renards = 0 (cas (a) et (c)).

---

<sup>8</sup>matrice représentant la linéarisation du système en un point donné.

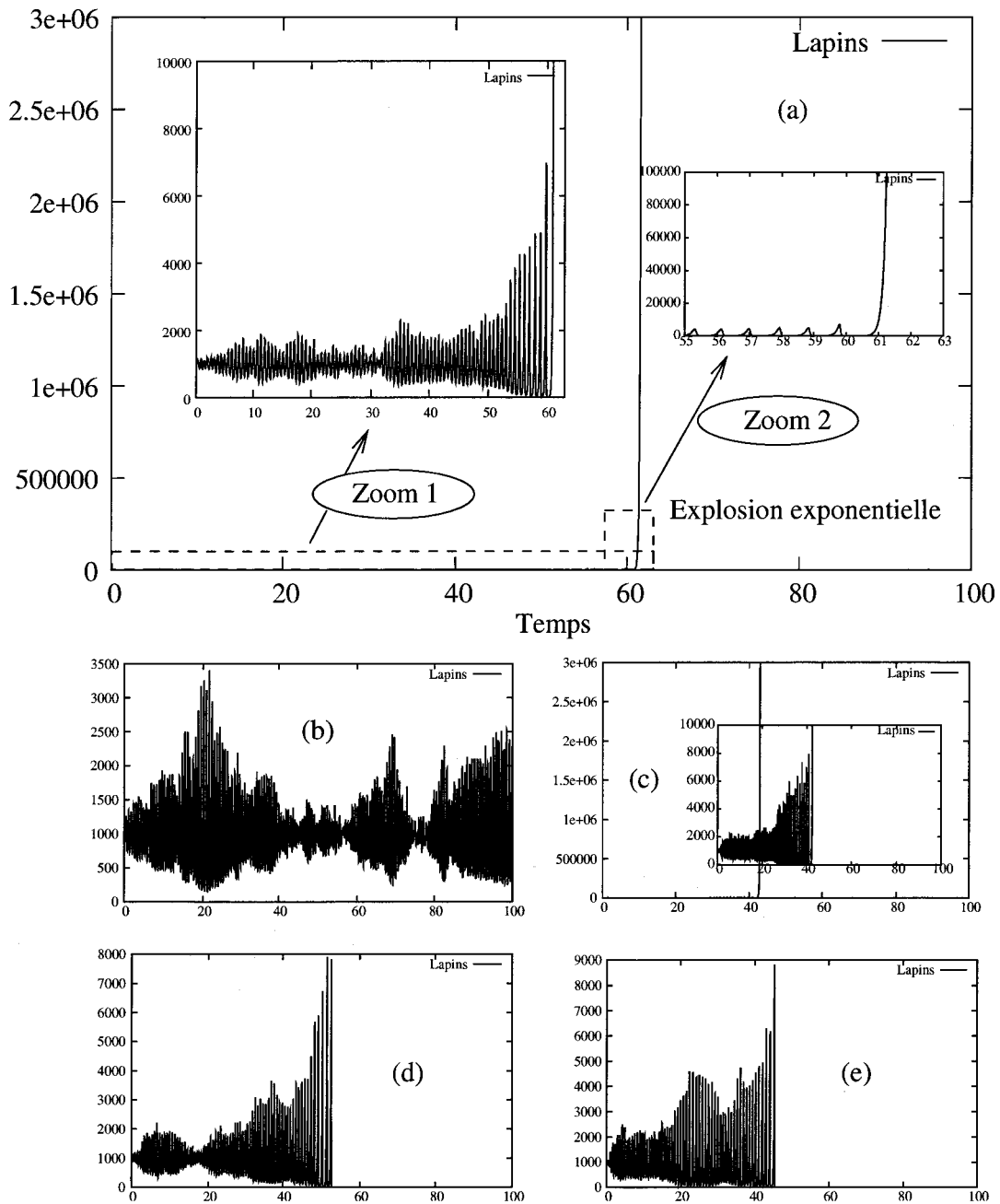


FIG. 3.5 – Evolutions possibles de la population de lapins pour  $c_1 X = 10$ ,  $c_2 = 0,01$ ,  $c_3 = 10$ , 1000 lapins et 1000 renards à l'origine (position du "point d'équilibre"). Seule la génération de nombres aléatoires change. Deux comportements asymptotiques sont possibles : explosion exponentielle du nombre de lapins après disparition des renards (a) et (c) ou disparition des lapins (d) et (e). La simulation (b) est en phase transitoire.

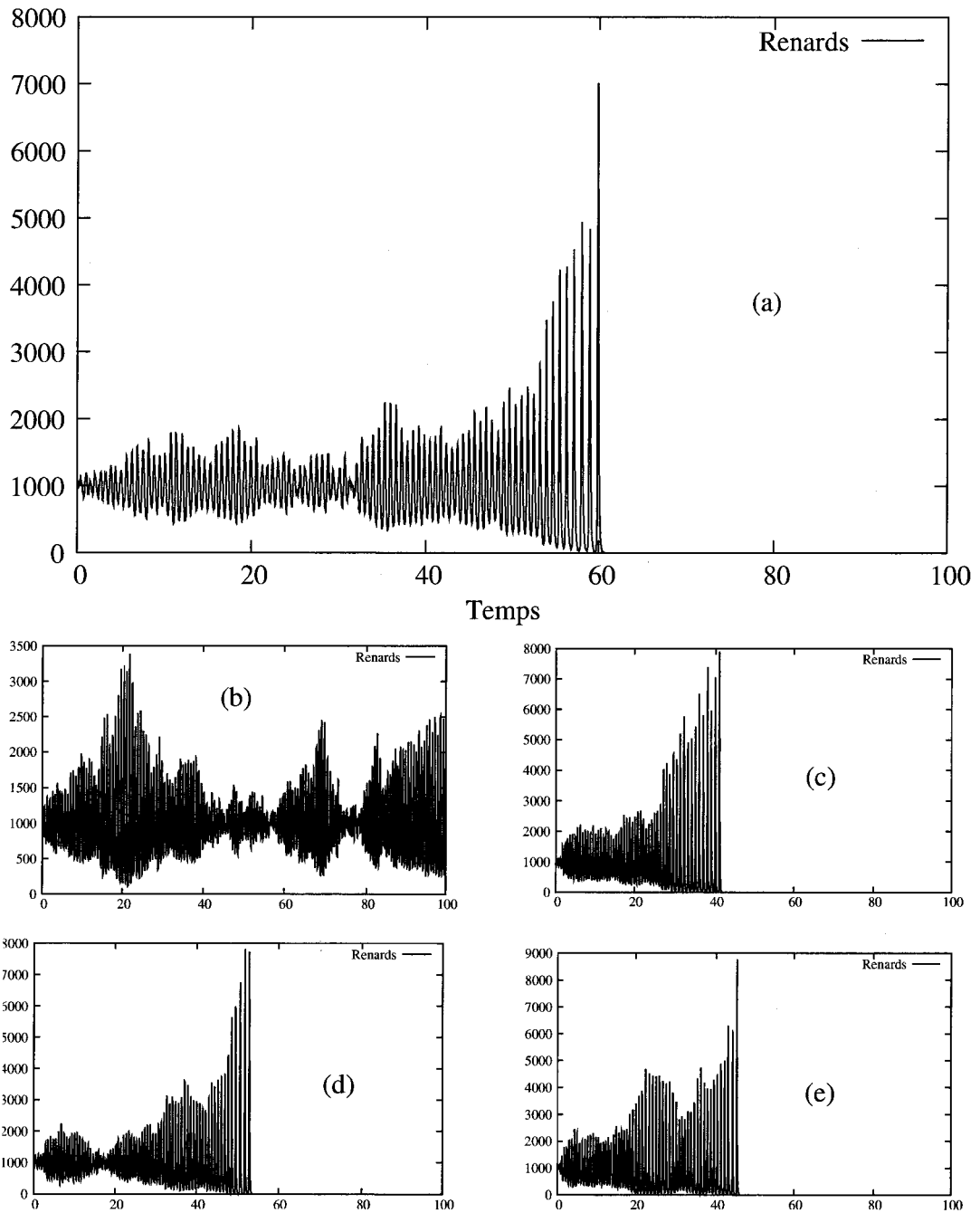


FIG. 3.6 – Evolutions possibles de la population de renards pour  $c_1 X = 10$ ,  $c_2 = 0,01$ ,  $c_3 = 10$ , 1000 lapins et 1000 renards à l’origine (position du “point d’équilibre”). Les renards finissent toujours par disparaître, soit d’eux-mêmes (en (a) et (c)), soit après disparition des lapins (en (d) et (e)). La simulation (b) est en phase transitoire.



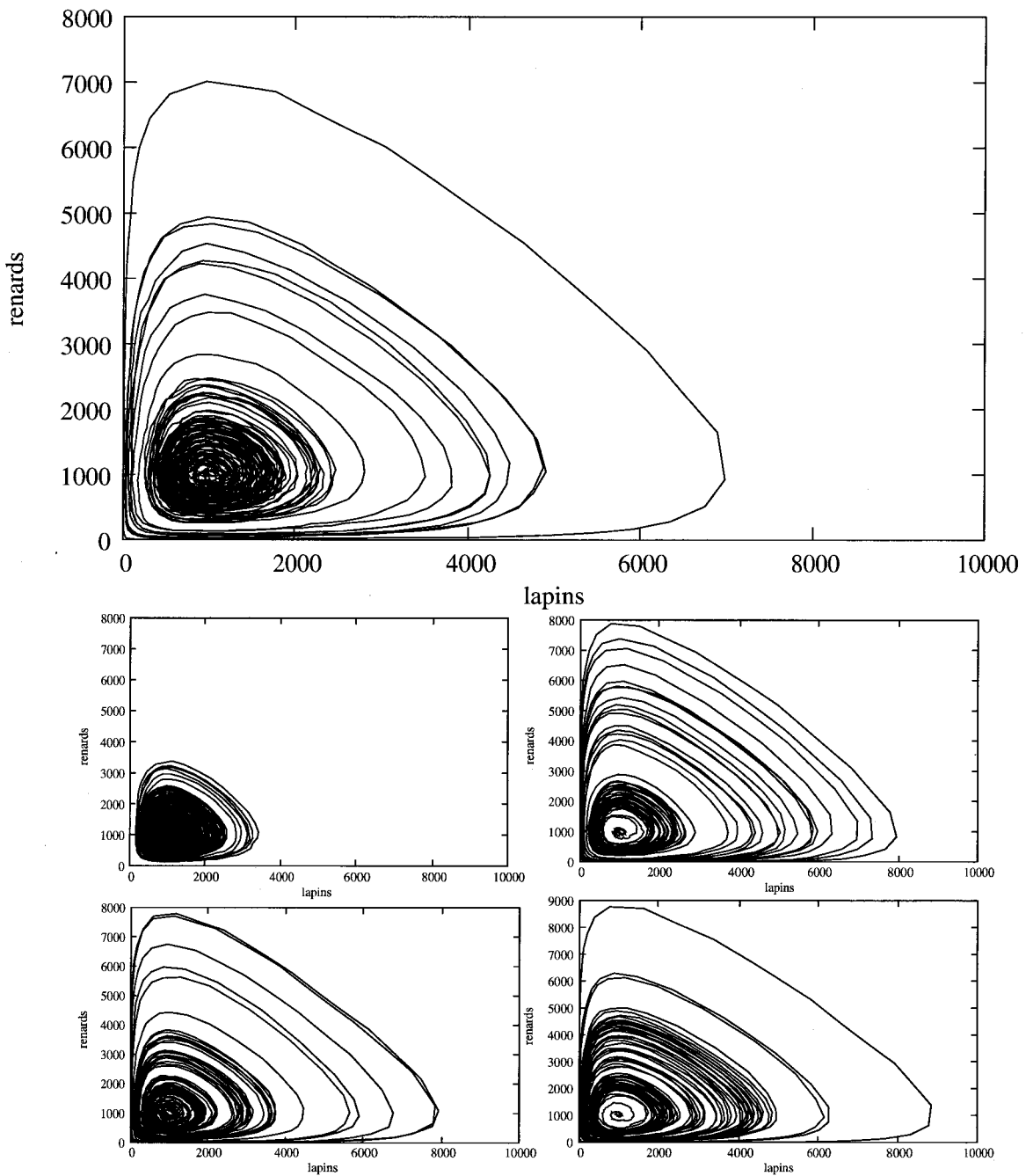


FIG. 3.7 – Portrait de phase avec  $c_1X = 10$ ,  $c_2 = 0,01$ ,  $c_3 = 10$ , 1000 lapins et 1000 renards à l'origine (position du “point d'équilibre”).

Le comportement du système stochastique présente donc des points communs avec le système déterministe (les oscillations), mais également des différences : l'amplitude des oscillations est aléatoire au cours du temps et finit par déboucher sur la destruction de "l'écosystème". La non-validité de ce système de réactions comme modèle d'un écosystème simple est donc dans ce cas flagrante, puisque celui-ci est amené à disparaître à relativement court terme.

Cette non-validité doit être mise en relation avec la non-robustesse du modèle différentiel : si l'on rajoute une petite perturbation structurelle, la dynamique en cycles concentriques autour du point d'équilibre disparaît, laissant place à un équilibre stable ou à une divergence des trajectoires. Or, le caractère "perturbé" est propre, à la fois aux systèmes stochastiques et aux systèmes naturels. Le modèle stochastique met donc ici simplement encore un peu plus en avant l'inadéquation du modèle de Lotka pour représenter un modèle (même très simplifié) d'écosystème réel. Mais, si, pris comme tel, cet exemple ne nous apprend que des choses très intuitives, il valide notre technique de simulation et son implémentation.

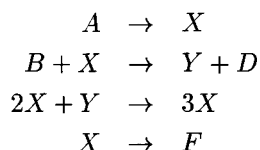
### Le modèle de l'*Oregonator* de Field et Noyes

Après cet exemple introductif, nous allons passer maintenant à des simulations plus riches. Il existe ainsi des systèmes de réactions possédant d'autres propriétés dynamiques particulières comme des cycles limites. Le premier système de réactions chimiques possédant ce type de propriété est le modèle du *Brusselator*<sup>9</sup> ainsi nommé parce que découvert au sein d'une équipe de l'Université Libre de Bruxelles (ULB). Ce système possédait néanmoins un gros défaut, celui de présenter une réaction contenant trois réactifs, ce qui est cinétiquement improbable.

En 1973, deux chercheurs de l'Oregon, Field et Noyes, ont présenté un modèle<sup>10</sup> exhibant également un cycle limite, mais ne comportant que des réactions à deux réactifs :

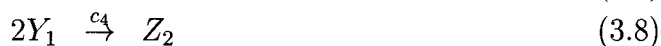
---

<sup>9</sup>ou dans sa version francisée le modèle du Brusselateur, dont les réactions sont :



avec  $A$  et  $B$  contrôlés de l'extérieur (voir [56] ou [57] pour la version française).

<sup>10</sup>conçu au départ comme une idéalisation poussée des réactions de Belousov-Zhabotinskii, connues pour leur caractère oscillatoire spectaculaire, réactions pouvant même devenir chaotiques sous certaines conditions.



par analogie avec le *Brusselator*, ce système a été baptisé modèle de l'*Oregonator*.

Lors de la simulation particulière, on observe des oscillations, très régulières (pour un système stochastique) autour des valeurs d'un cycle limite. Là encore, il existe une zone de contraction des trajectoires près de 0, qui amplifie les petits écarts. Mais la différence principale avec les réactions de Lotka est la présence de "forces" (en termes de différences de probabilités d'apparition) qui "ramènent" la trajectoire constamment sur le cycle limite (attractivité).

Ainsi, les figures 3.8 et 3.9, bien que correspondant à des tirages différents, présentent des trajectoires conservant un même cycle limite (intrinsèque au système) et ne s'en écartant que très peu, au contraire du modèle de Lotka, sensible aux aléas des tirages.

On voit donc que la méthode des RCG permet de modéliser des dynamiques assez complexes (les cycles limites). Néanmoins, les interactions modélisées sont encore trop simples pour représenter des modèles biologiques de croissance et de prédation tels que ceux dont nous aurons besoin par la suite. En effet, les termes mis en œuvre dans les mécanismes réactionnels précédents, ne permettent pas de représenter des phénomènes biologiques très courants comme par exemple des phénomènes de saturation.

Nous montrons dans l'exemple original qui suit que ce type d'interactions peut également être représenté par un tel mécanisme, étendant ainsi le champ d'application des RCG.

### Modèle de Tsuchiya en microbiologie

Nous avons vu la capacité des modèles particuliers stochastiques pour la représentation des réactions chimiques. Nous allons maintenant voir que ce type de modèle peut être également utilisé pour représenter certains modèles biologiques complexes<sup>11</sup>.

Si le modèle de l'Oregonateur constituait une réponse à la possibilité d'apparition d'un cycle limite dans un modèle de réactions chimiques, de nombreux modèles ont également été proposés en biologie pour améliorer le modèle issu de l'interprétation biologique des travaux de Lotka et Volterra. Un des défauts de

<sup>11</sup>Même si les équations de Lotka-Volterra ont une interprétation biologique (dont nous avons discuté les limites), elles représentent avant tout un ensemble de réactions chimiques.

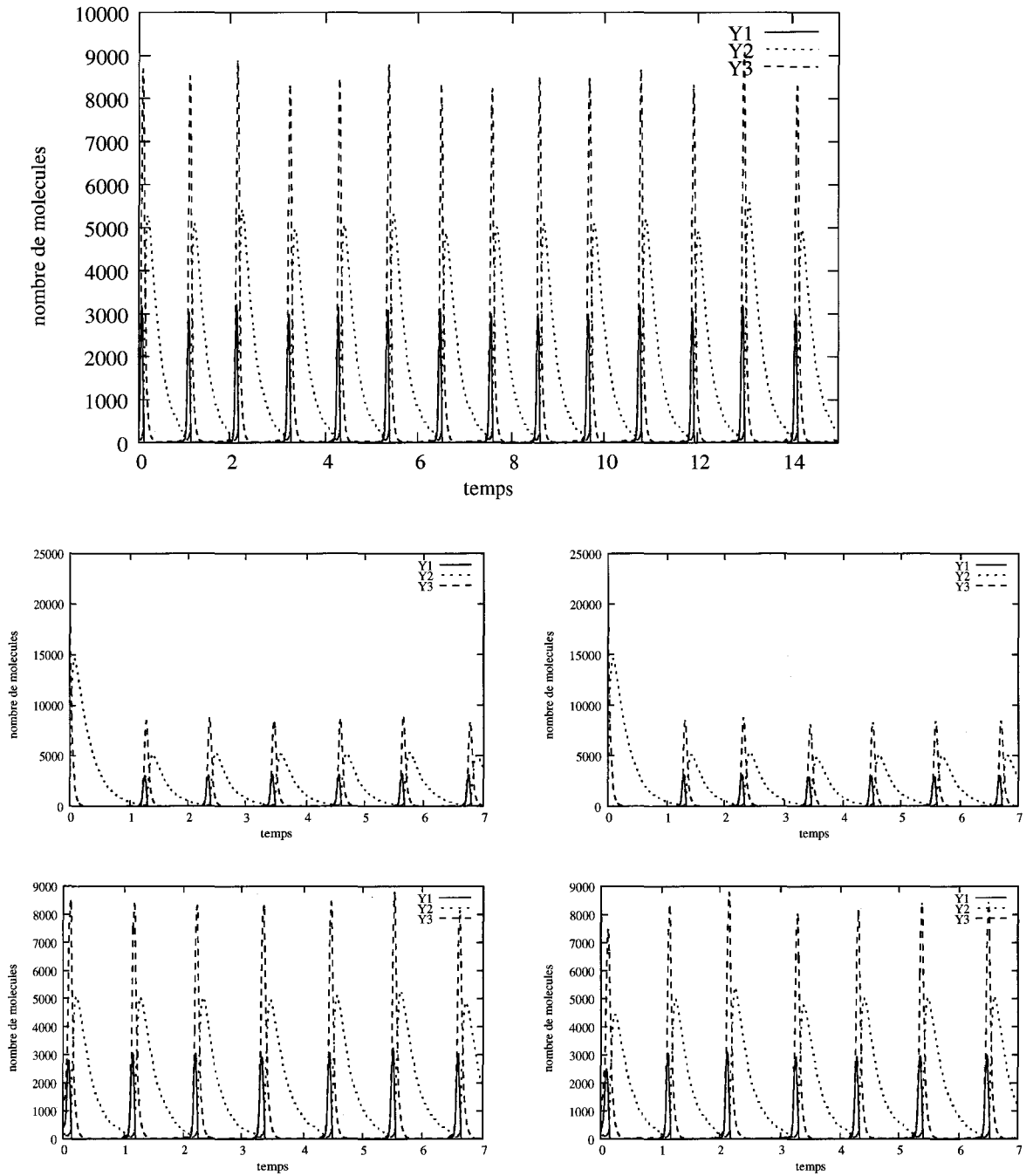


FIG. 3.8 – Evolutions temporelles des variables  $Y_1$ ,  $Y_2$  et  $Y_3$  pour  $c_1 \overline{X_1} = 2$ ,  $c_2 = 0.5$ ,  $c_3 \overline{X_2} = 104$ ,  $c_4/2 = 0.008$  et  $c_5 \overline{X_3} = 26$ . Figure principale :  $Y_1 = 500$ ,  $Y_2 = 500$  et  $Y_3 = 50$ . Pour les quatre petites figures, chaque ligne correspond à des conditions initiales données (première ligne  $Y_1 = 20000$ ,  $Y_2 = 20000$  et  $Y_3 = 20000$ , deuxième ligne  $Y_1 = 1000$ ,  $Y_2 = 1000$  et  $Y_3 = 1000$ ) et chaque colonne à une initialisation du générateur de nombre aléatoire différente.

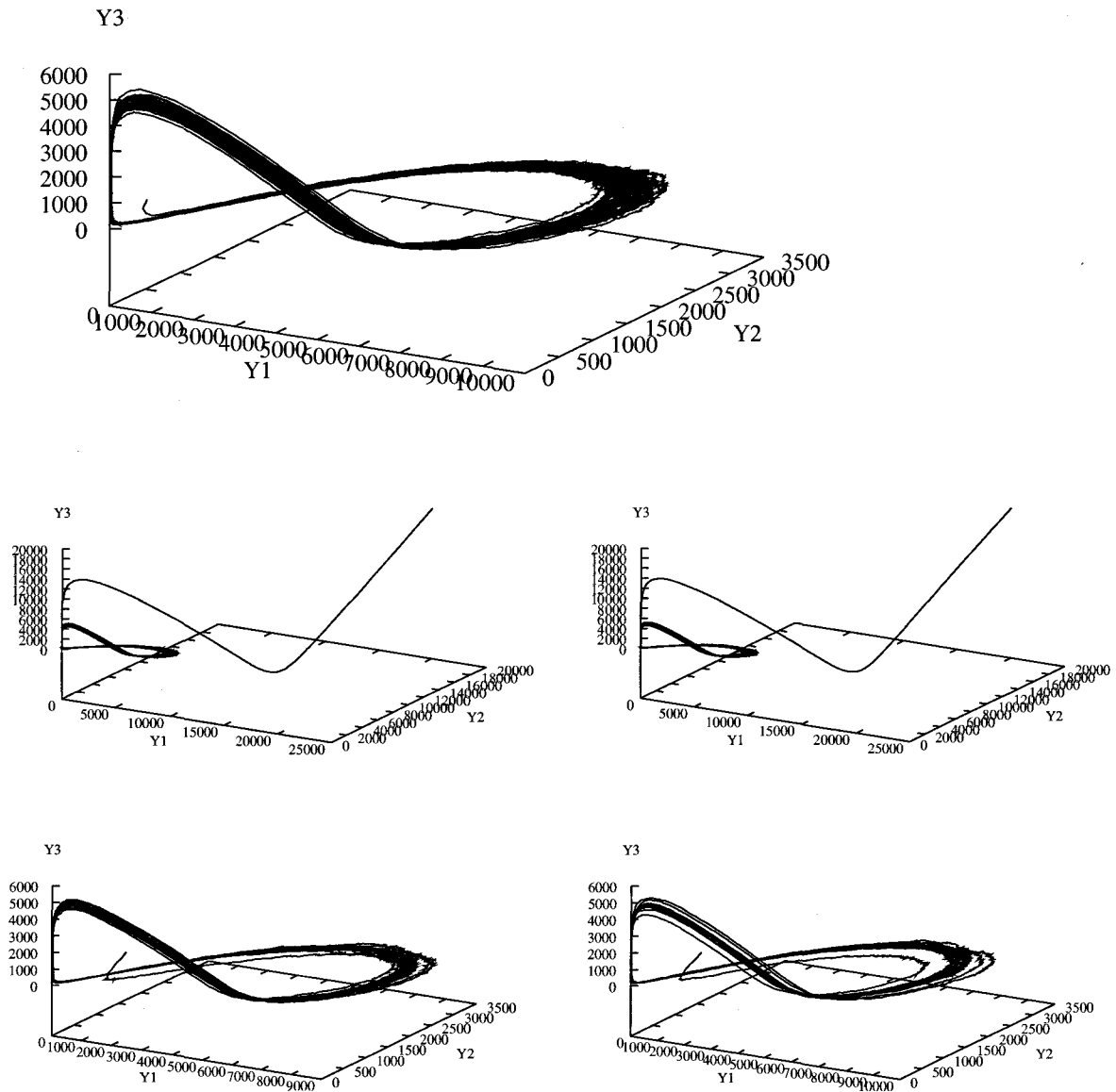


FIG. 3.9 - Cycle limite de l'oregonator pour  $c_1 \bar{X}_1 = 2$ ,  $c_2 = 0.5$ ,  $c_3 \bar{X}_2 = 104$ ,  $c_4/2 = 0.008$  et  $c_5 \bar{X}_3 = 26$ . Figure principale :  $Y_1 = 500$ ,  $Y_2 = 500$  et  $Y_3 = 50$ . Pour les quatre petites figures, chaque ligne correspond à des conditions initiales données (première ligne  $Y_1 = 20000$ ,  $Y_2 = 20000$  et  $Y_3 = 20000$ , deuxième ligne  $Y_1 = 1000$ ,  $Y_2 = 1000$  et  $Y_3 = 1000$ ) et chaque colonne à une initialisation du générateur de nombre aléatoire différente.

ce modèle était de présenter en l'absence de prédateur, une population de proies croissant de manière non bornée vers l'infini. Ce trait irréaliste du modèle peut être pallié en tenant compte de l'utilisation limitée des ressources (le substrat).

Les études expérimentales montrent également que les taux de croissance des prédateurs ne varient pas linéairement avec la concentration de proie comme on le suppose dans l'analyse faite originellement par Lotka-Volterra. On considère généralement que le taux de croissance d'un prédateur est mieux représenté par l'équation dite de Monod :

$$\mu_p = \frac{\mu_{p,\max} n_1}{K_p + n_1},$$

où  $\mu_p$  est le taux de croissance du prédateur,  $\mu_{p,\max}$  le taux de croissance maximum,  $K_p$  une constante de saturation et  $n_1$  la quantité de proies. En effet,  $\mu_p$  présente une allure sigmoïdale permettant de rendre compte d'une croissance quasi-linéaire pour des quantités de  $n_1$  faibles et d'un plateau de saturation ou de satiété pour des valeurs de  $n_1$  grandes.

De plus, dans un écosystème macroscopique, il existe une myriade d'autres phénomènes que les naissance, mort naturelle et prédation prises en compte par Lotka. On peut par exemple avoir à considérer les classes d'âges, des chaînes trophiques, des effets spatiaux, des épidémies, des effets environnementaux. Suite à ces considérations, certains auteurs ont étudié des systèmes expérimentaux proie-prédateur où tous ces effets sont absents ou minimisés. Pour se faire, ils se sont tournés vers la microbiologie et ont placé sur des substrats d'Agar contenu dans une boîte de Pétri, une culture bactérienne (proie) soumise à prédation par une population d'amibes. Ces expériences ont amenées Tsuchiya et co-auteurs [9] à créer un modèle plus réaliste avec pour proies les bactéries *E. Coli*<sup>12</sup> et pour prédateurs les amibes *Dictyostelium discoideum*.

En tenant compte des raffinements que l'on vient de voir et en supposant constant les facteurs de production  $Y_S$  et  $Y_B$  (respectivement pour la croissance des bactéries qui consomment le substrat et celle des amibes par prédation des bactéries), on obtient pour un chémostat<sup>13</sup> :

$$\begin{cases} \frac{dS}{dt} = \rho(S_0 - S) - \frac{1}{Y_S} \frac{\mu_{s,\max} SB}{K_S + S} \\ \frac{dB}{dt} = \frac{\mu_{s,\max} SB}{K_S + S} - \frac{1}{Y_B} \frac{\mu_{B,\max} BA}{K_B + B} - \rho B \\ \frac{dA}{dt} = \frac{\mu_{B,\max} BA}{K_B + B} - \rho A \end{cases}$$

avec :

–  $\rho S_0$ , l'alimentation du chémostat (débit  $\rho$ , concentration  $S_0$ ),

<sup>12</sup>La "star" des bactéries, de son nom complet *Escherichia Coli* du nom du Dr. Escher, le biologiste qui l'a découverte. Intensément étudiée, elle a fait dire à Jacques Monod cette phrase célèbre : "tout ce qui est vrai pour *E. Coli* est aussi vrai pour l'éléphant".

<sup>13</sup>Un chémostat (voir 4.1.1 pour plus de détails) est une cuve alimentée en substrat et présentant en sortie un débit équivalent à l'entrée (volume constant). La concentration d'entrée est une concentration de substrat fixée. En revanche en sortie, les concentrations sont celles en présence dans la cuve à l'instant considéré.

Organisme	Taux de croissance spécifique max $\mu_{\max}, h^{-1}$	Constante de saturation $K$	Coefficient de production $Y$
<i>D. discoideum</i>	0,24	$4 \times 10^8$ bactéries/mL	$1,4 \times 10^3$ bactéries/amibes
<i>E. Coli</i>	0,25	$5 \times 1^{-4}$ mg glucose/mL	$3,3 \times 10^{-10}$ mg glucose/bactérie

TAB. 3.2 – Valeurs des paramètres du modèle proie-prédateur de Tsuchiya et co-auteurs (d'après H. M. Tsuchiya et al, J. Bacteriol., **110** :1151 (1972)).

- $-\rho X$  avec  $X = S, B$  ou  $A$ , la dilution équivalente,
- $\frac{1}{Y_S} \frac{\mu_{S,\max} SB}{K_S + S}$ , la consommation de sucre par les bactéries et donc, à un facteur près, le terme de croissance pour ces mêmes bactéries,
- $\frac{1}{Y_B} \frac{\mu_{B,\max} BA}{K_B + B}$ , le terme de prédation pour les bactéries, de croissance pour les amibes.

Après avoir déterminé les constantes cinétiques du système (voir tableau 3.2), Tsuchiya et co-auteurs ont obtenu par simulation les courbes présentées en 3.10. On voit que, non seulement le modèle prédit des oscillations dans les conditions de l'expérience ( $D = 0,0625/h$ ,  $S_0 = 0,5$  mg/mL), mais il reproduit également bien les périodes et amplitudes expérimentales, ainsi que les relations entre les phases des trois variables en jeu dans le système.

De plus, d'un point de vue théorique, les oscillations de type cycle limite présentes dans ce modèle sont beaucoup plus réalistes que les oscillations de type "centre" présentes dans le modèle de Lotka-Volterra. En effet, la moindre petite perturbation structurelle détruit les oscillations du modèle de Lotka-Volterra (alors qu'un système naturel est sujet à de nombreuses perturbations), ce qui n'est pas le cas pour un cycle limite plus robuste. D'autre part, d'un point de vue expérimental, l'observation des oscillations que l'on trouve généralement en biologie est caractérisée par une amplitude et une période relativement constante et ne dépendant pas des conditions initiales du système, ce qui n'est pas le cas du modèle de Lotka-Volterra (une infinité de périodes et d'amplitudes possibles dépendant des conditions initiales dans le cas différentiel, une amplitude aléatoire dans le cas stochastique).

**Remarque 30** On peut déduire intuitivement à l'aide des considérations suivantes (qui, bien que relativement simples, sont originales) que l'apparition d'un cycle limite est le comportement le plus "complexe" que peut produire le modèle de Tsuchiya. En effet, si le modèle de Tsuchiya est de dimension 3 (trois variables d'état), on peut remarquer que, de par les relations étroites qui couplent les trois variables,

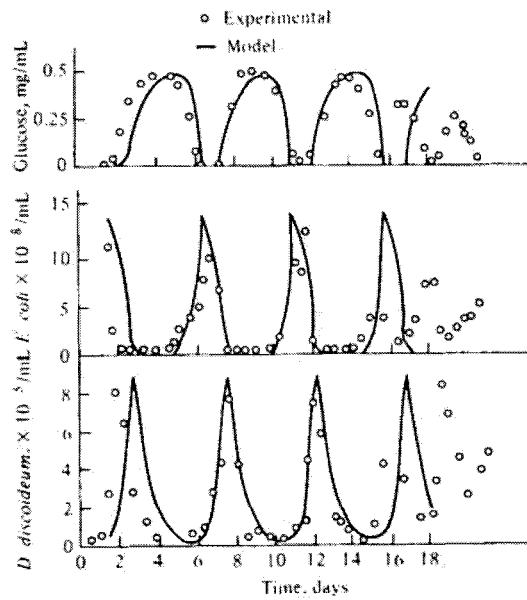


FIG. 3.10 – Comparaison entre les résultats du modèle (-) et expérimentaux (O) pour le système proie-prédateur *D. discoideum*-*E. coli* en culture continue (25°C,  $D=0,0625h^{-1}$ ). (tiré de H. M. TSUCHIYA et al., "Predator-Prey Interactions of Dictyostelium discoideum and Escherichia coli in Continuous Culture," J. Bacteriol., vol. 110, p.1147, 1972.)



on tend très vite vers un système “quasiment de dimension 2”. En effet, posons

$$T = S + \frac{B}{Y_S} + \frac{A}{Y_S Y_B},$$

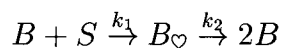
on a la relation  $\dot{T} = D(S_0 - T)$ , d'où  $T(t) = S_0 + \lambda e^{-Dt}$  avec  $\lambda = T(0) - S_0$ .

Après un transitoire, les variables  $S$ ,  $B$  et  $A$  se retrouvent donc sur le plan  $S + \frac{B}{Y_S} + \frac{A}{Y_S Y_B} = S_0$ . Or, d'après le théorème de Poincaré-Bendixon, le comportement le plus “complexe” possible pour un système d'équations différentielles bornées sur un plan est un cycle limite (cela veut dire qu'un tel système ne peut pas présenter de chaos par exemple). Pour obtenir des comportements plus compliqués, il faut rajouter d'autres variables, comme par exemple la prise en compte du temps d'assimilation ou encore une alimentation variable. De telles modifications engendrent des comportements dynamiques plus complexes (comme par exemple l'évolution sur un tore [3, 5]).

On peut se poser la question de savoir si, comme les systèmes étudiés précédemment, le système modélisé par Tsuchiya sous forme d'un système d'ODE, peut-être également modélisé également par un ensemble de RCG. Pour répondre à cette question, il est tout d'abord nécessaire d'essayer de comprendre les mécanismes de la saturation. En effet, l'équation cinétique de Monod ne correspond à aucun des mécanismes directs et élémentaires de type “action des masses” décrits plus haut (transformation, réaction bimoléculaire, dimérisation, création et dégradation simple).

Une bactérie qui a consommé du sucre comme nutriment nécessaire à sa division, ne donne pas immédiatement deux bactéries opérationnelles. Il existe un temps de latence (fabrication de protéines, réplication de l'ADN...) avant que la bactérie mère ne donne naissance à deux bactéries filles qui consommeront à leur tour du sucre.

