

ANALYSE DES SYSTEMES BIOLOGIQUES:
CERTAINS ASPECTS METHODOLOGIQUES LIES
A LA MODELISATION

o o o o

ARLETTE CHERUY (1), CHRISTIAN GAUTIER (2), ALAIN PAVÉ (2)

(1) LABORATOIRE D'AUTOMATIQUE DE GRENOBLE, Institut National Polytechnique
(LA 228) 38400 St-MARTIN-D'HERES.

(2) LABORATOIRE DE BIOMETRIE, Université Cl. BERNARD
(LA 243) 69621 VILLEURBANNE



SOMMAIRE

	<u>PAGES</u>
I - <u>INTRODUCTION</u>	1
II - <u>PRINCIPAUX TYPES DE MODELES</u>	4
II.1 <u>Analyse non dynamique</u>	5
II.1-1 Comparaison avec des modèles de références	5
<u>Exemple 1</u> : Production de soie par Bombyx Mori	5
<u>Exemple 2</u> : Structure d'un gène	7
<u>Exemple 3</u> : Rôle des granules du myocarde	9
II.1-2 Analyse de données	12
<u>Exemple 4</u> : Fréquence des codons dans les gènes	12
II.1-3 Modèles propres au phénomène étudié	16
<u>Exemple 5</u> : Modélisation des chromosomes	17
II.2 <u>Analyse dynamique</u>	18
II.2-1 Modèles déterministes en temps continu	20
a) Modèles régis par des équations différentielles ordinaires	20
<u>Exemple 6</u> : Transport intracellulaire de la fibroïne dans la glande séricigène de bombyx mori	22
<u>Exemple 7</u> : Analyse du système coagulolytique	25
<u>Exemple 8</u> : Une approche de la dynamique des macromolécules	27

	<u>PAGES</u>
b) Modèles régis par des équations aux dérivées partielles	30
II.2-2 Modèles déterministes en temps discret	31
<u>Exemple 9</u> : Fonctions de croissance dans les systèmes de Lindenmayer	32
II.2-3 Modèles stochastiques en temps continu	34
<u>Exemple 10</u> : Modèle de la cible : application à l'étude de la synthèse des chloroplastes	35
<u>Exemple 11</u> : Analyse des courbes de croissance d'individus ou de populations	37
II.2-4 Modèles stochastiques en temps discret	38
i) Processus de Markov	38
ii) Processus de ramification	39
II.3 <u>Conclusion</u>	41
III - <u>PROBLEMES LIES A LA CONSTRUCTION DES MODELES</u>	41
III.1 <u>Problèmes liés à la délimitation et à la caractérisation du système</u>	44
III.1-1 Délimitation du système	44
III.1-2 Caractérisation	44
III.2 <u>Problèmes liés à la simulation et à l'identification des paramètres</u>	45
III.2-1 Simulation	46
III.2-2 Estimation des paramètres	47

	<u>PAGES</u>
III.2-3 <i>Validation</i>	48
III.2-4 <i>Identifiabilité des systèmes</i>	49
III.3 <u><i>Problèmes liés à l'expérimentation</i></u>	51
IV - <u>CONTRIBUTION DE LA MODELISATION A L'ETUDE DES SYSTEMES BILOGIQUES</u>	52
IV.1 <u><i>Analyse de structures biologiques</i></u>	53
IV.2 <u><i>Analyse et test de mécanismes</i></u>	53
<u>Exemple 12</u> : Détermination du mécanisme de réaction d'une enzyme à partir de l'analyse de ses cinétiques	54
IV.3 <u><i>Prédiction et simulation de systèmes biologiques</i></u>	56
IV.4 <u><i>Estimation de paramètres non directement accessibles</i></u>	57
IV.5 <u><i>Optimisation de protocoles expérimentaux</i></u>	59
IV.6 <u><i>Contrôle de procédés biologiques</i></u>	61
V - <u>CONCLUSION</u>	63



I - INTRODUCTION

La pratique des biologistes est essentiellement fondée sur une approche analytique de l'exploration des systèmes vivants, laquelle est conduite de façon analogue à une dissection, même s'il s'agit de "biologie cellulaire ou moléculaire" (les techniques biochimiques tiennent lieu de bistouri et de microscope). Ainsi les résultats sont souvent plutôt qualitatifs que quantitatifs et au niveau de l'expérimentation, les problèmes méthodologiques sont, la plupart du temps, soit de définir la meilleure expérience pour observer une structure ou pour mettre en évidence un mécanisme, soit d'utiliser au mieux les données expérimentales accessibles.

Outre les techniques purement expérimentales, l'usage d'outils fondés sur la notion de modèle, notamment celle de modèle mathématique sur laquelle nous nous arrêterons, entre de plus en plus fréquemment dans la pratique de l'analyse des systèmes biologiques. En plus d'une aide à l'expérimentation classique, on a vu se développer la nécessité d'une quantification de plus en plus importante et la possibilité d'analyser de façon plus précise certains phénomènes, notamment la dynamique des processus biologiques.

Bien qu'il ne soit pas de notre propos de présenter l'aspect strictement expérimental, il n'empêche que nous y ferons référence car il nous semble que toute approche formelle d'un système "naturel" ne peut se détacher de l'expérimentation dans la mesure où celle-ci est possible [LEGAY, (1973)]. Il s'agit, en fait d'outils complémentaires qui permettent de prolonger, ou même d'optimiser, cette expérimentation.

Brièvement, nous signalons qu'au plan historique, les approches formelles ont souvent été tirées de domaines voisins (modèles de la biochimie, de la biophysique...). La génétique se distingue car, dans ce cadre, une formalisation propre a été développée [FISHER (1950), MALECOT (1966) ...]. Parallèlement, il nous semble que certains aspects de la dynamique des populations puisse fournir des outils précieux au niveau biologique (cf. par exemple la dynamique des macromolécules). Enfin, d'autres approches originales sont en gestation notamment en ce qui concerne les problèmes de forme : morphologie

et morphogénèse (théorie des catastrophes, modèles de la théorie des langages, objets fractals).

L'un des premiers apports de l'approche formelle est d'amener le biologiste à spécifier, i.e. définir exactement le système qu'il cherche à étudier, et à préciser les hypothèses et les objectifs avant toute tentative de modélisation. En fait, un modèle est une image simplifiée, sinon simpliste, de la réalité, ou un outil pour simplifier la réalité. Son élaboration et son utilisation dépendent d'objectifs que l'on cherche à atteindre. Ainsi, il n'est pas impossible que pour un même système on soit conduit à plusieurs modèles différents suivant les objectifs poursuivis. Les principaux objectifs auxquels doit répondre une modélisation, dans l'état actuel de cette méthodologie, peuvent être cités brièvement :

- . analyse de données, dans un but de réduction de données ou de détection de structures,
- . estimation d'un ou plusieurs paramètres,
- . prédiction, c'est-à-dire prédire l'évolution future du système en fonction de son passé, en tenant compte ou non des perturbations (cet objectif inclut la prévision : prédiction en l'absence de perturbations),
- . optimisation d'expériences, c'est-à-dire de proposer des procédures expérimentales qui donnent le maximum d'informations au moindre coût,
- . visualisation de variables intermédiaires par la simulation du système,
- . commande d'un système, qui consiste à définir une action sur un système pour contrôler son comportement.

Un certain nombre de solutions acceptables ont été trouvées pour certains de ces problèmes dans le cas de systèmes physiques ou technologiques. Mais, lorsque la transposition ou l'adaptation de ces méthodes est tentée dans le domaine biologique, on est confronté à certaines difficultés résultant essentiellement de la complexité des systèmes vivants :

- La première provient de la délimitation du système lui-même et de la caractérisation de son environnement. En effet, les systèmes biologiques sont fortement interconnectés, "isoler" l'un d'entre eux est rarement une opération simple.

- La seconde vient des nombreuses interactions qui se traduisent par des non-linéarités induisant des problèmes mathématiques au niveau de l'utilisation du modèle.

- La troisième est en rapport avec la métrologie des variables biologiques ; elles sont souvent inaccessibles ou non directement mesurables, particulièrement in vivo (problème de "capteurs"), et lorsqu'elles sont mesurées, elles sont souvent très bruitées. De plus, les expériences sont coûteuses et difficiles à mettre en oeuvre. En général, on est confronté à des ensembles limités de données, la plupart du temps peu précises (variabilité "bien connue" des variables biologiques). Cette variabilité intervient à deux niveaux : le niveau individuel, conséquence de la précision des mesures et de la non stationnarité des phénomènes, le niveau "population", conséquence des variations interindividuelles. Signalons qu'une source d'erreur souvent confondue ou intégrée dans la variabilité interindividuelle concerne le processus d'échantillonnage dans la population.