Si on appelle  $B_{\heartsuit}$  la bactérie occupée par l'assimilation du sucre et la division cellulaire, on peut alors modéliser cette latence par l'équation suivante :



De la même façon, que pour  $B_{\heartsuit}$ ,  $A_{\heartsuit}$  désigne l'amibe occupée par la consommation d'une bactérie. Quand au symbole  $*$ , il représente la notation que nous avons adoptée pour décrire une quantité qui ne varie pas (voir section 3.2.3) et qui est utilisée pour représenter une création à taux constant (ici, l'alimentation en nutriment  $S$ ) ou une disparition également à taux constant (ici, celle de tous les composés présents dans la solution, par dilution). Les réactions modélisant le système sont alors :



avec  $X = S, B, B_{\heartsuit}, A$  ou  $A_{\heartsuit}$ .

On peut donner à chaque équation un sens biologique ou physique particulier :

- l'équation 3.10 représente l'alimentation du chemostat,
- l'équation 3.11 représente l'assimilation du substrat par les bactéries,
- les équations 3.12 et 3.13 représentent la prédation sur les bactéries,
- l'équation 3.14 représente la vidange du chemostat.

Nous avons donc ramené le mécanisme de Tsuchiya avec saturation à une suite de mécanismes élémentaires de type action des masses<sup>14</sup>. Ce système est donc maintenant simulable par notre approche particulière.

La simulation stochastique de ce mécanisme réactionnel permet de retrouver pour une certaine plage de paramètres l'équivalent d'un cycle limite (c'est-à-dire, dans le cas stochastique, une trajectoire périodique aux petites fluctuations près, qui ne dépend pas des conditions initiales), avec une alternance de phases et de périodes de faible effectif, en concordance avec les comportements obtenus par Tsuchiya par l'expérience et la simulation (voir figures 3.11 et 3.12 à comparer à 3.10).

Nous avons donc montré que l'on peut recréer le comportement modélisé par le système de Tsuchiya par un modèle composé d'un ensemble de RCG simulées stochastiquement. Contrairement au cas différentiel, la justification biologique se fait ici au niveau élémentaire (la saturation s'exprime par le fait que la plupart des bactéries sont "occupées" à assimiler le substrat) et non au niveau global (par une modification du taux de croissance).

L'intérêt d'un tel résultat tient au fait que le modèle de Tsuchiya et co-auteurs représente l'embryon de systèmes plus complexes ou présentant une réalité biologique connexe (autres couples proie/prédateur par exemple). Parmi ces modèles, citons le modèle de Mosekilde décrivant la cohabitation dynamique entre les bactéries et leurs virus, les bactériophages [55], modèle qui sera la base du modèle développé au chapitre 4.

Ainsi, s'il est possible d'obtenir une contrepartie stochastico-chimique au modèle différentiel de Tsuchiya, il est également possible sans difficultés théoriques

<sup>14</sup>En fait, d'un point de vue chimique, nous avons introduit des "complexes"  $A_{\heartsuit}$  et  $B_{\heartsuit}$  analogues à ceux introduits en cinématique enzymatique de Michaëlis-Menten [1]. On sait en effet que la limite de ce mécanisme de Michaëlis-Menten est une courbe cinétique sigmoïde.

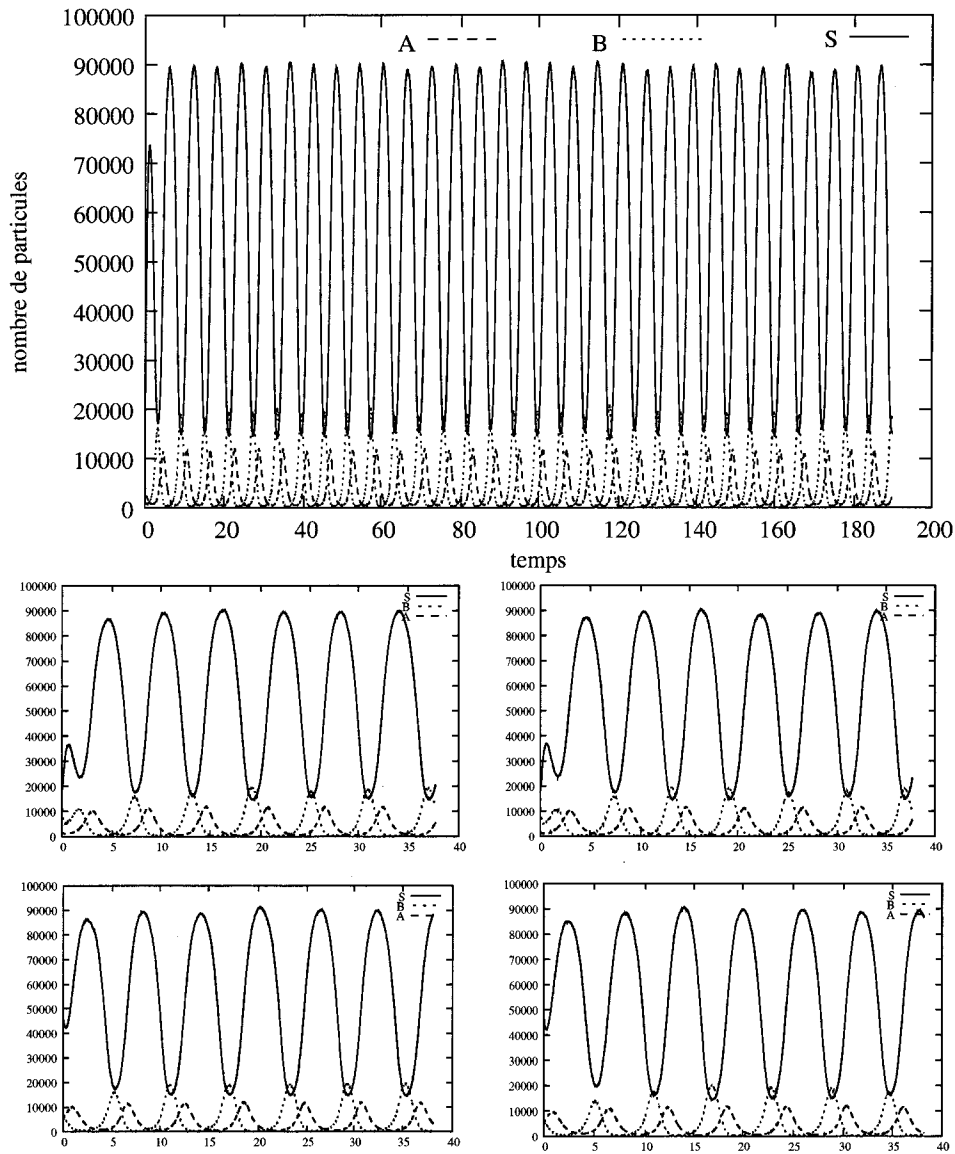


FIG. 3.11 – Concentrations de sucre, bactéries et amibes libres en fonction du temps du modèle de Tsuchiya stochastique pour  $D = 2$ ,  $S_0 = 100000$ ,  $\alpha_1 = 6.10^{-4}$ ,  $\alpha_2 = 3.10^{-4}$ ,  $\beta_1 = 4$  et  $\beta_2 = 6$ . Concernant les 4 simulations du bas, chaque ligne représente un même état initial, mais avec des tirages différents.

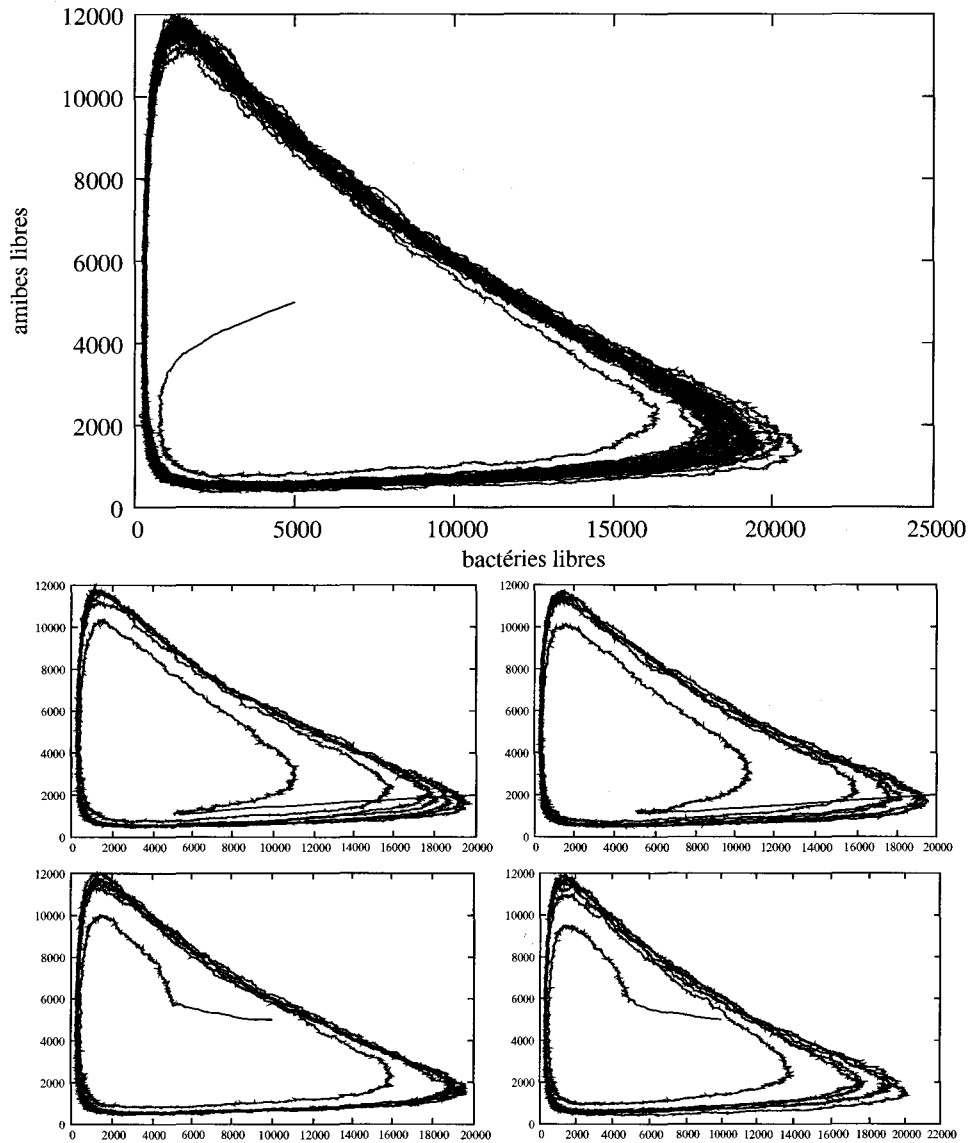


FIG. 3.12 – Cycle limite du modèle de Tsuchiya stochastique (bactéries libres vs. amibes libres) pour  $D = 2$ ,  $S_0 = 100000$ ,  $\alpha_1 = 6 \cdot 10^{-4}$ ,  $\alpha_2 = 3 \cdot 10^{-4}$ ,  $\beta_1 = 4$  et  $\beta_2 = 6$ . Concernant les 4 simulations du bas, chaque ligne représente un même état initial, mais avec des tirages différents.

particulières de le faire pour des modèles plus complexes présentant les mêmes types de termes (sommés, produit de deux variables et saturation).

Néanmoins, d'un point de vue pratique ce type de modèle devient vite difficilement simulable et on préfère alors utiliser un modèle différentiel approximant : en effet si les simulations présentées ici sont extrêmement rapides (autour de 40s pour les simulations de haut de pages, de 9s pour les petites), elles ne prennent en compte qu'un tout petit nombre de particules (nous reviendrons plus précisément sur la comparaison entre la complexité numérique des ODEs et de la simulation stochastique des RCG à la section 3.3.1).

Nous avons comparé dans cette section, la modélisation particulaire stochastique et la modélisation différentielle de mêmes systèmes. On peut se poser la question de savoir quand il est préférable de choisir l'une ou l'autre des modélisations. On considère généralement dans ce type de situation (voir par exemple [63]) que le choix du type de modèle dépend de l'effectif de la population. Ainsi, pour une population de petite taille, on privilégiera un modèle stochastique et pour un modèle de grande taille, un modèle différentiel. Cela se justifie par le fait que, dans une population de faible effectif, chaque événement "compte" et peut changer fondamentalement la physionomie du système, par exemple, dans le cas limite d'une population constituée d'un seul individu, la disparition accidentelle de cet individu entraîne la disparition de la population entière. En revanche, quand la population est de grande taille, de petites oscillations en grand nombre peuvent ne perturber que légèrement un mouvement plus général. La modélisation différentielle concentre alors les moyens de calcul sur ces événements importants et devient alors beaucoup plus efficace.

Le but de la section qui suit est de montrer que, dans certains cas (quand certains événements sont beaucoup plus rares que d'autres), l'approximation classique consistant à modéliser le système par un ensemble d'équations différentielles quand les effectifs des populations sont importantes, doit être remplacée par un SAM à base continue.

### 3.3 Du modèle particulaire au SAM à base continue

Avec la simulation stochastique de RCG, nous disposons d'un outil dont la finesse de description est intermédiaire entre les modèles agents et les modèles continus de type équations différentielles. La section précédente a montré que l'on pouvait retrouver des équivalents stochastiques aux modèles différentiels classiques déjà étudiés et que cela permettait de simuler de façon stochastique de nombreux systèmes étudiés en biologie théorique où ils sont décrits par des systèmes d'ODE (modèles de Tsuchiya ou Mosekilde par exemple).

Nous allons donc maintenant montrer que, lorsqu'il existe une hétérogénéité

de probabilité dans un système particulière, plusieurs échelles de temps peuvent apparaître naturellement et donner naissance à un système modélisable par un SAM. Nous utilisons pour cela une étude détaillée d'un modèle simple de réacteur bactérien qui, en l'absence de mutation, se comporte quand le nombre de particules est grand comme un système d'équations différentielles. En revanche quand on introduit dans le modèle la possibilité de mutation, ce passage à la limite n'est plus possible et on doit alors utiliser le formalisme des SAM (avec une base continue).

Ceci nous permettra, dans le chapitre 4, de traiter un exemple encore plus général, que ceux abordés jusqu'ici. En tant que SAM, il comprendra en effet un système d'ODE manipulé par une métadynamique discrète. Sa version particulière nécessitera des développements autres que ceux que nous venons de voir.

### 3.3.1 Un modèle non évolutif simple de réacteur bactérien

Nous commencerons par donner un modèle simple, non évolutif, de réacteur bactérien sous forme de RCG. Nous analyserons ensuite le comportement de ce système et montrerons comment il peut être relié aux équations différentielles.

#### Modèle particulière

Imaginons un bioréacteur contenant deux « substances » :

- un nutriment appelé  $S$ ,
- une souche bactérienne  $B$ .

Nous supposons que nous pouvons considérer ces substances comme des particules indistingables et nous allons à ce titre modéliser notre bioréacteur par un système d'équations stochastiques.

On suppose, de manière très fruste, que les « bactéries » se reproduisent en consommant une unité de nutriment  $S$ , ce qui nous donne la réaction suivante :



Notre réacteur est alimenté par un certain débit, ce qui se traduit par une certaine probabilité d'apparition de nutriment dans le bioréacteur :



avec  $k_a = \rho\sigma$ ,  $\rho$  le débit du réacteur et  $\sigma$  la teneur en nutriment de l'alimentation.

D'autre part, comme le volume est constant, il existe un débit de vidange équivalent, ce qui nous donne deux réactions supplémentaires :



Nous venons de modéliser grâce à ses quatre réactions, un réacteur bactérien élémentaire. Nous allons maintenant analyser ses propriétés dynamiques.

### Analyse et simulation de ce système

On s'intéresse au comportement asymptotique du système, comme l'apparition d'oscillations ou d'un équilibre (l'équilibre est à comprendre au sens de variations aléatoires de faibles amplitudes autour d'une valeur donnée et ayant pour valeur moyenne cette valeur). S'il existe un équilibre, alors il intervient quand les probabilités de création et de disparition des particules sont égales, soit :

- pour  $S$  :  $\rho\sigma = \rho S + k_1 SB$ ,
- pour  $B$  :  $\rho B = k_1 SB$ ,

d'où deux points d'équilibre possibles :

1.  $B = 0$ ,  $s = \sigma$ ,
2.  $B = \sigma - \frac{\rho}{k_1}$ ,  $S = \frac{\rho}{k_1}$ .

Nous avons simulé ce système en utilisant l'algorithme de Gillespie (voir section 3.2.3) et ceci montre (figure 3.13) que :

1. Si  $\frac{\rho}{k_1} > \sigma$ , le système est en situation de *washout*<sup>15</sup>. Les bactéries disparaissent petit à petit de la cuve et la quantité de sucre se stabilise autour de la valeur de l'alimentation (état d'équilibre 1).
2. Si  $\frac{\rho}{k_1} < \sigma$ , les bactéries se développent jusqu'à atteindre la valeur d'équilibre prédite  $\sigma - \frac{\rho}{k_1}$  et le sucre se stabilise à  $\frac{\rho}{k_1}$  (état d'équilibre 2).

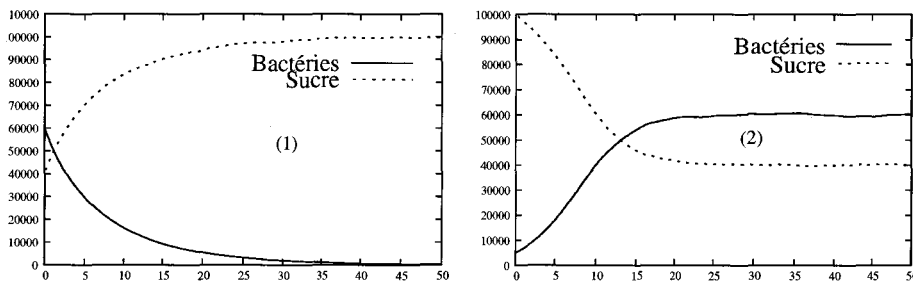


FIG. 3.13 – Deux situations possibles selon la valeur de  $k_1$  pour  $\rho = 0.2$  et  $\sigma = 100000$ . (1) Etat d'équilibre 1,  $k_1 = 10^{-6}$ , situation dite de *washout*, les bactéries disparaissent et le sucre se stabilise autour de la valeur de l'alimentation ( $S_0 = 40000$  et  $B_0 = 60000$ ). (2) Etat d'équilibre 2,  $k_1 = 5 \times 10^{-6}$ , la valeur du sucre se stabilise autour de  $\frac{\rho}{k_1} = 40000$ , les bactéries autour de  $\sigma - \frac{\rho}{k_1} = 60000$  ( $S_0 = 100000$  et  $B_0 = 5000$ ).

**Remarque 31** La présence de ces deux équilibres semble, au premier abord, contradictoire avec la théorie des chaînes de Markov. En effet, cette théorie nous dit que l'état 2 ne peut être que transitoire, puisque toutes les trajectoires doivent finir sur l'état absorbant 1, étant donné que,

$$P(X_{n+1} = \text{état 1} | X_n = \text{état 1}) = 1,$$

<sup>15</sup>vidange

avec  $X_n$  l'état du système après la  $n^{\text{ième}}$  réaction, (ce qui se traduit par  $\lim_{n \rightarrow \infty} P(X_n = \text{état } 1) = 1$ ).

Le "paradoxe" vient de ce que cette propriété est vraie à l'infini et qu'il peut s'écouler pas mal de temps avant d'y arriver. La véritable question est donc en fait : pour quelle valeur de  $N$ , pour un  $p$  donné pas trop petit, on a  $\forall n > N$ ,  $P(X_n = \text{état } 1 | X_0) > p$ , sachant que la suite  $P(X_n = \text{état } 1 | X_0)$  est croissante. Si pour un  $p$  petit,  $t_N$  est égal à plusieurs fois l'âge de l'univers, on peut considérer qu'en pratique, le système tend vers l'état d'équilibre 2.

### Passage aux équation différentielles

Nous allons maintenant étudier les liens théoriques qui existent entre les équations différentielles ordinaires (ODE) et les modèles stochastiques. Cette étude se terminera sur une propriété de convergence du système étudié vers un système d'ODE quand le nombre de particules dans le système tend vers l'infini.

Notre modèle est un modèle à temps continu, mais à variables discrètes. L'évolution des variables d'état est représentée par des fonctions en escalier. Le changement est décidé de manière stochastique : néanmoins, quand le nombre de particules augmente, on observe l'apparition d'une certaine régularité (figure 3.14).

Ce qui se produit est que, quand le nombre de particules est grand<sup>16</sup>, on ne s'intéresse plus à un nombre exact de particules, mais aux proportions des différentes espèces dans la "solution". Les petites variations aléatoires se compensent et deviennent imperceptibles à moins de regarder la courbe avec un fort grossissement (voir figure 3.15).

Si l'on ramène la "masse" totale de particules à 1, le "poids" d'une particule est de l'ordre de  $\frac{1}{N}$  avec  $N$  le nombre total de particules. La variation relative lors de la disparition ou de l'apparition d'une particule dépend donc du nombre total de particules, la création d'une particule parmi  $10^{10}$  est imperceptible alors qu'elle est notable parmi  $10^2$ . On assiste donc à un passage à la continuité des variables quand le nombre de particules tend vers l'infini dans le système.

A cela vient s'ajouter un autre phénomène pour le temps. Le grand nombre de particules augmente le coefficient de la loi exponentielle décidant de l'exécution de la prochaine réaction. Cela a pour effet de resserrer la fonction de répartition du temps entre deux réactions en un pic autour de 0 quand le nombre de particules est très grand. A des variations de temps infiniment rapides correspondent donc des variations d'état infiniment petites quand le système est de grande taille. Intuitivement, on peut supposer que la limite d'un tel système est un système différentiel. Ceci permet de formuler la conjecture suivante :

**Propriété 32 (Conjecture)** Pour le système correspondant aux réactions 3.15,

<sup>16</sup>le grand nombre de particules est garanti par les probabilités en œuvre dans le système (l'apport du sucre a une probabilité qui devient proche de 1 quand le nombre de variables est proche de 0).



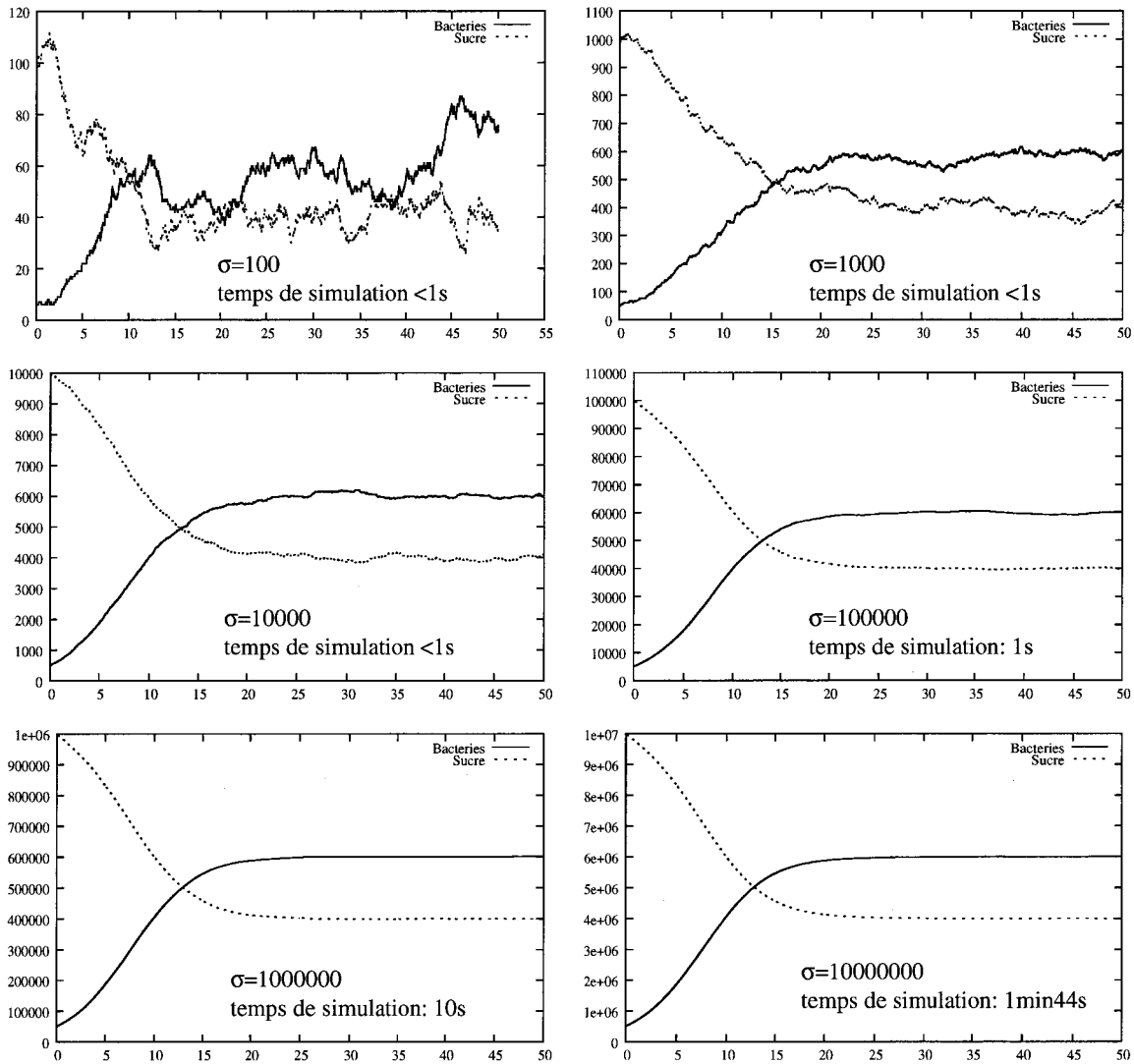


FIG. 3.14 – Évolution du système stochastique vers un système différentiel : six simulations avec  $\sigma$  croissant et  $\alpha = k_1\sigma = 0,5$  constant pour  $\rho = 0,2$ ,  $B_0 = 5 \times 10^{-2} \times \sigma$  et  $S_0 = \sigma$ . On observe également que le temps de simulation augmente linéairement avec  $\sigma$  (un temps de simulation < 1s indique que le temps de simulation est inférieur à la seconde et n'a pu être mesuré plus précisément).

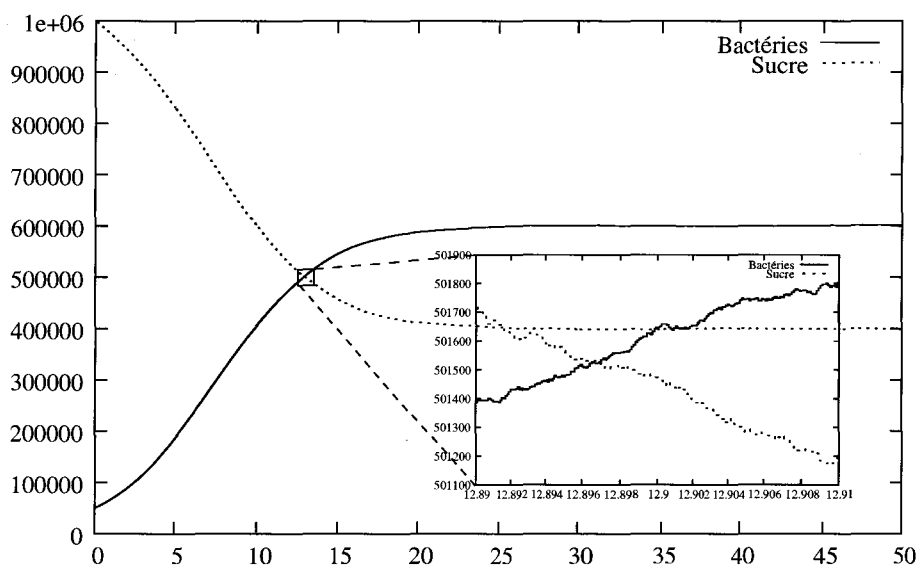


FIG. 3.15 – Zoom sur une partie en apparence lisse : les fluctuations aléatoires du modèle stochastiques apparaissent.

3.16, 3.17 et 3.18, auand  $\sigma \rightarrow \infty$ , les fonctions en escalier  $X_{\omega,\sigma} : t \mapsto \frac{n_B(t,\omega)}{\sigma}$  et  $Y_{\omega,\sigma} : t \mapsto \frac{n_S(t,\omega)}{\sigma}$ , avec  $k_1\sigma = \alpha$  donné, tendent pour presque tous les  $\omega$  vers les solutions du système d'équations différentielles

$$\dot{x} = \alpha xy - \rho x \quad (3.19)$$

$$\dot{y} = \rho(1 - y) - \alpha xy \quad (3.20)$$

avec  $x(0) = X_{\omega,\sigma}(0)$  et  $y(0) = Y_{\omega,\sigma}(0)$ .

Cette propriété stipule donc que quand le nombre de particules dans le système devient grand, la trajectoire engendrée se rapproche de celle des solutions du système différentiel composé des équations 3.19 et 3.20. Cela peut se voir graphiquement en comparant la figure 3.14 et 3.16.

### Complexité du calcul numérique

Il est intéressant de comparer le système d'équations stochastiques avec le système d'ODE du point de vue de la complexité du calcul numérique engendré. Pour simuler le système de la figure 3.16, on a pris un pas de  $5 \cdot 10^{-3}$ , soit  $10^4$  itérations (eulériennes), ce qui donne une résolution parfaite. Le temps de simulation est inférieur à la seconde. Pour ce qui concerne le système d'équations stochastiques, ce nombre d'opérations dépend de beaucoup de facteurs, en particulier de  $\omega$ , la suite de tirages aléatoires. On peut néanmoins l'évaluer : il se passe en moyenne une transition tous les  $\frac{1}{\lambda}$  avec à l'équilibre :

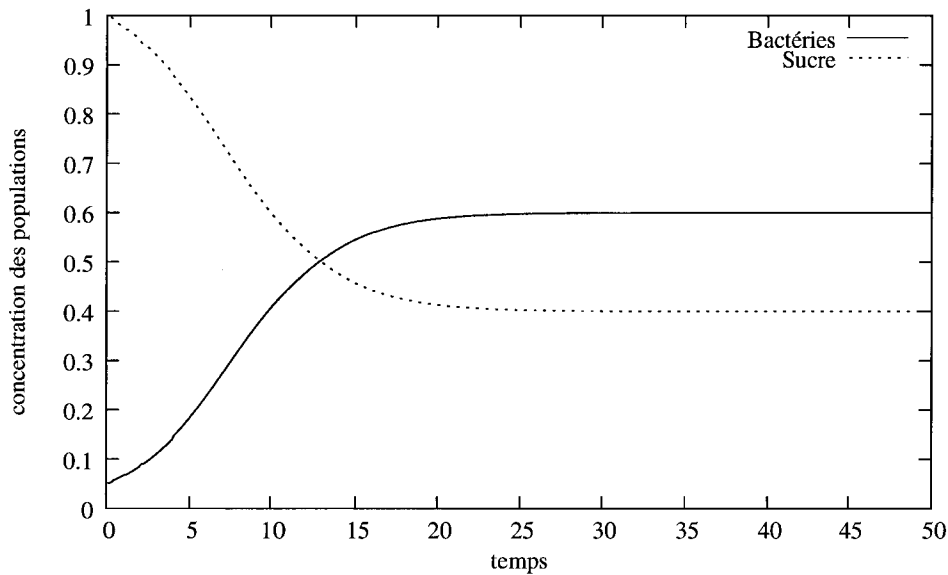


FIG. 3.16 – Simulation du système différentiel constitué des équations 3.19 et 3.20 pour  $\alpha = 0,5$  et  $\rho = 0,2$  par la méthode d'Euler. Le pas d'intégration pris est de  $5 \cdot 10^{-3}$ , le nombre d'itérations effectué lors de la simulation est donc de  $10^4$ . Le temps de simulation est inférieur à la seconde.

$$\begin{aligned} \lambda &\approx k_1 SB + \rho(S + B + \sigma) \\ &\approx k_1 \frac{\rho}{k_1} \left( \sigma - \frac{\rho}{k_1} \right) + \rho \left( \frac{\rho}{k_1} + \sigma - \frac{\rho}{k_1} + \sigma \right) = \left( 3 - \frac{\rho}{\alpha} \right) \rho \sigma. \end{aligned}$$

Le nombre d'itérations pendant le temps  $t$  est de l'ordre de  $\lambda t$  et est donc linéaire en  $\sigma$ . On retrouve bien une loi cohérente avec les temps de simulation donnés dans les exemples de la figure 3.14.

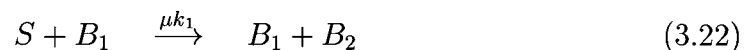
Quand le système est de grande taille (ce qui est par définition le cas dans les systèmes complexes) le temps de simulation par l'algorithme de Gillespie devient très long, même pour un système relativement simple (mais de grande taille). Un comportement typique (surtout dans des situations proches de l'équilibre) est la consommation par le système de la quasi totalité du temps de calcul pour produire de petites oscillations non perceptibles, très rapides et de faibles amplitudes. Les réactions "intéressantes" faisant réellement "bouger" le système ne se produisent que beaucoup plus rarement et consomment en comparaison très peu de temps de calcul.

### 3.3.2 Ajout de l'aspect évolutif (Modèle particulaire évolutif)

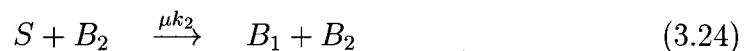
Après avoir analysé le comportement du modèle non-évolutif, nous allons maintenant ajouter l'aspect évolutif. Nous allons montrer que, dans ce cas, quand le nombre de particules augmente, le système n'est alors plus réductible à un système d'ODE, à moins de lui donner une structure variable. Il fournit alors un exemple simple de SAM.

#### Modèle particulaire avec mutations

On suppose maintenant que la réplication de  $B$  n'est pas parfaite et que survient avec une probabilité  $\mu$ , une erreur de recopie ou mutation. On différencie les deux espèces par  $B_1$  (l'espèce d'origine) et  $B_2$  (l'espèce mutante), l'équation 3.15 devient alors



On suppose dans un premier temps, pour simplifier le problème, que la bactérie  $B_2$  possède une même probabilité  $\mu$  d'erreur de recopie et que la copie redonne une bactérie  $B_1$ . On obtient alors la réaction suivante :



On suppose qu'au début de la simulation, seules les bactéries  $B_1$  sont présentes dans le système. Au bout d'un certain temps, les bactéries mutantes  $B_2$  sont susceptibles d'apparaître.

#### Analyse et simulation

La seule différence entre les bactéries  $B_1$  et  $B_2$  est la valeur de leur coefficient de croissance (respectivement  $k_1$  et  $k_2$ ). Dans un écosystème réel, une mutation peut avoir trois conséquences possibles :

1. la nouvelle bactérie est moins adaptée, on parle alors de **mutation défavorable**, cela correspond ici à  $k_2 < k_1$ ,
2. elle ne modifie pas le fonctionnement<sup>17</sup>, on parle alors de **mutation neutre**, ceci se traduit dans notre exemple par  $k_2 = k_1$ ,

<sup>17</sup>on rappelle qu'une mutation change le génotype (l'information contenue dans l'ADN), mais pas forcément l'expression de ce génotype, le phénotype.

3. la mutation améliore l'efficacité de la bactérie, on parle alors de **mutation favorable**, ce qui donne  $k_2 > k_1$  pour notre système.

Toujours en utilisant l'algorithme de Gillespie, on simule le système, afin d'analyser le comportement dynamique de notre système en fonction de la valeur de  $k_2$ .

**Mutation défavorable** ( $k_2 < k_1$ ) Rapidement la bactérie mutée disparaît. Elle ne dépasse jamais un nombre confidentiel.