La prise en compte d'un type de variabilité dépend principalement de l'objectif, par exemple "si le but est de prescrire un traitement médical, la posologie efficace pour un patient donné peut être catastrophique pour un autre si la variabilité de la population est négligée" [BOISVIEUX (1978)]. On peut noter en outre que la variabilité peut être une source importante d'informations : la décomposition de la variabilité dépend aussi de l'objectif choisi. Ainsi, par exemple, dans une étude génétique, la variabilité entre individus pourra être considérée comme somme des variabilités entre famille et entre les individus d'une même famille.

Dans la suite du texte, après une présentation brève de quelques types de modèles rencontrés en Biologie, nous détaillerons certains problèmes importants liés à leur construction et à leur utilisation (problèmes expérimentaux, théoriques, numériques...), notamment ceux auxquels nous avons été confrontés dans nos études personnelles. Un certain nombre de points

sont illustrés par des exemples précis que nous avons développés de préférence à des travaux tirés de la littérature. Bien que nos références soient essentiellement empruntées au domaine de la biologie cellulaire, certains aspects seront tirés d'autres secteurs de la biologie, ou même de l'écologie, dans la mesure où ils nous semblent complémentaires et pouvant présenter un certain intérêt pour de futurs développements en biologie cellulaire.

Ce dernier aspect nous paraît important. En effet, ce qui semble, à l'heure actuelle, constituer une base de "l'approche Analyse des Systèmes" revient à tenter d'adapter des approches, essentiellement formelles, venant d'autres domaines et de les utiliser pour l'élaboration de modèles efficaces dans le secteur qui nous intéresse. Remarquons que cette démarche et cette analogie entre modèles à plusieurs niveaux de l'organisation des systèmes naturels ont été soulignées par L. VON BERTALANFFY (1973) qui parle "d'isomorphisme dans la science". Notons que ce concept, encore imprécis, est à rapprocher de la notion d'isométrie à laquelle se réfère MANDELBROT pour la construction des objets fractals [MANDELBROT, (1978)].

II - PRINCIPAUX TYPES DE MODELES

Il n'est pas de notre propos de dresser une classification générale des modèles employés en biologie, ni même de donner une liste exhaustive des méthodes de modélisation ou des types de modèles. En effet, d'une part une classification impliquerait une intersection vide entre sous-ensembles et actuellement, il nous semble qu'aucun critère ne permet une telle propriété, et d'autre part une "liste exhaustive" conduirait presque sûrement à une pondération uniforme alors qu'il nous semble plus important d'insister sur les types de modèles ou très utilisés, ou très efficaces, ou encore dont le développement semble prometteur. Pour chaque exemple, nous indiquerons clairement les objectifs visés par la modélisation et les résultats qui en ont découlé.

Ainsi, nous proposons de distinguer les modèles d'analyse dynamique des autres sur la prise en compte du caractère spécifique de la variable temps. Nous traiterons les autres types de modèles dans un paragraphe

intitulé 'modèles non dynamiques,' où nous expliciterons le sens que nous attribuons à ce terme.

II.1 Analyse non dynamique

L'analyse non dynamique inclut plusieurs méthodologies qui seront distinguées suivant que l'analyse est réalisée par comparaison à des modèles de référence (approche statistique), par un résumé des données disponibles (analyse de données). par la construction d'un modèle propre au système étudié. Il ne peut être question d'ébaucher ici une description systématique des techniques employées, nous présenterons donc simplement les concepts essentiels en les illustrant par des exemples.

II.1-1 Comparaison avec des modèles de référence

Cette démarche caractérise l'approche statistique : l'analyse des propriétés d'un système consiste à tester si oui ou non le système suit un modèle fixé a priori (c'est-à-dire avant l'obtention des résultats qui serviront à la prise de décision). Différentes facettes de cette méthodologie seront présentées dans les 5 exemples suivants :

EXEMPLE 1 : PRODUCTION DE SOIE PAR BOMBYX MORI

Les données présentées ici s'inscrivent dans le cadre de l'analyse du système de production de soie chez le ver à soie (Bombyx Mori) [JM FAYARD, (1977), (1978)]. L'auteur compte les nombres de cellules dans deux parties d'une glande séricigène de ver à soie : le tube sécréteur et le réservoir. Chaque ver ayant une glande droite et une glande gauche, 4 mesures sont réalisées par individus notées respectivement S_D , S_G , R_D , R_G (dans cette étude, 466 vers ont été ainsi étudiés). L'objectif précis de l'analyse rapportée ici est le test de l'identité des deux glandes pour les variables mesurées. En fait, l'auteur s'est intéressé aux deux variables $S + R$ (nombre total de cellules comptées par glande) et $\frac{R}{S}$. Nous allons analyser séparément ces deux cas.