**Mutation neutre** ( $k_2 = k_1$ ) Le cas où la mutation est neutre est particulièrement intéressant. On se ramène alors au cas précédent en posant  $B = B_1 + B_2$  et le comportement de la variable "agrégée" est, pour  $\sigma$  grand, proche de la solution du système différentiel.

En revanche, on peut s'intéresser à la proportion relative de  $B_1$  et de  $B_2$  au sein de  $B$ .

On a alors un modèle très proche de celui de la dérive génétique présenté au chapitre 1 en 1.2.2. Les bactéries  $B_1$  et  $B_2$  décrivent un mouvement brownien. Comme pour le cas de la dérive, la bactérie mutante peut remplacer l'ancienne, mais a en fait très peu de chances d'y arriver en un temps raisonnable si la population totale d'équilibre est grande.

**Mutation favorable** ( $k_2 > k_1$ ) En simulation, on remarque que  $B_2$  remplace  $B_1$  au sein du système, au bout de quelques mutations (souvent une suffit). Un nouvel équilibre s'établit alors, plus favorable aux bactéries  $B_2$  (voir figure 3.17).

Du point de vue des variables en jeu, on distingue trois périodes dans le processus de remplacement de  $B_1$  par  $B_2$  :

- des  $B_1$  seuls ou avec des  $B_2$ , mais en nombre insignifiant (période 1),
- puis des  $B_1$  et des  $B_2$  en compétition (période 2),
- et finalement une présence exclusive de bactéries  $B_2$  (période 3).

Il est à remarquer que dans chacune de ces trois phases, le comportement est quasi continu (grand nombre de particules).

Comme pour le réacteur bactérien non évolutif, l'augmentation du nombre de particules entraîne un "lissage" des trajectoires et un ralentissement du temps de simulation. En revanche, le passage à la limite vers un système d'ODE n'est plus possible, puisque le caractère déterministe n'est plus valable que temporairement en considérant un système différentiel sur l'espace d'état comprenant les variables présentes au sein du système en nombre suffisant. L'apparition d'une population macroscopique de bactérie mutante engendre une nouvelle dynamique.

On a donc une suite de systèmes dynamiques différents dont le passage de l'un à l'autre s'effectue selon des règles précises : il s'agit là d'un SAM tel que défini par le formalisme du chapitre 2.

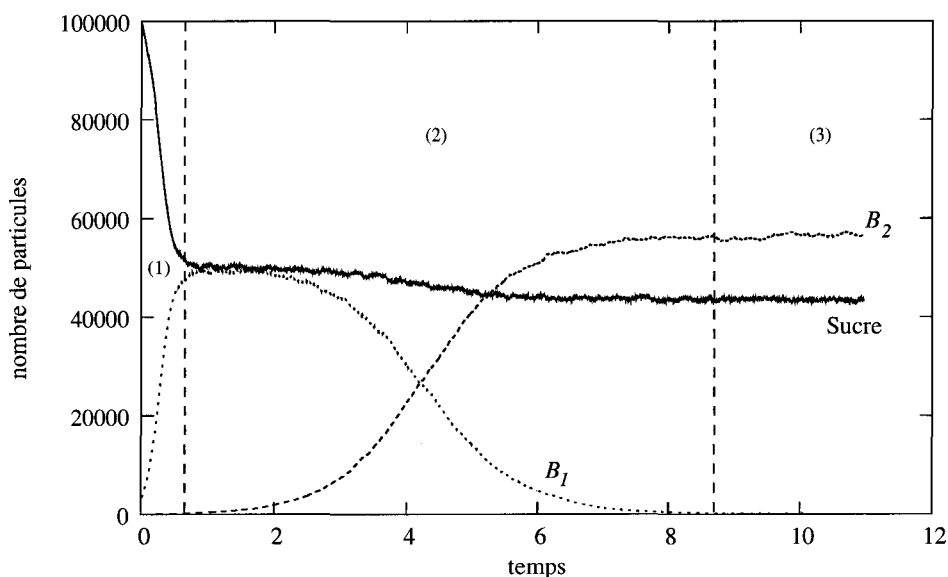


FIG. 3.17 – Échange de bactéries dans le système stochastique,  $\sigma = 100000$ ,  $k_1 = 2e - 4$ ,  $k_2 = 2, 3e - 4$ ,  $\rho = 10$ . On distingue clairement trois périodes : (1) des  $B_1$  et  $S$ , (2) compétition entre  $B_1$  et  $B_2$  avec toujours du  $S$ , (3) plus que des  $B_2$  et du  $S$ .

### 3.3.3 Etablissement du SAM équivalent

Définir les SD locaux suit une même démarche que celle mise en œuvre pour le réacteur non évolutif. En revanche, l'établissement des règles de transition entre ces SD (la règle métadynamique) nécessite quelques développements.

#### Etablissement des transitions métadynamiques

Nous n'étudions que les transitions de la période 1 vers la période 2 et de la période 2 vers la période 3. Nous n'étudions donc pas les transitions "retour" de 3 vers 2 et 2 vers 3, car nous avons précédemment vu que de telles transitions étaient très improbables (cas de la mutation défavorable).

**Transition de 1 vers 2** Le passage de la période 1 à la période 2 se fait quand le nombre de  $B_2$  dépasse un certain seuil, à partir duquel on le considère comme macroscopique. Ceci signifie qu'à partir de ce moment, on considère que les bactéries  $B_2$  sont en nombre suffisant pour :

- influencer la dynamique du système,
- pouvoir être considérées comme étant soumises elles-mêmes à une dynamique de type différentiel.

Ce "temps d'apparition d'un seuil macroscopique", que l'on note  $t_{Mac(B_2)}$  (figure 3.18), dépend de  $\mu$  et de la suite de tirages aléatoires effectuée lors de la simulation.

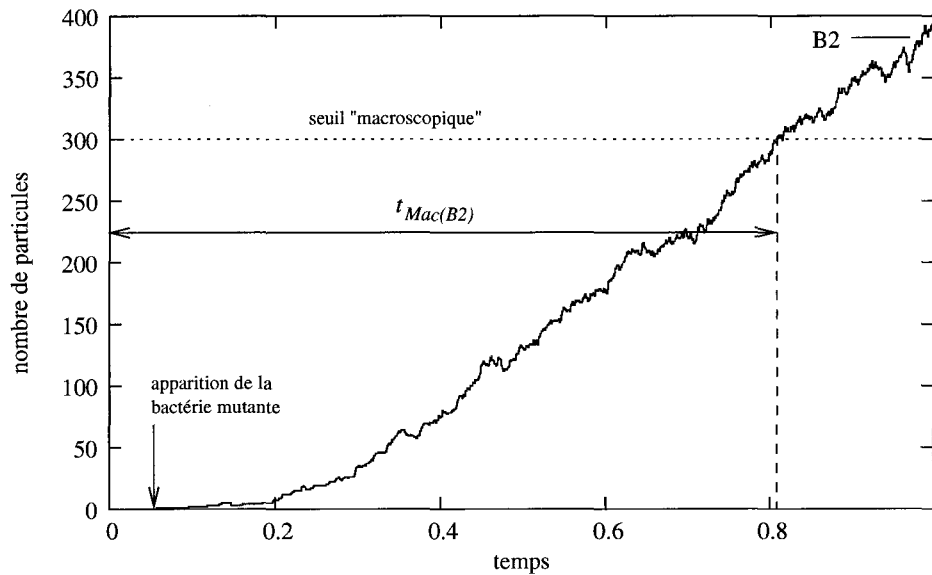


FIG. 3.18 – Dépassement du seuil macroscopique

Dans le cas de l'exemple, on commence par faire une analyse statistique pour trouver la distribution de  $t_{Mac(B_2)}$ . Pour cela on a réalisé  $10^5$  simulations du système avec une probabilité de mutation  $\mu = 0.01$ , et on a gardé en mémoire les  $t_{Mac(B_2)}$  de chaque simulation. Le diagramme de la distribution en question est donnée figure 3.19.

On remarque que la distribution du temps d'apparition du seuil macroscopique est très régulière et permet de donner une loi de distribution de l'apparition de  $t_{Mac(B_2)}$ . Bien sûr, dans le cas d'un système complexe réel, on devra évaluer la distribution par des expériences ou à l'aide d'un modèle issu de la connaissance du système.

**Transition de 2 vers 3** Dans le système, supposé d'état continu, la disparition d'une population se caractérise par une convergence de la trajectoire de cette population vers 0. Néanmoins, même si celle-ci devient de plus en plus petite, elle ne s'annule jamais. Dans le système stochastique premier, une population décroît jusqu'à atteindre un niveau très bas : elle finit par disparaître quand les évènements aléatoires provoquent la disparition du dernier élément de la population. Nous faisons l'hypothèse qu'il existe, dans le modèle différentiel, un seuil au delà duquel, on peut considérer comme certain que la population a disparu dans le modèle stochastique correspondant. On modélise donc la transition de 2 vers 3 par un seuil : quand la population atteint un seuil inférieur donné, elle disparaît.

On a donc deux transitions de nature différente, une stochastique de 1 vers 2, l'une déterministe de 2 vers 3. L'on peut résumer cela en présentant les transitions sous une forme type chaîne de Markov (figure 3.20).

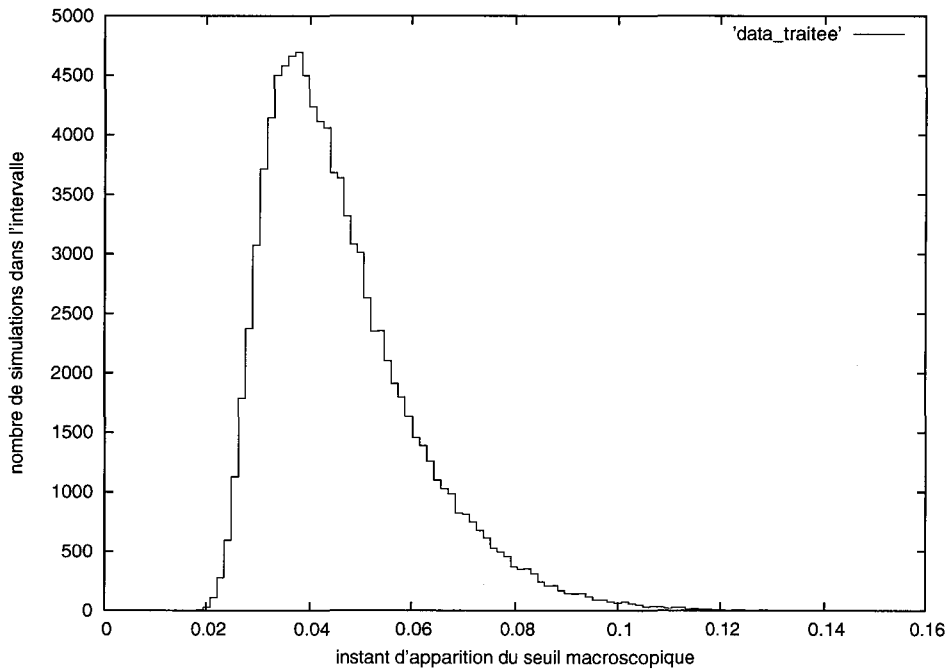


FIG. 3.19 – Distribution statistique des instants d’apparition d’un seuil macroscopique de  $B_2$  pour  $10^5$  simulations, avec une probabilité de mutation  $\mu = 0,01$ .

### SAM équivalent du réacteur bactérien évolutif

Maintenant que nous connaissons les transitions entre les systèmes dynamiques, nous disposons de tous les éléments pour implémenter le SAM équivalent au réacteur bactérien évolutif. Les trois systèmes dynamiques du niveau métadynamique correspondent à trois approximations différentielles du système particulaire. Le changement dans le temps est donc donné par l’intégration d’un système d’équations différentielles (le temps reste continu, mais on suppose maintenant que les changements sont également continus, les populations n’étant plus dénombrées par  $\mathbb{N}^n$ , mais par  $\mathbb{R}^n$ ).

Nous allons donc maintenant expliciter les différentes parties du SAM équivalent, ainsi que les ensembles, équations et fonctions associés, en nous référant aux notations introduites dans le chapitre 2.

**Partie dynamique** Nous ne considérons pas d’entrées particulières<sup>18</sup>, en revanche la sortie pourrait par exemple être dans  $\mathbb{R}^3$  avec, pour variables, les concentrations de  $S$ ,  $B_1$  et  $B_2$  avec ( $B_i = 0$  si  $B_i$  n’est pas dans le système) ou encore dans  $\mathbb{R}^2$  avec pour variables  $S$  et  $B = B_1 + B_2$ , etc. L’ensemble temps dynamique considéré est  $T_0 = [0, +\infty[$ . Il reste donc à décrire pour chacun des trois systèmes

<sup>18</sup>Mais l’on peut imaginer des cas où l’environnement aurait son importance : variation de la température ou des ressources fournies au système, par exemple par un contrôle de l’alimentation.



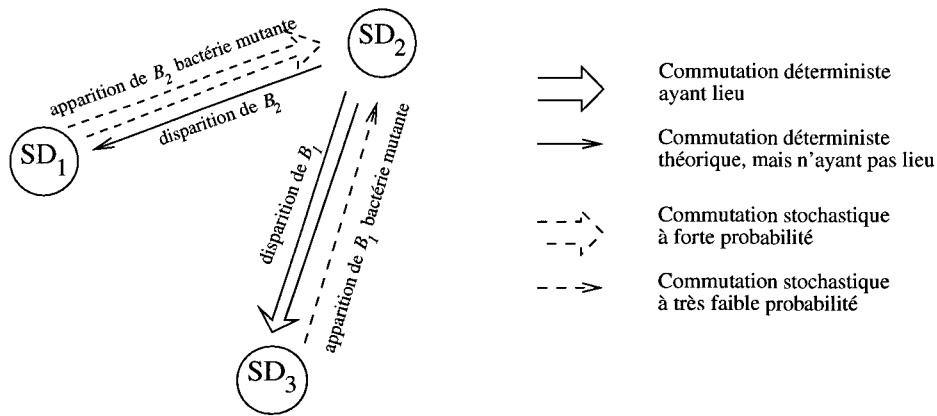


FIG. 3.20 – Représentation dans le style “chaîne de Markov” de la métadynamique du modèle de transition, on voit que le méta-état est quasi absorbant

possibles :

- l’espace d’état,
- la fonction de transition, que l’on présente sous la forme d’un système d’ODE, dont l’intégration donne la fonction de transition recherchée.

Comme pour le cas non-évolutif, on définit  $\alpha_i = k_i \sigma$ .

**Premier système dynamique** Espace d’état 1 :  $\mathbb{R}^2$  avec  $x_1$  et  $y$  décrivant les concentrations de  $B_1$  et de  $S$ .

Fonction de transition 1 donnée par les solutions du système différentiel suivant :

$$\begin{cases} \dot{x}_1 = \alpha_1 x_1 y - \rho x_1 \\ \dot{y} = \rho(1 - y) - \alpha_1 x_1 y \end{cases}$$

**Deuxième système dynamique** Espace d’état 2 :  $\mathbb{R}^3$  avec  $x_1, x_2$  et  $y$  décrivant les concentrations respectives de  $B_1, B_2$  et  $S$ .

Fonction de transition 2 donnée par les solutions du système différentiel suivant :

$$\begin{cases} \dot{x}_1 = \alpha_1 x_1 y - \rho x_1 \\ \dot{x}_2 = \alpha_2 x_2 y - \rho x_2 \\ \dot{y} = \rho(1 - y) - (\alpha_1 x_1 + \alpha_2 x_2) y \end{cases}$$

**Troisième système dynamique** Espace d’état 3 :  $\mathbb{R}^2$  avec  $x_2$  et  $y$  décrivant les concentrations de  $B_2$  et de  $S$ .

Fonction de transition 3 donnée par les solutions du système différentiel suivant :

$$\begin{cases} \dot{x}_2 &= \alpha_2 x_2 y - \rho x_2 \\ \dot{y} &= \rho(1 - y) - \alpha_2 x_2 y \end{cases}$$

**Métadynamique** Le “temps métadynamique” est considéré comme continu  $T_1 = [0, +\infty[$ , c’est-à-dire que les évènements métadynamiques (mutations et disparition d’une espèce) sont susceptibles de se produire à n’importe quel moment.

**Commutation de 1 vers 2** Cette commutation est donnée par la loi de distribution de la figure 3.19 et ne dépendant donc que du temps.

Les nouveaux états sont alors

$$B_1 \leftarrow B_1 - \text{seuil macroscopique}$$

$$B_2 \leftarrow \text{seuil macroscopique.}$$

**Commutation de 2 vers 3** Quand  $B_2 \leq \text{seuil de disparition}$ ,  $B_2$  disparaît.

Après simulation du SAM, on obtient les courbes de la figure 3.21 qui correspondent fidèlement à une idéalisation du comportement stochastique original (à comparer avec la figure 3.17).

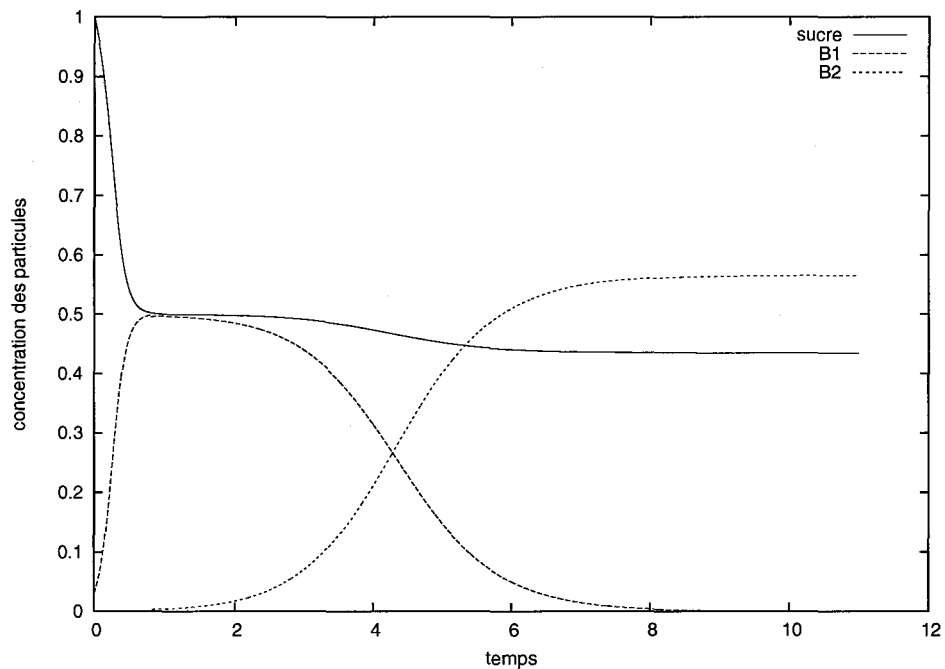


FIG. 3.21 – Échange de bactéries dans le SAM à base continue

**Remarque 33** *Il existe sans aucun doute des liens entre ce type de simplification et la théorie de l'information. En effet, d'un point de vue intuitif, on peut noter que la métadynamique met en valeur les événements aux probabilités faibles, ceux qui d'après la théorie de l'information contiennent le plus d'information.*

**Remarque 34** *Dans le SAM qui vient d'être décrit, on définit trois systèmes dynamiques continus sur trois espaces d'état et deux règles de commutation. Il nous a donc été possible de décrire cet exemple explicitement. Il n'en est pas de même dans les SAM complexes où le nombre de systèmes dynamiques et de règles de commutation devient énorme. On ne peut alors décrire de tels systèmes qu'en intention. Ce sera le cas du SAM décrit au chapitre 4 où le nombre de SD possibles sera égale à  $2^{2^{n_c}+1}$  avec  $n_c$  de l'ordre de la dizaine... Nous proposerons alors des solutions informatiques permettant la simulation d'un système décrit en intention.*

On peut résumer la modélisation des systèmes particulières à probabilités hétérogènes par le tableau 3.3.

type de représentation	organisation du modèle	équations d'évolution	avantage
système particulière	pas de hiérarchie, priorités données par les probabilités	discrète	plus fin et plus réaliste
SAM	2 niveaux : dynamique et métadynamique	continue	plus rapide

TAB. 3.3 – Comparaison des mérites respectives des modélisations particulières et par SAM.

## 3.4 Aspects de simulation

### 3.4.1 Algorithmes et structure de données

Les SAM permettent de décrire un grand nombre de systèmes complexes, parmi lesquels les systèmes hybrides ou les algorithmes génétiques. Si les SAM bénéficient de la puissance des différentes approches qu'ils synthétisent, ils héritent également de leurs difficultés de simulation. Nous ne présentons donc pas ici d'algorithme générique "miracle". Nous nous contentons de donner une structure de simulation très générale, esquissant les grandes lignes d'un programme de simulation sans résoudre les problèmes spécifiques liés à un système en particulier.

L'algorithme générique de base est représenté figure 3.22

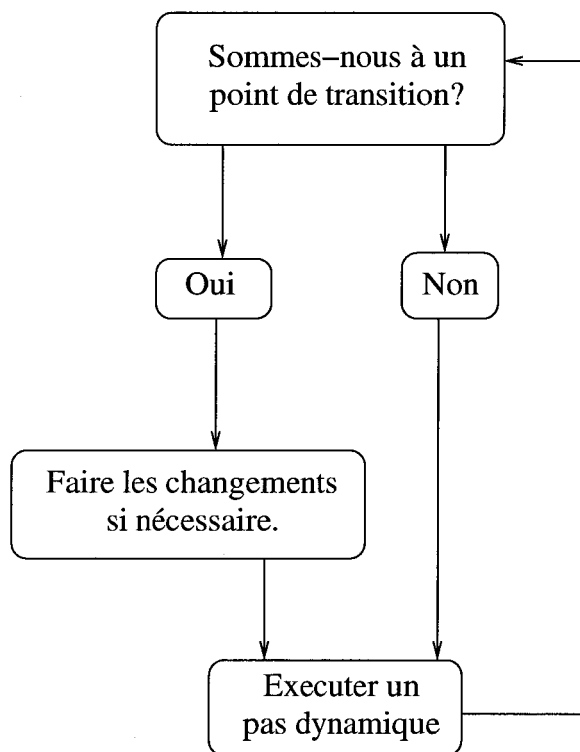


FIG. 3.22 – Algorithme de simulation d’un SAM

Dans le cas des systèmes continus, l’obtention d’un pas dynamique est un problème classique : celui de l’intégration d’un système d’équations différentielles. Le lecteur intéressé par la théorie de ce problème d’analyse numérique pourra se référer à [38] par exemple. Citons parmi les méthodes les plus couramment utilisées, l’algorithme de Runge-Kutta qui donne dans la plupart des cas d’excellents résultats.

Nous avons au chapitre précédent (section 8), décrit l’intérêt des SAM dans les situations où le système est décrit en intention, c’est-à-dire lorsqu’il se construit durant la simulation. Dans de tels cas, il est nécessaire d’adopter une structure de données flexible. Une structure se révèle en particulier très efficace : les listes chaînées. L’utilisation de telles structures de données sera développée à l’aide de l’application présentée au chapitre suivant.

### 3.5 Méthode numérique d’analyse des données

Le problème d’analyse de données revient à trouver un ensemble de sortie  $Y$  commun aux systèmes dynamiques du SAM et les  $\eta_i$ , fonctions de lecture de ces systèmes qui “convertissent” l’état du système en un élément de cet ensemble. Les difficultés résident dans le grand nombre de variables en présence dans le système

durant la simulation (souvent plusieurs milliers) et dans la taille énorme de l'espace d'exploration possible (quand on travaille sur des chaînes binaires de taille  $n$ , on a des espaces de l'ordre de  $2^{2^n}$  possibilités). On doit donc recourir à des méthodes statistiques (comme l'analyse en composantes principales) ou des variables agrégées de l'état du système comme l'entropie.

Nous verrons plus en détails l'implémentation de ces techniques lors du prochain chapitre consacré à l'étude d'un exemple biologique de SAM.

## 3.6 Conclusion

Nous avons donc justifié dans ce chapitre l'utilisation des SAM à base continue pour modéliser des populations de grande taille soumises à l'évolution de leurs caractéristiques.

Nous avons commencé par présenter un type de modèle stochastique particulier basé sur la simulation de réactions chimiques généralisées. Nous avons montré que ce type de modèle pouvait permettre de représenter des phénomènes biologiques complexes et constituait un niveau de description intermédiaire entre :

- les modèles d'agents,
- les modèles différentiels classiques.

Nous avons finalement montré que, lorsque les probabilités des événements élémentaires sont hétérogènes, il n'est plus possible d'utiliser la modélisation différentielle classique : on peut alors faire appel à une structure de type SAM, qu'elle soit explicite et en extension pour des tailles faibles ou implicite et en extension pour des tailles énormes telles que celles rencontrées en biologie de l'évolution. Nous développerons dans le chapitre qui suit un exemple d'un tel modèle biologique complexe.



# Chapitre 4

## Modèle Attaque-Défense

“La goutte toute entière paraissait grouiller de petits animaux [...] se déplaçant les uns parmi les autres, chacun animé de son mouvement propre.”

ANTOINE VAN LEEUWENHOEK

“Un biologiste numéricien devrait pouvoir en principe être capable d’étudier ce type de système en utilisant les plus lourdes des méthodes d’équations différentielles couplées de dynamique de population, appelées “équations de Lotka-Volterra” [...] Mais les temps de calcul sont si gigantesques que cela limite les systèmes que l’on peut étudier à un très petit nombre d’espèces, disons deux ou trois.”

PER BAK [10]

Après avoir abordé le problème de la modélisation des systèmes dynamiques adaptatifs complexes et proposé la structure de système à métadynamique comme outil de modélisation et d’étude théorique (chapitre 2), nous avons vu en détail comment cette approche pouvait se justifier comme approximation de systèmes particuliers de grande dimension (au sens du nombre de particules présentes dans le système), quand les probabilités des événements modifiant le système sont hétérogènes (chapitre 3). Nous allons maintenant présenter un modèle de la coévolution des bactéries et de leurs virus, les phages, utilisant ce concept de SAM et nous permettant d’en dégager diverses facettes.

### 4.1 Présentation générale

Nous cherchons à modéliser l’influence des interactions bactérie-phage sur la coévolution de ces deux espèces, c’est-à-dire en quoi le couplage entre les dynamiques des bactéries et des phages influencent le changement de composition génétique de leurs populations respectives.

Nous allons, dans cette présentation générale, commencer par décrire le contexte biologique du problème et en particulier les mécanismes en jeu. Dans la deuxième partie de cette présentation, nous exposerons la structure générale du modèle que nous en déduisons. Cette section 4.1 a pour objectif la compréhension des phénomènes biologiques et de la structure du modèle, sans donner pour le moment d'équations.

### 4.1.1 Mécanismes physiques et biologiques

Pour présenter les aspects physiques et biologiques, nous allons décrire le lieu, les acteurs et l'intrigue qui prennent place au sein de notre modèle.

#### Le lieu

Les bactéries constituent les plus petits êtres vivants sur terre et également les plus répandus. La diversité des lieux où on les rencontre est impressionnante, parmi ceux-ci citons :

- les fonds marins, où elles peuvent survivre à des pressions très importantes (*Escherichia Coli* supporte jusqu'à 300 atmosphères...),
- les milieux soufrés,
- les marais salins (bactéries halophiles),
- les réacteurs nucléaires, où l'on retrouve *Deinococcus Radiodurans*,
- le corps humain : elles sont essentielles à la digestion par exemple, mais peuvent également être pathogènes (porteuses de maladie), comme *Yersinia pestis* responsable de la peste,
- les réacteurs bactériens ou chémostats, installations spécialement dédiées à leur culture en vue d'exploiter les produits qu'elles synthétisent.

C'est dans ce dernier type de milieu, le milieu des chémostats que l'histoire de notre modèle se déroule. Les chémostats sont à la base de nombreux process biotechnologiques comme par exemple dans l'industrie du fromage [55]. Ce sont pour les bactéries des endroits idéaux : les nutriments leur sont fournis de façon continue et en quantité suffisante, la régulation de la température leur permettant des conditions optimales de croissance. Un chémostat<sup>1</sup> se compose d'une cuve contenant un milieu de croissance liquide et une population bactérienne appelée **culture** bactérienne. Cette cuve maintenue à température et à pH constants est mélangée de manière à garantir un milieu homogène.

L'alimentation se fait avec un certain débit d'entrée et peut contenir plusieurs nutriments (source de carbone) et un certain nombre d'ions, nécessaire au méta-

---

<sup>1</sup>Il existe également un autre type d'installation très proche du chémostat pour la culture bactérienne : le turbidostat. La différence entre les deux systèmes réside dans le fait que pour le chémostat, le débit de l'alimentation est fixé au préalable, alors que pour le turbidostat, il est régulé en fonction de la turbidité de la solution de sortie mesurée par exemple par une cellule photovoltaïque.



bolisme de la bactérie. Un **milieu minimum** est un mélange simple ne contenant pas de composés organiques autre qu'une source de carbone. Ainsi, un milieu minimum typique pourrait être une solution contenant des ions  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^+$ ,  $\text{NH}_4^+$ ,  $\text{Cl}^-$ ,  $\text{PO}_4^{3-}$  et  $\text{SO}_4^{2-}$  ainsi qu'une source de carbone qui pourrait être du glucose, du glycérol ou du lactate. Le choix de l'alimentation dépend de la bactérie concernée, certaines pouvant se contenter d'un tel milieu minimum (bactéries dites prototrophes), d'autres exigent pour leur croissance d'autres substances organiques que la source de carbone (bactéries auxotrophes).

Pour garder son volume constant, le chémostat dispose d'un système de vidange de débit équivalent à l'alimentation. La composition du débit de sortie (que nous appellerons également par la suite **dilution**) est celle de la solution contenue dans la cuve, elle se compose des constituants de l'alimentation non consommés (sucre(s), ions...), de bactéries et de produits synthétisés par ces dernières. Le principe du chémostat est résumé à la figure 4.1.

### Les acteurs

Parmi les éléments composant la solution du chémostat, nous nous intéressons aux populations "vivantes" : les phages et les bactéries.

**Les Bactéries** Nous avons déjà beaucoup évoqué la bactérie, principale héroïne de notre modèle. Voici venu le moment de donner plus de détails sur ses caractéristiques et son métabolisme.

On considère habituellement [62] que les bactéries forment un des cinq groupes de microorganismes (pour information les quatre autres sont : les algues, les protozoaires, les mycètes et les virus dont nous parlons juste après). Les microorganismes sont caractérisés par :

- leur organisation cellulaire (**procaryotes** ou **eucaryotes**)<sup>2</sup>,
- l'unicellularité,
- leur taille microscopique,
- leur potentiel métabolique et de reproduction,
- l'omniprésence et l'abondance.

Le terme de procaryote ou d'eucaryote désigne l'absence ou non de noyau au sein de la cellule, c'est-à-dire d'un endroit à l'intérieur de cette cellule fermé par une membrane et où est stockée l'information génétique. Ainsi les cellules procaryotes (sans noyau), dont font partie les bactéries, possèdent un seul chromosome libre à l'intérieur de la cellule au milieu de quelques organites spécialisées (l'organisation d'une cellule eucaryote est quant à elle beaucoup plus compliquée). Le diamètre moyen d'une bactérie est compris entre 0,5 et 1  $\mu\text{m}$ .

---

<sup>2</sup>Certains distinguent deux groupes de procaryotes :

- les "archéo" (anciennes bactéries),
- les bactéries.

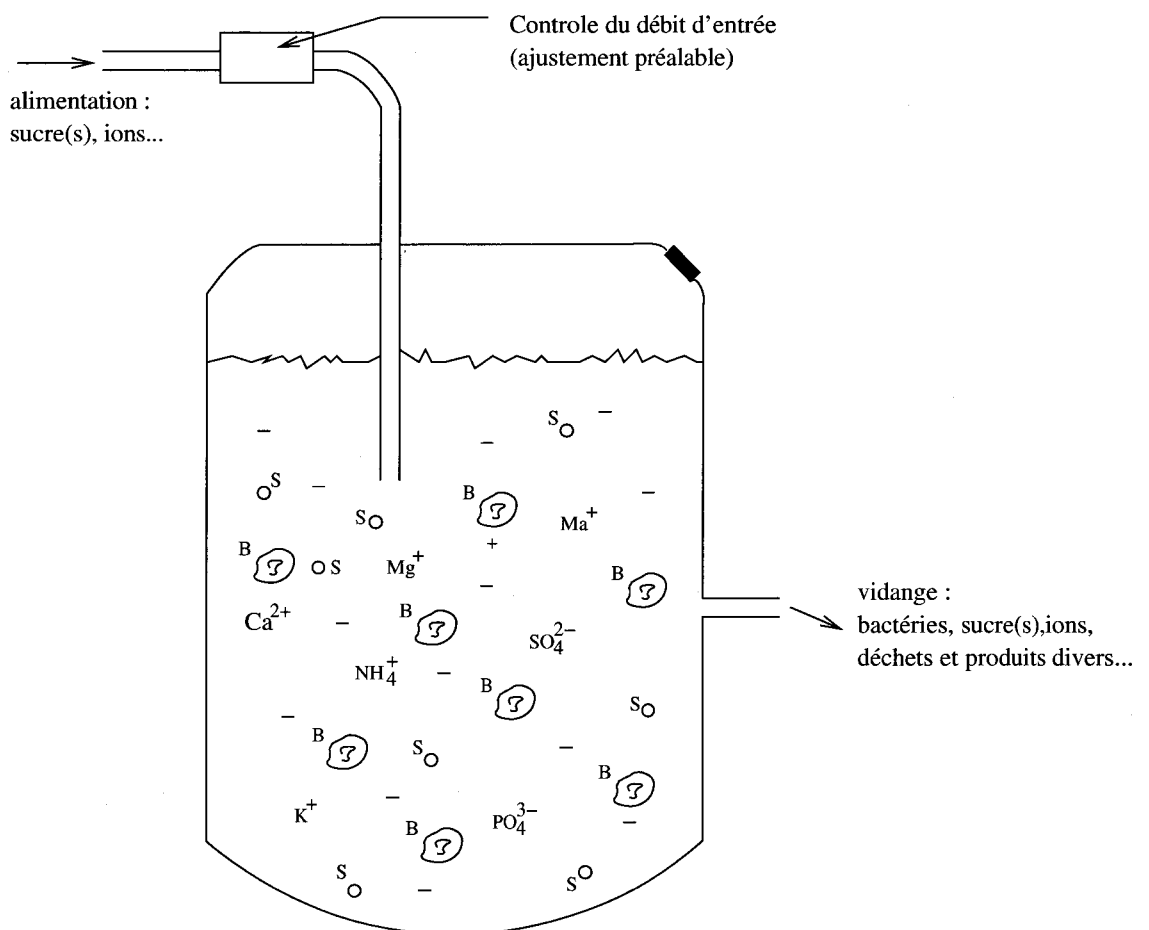


FIG. 4.1 – Schéma d'un chémostat alimenté par une solution contenant une ou plusieurs sources de carbone et des ions nécessaires à la croissance des bactéries présentes dans la solution. Un débit de vidange équivalent permet de garder le volume constant.

L'un des traits caractéristiques des bactéries est leur potentiel de reproduction incroyablement élevé. Le temps entre deux divisions cellulaires est de l'ordre de quelques dizaines de minutes. Pour donner une image de cette capacité de reproduction incroyable, si l'on considère la bactérie *Staphylococcus aureus* (masse de  $5.10^{-3}g$  et temps de génération<sup>3</sup> de 20 min.), il faudrait moins de deux jours à cette bactérie pour produire une culture de la masse de la terre ( $6.10^{27}g$ ), si bien sûr, elle pouvait disposer de suffisamment d'éléments nutritifs<sup>4</sup>.

**Les bactériophages** Les bactériophages ou de manière plus courte simplement phages sont les virus des bactéries. Leur présence dans une culture bactérienne n'est généralement pas volontaire, mais provient la plupart du temps, d'un apport accidentel de l'extérieur, par le biais de l'alimentation par exemple.

Les virus et en particulier les phages se distinguent des autres microorganismes par leur structure non cellulaire (acaryotes), ils sont en fait constitués d'une enveloppe protéique (la capsid) entourant une molécule d'acide nucléique porteuse de leur information génétique. Un bactériophage est constitué typiquement d'une tête icosaédrique attaché à une queue creuse entourée d'une gaine contractile et se terminant par une ou plusieurs fibres flexibles. La longueur totale de cette structure est comprise entre 0,1 et 0,3  $\mu m$ .

Les virus sont des microorganismes à la frontière du vivant. En effet, ils leur manquent deux aspects essentiels du vivant : pouvoir s'alimenter et se reproduire ; ils ont besoin d'autres organismes pour assurer ses fonctions. Dans le cas des phages, ce sont les bactéries qui jouent ce rôle. C'est au travers de ce processus que se joue l'interaction bactérie-phage. Nous allons maintenant voir comment le phage prend le contrôle de la bactérie qu'il infecte pour assurer sa réplication<sup>5</sup>.

**L'intrigue** Quand le mouvement brownien à l'intérieur de la solution du chémostat provoque la rencontre d'un phage et d'une cellule bactérienne, les fibres de la queue du phage balayent la surface de la cellule et accrochent le phage à des récepteurs insérés dans la paroi cellulaire (figure 4.2). Le nombre de tels sites d'accrochages dépend des conditions de croissance de la bactérie, mais est typiquement de l'ordre de 100 à 1000 par cellule. Une fois adsorbé à la surface de la bactérie, la queue se contracte, faisant pénétrer l'ADN viral à l'intérieur de la cellule. En détournant les ressources et la machinerie cellulaire de la bactérie, l'ADN nouvellement arrivé peut alors commencer à se développer et après une période

<sup>3</sup>Le temps de génération est le temps entre deux divisions cellulaires

<sup>4</sup>La masse de la terre est équivalente à celle de  $\frac{6.10^{27}}{5.10^{-3}} = 1,2.10^{30}$  bactéries. Considérant qu'à chaque génération le nombre de bactéries présentes dans la culture double, on a, au bout de  $n$  générations et à partir d'une seule bactérie,  $2^n$  bactéries. On veut avoir ici  $n$  tel que  $2^n > 1,2.10^{30}$ , d'où  $n > \log_2(1,2.10^{30}) = 99,92$  et donc  $n = 100$ . Le temps qu'il faut à la culture pour atteindre la masse de la terre est donc de  $100 \times 20min = 2000min = 33h20min$ .

<sup>5</sup>On parle dans ce cas de réplication pour distinguer avec la reproduction assurée de façon autonome par un organisme.

de latence de l'ordre de 30 minutes provoquer la "lyse" de la cellule (la cellule bactérienne explose), un grand nombre de copies du phage est alors libéré dans la solution. Ce nombre est typiquement de l'ordre de 100 (mais peut aller jusqu'à 1000 pour certaines espèces de phages) (figure 4.3).

Certaines cellules peuvent être résistantes à l'attaque virale, soit parce qu'elles n'ont pas de récepteurs spécifiques, soit parce qu'elles ont développées des parades : sécrétion d'une couche protectrice qui empêche l'adsorption des phages par exemple, ou encore un mini système immunitaire qui les protège des molécules d'ADN étranger : des enzymes viennent couper l'ADN envahisseur en certains sites spécifiques de la molécule. Les phages également disposent de contre-parades pour échapper à ces stratégies : ils peuvent par exemples essayer de se faire reconnaître comme de l'ADN de la cellule [55].

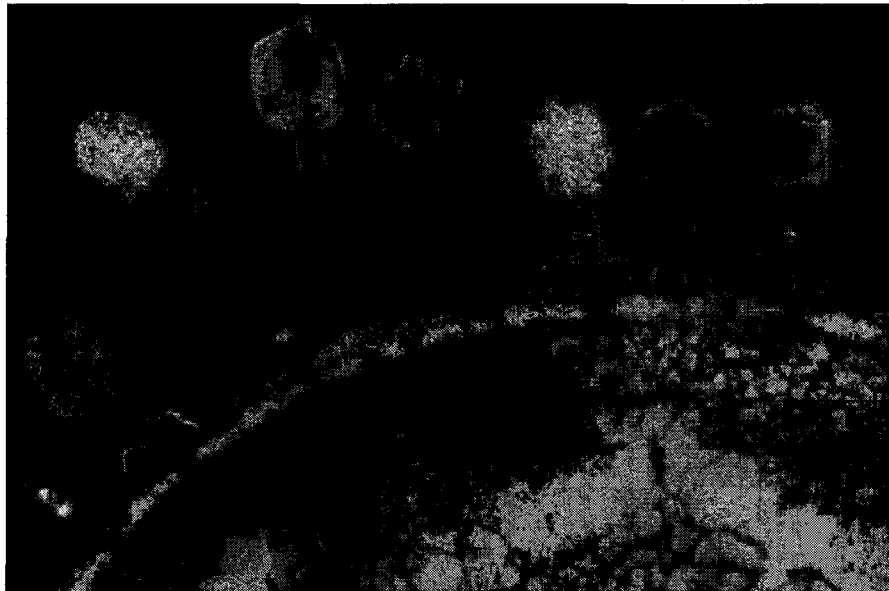


FIG. 4.2 – Phages accrochés sur une cellule bactérienne

Nous venons donc de spécifier dans cette section, les mécanismes biologiques présents dans le modèle. Nous allons maintenant présenter la structure générale de ce modèle, c'est-à-dire le type de représentation utilisée, justifiée par les travaux présentés dans les chapitres précédents.

#### 4.1.2 Structure du modèle

De la même manière que pour le réacteur bactérien évolutif de la section 3.3.2, on peut établir une liste des événements susceptibles de modifier l'état du système et les classer selon leur fréquence. Parmi les événements susceptibles de se produire à l'intérieur du chémostat, deux sont beaucoup plus rares que les autres :

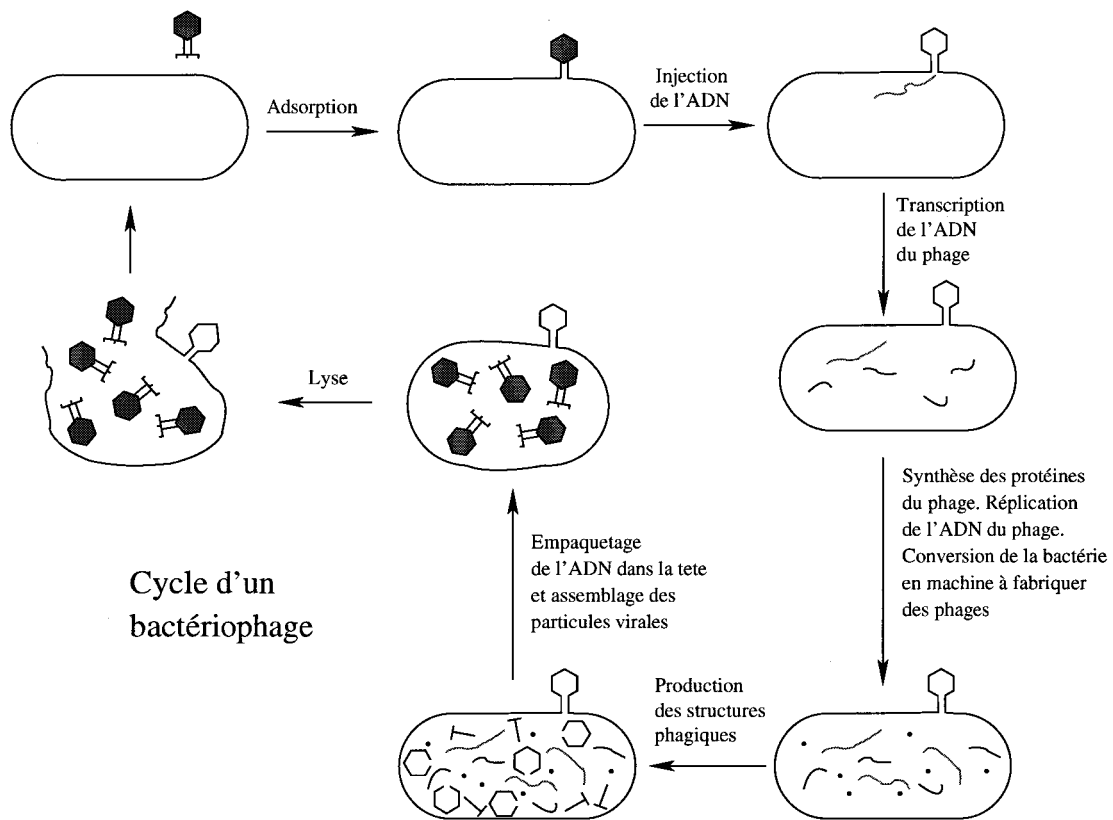


FIG. 4.3 – Cycle lytique d'un bactériophage (d'après [26]).

- une mutation qui donnerait naissance à une population de phages ou de bactéries possédant des caractéristiques nouvelles vis-à-vis de l'interaction bactéries-phages,
- la disparition complète d'une population de phages ou de bactéries ayant certaines caractéristiques particulières (voir tableau 4.1).

Evènements fréquents	Evènements rares
-division cellulaire quand la bactérie a absorbé tous les éléments nutritifs nécessaires -accrochage d'un phage sur une cellule bactérienne -infection d'une cellule bactérienne par un phage -arrivée de nutriment dans le chémostat -sortie d'un élément du chémostat -lyse d'une cellule infectée	-mutation (bactéries ou phages) donnant naissance à une population possédant des caractéristiques nouvelles vis-à-vis de l'interaction bactéries-phages -disparition d'une population

TAB. 4.1 – Fréquence des mécanismes biologiques dans un écosystème bactéries-phages-nutriment.

Selon la méthodologie établie à la section 3.3.2, on peut modéliser notre système par un SAM à base continue tel que défini au chapitre 2. On modélise donc les évènements fréquents par un système dynamique différentiel (comme discuté en 3.3.1, cette hypothèse est valable quand le nombre d'individus de la population est grand) que viennent modifier des évènements rares modélisés par une métadynamique discrète. Comme les mutations et disparitions dépendent des concentrations présentes dans le système, la métadynamique dépendra de l'état du système. D'autres part, l'existence de "caractéristiques" liées à une population suppose l'existence d'une structure de données "attachée" à chaque population.

Suivant notre promesse de développer cette section sans "rentrer dans les équations" afin de permettre une compréhension d'ensemble des mécanismes biologiques en présence et de la structure du modèle, nous ne rentrerons pas plus dans les détails et nous pouvons d'ores et déjà présenter la structure générale du modèle à la figure 4.4.

## 4.2 Description du modèle attaque-défense

Nous commençons par décrire les variables retenues pour modéliser le système. Nous décrivons ensuite les hypothèses qui nous permettent de modéliser les rela-

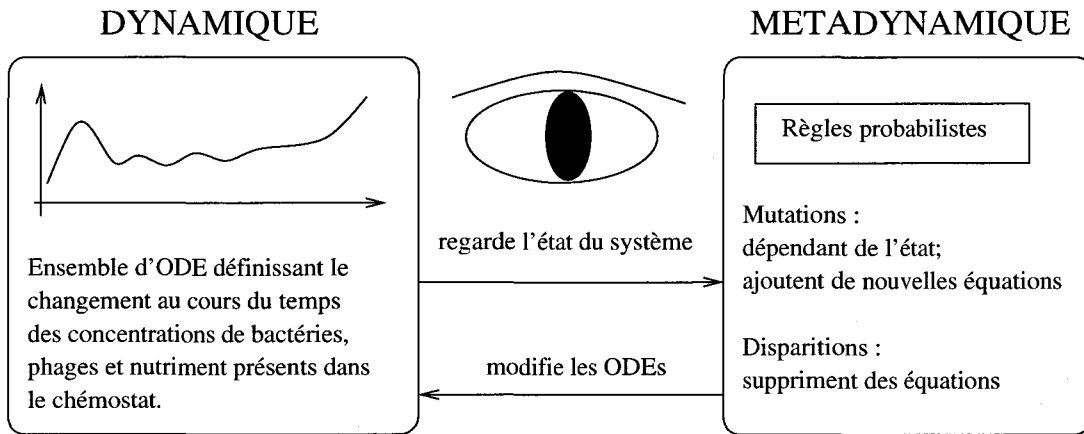


FIG. 4.4 – Structure générale du modèle attaque-défense.

tions entre les variables. Nous donnons finalement la description du modèle sous forme d'un SAM.

### 4.2.1 Composition du système

Nous présentons ici les variables choisies pour décrire le système, le choix de certaines variables ne pourra être pleinement perçu qu'après avoir pris connaissance de la section suivante. Nous avons placé cette section avant la description des hypothèses afin de faciliter leur compréhension par une vue d'ensemble des variables du système.

On suppose qu'il est suffisant pour décrire l'état de notre système, de considérer :

- une concentration de nutriment que nous notons  $S$  (substrat),
- $n_B$  concentrations de bactéries  $B_i$  auxquels sont attachées des chaînes binaires (CB)  $s_i^B$  décrivant les caractéristiques de cette population face à l'interaction bactéries-phages,
- $n_P$  concentrations de phages  $P_j$  et de bactéries  $I_{j_1}$ ,  $I_{j_2}$  et  $I_{j_3}$  détournées par ce même phage à trois stades d'infection. On attache également au phage  $P_j$  une CB  $s_j^P$  codant pour ses caractéristiques d'attaque des bactéries.

Ces concentrations sont exprimées en  $\mu g.l^{-1}$ .

### 4.2.2 Hypothèses

Nous présentons ici, pour chaque mécanisme biologique, les hypothèses que nous faisons et les équations qui en résultent. Nous donnons plus de détails pour les mécanismes qui ne sont pas classiques.

### Alimentation

On suppose qu'il n'y a qu'une seule source de nutriment et que l'alimentation à taux constant peut-être modélisée par l'équation différentielle suivante :

$$\frac{dS}{dt} \Big|_{\text{alimentation}} = \rho S,$$

où  $\rho$  est le débit du réacteur et  $\sigma$  le taux d'alimentation.

### Dilution

Le débit équivalent en sortie ou dilution peut-être modélisé par l'équation différentielle suivante :

$$\frac{dX}{dt} \Big|_{\text{dilution}} = -\rho X,$$

où  $X = S, B_i, P_j$  ou  $I_{j_k}$  et  $\rho$  le débit du réacteur.

### Croissance bactérienne

Comme discuté à la section 3.2.4 pour le modèle de Tsuchiya, on suppose que la croissance bactérienne suit une cinétique de Monod, c'est-à-dire est de la forme :

$$\frac{dB_i}{dt} \Big|_{\text{croissance}} = \frac{\nu S B_i}{\kappa + S},$$

où  $\kappa$  et  $\nu$  respectivement le terme de saturation et le coefficient de croissance de l'équation de Monod.

### Accrochage des phages

On considère que la formation d'un complexe "phage accroché-bactérie", résulte de la rencontre aléatoire d'un phage et d'une bactérie et peut donc être modélisé par une cinétique proportionnelle ne dépendant que d'une constante  $\alpha$  appelée constante d'adsorption théorique [55]. Sous cette hypothèse, on a :

$$\frac{dP_j}{dt} \Big|_{\text{accrochage}} = -P_j \left( \sum_{i=1}^{n_b} \alpha B_i + \sum_{j=1}^{n_p} \sum_{k=1}^3 \alpha I_{j_k} \right).$$

Les phages peuvent se fixer sur n'importe quelle bactérie, qu'elle soit déjà infectée ou non. L'accrochage du phage n'implique pas obligatoirement l'aboutissement du processus de lyse. Si la bactérie est déjà infectée, l'attaque échoue toujours, en revanche, dans le cas contraire, la probabilité de réussite de l'attaque dépend du phage et de la bactérie.



### Infection

Suivant une suggestion de LeFèvre [49], on suppose que la probabilité d'infection de la bactérie par un phage dépend d'un certain nombre de caractéristiques portées par le phage et la bactérie et que l'on peut traduire celles-ci par une chaîne binaire associée à l'individu<sup>6</sup>. Un bit de cette chaîne binaire peut représenter la présence ou non d'un récepteur par exemple ou encore l'acquisition ou non d'un mécanisme de résistance, etc. La probabilité d'infection dépend donc de la chaîne binaire traduisant les caractéristiques de défense de la bactérie et de la chaîne binaire traduisant les caractéristiques d'attaque du phage. On suppose qu'à chaque défense de la bactérie correspond une parade évolutive du phage. Chaque bit des deux chaînes est donc en correspondance et la fonction d'infection est une fonction croissante de la similitude entre les deux chaînes : si le phage possède toutes les parades aux défenses de la bactérie, l'attaque est maximale. Etant donné le caractère métaphorique de cette modélisation (on n'attache pas de mécanisme précis à chacun de ces bits), la fonction à choisir pour définir l'infection est donc relativement arbitraire. Nous avons choisi une fonction symétrique pour chaque bit et non linéaire pour essayer de traduire le fait qu'il existe généralement une grande différence d'efficacité entre un phage spécifique et un phage non-spécifique.

La probabilité d'infection  $\omega_{ij}$  de la bactérie  $i$  par le phage  $j$ , dépend de la ressemblance entre les chaînes d'attaque-défense  $s_i^b$  et  $s_j^p$  respectivement de la bactérie et du phage concernés :

$$\omega_{ij} = \left( 1 - \frac{d_H(s_i^b, s_j^p)}{n_c} \right)^2$$

avec  $d_H(s_i^b, s_j^p)$  la distance de Hamming (le nombre de bits différents) entre  $s_i^b$  et  $s_j^p$  et  $n_c$  la taille de la CB.

Sous ces hypothèses, l'équation différentielle des bactéries qui disparaissent par prédation est :

$$\frac{dB_i}{dt} \Big|_{\text{prédation}} = -B_i \sum_{j=1}^{n_p} \alpha \omega_{ij} P_j.$$

**Exemple 35** *Quelle est la probabilité d'infection de la bactérie  $B_1$  : [011100110101] par le phage  $P_1$  : [011000010101] ?*

*Les deux CB diffèrent de deux bits (le quatrième et le septième) comme le nombre de bits de la chaîne est de 12, la probabilité d'infection de la bactérie  $B_1$*

---

<sup>6</sup>Cette chaîne binaire sera le seul élément évolutif de notre bactérie. La génétique microbienne que nous modélisons est donc extrêmement simpliste et purement métaphorique. En fait, au niveau de généralité où nous nous plaçons, nous avons seulement besoin d'une structure de données simple, manipulable par la métadynamique évolutive. Notre chaîne binaire en est l'expression la plus aisée.

par le phage  $P_1$  est :

$$\begin{aligned}\omega_{1,1} &= \left(1 - \frac{2}{12}\right)^2 \\ &\approx 0,694.\end{aligned}$$

## Lyse

Le cycle lytique n'est pas instantané et se déroule en plusieurs étapes (voir la section 4.1.1 et la figure 4.3). On choisit de le modéliser en utilisant trois niveaux d'infection :  $I_{j_1}$ ,  $I_{j_2}$  et  $I_{j_3}$  (voir [55]).  $I_{j_k}$  représente le  $k^{\text{ième}}$  stade d'infection d'une bactérie par le phage  $j$ . On a donc les relations suivantes :

$$\frac{dI_{j_1}}{dt} \Big|_{infection} = P_j \sum_{i=0}^{n_p} \alpha \omega_{ij} B_i, \quad (4.1)$$

$$\frac{dI_{j_1}}{dt} \Big|_{incubation} = -3 \frac{I_{j_1}}{\tau}, \quad (4.2)$$

$$\frac{dI_{j_2}}{dt} \Big|_{incubation} = \frac{3}{\tau} (I_{j_1} - I_{j_2}), \quad (4.3)$$

$$\frac{dI_{j_3}}{dt} \Big|_{incubation} = \frac{3}{\tau} (I_{j_2} - I_{j_3}), \quad (4.4)$$

$$\frac{dP_j}{dt} \Big|_{lyse} = 3 \frac{\beta I_{j_3}}{\tau}, \quad (4.5)$$

avec  $\tau$  une constante de temps et  $\beta$  le nombre de copies de phages libérées.

## Mutation

Etant donné la structure que nous avons adopté pour ce modèle (voir section 4.1.2), nous ne nous intéressons pas à des mutations individuelles, mais à l'apparition d'un petit embryon de population macroscopique. Nous avons souligné à la section 3.3.3 que la modélisation de la formation d'un embryon de population macroscopique représentait un point délicat du passage d'une vision microscopique particulière à une vision différentielle macroscopique. Nous avons pu, à ce moment, tracer la distribution du temps d'apparition d'une telle population. Etant donné que l'on ne dispose pas, pour notre modèle de données "particulaires", on doit modéliser l'apparition des populations mutantes sur base de considérations biologiques.

Nous avons raisonné sur un petit intervalle de temps  $\Delta t$ . Sur ce petit intervalle de temps, nous considérons la probabilité  $P(B_i, t, k)$  que la population  $B_i$  donne naissance à  $k$  souches mutantes entre  $[t - \Delta t, t]$ . Pour une population  $B_i$  donnée, la probabilité doit décroître très vite avec  $k$ . Nous considérons la probabilité  $p_e$

qu'un petit groupe de bactéries mutantes de taille  $B_e$  prennent la place de bactéries de la population mère. Un tel raisonnement nous donne une forme binomiale pour  $P(B_i, t, k) = C_n^k p_e^k (1 - p_e)^{n-k}$  avec  $n = \frac{B_i(t)}{B_e}$ . Quand  $n$  est grand, on peut approximer cette distribution par une loi de Poisson :

$$p(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (4.6)$$

avec  $\lambda = \frac{B_i(t)}{B_e} p_e$ . On considère que l'apparition de toutes les populations<sup>7</sup> à une distance de Hamming de 1 (1 bit de différent) est équiprobable. L'apparition de la population mutante se traduit par la "transformation" d'un groupe de bactéries de la population mère de taille  $B_e$ , en un groupe avec de nouvelles caractéristiques. Si la population existe déjà, cette population mutante vient simplement s'ajouter à celle déjà existante. On emploie également le même mécanisme pour modéliser les mutations des phages.

**Exemple 36** *Supposons qu'à l'instant  $t$ , on ait deux populations bactériennes différentes et une population virale. Supposons également que la simulation utilise des chaînes binaires à trois bits (dans nos simulations, il y a beaucoup plus de variables et la taille des chaînes est plus grande).*

Nom	Type	Concentration ( $\mu\text{g.l}^{-1}$ )	CB
$B_1$	Bactéries	$1.72 \times 10^5$	[100]
$B_2$	Bactéries	$3.45 \times 10^5$	[001]
$P_1$	Phages	$1.24 \times 10^7$	[111]

On utilise la formule (4.6) pour calculer la probabilité  $p(k)$  qu'une population donnée donne naissance à  $k$  souches mutantes. Quand  $p(k)$  est trop faible on la considère comme nulle.

Nom	$p_e$	$B_e$ ou $P_e$	$p(k=0)$	$p(k=1)$	$p(k=2)$	$p(k=3)$
$B_1$	$10^{-5}$	100	0.983	0.0169	$1.454 \times 10^{-4}$	$8.336 \times 10^{-7}$
$B_2$	$10^{-5}$	100	0.966	0.0333	$5.75 \times 10^{-4}$	$6.612 \times 10^{-6}$
$P_1$	$10^{-4}$	10000	0.883	0.11	$6.8 \times 10^{-3}$	$2.807 \times 10^{-4}$

Maintenant que nous avons déterminé les distributions de probabilité, on peut décider de l'avenir de notre "simulation". Pour chaque population, on tire un nombre aléatoire dans  $[0, 1]$ . Les résultats de ce tirage sont représentés sur la figure 4.5. La seule population donnant naissance à une souche mutante est  $P_1$ . Nous devons donc maintenant choisir le bit qui va muter, par exemple le premier. Le nouveau système a donc maintenant quatre équations supplémentaires (rattaché à la CB [011]) :  $P_2, I_{21}, I_{22}$  et  $I_{23}$  dont les concentrations sont respectivement  $P_e, 0, 0$  et  $0$ . On doit également retirer  $P_e$  à  $P_1, P_1(t + \Delta t) \leftarrow P_1(t + \Delta t) - P_e$ .

<sup>7</sup>Pour être plus rigoureux du point de vue biologique, insistons sur le fait que ce que nous appelons deux populations différentes représente en fait deux sous-espèces d'un même "type sauvage" ne différant que par ces chaînes attaque-défense et non-nécessairement deux espèces de bactéries classées taxonomiquement comme différentes.

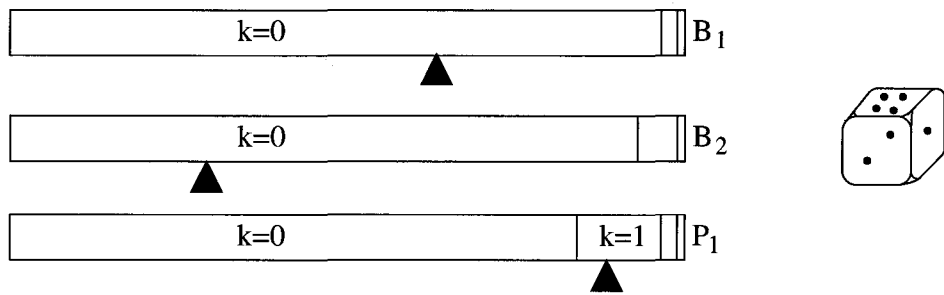


FIG. 4.5 – Modélisation de l’apparition des populations mutantes : tirage des mutations après avoir calculé les distributions de probabilité

**Remarque 37** Afin d’éviter toute confusion avec les  $P_i$ , concentrations des populations phagiques, on note  $\mu_B$  le coefficient  $p_e$  lié à la loi de Poisson de mutation des bactéries et  $\mu_P$  celui lié à la loi de mutation des phages.

### Disparition

On fait la même hypothèse que celle faite pour la transition 2 vers 3 du modèle évolutif de la section 3.3.3, c’est-à-dire, qu’il existe un seuil au dessous duquel la disparition de la population est considéré comme certaine. Si la population passe en dessous de ce seuil, elle est simplement supprimée.

### 4.2.3 Définition du MAD sous forme de SAM

Maintenant que nous avons spécifié les hypothèses et les composants de notre modèle, nous pouvons passer à la définition de notre modèle sous forme d’un SAM, en commençant par le niveau dynamique, constitué d’un ensemble de systèmes dynamiques différentiels continus.

#### Equations de Mosekilde et niveau dynamique

Pour une taille donnée de CB, il existe un nombre fini de systèmes dynamiques possibles comme base du SAM. En effet, le nombre de populations différentes possibles de bactéries est égale au nombre de parties de l’ensemble des CB de taille  $n_c$ . Il y a donc  $2^{2^{n_c}}$  populations possibles de bactéries. Par le même raisonnement, on déduit le nombre possible de populations de phages et donc le nombre total d’espaces d’état possibles :  $2^{2^{n_c}} \times 2^{2^{n_c}} = 2^{2^{n_c}+1}$ . Si on prend, par exemple,  $n_c = 10$ , on obtient  $2^{1025} \approx 3,6 \cdot 10^{308}$  espaces d’état et donc SD possibles. On comprend alors la nécessité de décrire cet ensemble en intention et non en extension<sup>8</sup>.

<sup>8</sup>Si, au lieu de représenter les caractéristiques évolutives par une chaîne binaire, nous les avons modélisées par des nombres réels évoluant au moyen d’opérateurs génétiques spécialisés tels que ceux définis en “évolution différentielle” [?], nous aurions un ensemble de systèmes dynamiques ayant la puissance du continu. Notons que cet ensemble pourrait, comme l’a montré

Comme décrit à la section 4.1.2, le niveau dynamique est composé d'un ensemble de systèmes d'ODE, on a donc  $T_0 = \mathbb{R}$ . A l'écosystème

$$\{S, \{B_i\}_{0 \leq i \leq n_b}, \{P_j, I_{j_1}, I_{j_2}, I_{j_3}\}_{0 \leq j \leq n_p}\},$$

on associe l'espace d'état  $(\mathbb{R}^+)^{1+n_b+4n_p}$ . La fonction de transition quant à elle est donnée par l'intégration du système d'équations différentielles suivant :

$$\left\{ \begin{array}{l} \frac{dB_i}{dt} = \underbrace{\frac{\nu S B_i}{\kappa + S}}_{\text{croissance}} - \underbrace{B_i \sum_{j=1}^{n_p} \alpha \omega_{ij} P_j}_{\text{infection}} - \underbrace{\rho B_i}_{\text{dilution}} \\ \frac{dI_{j_1}}{dt} = \underbrace{P_j \sum_{i=1}^{n_b} \alpha \omega_{ij} B_i}_{\text{infection}} - \underbrace{3 \frac{I_{j_1}}{\tau}}_{\text{incubation}} - \underbrace{\rho I_{j_1}}_{\text{dilution}} \\ \frac{dI_{j_2}}{dt} = \underbrace{\frac{3}{\tau} (I_{j_1} - I_{j_2})}_{\text{incubation}} - \underbrace{\rho I_{j_2}}_{\text{dilution}} \\ \frac{dI_{j_3}}{dt} = \underbrace{\frac{3}{\tau} (I_{j_2} - I_{j_3})}_{\text{incubation}} - \underbrace{\rho I_{j_3}}_{\text{dilution}} \\ \frac{dP_j}{dt} = \underbrace{3 \frac{\beta I_{j_3}}{\tau}}_{\text{lyse}} - \underbrace{P_j \left( \sum_{i=1}^{n_b} \alpha B_i + \sum_{j=1}^{n_p} \sum_{k=1}^3 \alpha I_{j_k} \right)}_{\text{accrochage}} - \underbrace{\rho P_j}_{\text{dilu.}} \\ \frac{dS}{dt} = \rho \left( \underbrace{\sigma}_{\text{alimentation}} - \underbrace{S}_{\text{dilution}} \right) - \underbrace{\sum_{i=1}^{n_b} \frac{\nu \gamma S B_i}{\kappa + S}}_{\text{consommation}} \end{array} \right.$$

avec :

- .  $\rho$  le coefficient de dilution ( $\rho = 0.0045 \text{ min}^{-1}$ ),
- .  $\kappa, \nu$  respectivement le terme de saturation et le coefficient de croissance de l'équation de Monod ( $\kappa = 10 \mu\text{g.ml}^{-1}$ ,  $\nu = 0.024 \text{ min}^{-1}$ ),
- .  $\alpha$  la constante d'adsorption théorique dépendant de la taille des phages et des bactéries ( $\alpha = 10^{-9} \text{ ml. min}^{-1}$ ),
- .  $\tau$  une constante de temps ( $\tau = 30 \text{ min}$ ),
- .  $\beta$  le nombre de copies de phages libérées lors de l'éclatement de la cellule bactérienne infectée ( $\beta = 100$ ),
- .  $\sigma$  le taux d'alimentation ( $\sigma = 10 \mu\text{g.ml}^{-1}$ ),
- .  $\gamma$  la quantité de nutriment consommée pour chaque division cellulaire ( $\gamma = 0.01 \text{ ng}$ ),
- .  $\omega_{ij}$  la probabilité d'infection de la bactérie  $i$  par le phage  $j$ ,  $\omega_{ij}$  dépend de la ressemblance entre les chaînes d'AD  $s_i^b$  et  $s_j^p$  respectivement de la bactérie et

précédemment notre exemple du tracé du Lapin de Douady, être décrit par le formalisme général du chapitre 2.

du phage concernés.

$$\omega_{ij} = \left( 1 - \frac{d_H(s_i^b, s_j^p)}{n_c} \right)^2$$

avec  $d_H(s_i^b, s_j^p)$  la distance de Hamming (le nombre de bits différents) entre  $s_i^b$  et  $s_j^p$ .

**Remarque 38** *Les paramètres  $\rho$  et  $\sigma$  sont les deux seuls paramètres physiques et non biologiques. De plus, ils sont ici données à titre indicatif, mais peuvent être modifiés par la personne qui contrôle le chémostat. Nous donnerons à la propriété 41, la relation qui doit exister entre ces deux variables pour qu'une culture bactérienne puisse se développer (condition d'ailleurs respectée pour les deux valeurs données ici).*

**Remarque 39** *La partie dynamique de notre modèle est très proche du modèle présenté dans [55]. Il faut noter trois différences essentielles :*

- $n_b$  et  $n_p$  ne sont pas forcément égaux (et même différents la plupart du temps) ;
- nous ne considérons pas de source extérieure de phages ;
- $\omega_{ij}$  est fixé par nos chaînes d'AD.

### Niveau métadynamique

Nous avons défini (section 4.2.2) l'apparition macroscopique d'une population mutante comme un évènement arrivant sur un intervalle  $[t - \Delta t, t]$ , avec  $\Delta t$  petit. Bien que, logiquement, la disparition d'une population n'ait aucune raison d'être également définie seulement sur les  $t_0 + n\Delta t, n \in \mathbb{N}$ . Nous choisissons, afin de faciliter l'implémentation du modèle :

$$T_1 = t_0 + \mathbb{N}\Delta t.$$

Tous les  $\Delta t$  (en pratique un multiple du pas d'intégration des intégrations différentielles), le système est susceptible de commuter. Cette commutation dépend de l'état du système et des coefficients  $\mu_B$  et  $\mu_P$  respectivement coefficient de la loi de Poisson de la bactérie et du phage. Nous ne donnons pas de valeurs pour  $\mu_B$  et  $\mu_P$  ici car notre étude consistera justement à regarder l'influence de ceux-ci sur le comportement du système.

Au niveau du changement d'espace d'état, pour une commutation incluant un seul évènement, on a quatre changements possibles :

- Apparition d'une nouvelle souche bactérienne : une variable  $B_{n_b+1}$  est ajoutée, la dimension du système augmente de un.
- Apparition d'une nouvelle souche phagique : quatre nouvelles variables sont ajoutées :  $P_{n_p+1}$ ,  $I_{n_p+1_1}$ ,  $I_{n_p+1_2}$  et  $I_{n_p+1_3}$ . On met ainsi en place la structure de modélisation de l'infection du phage arrivant. La dimension du système augmente de quatre.

- Disparition d'une souche de bactéries : la variable concernée est supprimée, la dimension baisse de un.
- Disparition d'une souche de phages : les variables du phage concerné et de ses trois états d'infection est supprimée, la dimension du système baisse de quatre.

Les différents espaces d'état possibles issus d'une commutation sont représentés à la figure 4.6.

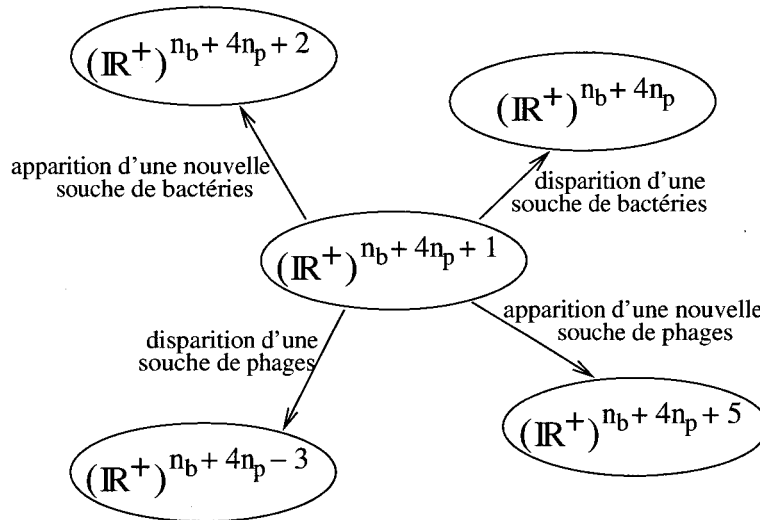


FIG. 4.6 – Les quatre espaces d'état atteignables par une commutation métadynamique.