1) S+R La première étape est la construction d'un modèle pour les 2 x 466 résultats. Ce modèle doit intégrer deux niveaux de variabilité, à savoir :

i) la variabilité du mécanisme biologique lui-même,

ii) la variabilité dite d'échantillonnage (la loi de la variable n'est atteinte que par la connaissance de 932 résultats : un autre échantillon tiré de la même population donnerait d'autres résultats).

Malheureusement, des raisons de techniques mathématiques ne laissent que peu de choix quant à ce modèle : la variable étudiée doit (à une transformation de variable près) suivre une loi normale. Ici le modèle a été : les 466 mesures de la glande droite (resp. gauche) constituent un échantillon aléatoire simple d'une variable normale de moyenne μ_D (resp. μ_G) et de variance σ . L'ensemble des décisions futures ne sera valable que dans la mesure où ce modèle est exact. Dans l'exemple cité, des analyses graphiques et des tests appropriés ont confirmé la validité du modèle.

La construction du modèle général étant réalisée, on définit un sous-modèle caractéristique de l'hypothèse à tester, ici : " $\mu_D = \mu_G$ ". Une étude statistique permet alors de décider si cette hypothèse est vraie avec un risque d'erreur connu en cas de rejet. Dans l'exemple présenté, l'hypothèse a été acceptée, on a donc décidé que l'égalité $\mu_D = \mu_G$ était vraie. Il reste alors à replacer ce résultat mathématique dans le contexte biologique. Pour cela, il faut déterminer de quelle population les 466 individus sont extraits (bien sûr en pratique ceci doit être examiné avant l'expérimentation !). Dans l'exemple choisi, les 466 vers étaient les descendants d'un couple de Bombyx Mori. On peut donc admettre qu'ils constituent un échantillon aléatoire simple du mécanisme biologique de reproduction de ce couple ; la conclusion biologique de l'étude doit donc être : le nombre de cellules dans les glandes droite et gauche des descendants potentiels de ce couple de papillons ont des distributions identiques.

2) $\frac{R}{S}$ Une analyse graphique a montré ici que cette variable ne suivait pas une loi normale. Une analyse du type précédent n'est donc pas possible. Un modèle plus général, indépendant de la distribution de $\frac{R}{S}$ doit

être construit. Il est plausible d'accepter, au vu d'une analyse graphique, que les lois de $\binom{R}{S}_D$ et de $\binom{R}{S}_G$ ne diffèrent que par une constante. On effectue alors une réduction des données : on classe les 932 résultats et on construit la suite de deux objets obtenue en remplaçant dans la suite croissante des résultats chaque valeur par le symbole G (resp. D) si elle provient d'une glande gauche (resp. droite). On démontre alors que l'égalité des moyennes équivaut à l'équidistribution des $\binom{932}{466}$ suites possibles obtenues par permutation de la suite observée. Une statistique permet alors de tester ce nouveau modèle (par exemple la somme des rangs de G). Une telle analyse est dite non paramétrique par opposition à l'analyse paramétrique de R+S.

Cet exemple a présenté une utilisation de deux branches importantes de la statistique: l'analyse paramétrique, qui est tributaire de l'adéquation d'un modèle précisant une classe de lois (dépendant de paramètres) à laquelle doit appartenir la variable étudiée et l'analyse non paramétrique qui n'impose aucune contrainte sur la nature des lois des variables. Dans les deux cas, l'analyse a consisté en une prise de décision associée à des notions précises de risques d'erreurs (une présentation formelle mais très claire de la notion de prise de décision et de risque peut être trouvée dans LEHMAN (1959)). Bien entendu, comme dans tout résultat d'une modélisation, cette décision s'exprime en termes mathématiques et doit être soigneusement transcrite au niveau du phénomène étudié, en particulier la définition de l'échantillon est primordiale.