De par la définition du SAM (principe hiérarchique de Weiss), les événements métadynamiques sont supposés rares et sont donc généralement uniques à chaque instant de  $T_1$ . Si plusieurs événements arrivent simultanément, il faut composer les changements d'état possible (par exemple, la disparition d'un phage associée à l'apparition d'une bactérie, diminue la dimension du système de  $4-1=3$ ; l'apparition d'un phage et de deux bactéries augmente sa dimension de  $4 + 2 \times 1 = 6$ , etc.).

La connaissance de la nouvelle fonction de transition est déduite de la composition de la nouvelle population, par l'intégration du système différentiel associé (dépendant en particulier des CB associées aux populations).

### 4.3 Détails sur l'implémentation

Dans cette partie sur l'implémentation informatique du MAD, nous ne donnons des détails que sur le moteur du logiciel développé et non sur ce qui a pu être également fait autour (présenté dans la présentation générale du logiciel en annexe ??).

### 4.3.1 Le choix de l'environnement

L'un des choix fait lors du début de ce travail de thèse a été le choix de l'environnement et du langage de développement pour la partie simulation du travail.

En ce qui concerne le choix du système d'exploitation, nous nous sommes tourné vers Linux et le logiciel libre, pour des raisons aussi bien scientifiques, qu'économiques et idéologiques. D'un point de vue scientifique, il existe une grande variété de logiciels (parfois hérités du monde Unix) à caractère scientifique que ce soit pour la partie graphique (Gnuplot et Xfig), que pour la partie calcul (SciLab) ou la programmation. D'un point de vue économique, le monde du libre<sup>9</sup> est pratiquement entièrement gratuit.

En ce qui concerne le choix du langage, nous avons opté pour le langage C pour plusieurs raisons :

- de part l'originalité de son approche, nous avons besoin d'une grande souplesse dans l'écriture de notre programme, et en particulier de pouvoir écrire toutes les fonctions possibles, sans autres limites que celles de l'habileté du programmeur et des performances de la machine,
- de part l'intensité de calcul de certaines simulations (intégration d'un système différentiel de dimension variable comprenant plusieurs milliers d'équations), nous avons besoin de rapidité, point souvent reproché à des logiciels comme Matlab,
- il s'agit d'un langage très universel, donc très documenté et avec un très bon interfaçage vers d'autres plates formes de simulation (20sim par exemple),
- c'est le langage sur lequel est basé Linux.

### 4.3.2 Les structures de données

Avant de parler de l'aspect des structures de données variables, commençons par remarquer que la structure de chaîne binaire a une représentation informatique très simple, puisque d'un point de vue informatique toute structure de données est elle même une chaîne binaire ! Le plus simple est de prendre un ou plusieurs entiers pour coder cette chaîne et de retrouver les bits quand le besoin est par des décalages de bits (opérations élémentaires) ou des opérations arithmétiques simples.

Nous avons vu dans la définition du modèle, que l'espace d'état varie au cours

---

<sup>9</sup>L'expression "Logiciel libre" [...] fait référence à quatre types de liberté pour l'utilisateur du logiciel :

- La liberté d'exécuter le programme, pour tous les usages (liberté 0).
  - La liberté d'étudier le fonctionnement du programme, et de l'adapter à vos besoins (liberté 1). Pour ceci l'accès au code source est une condition requise.
  - La liberté de redistribuer des copies, donc d'aider votre voisin (liberté 2).
  - La liberté d'améliorer le programme et de publier vos améliorations, pour en faire profiter toute la communauté (liberté 3). Pour ceci l'accès au code source est une condition requise.
- (extrait du site du projet GNU accessible à l'adresse <http://www.gnu.org/philosophy/freesw.fr.html>).



du temps et peut prendre un grand nombre de formes différentes au cours du temps. Il est donc impossible de décrire tous les espaces d'états possibles que le système pourrait atteindre durant la simulation. Même si cette structure varie au cours du temps, elle garde une structure semblable tout au long de la simulation :

- la concentration du nutriment  $S$ ,
- des variétés de bactéries  $B_i (i = 1 \dots n_B)$  avec une chaîne binaire  $s_i^B$  attachée à chacune d'elle,
- des variétés de phages  $P_j (j = 1 \dots n_P)$  avec une chaîne binaire également attachée et des bactéries infectées en train de produire des phages les  $I_{j_1}$ ,  $I_{j_2}$  et  $I_{j_3}$  selon le stade d'infection.

Nous avons choisi de représenter les éléments dont le nombre varie au cours de la simulation par une structure de liste chaînée. Une liste chaînée est une chaîne d'adresses mémoire avec sur chaque anneau (ou nœud) une variable attachée (par exemple une valeur numérique).

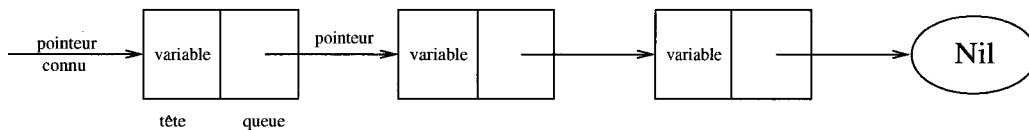


FIG. 4.7 – Structure de liste chaînée

Si l'on prend l'exemple de la liste chaînée associée à  $B_i$ , le programme connaît l'adresse mémoire où se trouve l'effectif de la dernière population bactérienne créée. A côté (en mémoire) de cette valeur est "collée" l'adresse de l'endroit (toujours en mémoire) où se trouve l'effectif de l'avant dernière population créée (et toujours en vie) et ainsi de suite jusqu'à la plus ancienne population encore en vie qui porte l'adresse "nulle part" (également appelé pointeur nul<sup>10</sup> ou NIL). Le programme sait alors qu'il s'agit de la dernière population et arrête son exploration.

On a deux groupes de chaîne, partageant leur indices respectifs, les chaînes liées aux bactéries (chaînes des effectifs représentées par des nombres à virgules flottantes de précision double et chaîne des chaînes binaires des bactéries) et les chaînes liées aux phages (effectif, stade d'infection 1, 2 et 3 et chaînes binaires associées). Aucun programme générique n'implémentant ces structures, nous fûmes forcés de développer notre logiciel "ab nihilo".

### 4.3.3 La partie dynamique

Comme expliqué à la section 3.4, l'implémentation de la partie dynamique correspond dans le cas où celle-ci est continue à l'intégration d'un système d'équations différentielles. Pour cela, de nombreuses méthodes existent (voir [38]) dont les plus connues sont sûrement les méthodes d'Euler et de Runge-Kutta. Le principe de

<sup>10</sup>parce qu'il porte par convention l'adresse mémoire 0

ces méthodes est le suivant : remplacer l'intégration de l'équation différentielle  $\dot{x} = f(x)$  par celle d'équations de récurrence approximantes. Ainsi, après avoir choisi un pas d'intégration (un intervalle de temps), on procède au calcul de :

$$x(t+h) = x(t) + kh$$

avec le calcul du coefficient  $k$  variant selon la méthode utilisée. Ainsi pour la méthode de Runge-Kutta que nous avons utilisé dans nos simulations,

$$k = \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

avec

$$\begin{aligned} k_1 &= f(x), \\ k_2 &= f\left(x + \frac{h}{2}k_1\right), \\ k_3 &= f\left(x + \frac{h}{2}k_2\right), \\ k_4 &= f(x + kh). \end{aligned}$$

La seule petite originalité de notre implémentation par rapport à une intégration classique, est l'utilisation de fonction d'intégration œuvrant sur des listes (fonctions prenant comme variables le pointeur de début de liste) et non sur des flottants.

#### 4.3.4 La partie métadynamique

La partie métadynamique se réduit à une série de test et de tirages et ne présente pas de difficultés particulières. Si l'espace d'état change, le nouvel espace d'état est obtenu par des opérations classiques sur les listes : suppression ou ajout d'éléments.

On rappelle que l'on a choisi, pour des raisons de facilité de programmation  $T_0 = t_0 + N\Delta t$  avec  $\Delta t = n_a h$ ,  $h$  le pas d'intégration du système différentiel et  $n_a$  l'accélérateur dynamique, qui permet comme son nom l'indique d'accélérer la simulation en réduisant le nombre d'appel à la règle métadynamique. Ainsi, on peut approcher le SAM par un autre SAM avec moins d'appel à la règle métadynamique, mais avec un taux de mutation moyen équivalent entre deux commutations (on suppose qu'une population faible le restera sur un intervalle relativement petit et donc que cela ne perturbe pas l'élimination des populations faibles).

### 4.4 Analyse du modèle

Nous développons dans cette section, quelques points théoriques du MAD. Nous commençons par quelques résultats théoriques sur le modèle, puis nous donnons des définitions utiles à son analyse.

### 4.4.1 Aspects globaux

Nous allons dans cette partie dégager des aspects globaux du système correspondant à des invariants des systèmes dynamiques composant le SAM.

#### Positivité des variables d'états

Nous avons défini notre modèle sur  $(\mathbb{R}^+)^{1+n_b+4n_p}$ , il est en effet logique que des valeurs de concentration soient positives. Nous montrons ici que cette propriété, de par la structure des équations, est vraie au cours du temps.

**Propriété 40** *Du moment que les conditions initiales sont positives les variables d'états le restent.*

#### Preuve.

Il suffit de montrer dans un premier temps que sur la face positive des hyperplans  $x_i = 0$ , on a  $\dot{x}_i \geq 0$

- sur  $B_i = 0$ , on a  $\dot{B}_i = 0$
- sur  $I_{j_1} = 0$ , on a  $\dot{I}_{j_1} = P_j \sum \alpha \omega_{ij} B_i \geq 0$
- sur  $I_{j_2} = 0$ , on a  $\dot{I}_{j_2} = \frac{3I_{j_1}}{\tau} \geq 0$
- sur  $I_{j_3} = 0$ , on a  $\dot{I}_{j_3} = \frac{3I_{j_2}}{\tau} \geq 0$
- sur  $P_j = 0$ , on a  $\dot{P}_j = \frac{3\beta_j I_{j_3}}{\tau} \geq 0$
- sur  $S = 0$ , on a  $\dot{S} = \rho\sigma > 0$

Et comme la règle métadynamique (mutation et extinction) conserve ces invariants, on en conclut que du moment que les CI sont positives les variables d'états le restent.

■

Désormais, on supposera  $B_i, P_j, I_{j_1}, I_{j_2}, I_{j_3}, S$  toujours positifs.

#### Aspect métadynamique du “wash out”

Si le débit  $\rho$  est trop élevé, les populations bactériennes et phagiques disparaissent et laissent place à une solution ne contenant que du sucre. D'un point de vue métadynamique, cela se traduit par la propriété suivante :

**Propriété 41** *Si  $\rho > \frac{\nu\sigma}{\kappa+\sigma}$ , le MAD est en situation de “washout” : quelle que soient les conditions initiales, le MAD converge vers le méta-état  $(\mathbb{R}, \dot{S} = \rho(\sigma - S))$ .*

#### Preuve.

On a l'inégalité différentielle suivante :  $\dot{S} < \rho(\sigma - S)$ ,

donc  $\forall S_0, \forall \epsilon_1, \exists t_1$  tel que  $\forall t \geq t_1, S(t) < \sigma + \epsilon_1$

(et ceci quelquesoit les commutations métadynamiques ayant lieu sur l'intervalle, puisque celles-ci n'altèrent ni l'inégalité différentielle, ni la valeur de  $S$ ).

$\frac{\nu x}{\kappa+x}$  est croissante si  $x \geq 0$ , car  $\left(\frac{\nu x}{\kappa+x}\right)' = \frac{\nu x}{\kappa+x^2}$   
d'où  $\forall \epsilon_1 > 0, \exists t_1 > 0$  tel que  $\forall t \geq t_1, \frac{\nu S}{\kappa+S} < \frac{\nu(\sigma+\epsilon_1)}{\kappa+(\sigma+\epsilon_1)} = \frac{\nu\sigma}{\kappa+\sigma} + \frac{\epsilon_1}{\kappa+\sigma+\epsilon_1} \left(\frac{\nu\kappa}{\kappa+\sigma}\right)$ .

On choisit  $\epsilon_1$  pour que  $\frac{\epsilon_1}{\kappa+\sigma+\epsilon_1} \left(\frac{\nu\kappa}{\kappa+\sigma}\right) < k = \left(\rho - \frac{\nu\sigma}{\kappa+\sigma}\right)/2$

donc,  $\forall S_0, \exists t_1$  tel que  $\forall t > t_1, \frac{\nu S}{\kappa+S} < \rho - k$ .

En reportant cela dans l'équation des bactéries,

$$\text{on a } \dot{B}_i = \left( \frac{\nu S}{\kappa+S} - \sum_{j=1}^{n_p} \alpha \omega_{ij} P_j - \rho \right) B_i < -k B_i.$$

$B_i$  strictement décroissante sur  $]0, +\infty[$  et positive,

donc  $B_i$  tend vers 0.

Et de la même façon,

$$\sum \dot{B}_i + \sum \dot{I}_{j_1} = \left( \frac{\nu S}{\kappa+S} - \sum_{j=1}^{n_p} \alpha \omega_{ij} P_j - \rho \right) \sum B_i - (\rho + 3/\tau) \sum I_{j_1} < -k \sum B_i.$$

Comme  $\sum \dot{B}_i + \sum \dot{I}_{j_1}$  strictement décroissante sur  $]0, +\infty[$  et positive,

donc  $\sum B_i + \sum I_{j_1}$  tend vers 0, donc  $I_{j_1}$  aussi.

On calcule de même  $\sum \dot{B}_i + \sum \sum \dot{I}_{j_k}$  et on trouve  $I_{j_k}$  tend vers 0.

$$\text{On calcule } \beta \left( \sum_i \dot{B}_i + \sum_j \dot{I}_{j_k} \right) + \sum_j \dot{P}_j = \left( \frac{\nu S}{\kappa+S} - \rho \right) \beta \sum_i B_i < -k \beta \sum_i B_i.$$

Toujours pour les mêmes raisons  $P_j$  tend vers 0.

Donc les  $B_i$  et les  $P_j$  tendent vers 0 et au bout d'un moment les populations disparaissent nécessairement en raison des seuils du modèle, quand toutes les populations ont disparu, on atteint le méta-état  $(\mathbb{R}, \dot{S} = \rho(\sigma - S))$ .

■

**Remarque 42** On peut ainsi vérifier que les valeurs numériques choisies pour  $\sigma$  et  $\rho$  sont bien en dehors de cette plage :

$$\frac{\nu\sigma}{\kappa+\sigma} = \frac{0,024 \times 10}{10+10} = 0,012 \text{min}^{-1},$$

or  $\rho = 0,0045 \text{min}^{-1}$ , donc  $\rho < \frac{\nu\sigma}{\kappa+\sigma}$  et le système n'est pas en situation de washout.

### Convergence de la concentration massique pondérée du système en régime permanent

Notre modèle impose tout d'abord quelques modifications (voir remarque 39) légères au modèle bactéries-phages non-évolutif de Mosekilde. Il complexifie alors grandement celui-ci en lui ajoutant la métadynamique. Or les études réalisées sur ce modèle montre qu'il est hyperchaotique, les variations des concentrations au sein du système peuvent donc être extrêmement compliquées lors d'une simula-

tion (comme des oscillations de type hyperchaotique<sup>11</sup>), la concentration massique totale pondérée se stabilise dans une certaine plage de valeur.

Le calcul de cette concentration massique totale pondérée<sup>12</sup> tient compte des transformations que subissent les éléments du système. Ainsi, il faut  $\gamma$  microgrammes de nutriment pour former une bactérie et une bactérie infectée donnera naissance à  $\beta$  phages (figure 4.8). Si on compte tout en "unité nutriment", une bactérie vaut donc " $\gamma$  nutriment" et un phage "vaut"  $\frac{1}{\beta}$  bactérie donc  $\frac{\gamma}{\beta}$  "unité nutriment".

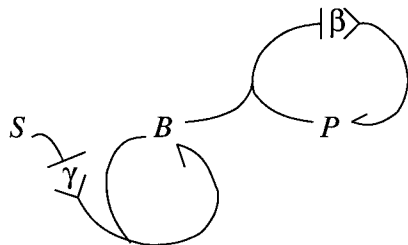


FIG. 4.8 – Chaîne de transformations au sein du chémostat.

Le fait que le débit d'alimentation soit égal au débit de sortie nous permet de déduire la propriété suivante :

**Propriété 43** *Les concentrations du MAD sont bornées et la concentration massique totale pondérée*

$$T = S + \gamma \left( \sum_i B_i + \sum_j \sum_{k=1}^3 I_{j_k} \right) + \frac{\gamma}{\beta} \sum_j P_j$$

converge quelle que soit la dynamique courante vers une zone attractrice

$$\mathcal{A} = [\sqrt{a^2 + 2a\sigma} - a, \sigma],$$

avec  $a = \frac{2\rho\gamma}{\alpha}$ .

<sup>11</sup>Les systèmes hyperchaotiques présentent les caractéristiques "classiques" du chaos, c'est-à-dire des trajectoires bornées et apériodiques et une sensibilité aux conditions initiales. Cette sensibilité aux conditions initiales a pour origine l'amplification exponentielle au cours du temps de petites différences de l'état. Contrairement au chaos "simple", où cette amplification se fait dans une seule dimension de l'espace, elle a lieu, dans le cas de l'hyperchaos, dans plusieurs dimensions. Les directions de l'espace en contraction ou en expansion sont données par les exposants de Lyapunov, caractérisant la vitesse à laquelle deux points proches du système se rapprochent ou se séparent localement. Dans le cas de l'hyperchaos, le nombre d'exposant de Lyapunov positifs (directions en expansion) est strictement supérieur à 1. On parle "d'oscillations de type hyperchaotique" car l'évolution du système est sûrement dans certains cas encore plus compliqué. En effet, aux effets dynamiques -pouvant présenter de l'hyperchaos- s'ajoutent les effets de la métadynamique qui change la structure du système.

<sup>12</sup>Nous entendons par concentration massique totale pondérée une combinaison linéaire des différentes concentrations massiques (les variables) du système.

**Preuve.**

On a,

$$\dot{T} = \dot{S} + \gamma \left( \sum_i \dot{B}_i + \sum_j \sum_{k=1}^3 \dot{I}_{jk} \right) + \frac{\gamma}{\beta} \sum_j \dot{P}_j \quad (4.7)$$

$$= \rho(\sigma - T) - \frac{\alpha\gamma}{\beta} \left( \sum_i B_i + \sum_j \sum_{k=1}^3 I_{jk} \right) \sum_j P_j \quad (4.8)$$

or par définition de  $T$ ,

$$0 \leq \gamma \left( \sum_i B_i + \sum_j \sum_{k=1}^3 I_{jk} \right) \leq T,$$

et

$$0 \leq \frac{\gamma}{\beta} \sum_j P_j \leq \gamma \left( \sum_i B_i + \sum_j \sum_{k=1}^3 I_{jk} \right) + \frac{\gamma}{\beta} \sum_j P_j \leq T,$$

d'où

$$0 \leq \frac{\gamma^2}{\beta} \left( \sum_i B_i + \sum_j \sum_{k=1}^3 I_{jk} \right) \sum_j P_j \leq \frac{T^2}{4}$$

et on a donc l'inégalité différentielle suivante :

$$\rho(\sigma - T) - \frac{\alpha}{4\beta} T^2 \leq \dot{T} \leq \rho(\sigma - T).$$

Par suite,  $\dot{T} < 0$  quand  $T > \sigma$  et  $\dot{T} > 0$  quand  $T < \sqrt{a^2 + 2a\sigma} - a$  avec  $a = \frac{2\rho\gamma}{\gamma}$  et donc  $T$  converge asymptotiquement vers  $\mathcal{A} = [\sqrt{a^2 + 2a\sigma} - a, \sigma]$  (voir figure 4.9).

■

**Remarque 44** L'égalité 4.8 traduit seulement que la variation de la concentration massique totale dans le réacteur est due :

- à l'alimentation :  $\rho\sigma$ ,
- à la vidange :  $-\rho T$

- aux phages qui se fixent sur les bactéries :  $-\frac{\alpha\gamma}{\beta} \left( \sum_i B_i + \sum_j \sum_{k=1}^3 I_{jk} \right) \sum_j P_j$ .

Tout se passe du point de vue du bilan, comme si le phage se "suicidait" en donnant l'ordre à la bactérie de se transformer en  $\beta$  phages.

On vient donc de démontrer que quelle que soit la configuration, le système dynamique évolue vers une zone attractrice  $\mathcal{A}$ , dépendant des paramètres du

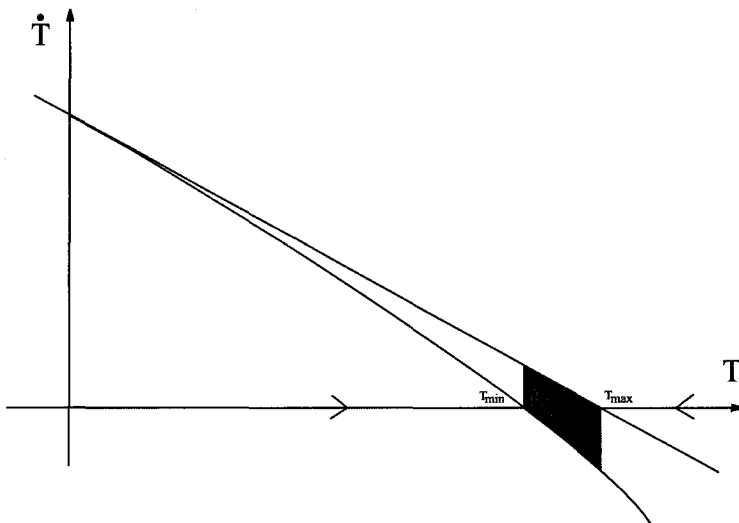


FIG. 4.9 – Convergence de  $T$  pour l'inégalité différentielle  $\rho(\beta\sigma - T) - \frac{\alpha}{4\gamma\beta}T^2 \leq \dot{T} \leq \rho(\beta\sigma - T)$ , les valeurs de  $\dot{T}$  sont comprises entre les deux courbes représentées sur le graphique et  $T$  converge vers  $\mathcal{A}$  (en noir sur la figure).

systemes. Comme on considère les paramètres  $\alpha$  (la constante théorique d'adsorption) et  $\gamma$  (la quantité de nutriment nécessaire pour une division bactérienne) comme des constantes biologiques,  $\mathcal{A}$  ne dépend que de  $\rho$  et  $\sigma$ . La propriété 43 nous dit donc que à  $\rho$  et  $\sigma$  fixés, quelquesoit la composition du système, sa concentration massique totale pondérée doit converger vers un intervalle relativement étroit ne dépendant que de  $\rho$  et  $\sigma$ .

**Exemple 45** On rappelle que  $\alpha = 10^{-9} \text{ml. min}^{-1}$  et  $\gamma = 10^{-5} \mu\text{g}$ .

- On réalise une première série de cinq simulations pour  $\rho = 0,0045$  et  $\sigma = 5$ , au cours desquelles on représente graphiquement les variations de  $T$ . Les résultats de ces simulations sont représentés à la figure 4.10.
- On réalise une deuxième série de cinq simulations pour  $\rho = 0,006$  et  $\sigma = 10$ , au cours desquelles on mesure  $T$ . Les résultats de ces simulations sont représentés à la figure 4.11.

**Remarque 46** En observant les résultats des simulations, on s'aperçoit que  $T$  reste confinée dans la partie supérieure de  $\mathcal{A}$ . Cela s'explique par le fait que la partie basse de l'intervalle attracteur ne serait visitée que pour des valeurs très faibles de concentration de  $S$ . Or, intuitivement, on perçoit que le niveau de  $S$  du système ne peut pas descendre trop bas, car un niveau de nutriment trop faible provoque la mort des bactéries, ce qui a pour effet d'augmenter  $S$ .

**Remarque 47** On montre la propriété 43, pour toutes les dynamiques possibles du MAD. Mais qu'en est-il de l'aspect métadynamique ? La mutation conserve  $T$

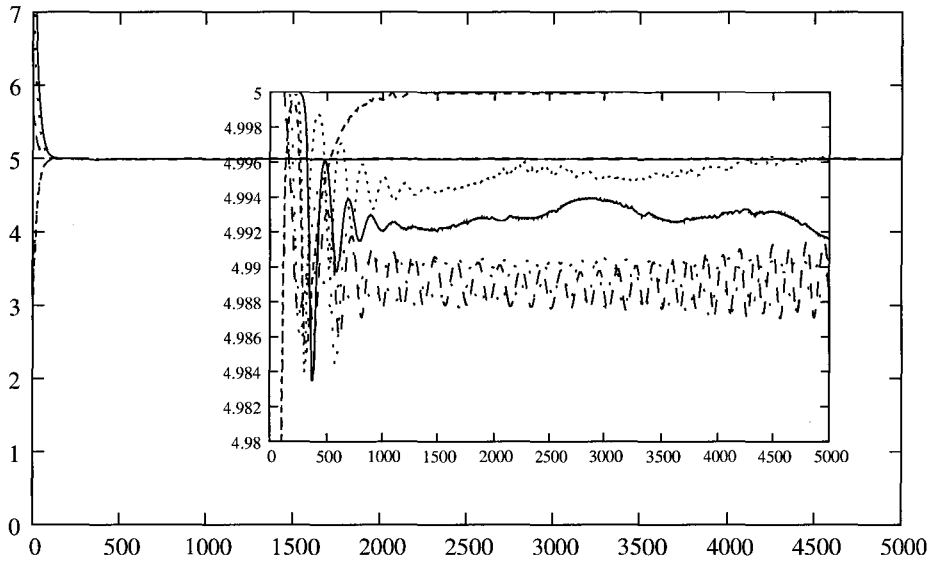


FIG. 4.10 – Tracé de  $T$  pour 5 simulations différentes avec  $\rho = 0.0045$  et  $\sigma = 5$ .  $T$  converge alors vers  $\mathcal{A} = [4, 32; 5]$ . Un agrandissement permet de voir que malgré cette convergence beaucoup de dynamiques différentes peuvent apparaître.

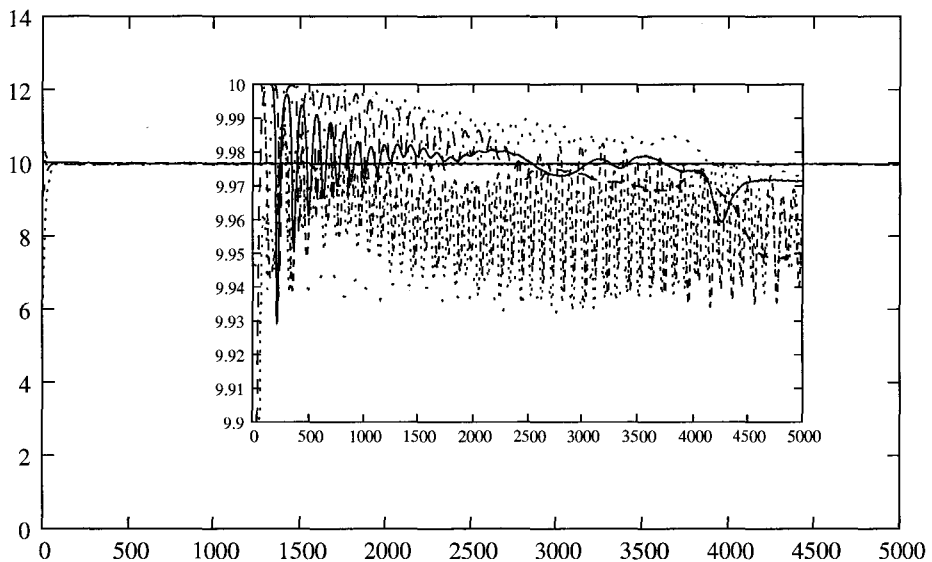


FIG. 4.11 – Tracé de  $T$  pour 5 simulations différentes avec  $\rho = 0.006$  et  $\sigma = 10$ .  $T$  converge alors vers  $\mathcal{A} = [9, 615; 10]$ . Un agrandissement permet de voir que malgré cette convergence beaucoup de dynamiques différentes peuvent apparaître.



(elle conserve la masse totale des phages ou des bactéries), en revanche la disparition d'une souche du chimostat fait baisser  $T$  d'un peu moins de  $\gamma \text{seuil}_B$  pour les bactéries et de  $\frac{\gamma}{\beta} \text{seuil}_P$  pour les phages. Ainsi, l'on devrait pouvoir par une extension soudaine et massive de population pouvoir descendre par une commutation en dessous de la limite inférieure de  $\mathcal{A}$ . Il est possible que l'on puisse démontrer théoriquement que cela n'est possible que pour des choix extrêmement bizarres des conditions initiales amenant à des situations qu'il n'est autrement pas possible d'atteindre par une évolution normale du système. D'un point de vue pratique en tout cas, ce type de situation n'arrive jamais, au contraire, les trajectoires se concentrent dans la partie supérieure de  $\mathcal{A}$ .

**Remarque 48** On a donc l'assurance que les grandeurs de notre système resteront bornées quels que soient les paramètres et les conditions de simulation. On peut également déduire de ces valeurs :

- la concentration maximale de sucre,  $S_{max} = \sigma$ ,
- la concentration maximale totale de bactéries,  $B_{max} = \frac{\sigma}{\gamma}$ ,
- la concentration maximale de phages  $P_{max} = \frac{\beta\sigma}{\gamma}$ .

Comme une population ne peut pas descendre au-dessous d'un certain seuil, ces valeurs donnent également le nombre d'équations maximum possibles lors d'une simulation :

$$n_{\text{éq\_max}} = \max\left(\frac{B_{max}}{\text{seuil}_B}, \frac{P_{max}}{\text{seuil}_P}\right) + 1$$

**Remarque 49** Cette propriété permet de mettre en lumière l'approximation classique consistant à considérer le nombre total d'individus d'un écosystème comme constant. On a obtenu ici avec un changement de variables un résultat montrant que cette approximation ne serait pas trop mauvaise d'un point de vue quantitatif, mais ferait disparaître l'aspect dynamique (grande variété de dynamiques différentes à l'intérieur du domaine d'attraction).

#### 4.4.2 Représentation des populations et définitions utiles

Comme, lors d'une simulation, le nombre de variables à intégrer peut être de quelques milliers, la sauvegarde du système nécessite alors plusieurs dizaines de kilo octets de mémoire. Cette opération doit en plus être répétée un grand nombre de fois si l'on veut être en mesure de produire de bonnes représentations du système continu. On atteint alors vite une masse de données gigantesques (d'autant plus que notre perspective évolutive nous oblige à regarder des échelles de temps longues).

La solution à ce problème est de ne sauvegarder que des valeurs significatives et non l'ensemble du système. Pour pouvoir dégager et définir de telles variables nous devons analyser la structure de nos populations.

Plus que les dynamiques individuelles, il est intéressant de regarder la dynamique collective, c'est-à-dire la dynamique du nuage dans l'espace des caractéris-

tiques (voir fig 4.12 et [22]), mais qu'entendons-nous par nuage dans l'espace des caractéristiques?

Il convient de préciser un peu les notions et les notations que nous allons utiliser par la suite.

### Espace des caractéristiques

Nous représentons les caractéristiques de nos populations<sup>13</sup> par des chaînes binaires de taille  $n_c$ . Nous commençons donc par définir une distance pour ces CB :

**Définition 50** On appelle *distance de Hamming* la fonction  $d_H$  définie par

$$d_H(s_1, s_2) = \sum_{k=0}^{n_c} |s_i^k - s_j^k|$$

avec  $s_i = s_i^1 s_i^2 \dots s_i^{n_c}$ ,  $i = 1, 2$  et  $s_i^k \in \{0, 1\}$ .

Il s'agit donc juste de compter le nombre de bits différents ou de manière équivalente le nombre de mutations minimum nécessaire pour passer d'une chaîne à l'autre.

L'espace des caractéristiques est représenté par l'ensemble des CB agencées de telle manière que les voisines de chaque CB soient les CB à une distance de Hamming de 1, c'est-à-dire l'ensemble des CB accessibles par une seule mutation. Les différentes CB se disposent alors sur les sommets d'un  $n_c$ -hypercube. Ainsi pour  $n_c = 1$ , la CB est égal à 0 ou 1 et l'espace des caractéristiques est donc réduit à un segment, pour  $n_c = 2$ , elles définissent un carré, pour  $n_c = 3$ , un cube, pour  $n_c = 4$  un 4-hypercube, etc (figure 4.12).

**Définition 51** On note  $S_N = \{0, 1\}^N$  muni de la distance de Hamming l'espace des chaînes binaires de taille  $N$ , on appelle  $G \equiv S_{n_c}$  l'espace des caractéristiques.

### Définitions liées aux populations

Le mot population peut désigner a priori dans notre cas aussi bien une variable  $P_i$  qui a pour valeur la concentration d'une population homogène partageant les mêmes caractéristiques que l'ensemble de ces variables et on peut encore imaginer d'autres sens qui resteraient valables d'un point de vue intuitif. Pour éviter les confusions par la suite, nous donnons donc un sens précis à ce mot.

<sup>13</sup>Ceci est bien entendu une caractéristique générale de nos systèmes complexes adaptatifs et un bon nombre des définitions présentées ici ont une portée plus générale que celle de notre exemple dont, l'un des buts est, en plus de son intérêt biologique, de les découvrir et de les tester.

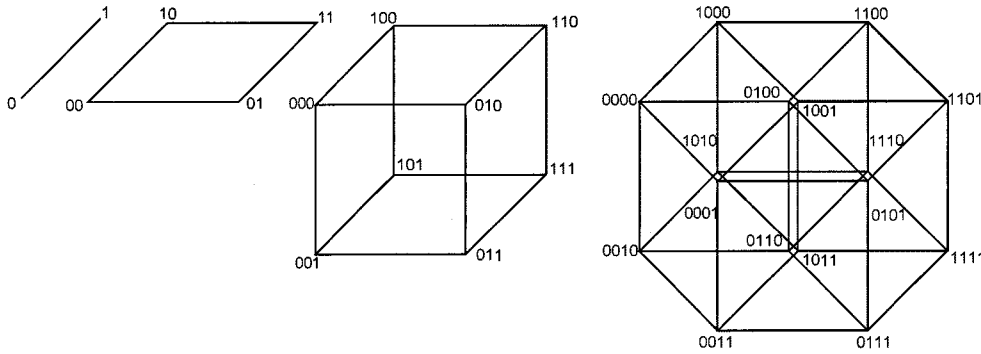


FIG. 4.12 – Organisation géométrique des CB sous forme d’une structure d’hypercube cas  $n_c = \{1, 2, 3, 4\}$  avec respectivement un segment de droite, un carré, un cube et un 4-hypercube comme structure d’espace.

**Définition 52** Une *population*  $P$  sur  $G$  est une fonction  $P : G \rightarrow \mathbb{R}^+$ <sup>14</sup>.

Par abus de notation, nous écrirons peut-être parfois  $g \in P$  pour  $P(g) > 0$ .

**Définition 53** La *taille* de  $P$  est

$$|P| \equiv \sum P(g).$$

**Définition 54** Le *support* de  $P$  est

$$\text{supp}(P) \equiv \{g \in G | P(g) > 0\},$$

c’est-à-dire que  $\text{supp}(P)$  est l’ensemble des CB présentes dans la population.

**Définition 55** La *CB dominante* de  $P$  est la CB  $g^*$  tel que  $P$  soit maximum en  $g^*$ .

**Définition 56** On appelle *nuage* de  $P$

$$\mathcal{N}(P) \equiv \{(g \in \text{supp}(P), P(g))\}$$

qui correspond à l’ensemble des CB présentes dans la population pondérées par leurs tailles respectives.

**Définition 57** L’*entropie* d’une population  $P$  est donné par :

$$E(P) \equiv - \sum_{g \in \text{supp}(p)} \frac{P(g)}{|P|} \log \frac{P(g)}{|P|}.$$

<sup>14</sup> Nous définissons une fonction à valeurs dans  $\mathbb{R}^+$  parce que nous travaillons sur un modèle continu à très grand nombre d’éléments, dans d’autres circonstances, il pourrait être plus approprié de définir cette fonction à valeurs dans  $\mathbb{N}^+$ .

L'entropie est une mesure de la dispersion de la population, en effet, elle a les propriétés suivantes :

- elle est nulle quand la population est concentrée en un point,
- pour un nombre donné de sites occupés, elle est maximale pour une distribution homogène (quand  $P(g)$  est constant  $\forall g$ ),
- à chaque division d'un groupe d'individus entre deux loci, l'entropie croît (ce qui se traduit par<sup>15</sup>  $E(p_1, p_2, p_3 \dots p_n) > E(p_1 + p_2, p_3 \dots p_n)$  avec  $p_i = \frac{P(g_i)}{|P|}$ ).

**Remarque 58** *Il existe un lien très étroit entre la théorie de l'information et l'entropie. On rappelle que l'information contenue par la connaissance d'un événement appartenant à un ensemble de possibilités  $\Omega$  est égale à  $\log_2(|\Omega|)$ . Sachant cela, on peut calculer la quantité d'information moyenne portée par un individu d'une population  $P$ .*

*On note  $N$  le nombre d'individus de la population  $P$ . Considérant que les individus ayant les caractéristiques  $g_i$  sont en nombre  $p_i N$ , le nombre de possibilités de répartir les  $N$  individus en respectant les proportions  $p_i$  est*

$$|\Omega| = \frac{N!}{(p_1 N)! (p_2 N)! \dots (p_k N)!}$$

*La quantité moyenne d'information portée par un individu est donc de<sup>16</sup> :*

$$S = \frac{1}{N} \log_2 |\Omega|,$$

*or, quand  $N$  est grand  $\log N! \sim N \log N!$ , d'où*

$$S \sim - \sum p_i \log p_i.$$

*On retrouve donc ici la définition de l'entropie d'une population. L'entropie est donc la quantité moyenne d'information donnée spécifiant l'état d'un individu.*

**Définition 59** *Le **centroïde** de  $P$  est le vecteur  $c \in [0, 1]^{n_c}$  défini par*

$$c_i \equiv \frac{1}{|P|} \sum_{g \in \text{gsupp}(p)} P(g) g_i,$$

*où  $c_i$  est la  $i^{\text{ème}}$  composante de  $c$  et  $g_i$  celle de  $g$ . Ce centroïde peut se voir comme le barycentre de  $\mathcal{N}(P)$ .*

<sup>15</sup>avec un léger abus de notation. Le lecteur pointilleux lira plutôt :  $\tilde{E}(p_1, p_2 \dots p_n) = - \sum_{i=0}^n p_i \log p_i$ .

<sup>16</sup>Dans le cas où toutes les probabilités sont égales à  $\frac{1}{|\Omega|}$ , au lieu de la formule ci-après, on obtient à un coefficient près, la célèbre formule  $S = k \log W$  avec  $k$  la constante de Boltzmann et  $W$  le nombre d'état possible, qui est gravée sur la tombe de Boltzmann.

Si l'on suit la trajectoire d'une population au cours du temps, on obtient une suite de points appartenant à  $[0; 1]^{n_c}$ . Il se pose alors le problème de la visualisation : comment projeter de la meilleure manière possible cet ensemble de points sur le plan ?

On cherche à conserver au maximum les distances entre les points, le critère de projection est donc de maximiser les distances entre les points projetés, sachant que ces distances projetées seront plus faibles que les distances originales.

On peut démontrer que rechercher le meilleur plan de projection du nuage est équivalent à rechercher d'abord la meilleure droite de projection pour le nuage, puis la meilleure droite de projection pour le nuage moins sa partie projetée. La direction de ces droites est donnée par les deux vecteurs propres de plus grandes valeurs propres de la matrice  $V$  (les "composantes principales"), où

$$V = X^T X - gg^T$$

avec

$$X = \begin{bmatrix} x_{1,1} & \dots & x_{1,j} & \dots & x_{1,n_c} \\ \vdots & \ddots & \vdots & & \vdots \\ x_{i,1} & \dots & x_{i,j} & \dots & x_{i,n_c} \\ \vdots & & \vdots & \ddots & \vdots \\ x_{m,1} & \dots & x_{m,j} & \dots & x_{m,n_c} \end{bmatrix},$$

$[x_{i,1}, \dots, x_{i,j}, \dots, x_{i,n_c}]$  étant le  $m^{\text{ième}}$  centroïde du nuage et

$$g = \left[ \sum_{i=1}^m \frac{x_{i,1}}{m}, \dots, \sum_{i=1}^m \frac{x_{i,j}}{m}, \dots, \sum_{i=1}^m \frac{x_{i,n_c}}{m} \right]$$

[65]. Cette technique est appelé "Analyse en Composantes Principales" (ACP).

## 4.5 Résultats de simulations

Nous présentons maintenant quelques résultats de simulation<sup>17</sup>. Ces résultats ont été publiés dans [5, 6, 7, 4, 8].

### 4.5.1 Aspect général de simulation

Le but de ce modèle est d'analyser l'évolution dynamique et métadynamique des populations en fonction de la variation et/ou de l'évolution de certains pa-

<sup>17</sup>La première étape de notre travail a bien entendu consisté à valider numériquement notre logiciel. Pour cela, nous avons inhibé complètement les mutations et nous avons réglé notre système de telle façon à supprimer les différences existant entre notre modèle différentiel (devenu sans métadynamique) et celui de Mosekilde. Comparant alors nos résultats avec une implémentation réalisée antérieurement sous 20-sim pour le modèle de Mosekilde à l'occasion de mon DEA, nous avons alors pu retrouver les mêmes valeurs numériques et valider ainsi notre implémentation.

paramètres caractérisant ces populations. Les paramètres susceptibles de varier ou d'évoluer dans ce modèle sont :

- $\rho$  la dilution du système,
- $\sigma$  l'alimentation,
- $n_c$  la taille des chaînes binaires associées aux bactéries et aux phages du système,
- $\mu_B$  et  $\mu_P$  les taux de mutations respectivement des phages et des bactéries.

On autorise, par définition, notre système à faire évoluer sa dimension durant la simulation. Du point de vue de la simulation, ceci peut représenter une contrainte forte. A titre informatif, nous donnons ici en exemple, deux comportements très différents du changement au cours du temps de la dimension (figure 4.13). Sur le panneau de gauche, on peut voir un cas pathologique. A cause d'une variabilité trop importante de la population de phages, la dimension du système explose atteignant la valeur de 10744. Au contraire, dans le cas représenté au panneau de droite, la dimension se stabilise assez rapidement. Un cas tel que le premier induit une augmentation très importante du temps de simulation qui peut-être rédhibitoire.

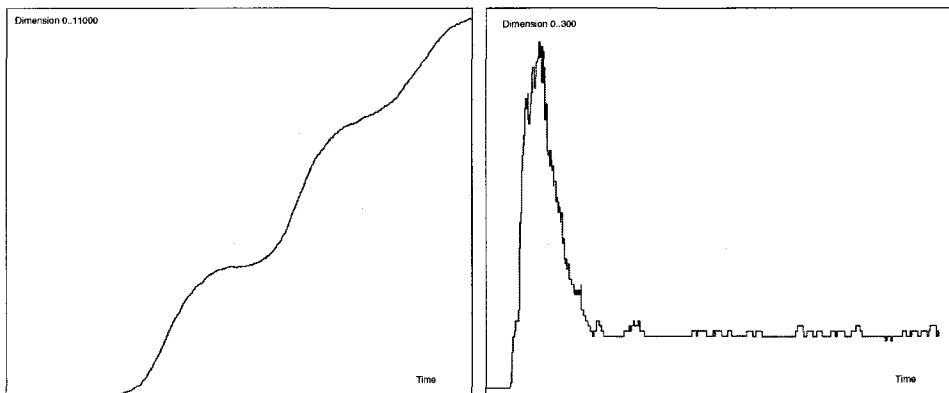


FIG. 4.13 – Deux évolutions très différentes de la dimension au cours du temps. A gauche, on est en présence d'un cas pathologique du point de vue de la simulation : la dimension du système explose et atteint la valeur de 10744. Au contraire, à droite, la courbe montre une dimension qui se stabilise autour d'une valeur basse.

### 4.5.2 Comportement du système sans mutation

Comme on l'a vu précédemment, l'expression des gènes, tant des phages que des bactéries, se fait au travers du coefficient de probabilité d'infection  $\omega_{ij}$ . Il semble donc intéressant pour cela de regarder quel peut être l'influence du changement de valeur de ce coefficient sur la dynamique du système le plus simple possible, c'est-à-dire deux populations homogènes de phages et de bactéries. Considérant

la forme de  $\omega_{ij}$  donnée précédemment, cela se fait en faisant varier  $d_H(s_i^b, s_j^p)$ . Les simulations ont été réalisées avec  $n_c = 32$  pour une meilleure précision.

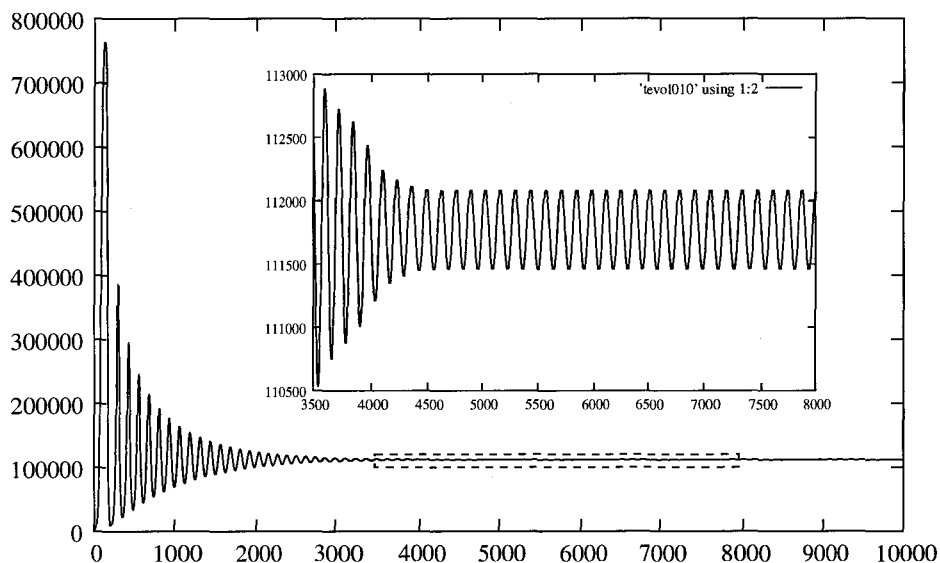


FIG. 4.14 – Evolution des bactéries et agrandissement pour  $n_c = 32$ ,  $d_H = 10$ . On observe la présence d'un cycle limite.

On observe une bifurcation, quand la distance passe de 15 à 16 (figure 4.14 et 4.15). C'est-à-dire pour une valeur de  $\omega_{ij}$  comprise entre 0,25 et 0,282. La dynamique passe d'un point fixe à une zone de périodicité/quasi-périodicité dont la complexité semble s'accroître avec la ressemblance entre les deux chaînes.

Ici le système est par définition de relativement basse dimension (six pour être précis), mais on sait que, pour des populations plus riches (un grand nombre de génotypes différents et fixés), on peut, au lieu d'oscillations périodiques se retrouver en présence de chaos, voire d'hyperchaos [55].

### 4.5.3 Mobilité des phages dans l'espace des CB

Cette expérience consiste en la mise en œuvre de l'aspect le plus simple de la métadynamique : on "autorise" les phages à muter. En revanche, on conserve  $\mu_B = 0$ . Cette situation correspond également de manière plus réaliste à une approximation du cas où  $\mu_P \gg \mu_B$ . Les paramètres que nous faisons varier ici sont  $\mu_P$  et la distance initiale (au sens de la distance de Hamming) entre la population du phage et de bactérie. D'un point de vue intuitif, deux possibilités apparaissent :

- les phages n'arrivent pas à assurer une prédation efficace des bactéries et disparaissent,
- les phages arrivent à produire une prédation efficace sur les bactéries, se reproduisent et assurent leur pérennité. On assiste à une évolution de la

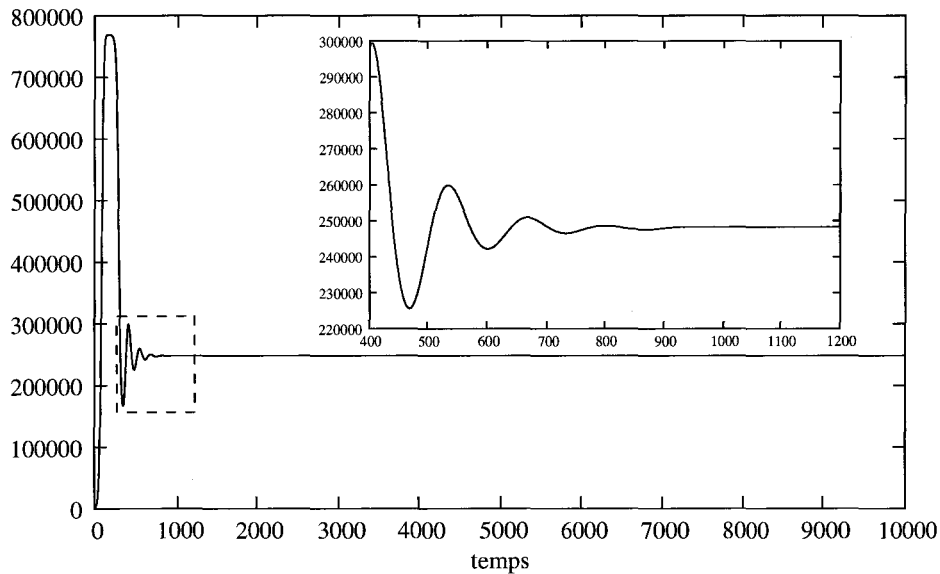


FIG. 4.15 – Evolution des bactéries et agrandissement pour  $n_c = 32$ ,  $d_H = 17$ . Le comportement asymptotique du système est un point du système.

population vers une meilleure efficacité dans la prédation (on appelle ce scénario “l’attaque évolutive des phages”). L’état final est un état proche de ceux étudié par Mosekilde.

**Problème 60** *Au delà de cet aspect intuitif, plusieurs questions se posent :*

- *Est-ce que l’on retrouve bien ces deux scénarii lors de la simulation ?*
- *Quels sont les facteurs qui favorisent l’un ou l’autre ?*
- *Existe-il un scénario générique de l’attaque des phages ? Quelle est l’influence de la variation de paramètres sur le scénario de l’attaque de phages ?*
- *Quel est l’état final de l’attaque évolutive de phage ? Existe-t-il un état métadynamique final ?*

*On devra également à chaque réponse, associer une “robustesse” des conclusions en regard de l’aspect stochastique des mutations.*

### Conformité des comportements attendus

On observe bien les deux types de comportements escomptés. En fait, dépendant de la distance de Hamming entre les deux populations initiales, on observe trois comportements possibles :

1. Le gène moyen de la population initiale de bactéries est trop éloigné de celui des phages dans l’espace génétique booléen ou la concentration initiale de phages est trop faible, les phages disparaissent alors assez rapidement.
2. Le gène moyen de la population initiale de bactéries est très proche de celui des phages, on observe directement des oscillations très compliquées évoquant



du chaos.

3. Le gène moyen de la population initiale de bactéries est éloigné, mais encore dans le « champ » de capture des phages, après une phase transitoire où les bactéries atteignent un équilibre de tranquillité, les phages se rapprochent<sup>18</sup> des bactéries, on observe une phase critique puis très rapidement des oscillations relativement régulières (figures 4.16).

D'un point de vue métadynamique, le seul comportement intéressant est le comportement 3, nous nous intéresserons donc tout particulièrement à celui-ci. Il est à également à noter que le deuxième cas correspond au comportement asymptotique du cas 3.

### Scénario de l'attaque des phages, influence des paramètres

On considère dans cette section, que l'on part de CI situées dans la "zone d'attaque évolutive". Le seul paramètre dont nous étudions l'influence est donc  $\mu_P$  le taux de mutation des phages.

Nous commençons par présenter le scénario "générique" de cette attaque qui a lieu quel que soit  $\mu$ . Nous verrons ensuite, quel est l'influence de  $\mu$  sur les variations possibles de ce scénario.

**Scénario générique** Quelle que soit la manière dont l'attaque a lieu, on peut diviser l'attaque en trois phases (voir figure 4.16) :

1. les bactéries se développent sans prédateurs, on observe l'apparition d'un équilibre,
2. les phages se développent et parallèlement commence à explorer l'espace génétique par mutations,
3. les phages se stabilisent (du point de vue de leur composition "génétique"), et rentrent dans un régime oscillatoire incluant les bactéries et le sucre.

On a donc un nuage qui se "déplace" dans  $(G \times \mathbb{R}^+)^{2n_c}$ . Comme on ne peut pas sauvegarder toutes les variables, on choisit de sauvegarder des grandeurs caractéristiques de ce nuage : son centre de gravité (le centroïde) et sa dispersion (l'entropie). La constante que l'on retrouve dans toutes les situations "d'attaque phagique" est le déplacement du nuage des phages de façon à "optimiser" son attaque. La direction générale est donc de la CB de départ vers la CB de la bactérie (on rappelle que, par définition de notre modèle, plus la similarité est grande entre les deux chaînes, plus la prédation est forte) (voir figure 4.17). Néanmoins, le chemin parcouru dépend fortement de  $\mu_P$ , ainsi que de l'aspect aléatoire (tirage lié aux mutations). Concernant l'entropie du nuage des phages, elle dépend également

---

<sup>18</sup>Tous les "mouvements" mentionnés ici se produisent bien entendu dans l'espace des CB et non dans l'espace physique (on rappelle que notre modèle est aspatial).

fortement des tirages aléatoires et de  $\mu_P$ , mais conserve également certaines caractéristiques : partant d'une entropie nulle (sans diversité au départ), la population de phages se disperse autour de cette position de départ, on a donc une phase d'expansion correspondant à une forte augmentation de l'entropie. Suite à cette phase d'expansion se produit une phase de contraction caractérisée donc par une baisse de l'entropie. Ce processus expansion/contraction peut se répéter dans certain cas, jusqu'à atteindre une valeur basse d'entropie qui se stabilise (figure 4.18).

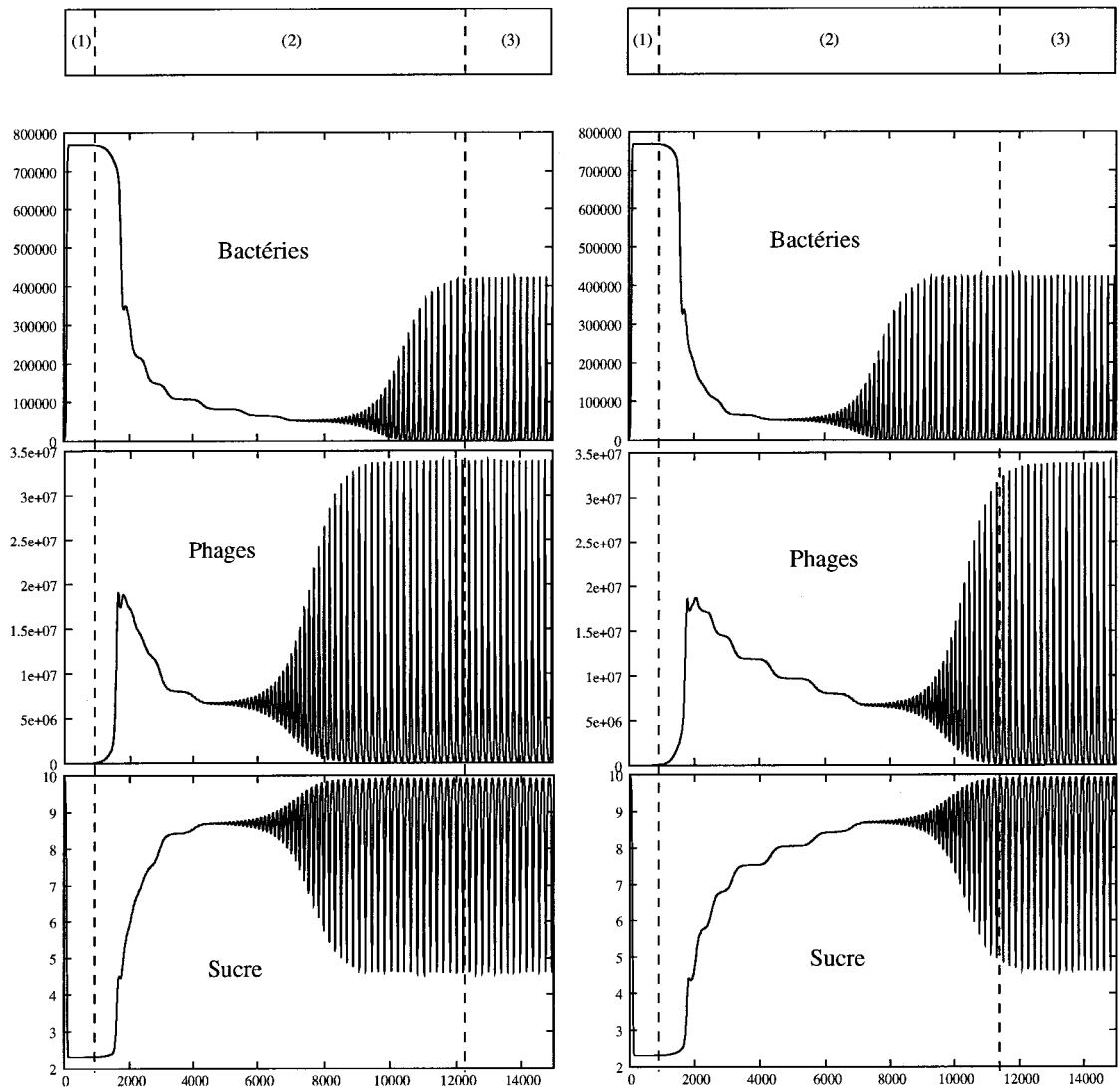


FIG. 4.16 – On observe quelle que soit l'attaque évolutive un scénario générique : (1) bactéries sans prédateurs (2) phase transitoire (3) état oscillatoire final. On a tracé ici les courbes des concentrations de bactéries, phages et sucre en fonction du temps pour deux valeurs de  $\mu$ ,  $\mu = 10^{-5}$  (à gauche) et  $\mu = 10^{-6}$  (à droite).

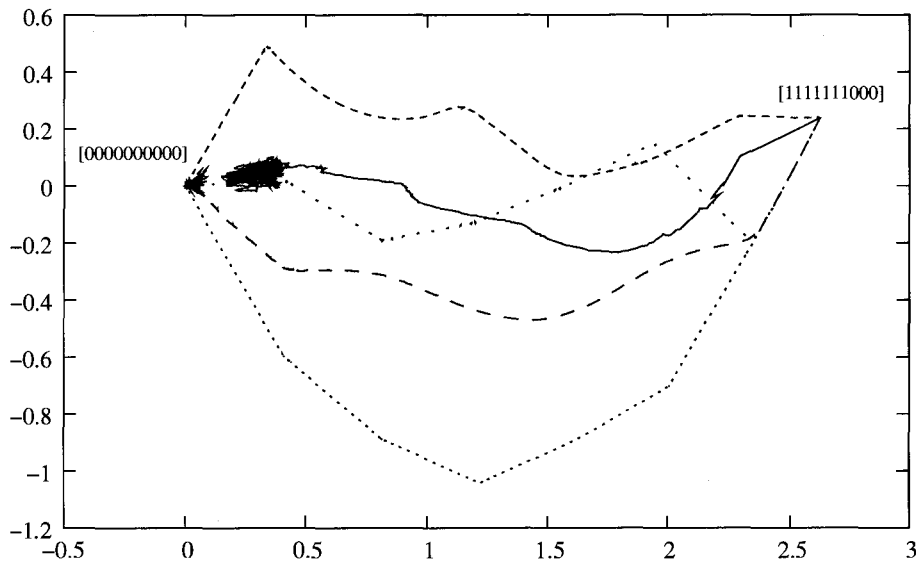


FIG. 4.17 – Direction générale de l'évolution des centroïdes lors de l'attaque phagique : ceux-ci évoluent de leur caractéristiques initiales  $[1111111000]$  vers les caractéristiques de meilleure prédation quand leur chaîne binaire est égale à  $[0000000000]$ . On obtient les projection de  $[0, 1]^{32}$  vers le plan

**Influence des paramètres** Si l'on observe un certain nombre de traits génériques dans l'attaque des phages, on observe également un certain nombre de différences, en fonction de la valeur de  $\mu$  la vitesse de mutation des phages. Ces différences concernent deux aspects de l'attaque :

- le déroulement de l'attaque en elle-même,
- l'organisation finale de la population.

En ce qui concerne l'attaque elle-même, plus la valeur de  $\mu$  est grande, plus la trajectoire du centroïde est lisse et rapide, au contraire quand la valeur de  $\mu$  est petite, la trajectoire est brisée : dès qu'un mutant plus efficace apparaît, il remplace l'espèce courante et ainsi, d'équilibre en équilibre, on atteint la valeur du phage optimum (voir figure 4.21). Du point de vue de l'entropie, on observe alors une succession de pics (voir figure 4.18). Si l'on trace les concentrations des phages à une distance donnée (au sens de la distance de Hamming) du phage  $[0000000000]$ , on observe des vagues successives (figure 4.19).

Nous arrivons maintenant au point le plus intéressant de cet exemple : l'organisation asymptotique de la population des phages. Revenons à la figure 4.20. On observe que, quand les valeurs de  $\mu$  sont grandes, le centroïde se concentre dans une région donnée de l'hypercube  $[0, 1]^{10}$ , proche du sommet  $[0000000000]$ . Si l'on regarde l'entropie (figure 4.18), on s'aperçoit que, à partir d'une certaine valeur de  $\mu$ , elle est non nulle, de plus celle-ci croît quand  $\mu$  augmente. Cela signifie, qu'à partir d'une certaine valeur de  $\mu$ , le méta-attracteur de la métadynamique n'est plus le système de méta-état ( $\{S, B = \{[0000000000]\}, P = \{[0000000000]\}\}$ ,  $\varphi$ ), avec  $\varphi$

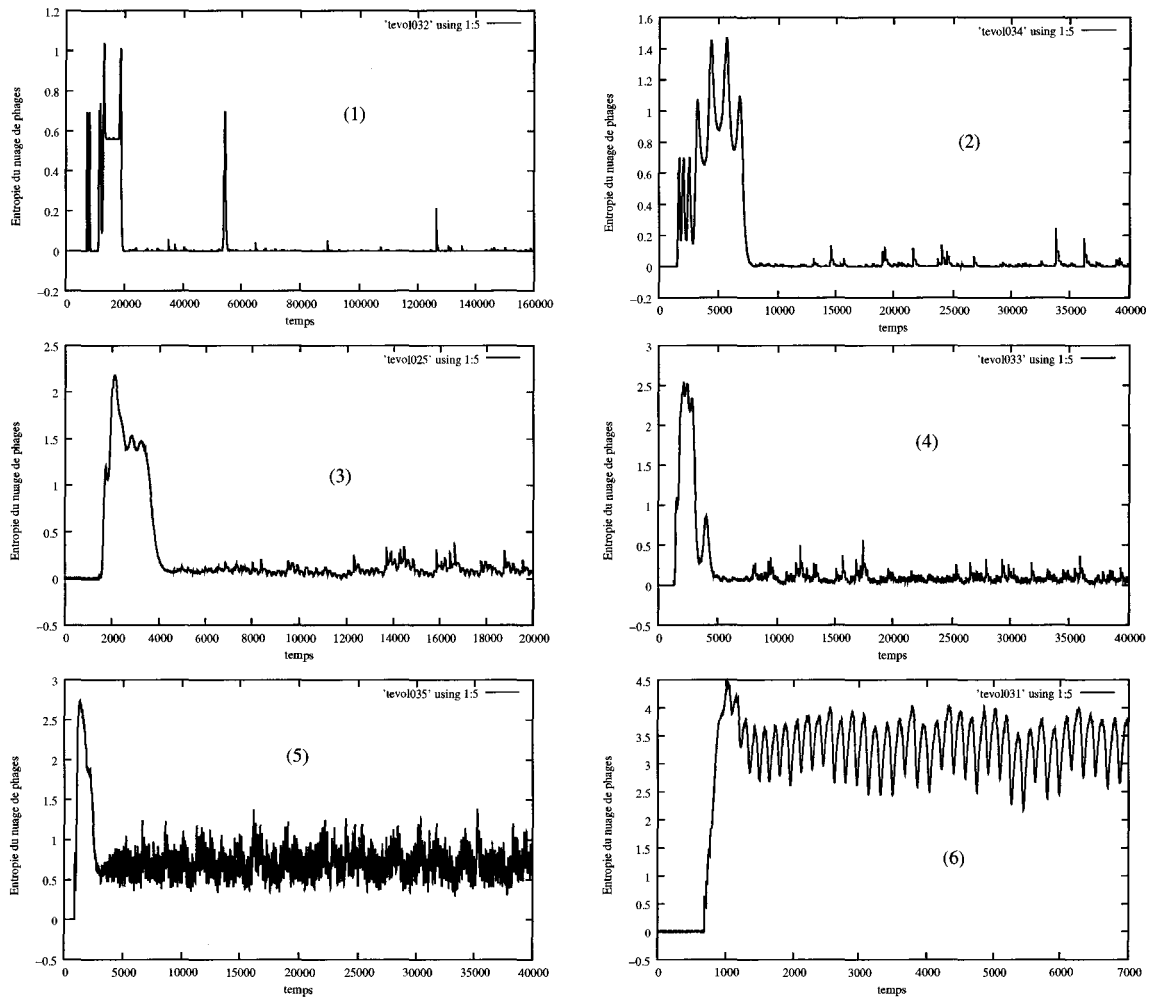


FIG. 4.18 – Tracés de différentes entropies correspondant à des valeurs de  $\mu$  différentes et/ou à des générateurs aléatoires différents (1)  $\mu = 10^{-7}$  (2)  $\mu = 10^{-6}$  (3)  $\mu = 10^{-5}$  (4)  $\mu = 10^{-5}$  (5)  $\mu = 10^{-4}$  (6)  $\mu = 10^{-3}$ .

donné par l'intégration de système d'équations différentielles correspondant, mais un domaine d'attraction plus large, qui croît avec  $\mu$ . Ces domaines correspondent à des "couronnes", autour du phage dominant [0000000000], plus ou moins étendues, et avec des densités qui oscillent et dont la répartition dépend de  $\mu$ . Ainsi, pour la valeur de  $\mu$  la plus grande ( $\mu = 10^{-3}$ ), on peut tracer les concentrations des  $[i]$ ,  $0 < i < 7$ , correspondant aux phages à la distance  $i$  de  $P(0)$ . On s'aperçoit que cinq classes d'équivalence de populations coexistent : [0], [1], [2], [3] et [4]. Si ces populations peuvent coexister, c'est que le taux de mutations compense les pertes dues à une moins bonne adaptation. La vitesse de mutation  $\mu$  engendre des probabilités de mutations de l'ordre des probabilités élémentaires du système, on retrouve alors un système quasi-différentiel que l'on peut comparer au modèle de quasi-espèce de Eigen et Schuster (voir section 1.2.2). Comme pour le modèle de quasi-espèce, ce cas limite engendre une population dont la composition relative est relativement constante.

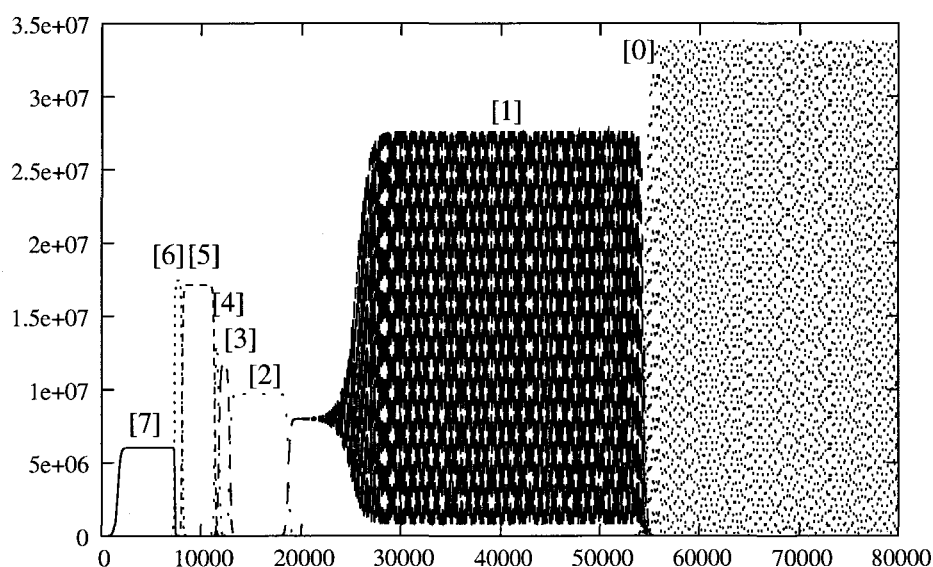


FIG. 4.19 – Tracé des populations de phages  $[i]$  avec  $[i]$  l'ensemble des phages à une distance de Hamming  $i$  du phage [0000000000] pour  $\mu = 10^{-7}$ . On observe une succession d'états d'équilibre avec une transition entre [2] et [1] avec le passage vers un cycle limite (bifurcation de Hopf au sens large, car "métadynamique").

**Remarque 61** La forme de la quasi-espèce engendrée est très dépendante de la "génétique" sous jacente. Ainsi dans notre cas, on considère que la CB code pour la défense de la bactérie ou l'attaque du phage, de façon symétrique par rapport à tous ses bits. Il n'y a donc pas de bit privilégié et la forme de quasi-espèce est une couronne autour du phage optimum. Pour une distribution plus compliquée des avantages dus aux caractéristiques, on pourrait avoir affaire à des formes de quasi-espèce beaucoup plus compliquées.

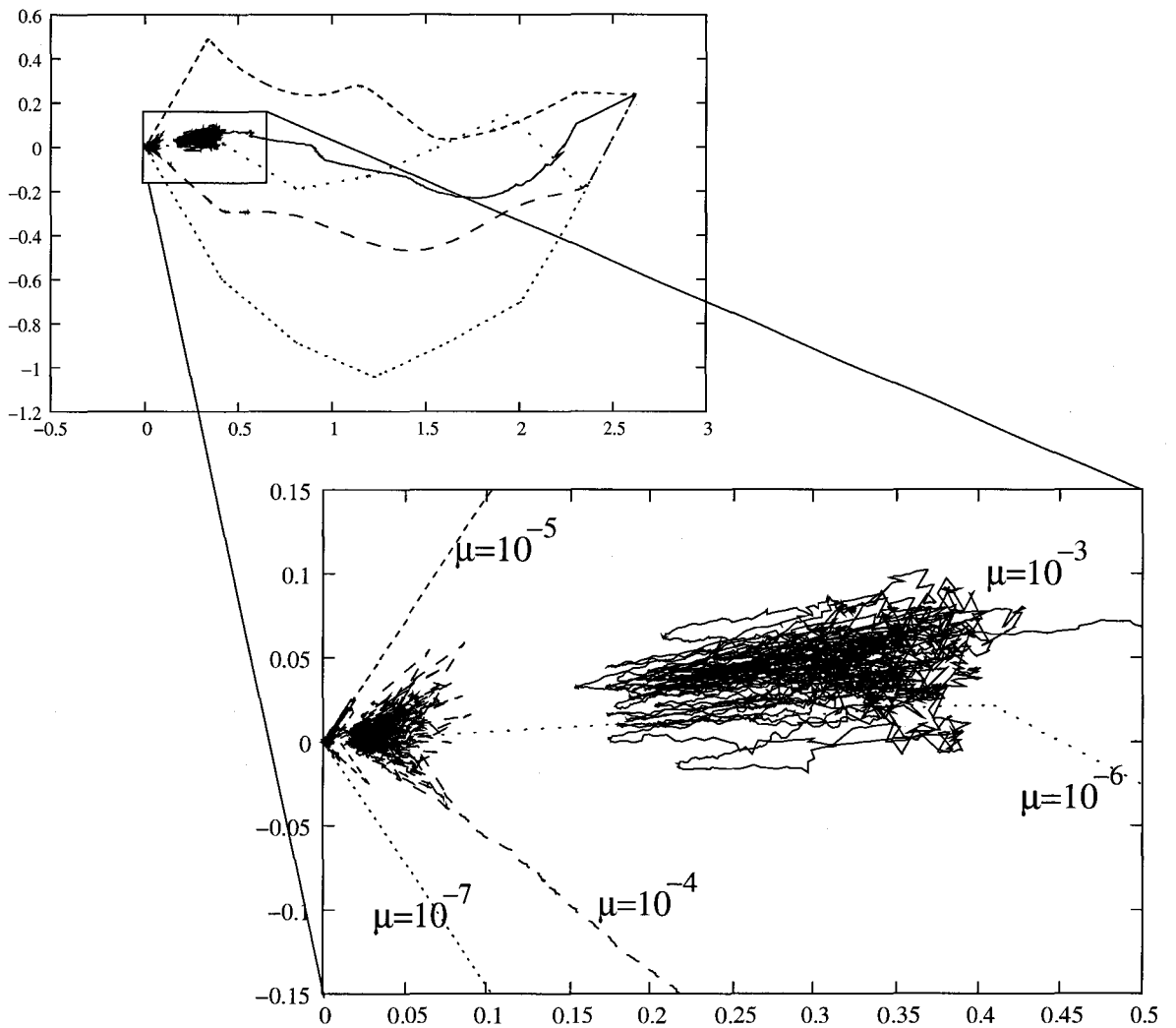


FIG. 4.20 – Projection de la trajectoire des centroïdes du nuage phagique de  $[0, 1]^{10}$  vers  $\mathbb{R}^2$  pour cinq valeurs de  $\mu$  différentes. A partir de  $\mu = 10^{-5}$ , on observe l'apparition d'une quasi-espèce, quasi-espèce incluant d'autant plus de phénotypes différents que  $\mu$  est grand.

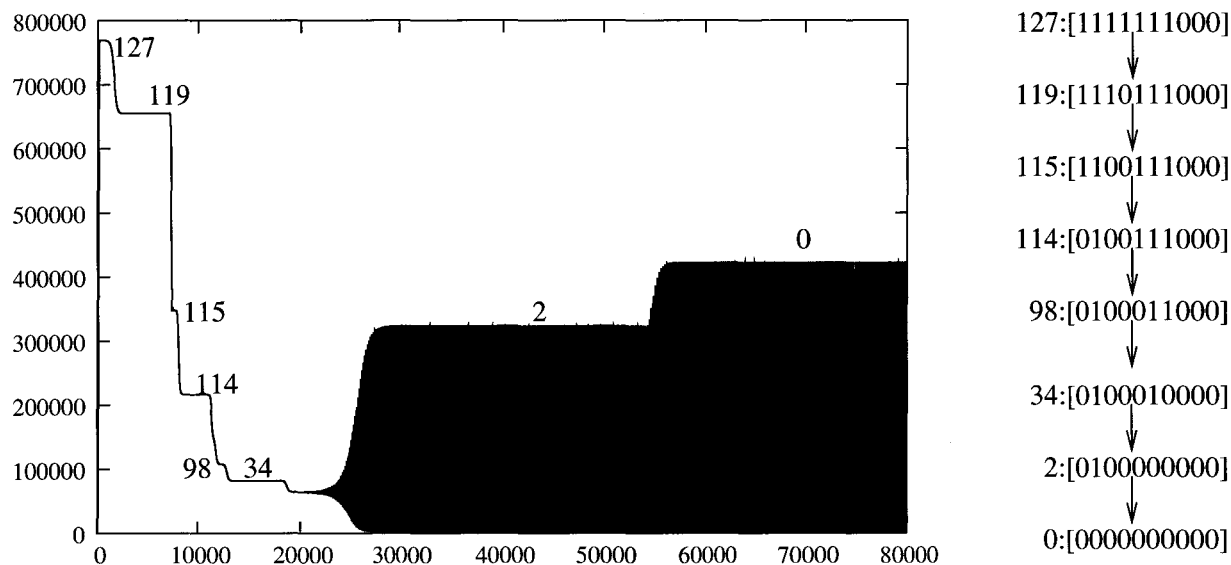


FIG. 4.21 – Pour des valeurs de  $\mu$  faibles, l'attaque se résume en une série d'équilibres (avec une bifurcation de Hopf, c'est-à-dire passage d'un point d'équilibre à un cycle limite) correspondant au basculement successif de la population vers des variants de mieux en mieux adaptés (les chiffres donnés correspondent aux valeurs décimales des CB).

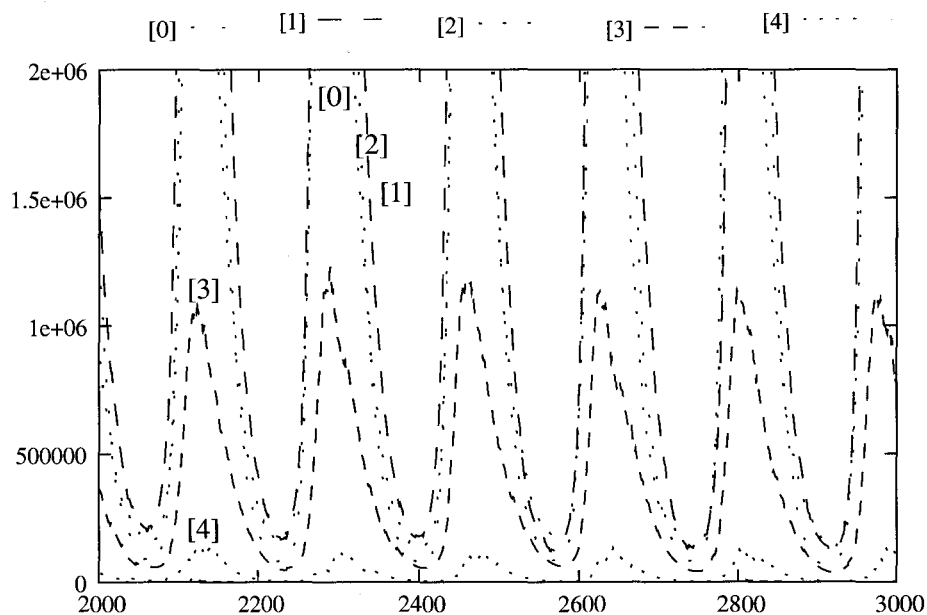


FIG. 4.22 – Zoom sur les changements au cours du temps des populations de phages à différentes distances du phage 0. Pour la figure présentée,  $\mu = 10^{-3}$  et on observe la coexistence de quatre classes d'équivalence de phages. On obtient une structure de quasi-espèce, dont la population totale oscille.

### Interprétation métadynamique

L'interprétation métadynamique de cette expérience est la présence d'un méta-état d'équilibre asymptotique ou d'un ensemble de méta-états attracteurs. Cet ensemble de méta-états attracteur dépend de  $\mu_P$  et sa taille augmente avec la valeur de celui-ci.

#### 4.5.4 Fuite des bactéries

Il s'agit là d'effectuer l'expérience théorique duale de la précédente. On soumet une population de bactéries pouvant muter à une attaque par une population de phages homogène et stable (toujours au sens de ses caractéristiques). On a donc, du point de vue des paramètres,  $\mu_P = 0$  et  $\mu_B > 0$ .

On peut essayer de déduire de la connaissance intuitive du système, son évolution : on s'attend à voir la population bactérienne fuir, c'est-à-dire que la disparition des populations ayant les caractéristiques les moins favorables (soit celles ayant leurs CB les plus proches de celles des phages existant) va créer un gradient "poussant" le nuage, vers les caractéristiques complémentaires de la population de phages (au sens d'une chaîne ayant tout les bits opposés). Cette fuite s'arrête quand la pression de prédation des phages devient faible et un équilibre s'établit alors entre les bactéries et le sucre.

**Problème 62** *Au delà de ce scénario relativement évident<sup>19</sup>, on peut de la même manière que nous l'avons fait dans la section précédente se poser des questions plus délicates :*

- *Retrouve-t-on bien pour le modèle le scénario annoncé ? Y-a-t-il des cas particuliers ?*
- *Peut-on trouver des valeurs théoriques à l'équilibre annoncé en tenant compte des simplifications engendrées par les hypothèses ?*
- *Quel est le comportement asymptotique du point de vue métadynamique ?*

On observe toujours le même scénario : la population de bactéries se stabilise de manière asymptotique autour d'une valeur après que sa composition génétique moyenne finale ait atteint une distance suffisante (de l'ordre de 7,3) du gène de la population virale (figure 4.23). Essayons de comprendre ce phénomène.

Quand les caractéristiques du nuage bactérien sont suffisamment distantes des caractéristiques des phages, on peut supposer que l'attaque est très faible, en effet  $\alpha = 10^{-9}$ , si  $d_H = 7$  avec  $n_c = 10$ , alors  $\omega_{ij} = (1 - 7/10)^2 = 0.09$ , comme  $P_j \approx 10^5$ , on peut considérer que  $\sum_{j=1}^{n_p} \alpha \omega_{ij} P_j + \rho \approx \rho$ , alors le système se simplifie en deux sous-systèmes déconnectés : la population de bactéries ne dépend plus des phages et le premier sous-système devient alors :

<sup>19</sup> de par ce qui a été déjà exposé auparavant.



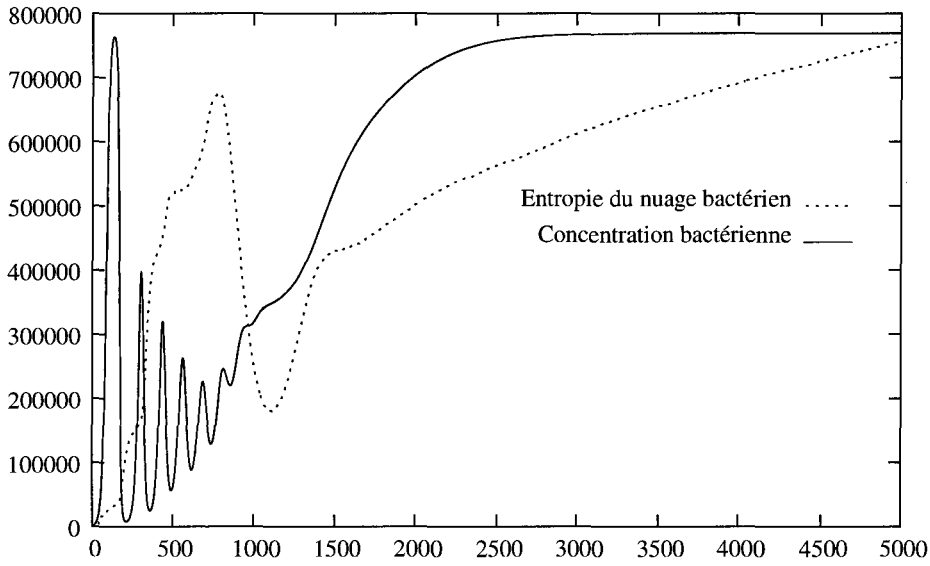


FIG. 4.23 – Population de bactéries en fuite et entropie du nuage bactérien

$$\begin{cases} \frac{dB_i}{dt} = \frac{\nu SB_i}{\kappa + S} - \rho B_i \\ \frac{dS}{dt} = \rho(\sigma - S) - \sum_{i=1}^{n_b} \frac{\nu \gamma SB_i}{\kappa + S} \end{cases}$$

avec  $n_b$  qui peut être grand (par exemple, à la fin de la simulation de la figure 4.23, on a  $n_b = 156$ ).

On pose alors  $B = \sum_{i=1}^{n_b} \gamma B_i$ , ce qui nous donne :

$$\begin{cases} \frac{dB}{dt} = \frac{\nu SB}{\kappa + S} - \rho B \\ \frac{dS}{dt} = \rho(\sigma - S) - \frac{\nu SB}{\kappa + S} \end{cases}$$

Considérons enfin la concentration bactéries+nutriment  $T = B + S$ , on a :

$$\dot{T} = \rho(\sigma - T),$$

donc  $T$  tend vers  $\sigma$ . Au bout d'un certain temps,  $T$  est suffisamment proche de  $\sigma$  pour que l'on puisse considérer que  $S = \sigma - B$ , en remplaçant dans l'équation de  $\dot{B}$  qui devient alors scalaire, on obtient que  $B$  converge vers 7,693. Si l'on considère que nos approximations sont raisonnables, cela explique l'allure de nos courbes. La valeur trouvée lors des simulations est de 7,687.

Ce raisonnement du type de ceux couramment fait en physique est assez intuitif, mais il est probable qu'il est possible de lui donner une formulation rigoureuse dans le cadre de l'analyse non standard (qui permet une utilisation rigoureuse des infinitésimaux, voir par exemple [51]).

Mais, si la forme générale de la « fuite » est toujours la même, d'une simulation à l'autre et à valeur de  $\mu$  égale, du fait de l'aspect probabiliste des mutations, on peut avoir des différences importantes sur plusieurs points :

- le moment où se produit la phase critique ;
- la valeur du gène moyen du nuage (en effet, quand l'« équilibre » s'installe, le nuage peut être n'importe où du moment qu'il est à une certaine distance du phage.

On a donc une contingence par rapport aux événements du passé, il y a donc un effet de mémoire.

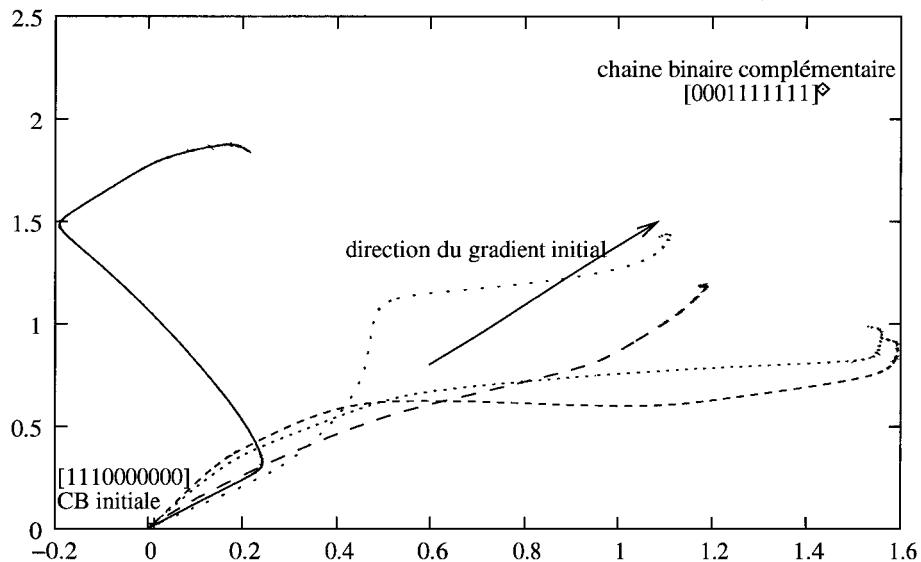


FIG. 4.24 – Projection par ACP des centroïdes des populations de bactéries en fuite. On s'aperçoit que, après être poussée par un gradient initial dans la direction de la CB complémentaire, les nuages commencent un processus de diffusion.

**Exemple 63** Gène du phage :  $[1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]$

On prend les mêmes conditions initiales pour les deux simulations, seul le générateur de nombre aléatoire change.

. Simulation 1 :

gène bactérien moyen final ( $d_H = 7,335$ )

$[0.03\ 0.053\ 0.968\ 0.969\ 0.068\ 0.972\ 0.454\ 0.966\ 0.967\ 0.948]$

. Simulation 2 :

gène bactérien moyen final ( $d_H = 7,308$ )

$[0.02\ 0.443\ 0.267\ 0.98\ 0.116\ 0.946\ 0.976\ 0.979\ 0.971\ 0.65]$

On a donc deux gènes à la même distance du gène des phages, mais néanmoins très différents.

### Interprétation métadynamique

On peut interpréter le méta-état de départ comme un méta-état répulsif. La diffusion finale avec augmentation de l'entropie à un méta-état de plus en plus

grande dimension. Les méta-états d'arrivée lors de l'arrêt des différentes simulations que nous avons effectuées sont tous très différents (voir exemple 63). L'on peut se demander ce qu'il arrive au bout d'un temps très long (en effet l'entropie ne peut augmenter de façon infinie). Y-a-t-il un méta-état ou un ensemble relativement petit de méta-états attracteurs ? Nos simulations ne nous permettent pas encore de donner des réponses certaines à cette question.

#### 4.5.5 Deux cas particulier du cas général

On laisse maintenant une totale liberté aux deux populations, on a donc  $\mu_B \neq 0$  et  $\mu_P \neq 0$ . Il est ici beaucoup plus difficile de décrire un scénario générique que dans le cas des simulations précédentes. En effet, le réglage simultané de  $\mu_B$  et de  $\mu_P$  laisse un grand choix de possibilités. Nous avons choisi deux expériences qui nous ont semblé représentatives des types de comportements qui peuvent apparaître alors.

La seule différence entre les deux expériences est la vitesse de mutation des bactéries. Dans la deuxième expérience, elle est cinq fois plus rapide que dans la seconde. La variation de ce paramètre entraîne deux comportements très différents du système.

#### Blocage des bactéries par les phages

Si l'on observe l'évolution de concentration de la population de bactéries (figure 4.25), on s'aperçoit que l'on a une alternance de phases oscillatoires très fortes (b, fin de c et d) et de phases plus calmes (a et début de c). Notre connaissance acquise sur le comportement du système, nous permet d'identifier les phases de fortes oscillations comme des phases, où les phages sont très bien adaptés à l'attaque des bactéries, et les phases intermédiaires (a et c) des phases, où ils sont plus éloignés (au sens de leurs caractéristiques).

Cette hypothèse est confirmée par le tracé de la distance entre les centroïdes des populations phagique et bactérienne (figure 4.26). On observe alors ces mêmes phases, les phases de fortes oscillations correspondant bien aux phases où les phages possèdent une adaptation maximum (distance 0). Regardons l'état du système pendant une de ces phases :

```
S 9.784805e+00
B 4071 1.943020e+04
P 4071 1.960451e+05 3.507994e+01 3.233019e+01 2.981813e+01.
```

On s'aperçoit que le système est bloqué dans un méta-état donné (ou autour) et n'en sort plus que très difficilement. Comment expliquer cela ? Comme nous l'avons vu auparavant, le moteur de l'évolution est assuré par les mutations favorables. On s'attendrait donc à voir les bactéries évoluer, d'autant plus, que dans

une telle situation, toutes les mutations sont favorables. En fait, les mutations ont bien lieu, mais, l'avantage donné par la mutation correspondante est de courte durée car les phages ayant une bien meilleure capacité d'adaptation, produisent rapidement un phage adapté à la nouvelle souche bactérienne, anihilant l'avantage de celle-ci. La plupart du temps, cette population naissante est en très faible proportion par rapport à l'ancienne population et est éliminée lors de la partie basse de l'oscillation<sup>20</sup>.

Les phases a et b, où la bactérie arrive à fuir quelque peu, correspondent donc à des phases de migrations, ainsi la migration b, fait passer la bactérie de caractéristiques [001111010101] à celle de caractéristiques [111001111111]. Le phage met un peu de temps à "réagir". Cette migration s'accompagne d'un accroissement de l'entropie (figure 4.27).

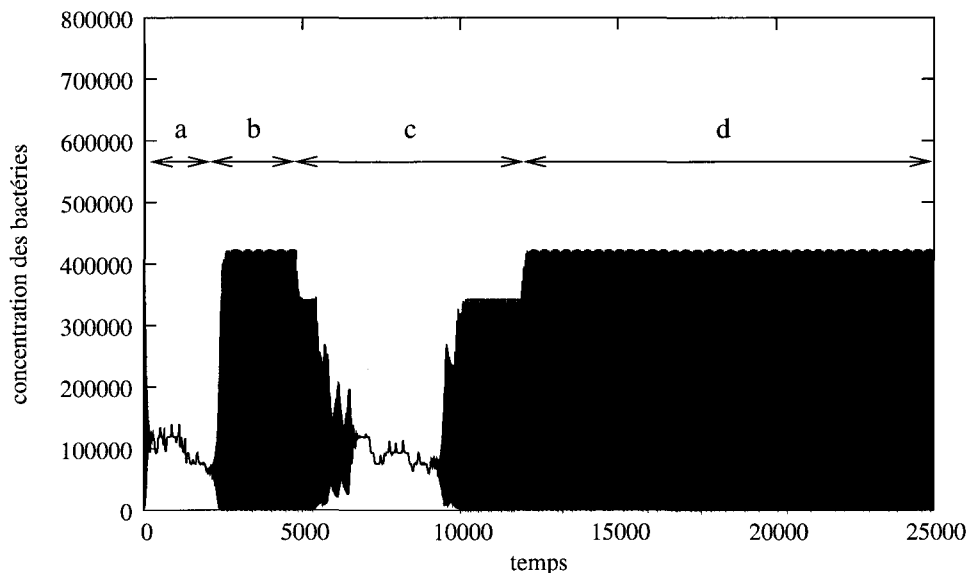


FIG. 4.25 – Evolution de la dynamique au cours du temps des populations de bactéries pour  $h = 0.1$ ,  $n_a = 100$ ,  $\mu_B = 10^{-6}$  et  $\mu_P = 10^{-7}$ . On observe une alternance de phases présentant de fortes oscillations (b et d) et des phases migratoires (a et c). Les phases à forte oscillations correspondent à une forte attaque phagique.

Il peut-être intéressant de regarder un diagramme de phase bactéries-phages (figure 4.28). On ne considère que la concentration totale des populations (leurs taille), indépendamment de leur composition. On n'a alors qu'un aspect dynamique. On peut observer alors sur la figure, une superposition d'attracteurs :

<sup>20</sup>Respectant ainsi le proverbe yiddish suivant :

ביז דער פעטער ווערט מאַגער, גייט דער מאַגערער צו גרונט.

[Biz der feter vert moger, git der mogerer tsu grunt.]

Avant que le gros ne maigrisse, le maigre va en terre.

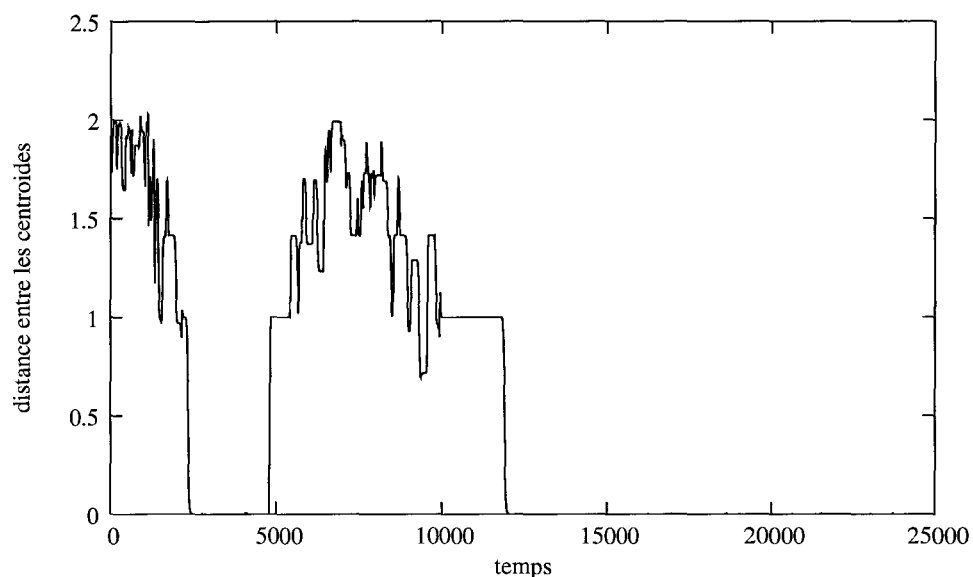


FIG. 4.26 – Distance entre les centroïdes des nuages des bactéries et des phages durant la même simulation que celle de la figure précédente ( $h = 0.1$ ,  $n_a = 100$ ,  $\mu_B = 10^{-6}$  et  $\mu_P = 10^{-7}$ ).

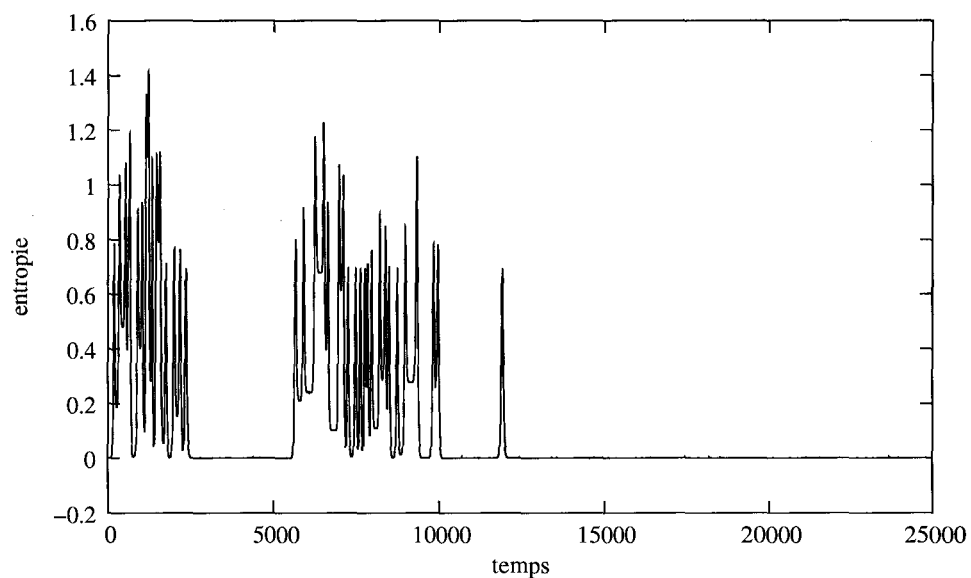


FIG. 4.27 – Tracé de l'entropie de la population bactérienne durant la même simulation que celle de la figure précédente ( $h = 0.1$ ,  $n_a = 100$ ,  $\mu_B = 10^{-6}$  et  $\mu_P = 10^{-7}$ ).

- le cycle limite correspondant au méta-état  $n_P = n_B = 1$ ,  $d_H(c_P, c_B) = 0$ , que nous appelons sur la figure “cycle 0”,
- le cycle limite correspondant au méta-état  $n_P = n_B = 1$ ,  $d_H(c_P, c_B) = 1$ , que nous appelons sur la figure “cycle 1”,
- le cycle limite correspondant aux méta-états proche de  $n_P = n_B = 1$ ,  $d_H(c_P, c_B) = 2$ , que nous appelons sur la figure “cycle 2” (moins nette que les précédents),
- la zone centrale correspondant aux migrations.

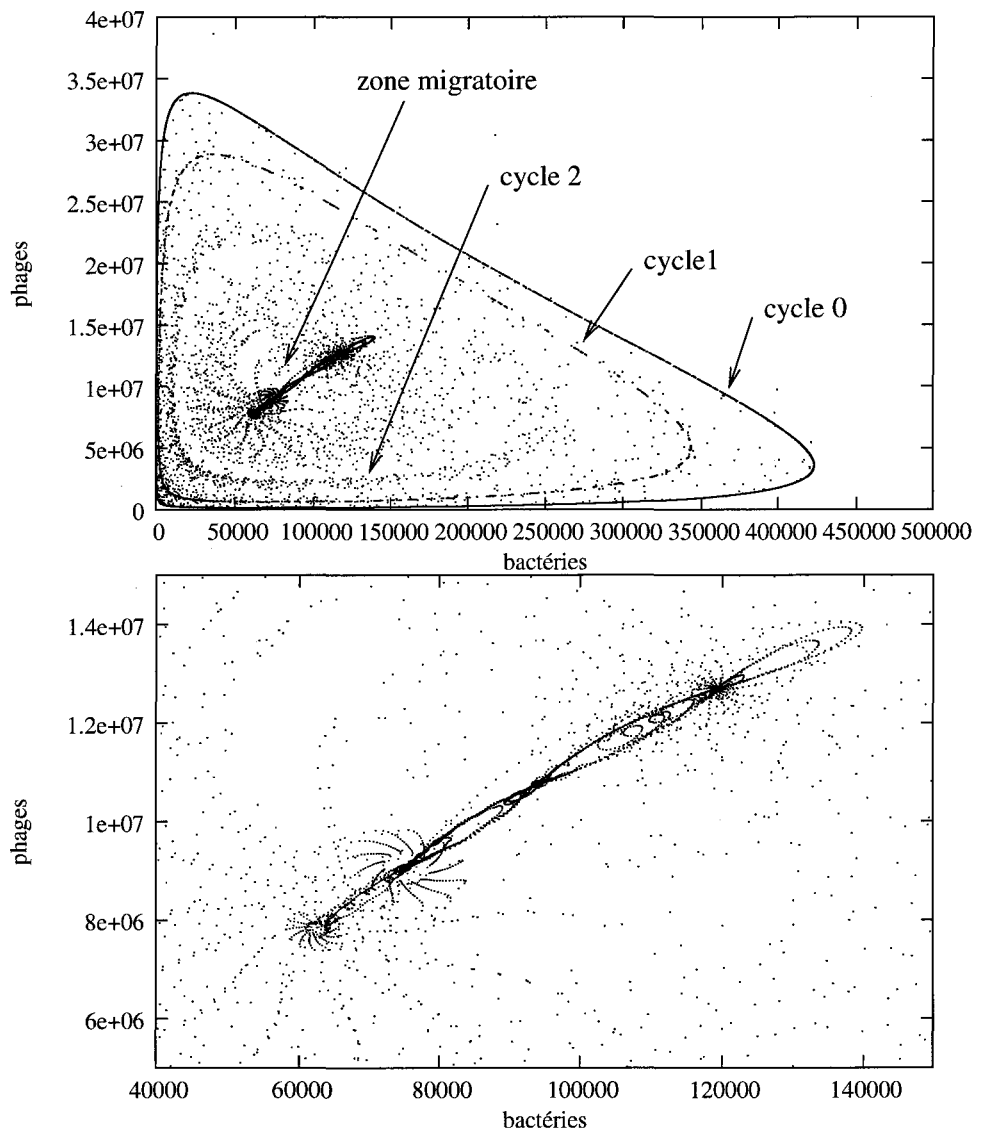


FIG. 4.28 – Tracé des couples  $(B(t), P(t))$  au cours du temps. On observe des zones de plus grandes densités correspondant aux différents attracteurs dynamiques traversés par le SAM. Agrandissement de la zone “centrale” du graphique du dessus.

D'un point de vue métadynamique, si l'on identifie les méta-états correspondants au cycle 0 (même espace d'état, même fonction de transition, on peut le faire car tous les bits des CB sont caractéristiques), on a une alternance entre le "méta-état cycle 0" et une "suite de méta-états de transition".

### Petite poursuite entre amis

La fréquence de mutation bactérienne est maintenant plus grande permettant aux bactéries d'échapper au joug des phages.

**Structure dynamique** De la même manière que dans le cas précédent, on peut observer l'évolution de la concentration totale de bactérie dans la solution (figure 4.29). Contrairement au cas précédent (figure 4.25), on ne voit pas d'apparition de motifs réguliers. Le tracé du diagramme de phases confirme cette impression (figure 4.30). Tout au plus peut-on dire que, le fait qu'il n'y ait pas de fortes oscillations nous permet de dire que les bactéries arrivent à maintenir un avantage évolutif sur les phages et le fait que cette évolution est très hachée, nous permet de dire que les phages restent, malgré tout, près (au sens de la distance des centroïdes).

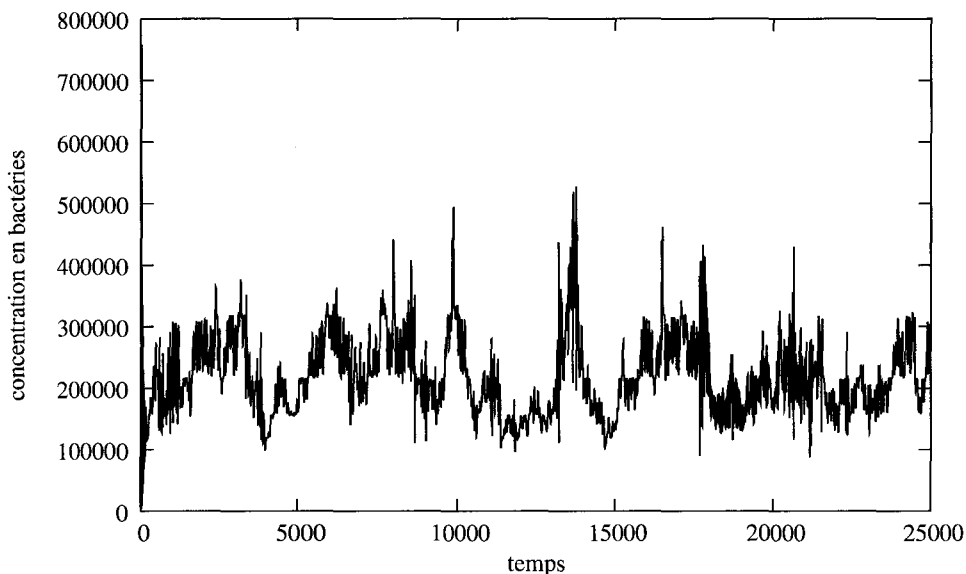


FIG. 4.29 – Evolution de la dynamique au cours du temps des populations de bactéries, avec  $h = 0.1$ ,  $n_a = 100$ ,  $\mu_B = 5.10^{-6}$ .

**Coévolution et effet Reine Rouge** Si l'on considère que la distance entre les centroïdes donne une mesure de l'adaptation des bactéries, alors on peut voir (figure 4.31) que cette adaptation est en perpétuel changement mais reste dans une plage de valeur relativement resserrée. De même, l'observation de l'entropie

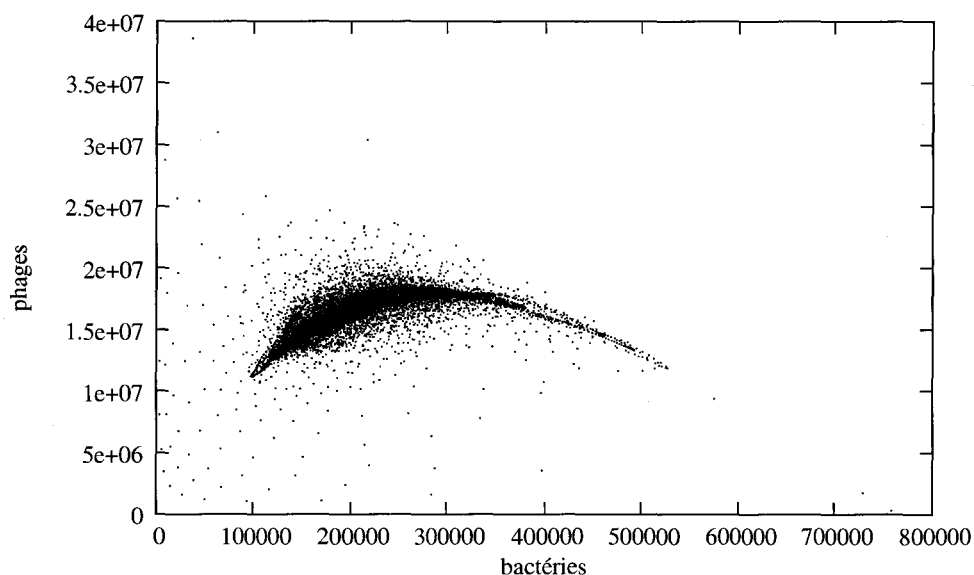


FIG. 4.30 – Tracé des couples  $(B(t), P(t))$  au cours du temps pour  $h = 0.1$ ,  $n_a = 100$ ,  $\mu_B = 5.10^{-6}$  et  $\mu_P = 10^{-7}$ . Contrairement à la figure 4.28, la structure des points semble ici beaucoup moins évidente.

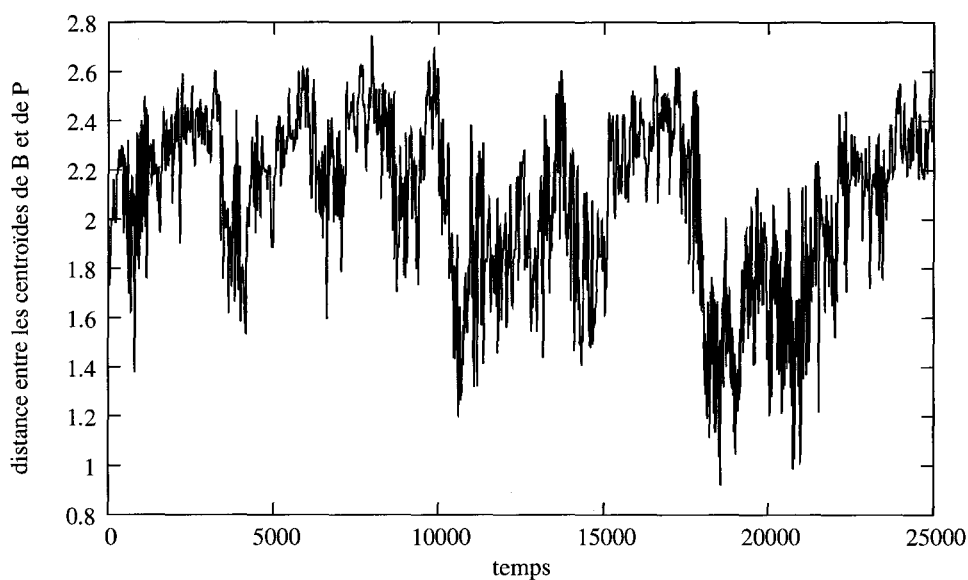


FIG. 4.31 – Distance entre les centroïdes des phages et des bactéries pour des valeurs de  $h = 0.1$ ,  $n_a = 100$ ,  $\mu_B = 5.10^{-6}$ .



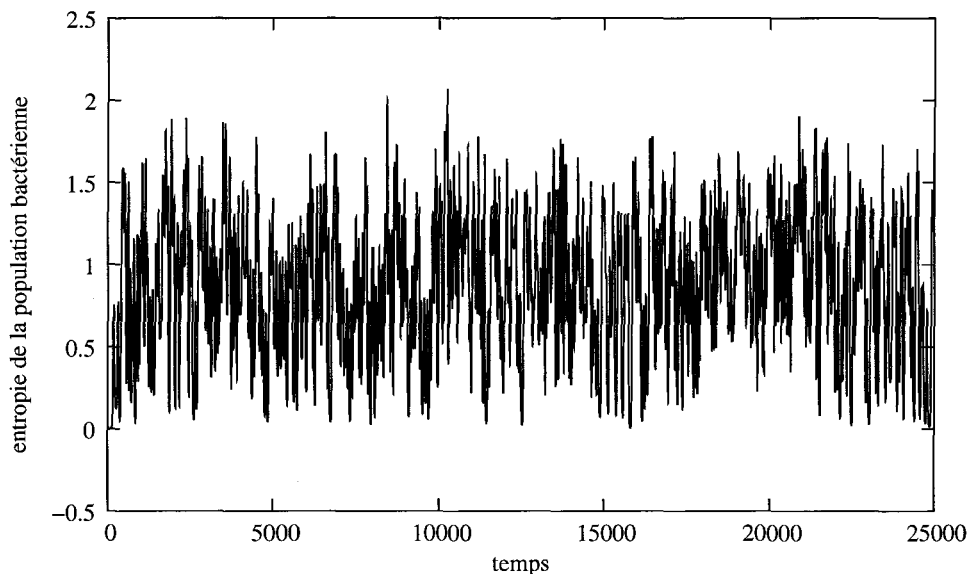


FIG. 4.32 – Entropie de la population bactérienne toujours avec  $h = 0.1$ ,  $n_a = 100$ ,  $\mu_B = 5.10^{-6}$ .

reflète ce changement perpétuel, avec une population dont la dispersion oscille fortement. Tous ces indices nous permettent de supposer l'existence de l'effet Reine Rouge. On rappelle que l'effet Reine Rouge consiste en une coévolution entretenue et simultanée de deux populations, dont les adaptations sont liées : l'amélioration de l'adaptation de l'une implique une baisse de l'adaptation de l'autre ; pour cela chacune des deux populations évoluent, mais conservent une adaptation relativement constante (voir 1.2.1). Nous devons pour cela observer ce qui se passe dans "l'espace des caractéristiques". Nous avons pour cela, en utilisant la technique de l'analyse en composantes principales (voir section 4.4.2), projeté l'évolution des centroïdes des bactéries (partie haute) et des phages (partie basse). Nous avons divisé la trajectoire en cinq étapes et nous avons tracé le rectangle correspondant à la zone dans lequel évolue la trajectoire pendant cette période (figure 4.33).

Ainsi au temps 1, les bactéries sont dans leur zone 1 et les phages dans la leur. On observe alors le comportement suivant : les phages semblent suivre à distance les bactéries qui bougent dans l'espace des caractéristiques. Bien que ce mouvement leur permettent une meilleure adaptation, la "chasse" effectuée par les phages condamne les bactéries à continuer à fuir sans augmenter leur adaptation (elles doivent "courir" pour "rester sur place") : on est donc en présence de l'effet Reine Rouge.

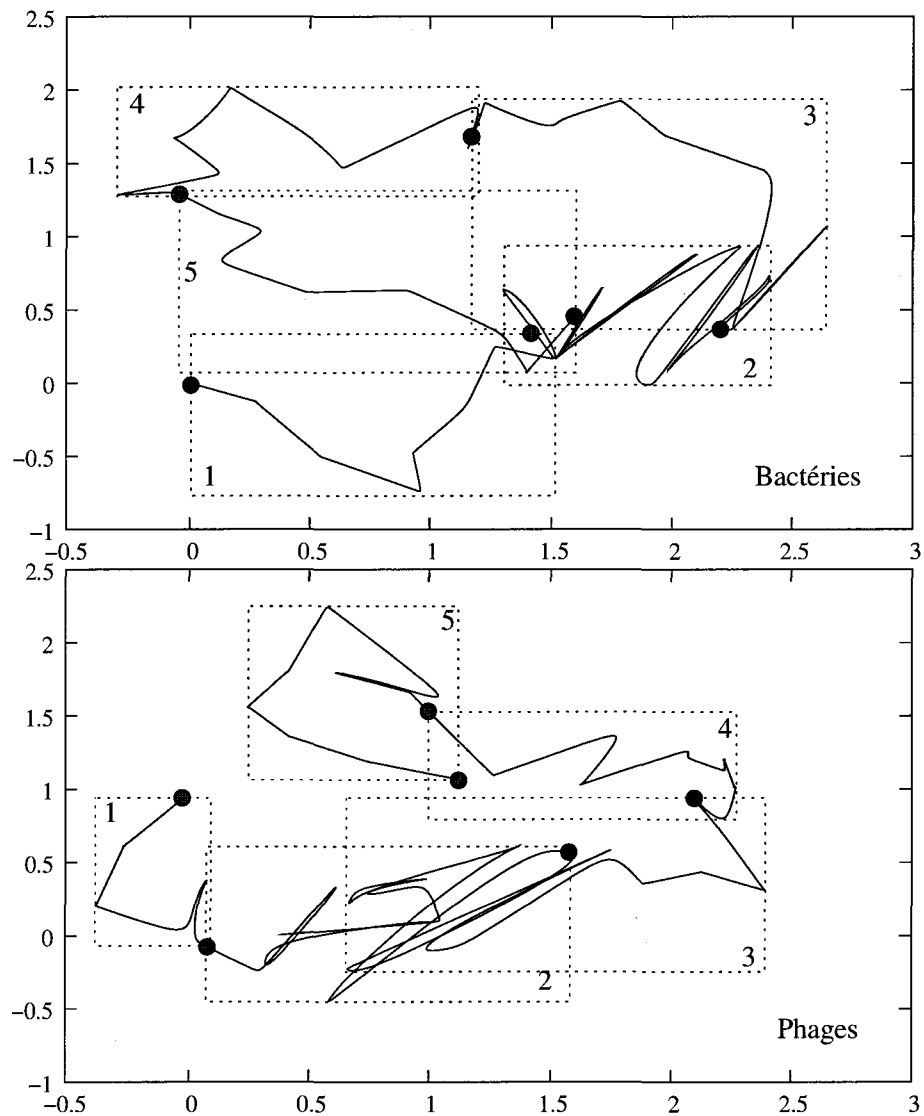


FIG. 4.33 – Poursuite entre phages et bactéries. Trajectoires des centroïdes et position respective (rectangles pointillés) pendant cinq périodes consécutives. Les petits disques indiquent les positions des centroïdes entre deux périodes. Les projections sont obtenues en utilisant la méthode de l'ACP.

## 4.6 Conclusions

Nous avons, dans ce chapitre, mis en œuvre les principes de modélisation vu au chapitre 3. Nous avons ainsi mis en place un modèle de coévolution bactéries-phages sous forme d'un SAM. Ce modèle nous a servi de support dans le but d'une meilleure compréhension des difficultés d'implémentation, d'analyse et de visualisation de la structure formelle de SAM vue au chapitre 2.

Nous avons également obtenu des résultats importants d'un point de vue de la modélisation biologique, nous avons en particulier montré comment, les modèles de quasi-espèce de Eigen pouvaient être vus comme un cas limite de notre modèle quand le taux de mutation des phages est très supérieur à celui des bactéries. Dans le cas général où les deux populations évoluent, nous avons montré que sous certaines conditions, on pouvait assister à un effet Reine Rouge.

Néanmoins, ce modèle ne représente qu'un premier pas vers d'autres modèles prenant plus en compte la réalité biologique qui reste dans notre modèle encore relativement métaphorique. Ainsi, les points suivants constituent des pistes intéressantes pour des perspectives futures :

- ajout d'interactions entre bactéries comme par exemple les sécrétions de toxines ou d'hormones,
- compétition pour différents nutriments comme c'est le cas dans la plupart des écosystèmes,
- coopération métabolique (les produits de certaines bactéries devenant les substrats d'autres espèces) et informationnelle,
- prise en compte du mécanisme de mutations dirigées [67],
- un plus grand réalisme des interactions bactéries-phages, en tenant compte par exemple de remarques formulée dans [50].

Nous obtiendrons ainsi un modèle beaucoup plus réaliste d'écosystème bactérien. En permettant alors à plus de caractéristiques (différents coefficient) d'évoluer génétiquement, nous pourrions par exemple évaluer de façon assez fidèles des résultats tels que ceux de Lenski [?] et caractériser les formes de connectivité du graphe évolutif dont les nœuds seraient les populations et les arcs les différents types d'interactions. En plus des arcs de type "chaîne alimentaire" un tel graphe présenterait-il une connectivité critique de type "*small world*" ?



# Chapitre 5

## Conclusions et perspectives

“Livet skal forst s baglaens, men leves forlaens”<sup>1</sup>

S REN KIERKEGAARD

Nous effectuons dans ce chapitre une synth se des principales notions et des r sultats les plus importants expos s en d tail dans ce m moire. Nous exposons ensuite les perspectives de ce travail, dont un grand nombre correspondent   des voies que nous avons d j  commenc    explorer : pour certaines, des r sultats pr liminaires ont d j   t  obtenus.

### 5.1 Synth se

A partir de la notion tr s classique de syst me dynamique (SD), nous avons d fini au chapitre 1 (de fa on relativement informelle), la notion de syst me dynamique adaptatif complexe (SDAC), bas e sur l’analyse de diff rents exemples de la litt rature (r seaux autocatalytique de Kauffmann, algorithmes g n tiques, chimies artificielles...).

Un SDAC est d fini de la fa on suivante :

- c’est une interconnexion d’ l ments caract ris s par diff rentes propri t s (un syst me),
- les propri t s de ses  l ments varient au cours du temps (caract re dynamique),
- les  l ments sont en grand nombre et interconnect s de mani re non-lin aire (caract re complexe),
- sa structure change au cours du temps de mani re constructive (caract re adaptatif).

Nous avons alors d fini les deux fils rouges de notre th se :

---

<sup>1</sup>“Life is to be understood backwards, but it is lived forwards”

- l’élaboration d’une définition formelle permettant de regrouper sous une même formulation les exemples de SDAC passés en revue et d’autres du même style,
- l’étude d’un type particulier de SDAC, les modèles de populations soumis à l’évolution darwinienne.

Le reste du chapitre a été consacré à exposer les approches existantes concernant ces deux problèmes, ainsi qu’à présenter les notions de base nécessaires à leur compréhension (par exemple quelques points importants de la théorie de l’évolution).

Le chapitre 2 traitait du premier problème, la caractérisation formelle des SDAC. Partant de la définition classique des systèmes dynamiques de Kalman, notre principal résultat est la définition axiomatique des systèmes à métadynamique (SAM) nous permettant de réunir un grand nombre de SDAC (publié dans [8]). Un SAM est défini par :

- deux échelles de temps imbriquées,
- un ensemble d’espaces d’état et de fonctions de transition, des fonctions d’entrées et de sorties, permettant de constituer un ensemble de systèmes dynamiques différents mais partageant les fonctions d’entrées et de sortie,
- une règle métadynamique qui vient, à son échelle de temps propre et en fonction de l’état du système, modifier la structure du système en le faisant changer de système dynamique courant,
- des axiomes permettant de déduire l’évolution au cours du temps d’un tel système et la possibilité de sa réalisation.

Nous avons ensuite discuté d’extensions possibles de cette définition de base et traité plus en détails le cas stochastique qui forme la base de notre étude du système évolutif du chapitre 4.

La portée des systèmes couverts par la définition des SAM est très grande et nous a permis de développer un exemple original donnant un algorithme pour tracer de manière efficace une fractale appelée ensemble de Julia. Nous avons enfin donné brièvement des définitions tendant à montrer qu’une étude théorique des “propriétés métadynamiques” devrait être intéressante.

Dans l’optique de la modélisation des systèmes de population nous avons centré notre étude sur un type plus particulier de SAM. Beaucoup des systèmes étudiés dans la littérature (et maintenant couverts par notre nouvelle définition) appartiennent à une classe particulière de SAM : les SAM à base continue. C’est ainsi le cas de l’exemple du chapitre 4. C’est pourquoi nous avons approfondi et justifié au chapitre 3 cette notion de SAM à base continue.

Nous avons commencé par considérer un certain type de modèles appelés “modèles particuliers” basés sur la simulation stochastique de réactions chimiques généralisées (RCG) à l’aide de l’algorithme de Gillespie. Nous avons montré alors que ce type de modèles, qui constituent un niveau de description intermédiaire entre ceux de la modélisation agents et des équations différentielles, permet de

modéliser les modèles cinétiques de populations les plus classiques en biologie. Nous avons pour cela créé un modèle original de l'écosystème bactéries-amibes de Tsuchiya par un ensemble de RCG. Ce résultat nouveau permet de montrer que, en plus des mécanismes de type action des masses, on peut modéliser avec les RCG des mécanismes de saturation, ouvrant la voie à la modélisation par cette méthode de systèmes biologiques complexes (comme par exemple le système bactéries-phages non-évolutif de Mosekilde).

Nous montrons dans une deuxième partie que, si l'approximation traditionnelle (consistant à remplacer le caractère discret d'une population par une modélisation continue en utilisant des grandeurs telle que la concentration quand le nombre d'éléments dans cette population est grand) aboutit pour des probabilités homogènes à un système d'équations différentielles, on doit, dans le cas où ces probabilités sont hétérogènes, utiliser autre chose, comme un SAM.

Les principaux résultats originaux de ce chapitre sont :

- la réalisation d'une contrepartie stochastique au modèle de Tsuchiya, modèle classique en microbiologie,
- la démonstration de la validité de la structure de SAM à base continue pour modéliser une population dont les événements internes sont soumis à des probabilités hétérogènes.

Le chapitre 4 a traité en détail un exemple biologique de SAM modélisant des populations en coévolution. Cet exemple est particulièrement important puisqu'il permet de mettre en pratique les notions développées au chapitre 2. En particulier, l'aspect informatique est détaillé, ainsi que les problèmes liés à l'analyse de tels systèmes (comme les variables à sauvegarder lors d'une simulation ou les traitements statistiques à effectuer pour retirer de l'information des données engendrées). Après une analyse théorique du modèle, nous avons utilisé la simulation pour dégager des résultats à portée biologiques; on a ainsi montré que selon les valeurs des paramètres, notre modèle pouvait présenter :

- l'émergence d'une structure de quasi-espèce,
- un effet Reine Rouge.

Les principaux résultats originaux de ce chapitre (les trois premiers points ont été publiés dans [4, 6, 7, 8] sont :

- l'élaboration d'un modèle de coévolution de populations bactériennes et phagiques sur la base d'un SAM,
- l'implémentation informatique de ce système,
- la mise en œuvre de techniques de visualisation et de condensation de l'information,
- la démonstration, au sein de notre système, de l'apparition d'une structure de type quasi-espèce et de l'effet Reine Rouge.

## 5.2 Perspectives

Sur la base des résultats de ce travail, il est possible d'envisager de nombreux prolongements. Nous avons déjà commencé à travailler sur certains d'entre eux.

### 5.2.1 Sur le formalisme des SAM

Nous avons présenté des définitions permettant de décrire formellement les SAM, ainsi que certaines propriétés caractérisant leur changement de structure. Nous avons appliqué ces concepts à l'exemple de populations bactériennes et phagiques en coévolution. Nous avons pu ainsi mettre en œuvre un certain nombre de points importants d'un point de vue pratique et théorique : implémentation informatique d'une telle structure de modélisation, méthodes de visualisation permettant de donner un caractère tangible à l'évolution des méta-états (en utilisant par exemple l'analyse en composantes principales ou des variables agrégées). Cet exemple, bien que très riche en lui même, ne représente qu'une petite portion des systèmes que l'on peut définir comme SAM (voir l'algorithme de tracé de Julia par exemple). Une perspective de notre travail consiste donc à étudier d'autres exemples de SAM, si possible paradigmatiques et les plus simples possibles, afin d'analyser les propriétés qui peuvent apparaître au sein de ces systèmes.

### 5.2.2 Sur le modèle attaque-défense

Le modèle attaque-défense présenté au chapitre précédent est très riche et peut déboucher sur de nombreuses simulations et perfectionnements théoriques autres que ceux déjà explicités.

Les contraintes de temps inhérentes à notre travail nous interdisent de présenter tous ces aspects en détail. Cependant, nos travaux continuent. Nous sommes actuellement en train d'étudier les divers points mentionnés ci-après :

- Dans l'expérience consistant à observer le comportement évolutif d'une population bactérienne face à une population de phages ayant des caractéristiques fixes données, on peut par injection de nouvelles populations phagiques obtenir une sorte de "contrôle" de la trajectoire évolutive du nuage bactérien. Nous avons déjà publié des résultats préliminaires d'un tel contrôle [7]. Néanmoins, la réalité biologique d'une telle procédure est loin d'être, aujourd'hui, évidente. Nous voudrions donc développer un tel type de contrôle, dans un cas plus général, afin d'offrir des perspectives pour l'ensemble des systèmes modélisables par des SAM.
- On a déjà mentionné, en conclusion du chapitre 4, l'intérêt de développer un modèle plus riche du point de vue des interactions entre les composantes de l'écosystème bactérien ou sous des aspects comme la présence de plusieurs sources de nutriment. Nous avons commencé des études sur l'organisation émergente d'un système du type de celui du chapitre 4, et comprenant, non



plus un seul, mais deux substrats. Nous avons discuté de cette expérience dans [6].

### 5.2.3 Sur la modélisation de la théorie de l'évolution

Nous présentons au chapitre 4, un modèle de coévolution bactéries-phages. Un des objectifs de futurs développements théoriques serait de formaliser de manière rigoureuse certains concepts de la théorie de l'évolution afin de pouvoir appliquer ces définitions aux modèles évolutifs de populations. De telles définitions permettraient également ainsi de fournir des méthodes pour visualiser ces propriétés évolutives de façon systématique, sans devoir passer par l'interprétation.

# Annexes

La version finale de ce mémoire comportera en annexe :

- un exposé développé des travaux de Bagley
- un bref aperçu de notre programme de simulation
- une publication en anglais de nos travaux

# Bibliographie

- [1] P. AUGER. *Dynamics and Thermodynamics in Hierarchically Organized Systems : Applications in Physics, Biology, and Economics*. IFSR International Series on Systems. Pergamon Press, octobre 1989.
- [2] R.J. BAGLEY, J.D. FARMER, S.A. KAUFFMAN, N.H. PACKARD, A.S. PERELSON, and I.M. STADNYK. Modeling adaptive biological systems. *BioSystems*, 23 :113–138, 1989.
- [3] M. BAGUELIN. Dynamique non-linéaire de modèles simples et complexes de compétition entre bactéries et phages. Master's thesis, DEA d'automatique et d'informatique industrielle, Université des Sciences et Technologies de Lille, 2000.
- [4] M. BAGUELIN. Un modèle hybride de coévolution dans un système bactéries-phages. In *Journées Doctorales d'Automatiques (JDA03)*, Valenciennes, France, 2003.
- [5] M. BAGUELIN and J. LEFÈVRE. Modelling the complex non-linear dynamics of bacterial populations with kinetic process graphs. In *Simulation in Industry*, pages 753–757, 2001.
- [6] M. BAGUELIN, J. LEFÈVRE, and J.P. RICHARD. Preliminary approach to the simulation of coevolution in an ecosystem : the bacterial-phage interactions. In *16<sup>th</sup> European Simulation Multiconference*, pages 466–471, Darmstadt, Deutschland, 2002.
- [7] M. BAGUELIN, J. LEFÈVRE, and J.P. RICHARD. Vers un modèle hybride de coévolution dans un système bactéries-phages. In *Congrès Internationale Francophone d'Automatique (CIFA)*, Nantes, France, 2002.
- [8] M. BAGUELIN, J. LEFÈVRE, and J.P. RICHARD. A formalism for models with a metadynamically varying structure. In *European Control Conference*, 2003.
- [9] J.E. BAILEY and D.F. ALLIS. *Biochemical Engineering Fundamentals*. Chemical Engineering. Mc Graw-Hill International editions.
- [10] P. BAK. *Quand la nature s'organise*. Flammarion, 1999.
- [11] H. BERSINI. Dynamics and metadynamics in biological and chemical networks. [www.info.fundp.ac.be/~pys/DEA/sems00/bersini.ppt](http://www.info.fundp.ac.be/~pys/DEA/sems00/bersini.ppt)

- [12] N. BOULEAU. *Probabilité de l'ingénieur, variables aléatoires et simulation*. Hermann, 1986.
- [13] J.R. BÜCHL. *Finite Automata, Their Algebras and Grammars. Toward a Theory of Formal Expressions*. Springer-Verlag, 1988.
- [14] A. CHURCH. A set of postulates for the foundation of logic i. *Ann. Math.*, 33 :346–366, 1932.
- [15] A. CHURCH. A set of postulates for the foundation of logic ii. *Ann. Math.*, 34 :839–664, 1932.
- [16] P. COQUILLARD and D.R.C. HILL. *Modélisation et simulation d'écosystèmes : des modèles déterministes aux simulations à événements discrets*. Masson, 1997.
- [17] G.A. COWAN, D. PINES, and D. MELTZER, editors. *Complexity : Metaphors, Models and Reality*. Perseus Books, 1994.
- [18] P. DAVID and S. SAMADI. *La théorie de l'évolution, une logique pour la biologie*. Champs Université. Flammarion, 2000.
- [19] R.J. DE BOER and A.S. PERELSON. Size and connectivity as emergent properties of a developing immune network. *J. theor. Biol.*, 149, 1991.
- [20] A. DOUADY and J.H. Hubbard. Étude dynamique des polynômes complexes. *Publ. Math. Orsay*, I & II, 1984-85.
- [21] B. DROSSEL. Biological evolution and statistical physics. *Advances in Physics*, 50 :209–295, Janvier 2001.
- [22] M. EIGEN. Viral quasispecies. *Scientific American*, pages 32–39, July 1993.
- [23] M. EIGEN and P. SCHUSTER. *The Hypercycle, A Principle of Natural Self-Organization*. Springer-Verlag, 1979.
- [24] P. FATOU. Sur les équations fonctionnelles. *Bull. Soc. Math. France*, 48 :208–314, 1920.
- [25] W. FONTANA and L. BUSS. The barrier of objects : From dynamical systems to bounded organizations.
- [26] E. FREIFELDER. *Biologie moléculaire*. Masson, 1990.
- [27] M. GELL-MANN. Complex adaptive systems. In G.A. COWAN, D. PINES, and D. MELTZER, editors, *Complexity : Metaphors, Models and Reality*. Perseus Books, 1994.
- [28] M. GELL-MANN. *Le quark et le jaguar*. Flammarion, 1997.
- [29] D.T. GILLESPIE. Exact stochastic simulation of coupled chemical reactions. *The J. of Phys. Chem.*, 81(25) :2340–2361, 1977.
- [30] R. GODEMENT. *Analyse mathématique I, Convergence, fonctions élémentaires*. Springer, 2<sup>ème</sup> édition, 2001.

- [31] D.E. GOLDBERG. *Genetic Algorithms in Search, Optimization, and Machine learning*. Addison-Wesley, 1989.
- [32] S. J. GOULD. *The Structure of Evolutionary Theory*. Harvard University Press, 2002.
- [33] S.J GOULD and N. ELDREDGE. Punctuated equilibria : the tempo and mode of evolution reconsidered. *Paleobiology*, 3 :115–151, 1977.
- [34] J. HOLLAND. Echoing emergence : Objectives, rough definitions, and speculations for echo-class models. In G.A. COWAN, D. PINES, and D. MELTZER, editors, *Complexity : Metaphors, Models and Reality*. Perseus Books, 1994.
- [35] J.H. HOLLAND. *Adaptation in Natural and Artificial Systems*. The University of Michigan, Ann Arbor, MI.
- [36] J.H. HOLLAND. *Hidden Order : How Adaptation Builds Complexity*. Helix Books. Addison-Wesley Publishing Company, 1995.
- [37] J.H. HOLLAND. *Emergence : From Chaos to Order*. Perseus Books Group, avril 1999.
- [38] J. HUBBARB and B. WEST. *Equations différentielles et systèmes dynamiques*. Cassini, 1999.
- [39] J. ISTAS. *Introduction aux modélisations mathématiques pour les sciences du vivant*, volume 34 of *Mathématiques et Applications*. Springer, 2000.
- [40] G. JULIA. Mémoire sur l'itération des fonctions rationnelles. *J. Math. Pures Appl.*, pages 47–245, 1918.
- [41] R.E. KALMAN, P.L. FALB, and M.A. ARBIB. *Topics in Mathematical System Theory*, chapter Reader's guide. International Series in Pure and Applied Mathematics. McGraw-Hill, 1969.
- [42] S.A. KAUFFMAN. Autocatalytic sets of proteins. *J. theor. Biol.*, (119) :1–24, 1986.
- [43] S.A. KAUFFMAN. *The Origins of Order : Self-Organization and Selection in Evolution*. Oxford University Press, mai 1993.
- [44] S.A KAUFFMAN and S. LEVIN. *J. Theor. Biol.*, 128 :11, 1987.
- [45] M. KIMURA. *Nature*, 27 :624, 1968.
- [46] D.E. KNUTH. *The Art of Computer Programming*, volume 2, Seminumerical Algorithms. Addison-Wesley, 3<sup>ème</sup> édition, 1998.
- [47] J.-U. KREFT, G. BOOTH, and J.W.T. WIMPENNY. Bacsim, a simulator for individual-based modelling of bacterial colony growth. *Microbiology*, 144 :3275–3287, 1998.
- [48] H. LE GUYADER, editor. *L'évolution*. Bibliothèque Pour La Science, 1998.
- [49] J. LEFÈVRE. Modélisation de systèmes complexes, rapport d'habilitation à diriger des recherches. Ecole Centrale de Lille, Janvier 2000.

- [50] R.E LENSKI. Coevolution of bacteria and phage : Are there endless cycles of bacterial defenses and phage counterdefenses. *J. Theor. Biol.*, 108 :319–325, 1984.
- [51] C. LOBRY. *et pourtant... ils ne remplissent pas N!* ALEAS Editeur, septembre 1989.
- [52] M. LUCK, M. MARIK, O. STEPANKOVA, and R. TRAPPL, editors. *Multi-Agent Systems and Applications by*. Springer-Verlag, décembre 2001.
- [53] J.M. MAYNARD-SMITH. *Evolution and the Theory of Games*. Cambridge University Press, 1982.
- [54] H.H. MCADAMS and A. ARKIN. Stochastic mechanisms in gene expressions. *Proc. Natl. Acad. Sci. USA*, 94 :814–819, 1997.
- [55] E. MOSEKILDE, H. STRANDORF, J. S. THOMSEN, and G. BAIER. *Topics in NonLinear Dynamics*, chapter A Hierarchy of Complex Behaviors in Microbiological Systems, pages 165–208. World Scientific, 1996.
- [56] G. NICOLIS and I. PRIGOGYNE. *Exploring complexity*. W. H. Freeman, 1989.
- [57] G. NICOLIS and I. PRIGOGYNE. *A la rencontre du complexe*. Philosophie d'aujourd'hui. Presses universitaires de France, 1992.
- [58] U. NIESERT, D. HANNASCH, and C. BRESCH. Origins of life between scylla and charybdis. *J. of Molecular Evolution*, page 348, 1981.
- [59] H.-O. PEITGEN and P.H. RICHTER. *The Beauty of Fractals*. Springer-Verlag, Heidelberg, 1986.
- [60] H.-O. PEITGEN and D. SAUPE. *The Science of Fractal Images*. Springer-Verlag, 1988.
- [61] J.C. PRINCIPE, N.R. EULIANO, and W.C. LEFEBVRE. *Neurals and Adaptive Systems, Fundamentals Through Simulations*. John Wiley & sons, 1999.
- [62] J.-P. REGNAULT. *Microbiologie générale*. Décarie-Vigot, 1990.
- [63] E. RENSHAW. *Modelling Biological Populations in Space and Time*. Cambridge University Press, 1993.
- [64] G.W. ROWE. *Theoretical models in Biology, the origin of life, the immune system and the brain*. Oxford Science Publications, Oxford, United Kingdom.
- [65] G. SAPORTA. *Probabilités, Analyse de données et Statistiques*. Editions TECHNIP, 1990.
- [66] P. SCHUSTER. Complex optimization in an artificial rna world. In *Artificial life II*, volume X of *Santa Fe Institute Studies in the Sciences of Complexity*, pages 277–291, Redwood City, CA, 1992. Addison-Wesley.
- [67] P.D SNIEGOWSKI, P.J. GERNISH, and al. The evolution of mutation rates : separating causes from consequences. *Bioessays*, 22(12) :1057–1066, 2001.
- [68] R. SOLÉ and B. GOOWIN. *Signs of Life : How Complexity Pervades Biology*. Basic Books, janvier 2002.

- [69] R.V. SOLÉ, R. FERRER, I. GONZÁLEZ-GARCIA, J. QUER, and E. DOMINGO. Red queen dynamics, competition and critical points in a model of rna virus quasispecies. *Submitted to J. of Mol. Bio.*
- [70] I. STENGERS and I. PRIGOGINE. *Entre le temps et l'éternité*, volume 262 of *Champs*. Flammarion, 1992.
- [71] S.H STROGATZ. *Nonlinear Dynamics and Chaos, with Applications to Physics, Biology, Chemistry, and Engineering*. Addison-Wesley, 1994.
- [72] M.M. WALDROP. *Complexity, the emerging science at the edge of order and chaos*. Simon & Schuster, 1992.
- [73] D.J WATTS. *Small Worlds*. Princeton University Press, 1999.
- [74] P.A. WEISS. *l'archipel scientifique. recherches interdisciplinaires*. Maloine éditeur, Paris, 1974.
- [75] B.P. ZEIGLER, H. PRAEHOFER, and T.G. KIM. *Theory of modeling and simulation, Integrating discrete events and continuous complex dynamic systems*. Academ Press, 2001.
- [76] J. ZIEGLER, P. DITTRICH, and W. BANZHAF. *Information Processing in Cells and Tissues*, chapter Towards a Metabolic Robot Control System. Plenum Press, NY, 1997.

# Index

- algorithme
  - de Gillespie, 73
  - et structure de données, 105
  - génétique, 16
- alimentation, 117
- analyse en composantes principales, 107, 138
- Auger (Pierre), 22
- bactériophage, 112
- Brusselator*, 80
- centroïde, 137
- chaînes de Markov, 93
- chemostat, 109
- chimies artificielles, 22
- cinématique enzymatique de Michaëlis-Menten, 88
- complexité du calcul numérique
  - comparaison particulière vs. continu, 96
- concentration massique pondérée (convergence), 129
- culture bactérienne, 109
- cycle limite, 81, 85
- dérive génétique, 33
- dilution, 110, 117
- distance de Hamming, 135
- écosystème proie-prédateur simple, 76
- entropie, 107, 136
- évènement cinétique élémentaire, 71
- évolution (théorie de l'), 27
- fonction de transition, 44
- hyperchaos, 130
- hypercycles de Eigen, 21
- infection d'une cellule bactérienne, 118
- Julia (ensemble de Julia)
  - tracé, 51
- Linux, 125
- logiciel libre, 125
- loi
  - cinétique d'une réaction, 71
  - exponentielle, 67
  - sans vieillissement, 67
- Lotka
  - réactions chimiques, 75
- Lotka-Volterra (équations de), 76
- lyse, 112, 113
- méta-état, 57
  - attracteur, 58
  - répulsif, 58
- méta-orbite, 57
- méthode numérique d'analyse des données, 106
- microorganisme, 110
- milieu minimum, 110
- modèle particulière évolutif, 98
- Monod (équation de), 84, 117
- Mosekilde (équations de), 122
- mutation
  - défavorable, 99
  - favorable, 99
  - neutre, 99
- nuage
  - définition, 136
- Oregonator*, 81



- population
  - définition, 136
- principe de Hardy-Weinberg, 29
- principe de Weiss, 47, 55
  
- quasi-espèce, 36, 146
  
- réaction chimique généralisée, 71
- réseaux autocatalytiques de Kauffman,  
19
- Reine Rouge (effet), 30, 156
- risque instantané, 67
  
- SAM
  - équivalent, 100
- satiété, 84
- saturation, 84, 87
- simulation
  - d'une loi exponentielle, 70
  - stochastique, 71
- small worlds*, 21
- Solé, 22, 34
- stratégies évolutivement stables, 32
- support d'une population, 136
- système, 9
  - à métadynamique
    - définition formelle, 45
  - dynamique, 9
    - adaptatif, 11
    - complexe, 9
    - définition de Kalman, 43-45
  
- transition métadynamique, 100
- Tsuchiya (modèle de), 81-91
  
- Volterra (Vito), 76