Dans cet exemple, les statistiques non paramétriques sont apparues comme un substitut aux statistiques paramétriques dans le cas où la normalité des variables considérées ne peut pas être admise. L'exemple suivant montre leur utilisation dans un autre contexte.

EXEMPLE 2 : STRUCTURES D'UN GENE

L'information nécessaire à la synthèse des protéines est stockée dans les gènes. Ces derniers sont des macromolécules constituées par l'enchaînement linéaire de molécules (les bases nucléotidiques) de 4 natures distinctes notées respectivement A, C, G, U. Ainsi, le début du gène cor-

respondant à la protéine capsulaire du virus MS2 est représenté par l'enchaînement suivant :

CGAGCUUU.....

La séquence totale comprend 1176 bases.

Dans le cadre plus général de l'analyse du système de traduction (passage du gène à la protéine), une analyse de cette séquence a été entreprise. Il est clair que la découverte de structures exceptionnelles ne peut se faire sans référence à un modèle. Plusieurs modèles sont envisageables ; leur discussion pourra être trouvée dans [C. GAUTIER, R. GRANTHAM (1976)]. Le modèle de l'équiprobabilité de $\binom{1176}{n_A, n_C, n_G, n_U}$ séquences obtenues par permutations (où n_A, n_C, n_G, n_U sont les nombres d'occurrence des bases A, C, G, U dans la séquence) a été le plus utilisé car il permet des analyses statistiques simples. Une statistique en ordonnant les séquences possibles permettra alors de décider que les séquences situées à l'extérieur d'un intervalle en contenant 95 %, sont exceptionnelles. Ainsi, la statistique "somme des rangs de A dans la séquence" permettra de détecter un groupement anormal (par rapport au modèle de référence) des A en début ou en fin de séquence. (L'utilisation d'un ensemble de statistiques permet alors une description de la séquence considérée).

Dans ce dernier exemple, le but de l'analyse est fondamentalement différent de celui de l'exemple 1. En effet, alors que dans l'exemple 1, on voulait, à partir d'un échantillon, déterminer certaines propriétés d'une population plus large, dans l'exemple 2 le but est la "simple" description d'un objet biologique sans volonté immédiate de l'étendre à une classe plus large d'objets. Les méthodologies peuvent parfois être complémentaires, une description du type de l'exemple 2 pouvant permettre un contrôle de la qualité de l'échantillon dont on envisage l'utilisation pour une étude inférentielle. Ce dernier type de raisonnement peut être rapproché d'analyses hiérarchisées de la variabilité dont nous donnons un exemple ci-dessous (un autre exemple du même type a été présenté par D. HEMON, N. ANGELIER, M. BOUTEILLE (1978)).

EXEMPLE 3 : RÔLE DES GRANULES DU MYOCARDE

Cette analyse statistique a été réalisée par D. CHESSEL sur des données de [S. BLAINEAU PEYRETTI (1976)].

Le but était de déterminer chez la souris le rôle des granules des cellules du myocarde dans la régulation des cations, dans ce tissu. Dans cet objectif, l'auteur a placé le tissu cardiaque dans des conditions de surcharge ionique par incubation dans des solutions enrichies en calcium ou en strontium (numérotées ici de 1 à 4)* et il examine les granules au microscope électronique. Cet examen nécessite une fixation : deux fixateurs ont été utilisés : ils seront notés ici A et B. Sous ces conditions, on remarque que certains granules sont altérés. L'analyse de ce système passe par une décomposition préalable de la variabilité. Les niveaux de variabilité envisageables sont ici relatifs :

- 1) à l'influence croisée de la solution d'incubation et du fixateur,
- 2) à une différence entre souris pour un type de traitement donné,
- 3) à une différence entre cellules d'une même souris pour un même traitement.

Le plan expérimental doit permettre de détecter chacune de ces variabilités. Sa construction doit donc tenir compte de la hiérarchie des variabilités d'une part et des outils statistiques existants d'autre part. Le plan expérimental choisi apparaît dans la structure du tableau de données de la figure 1. Le nombre de granules dénaturés a été déterminé parmi 15 granules tirés au hasard dans une cellule. Cette variable permet, en effet, l'utilisation de nombreux tests non paramétriques [D. CHESSEL, C. GAUTIER (1977)].

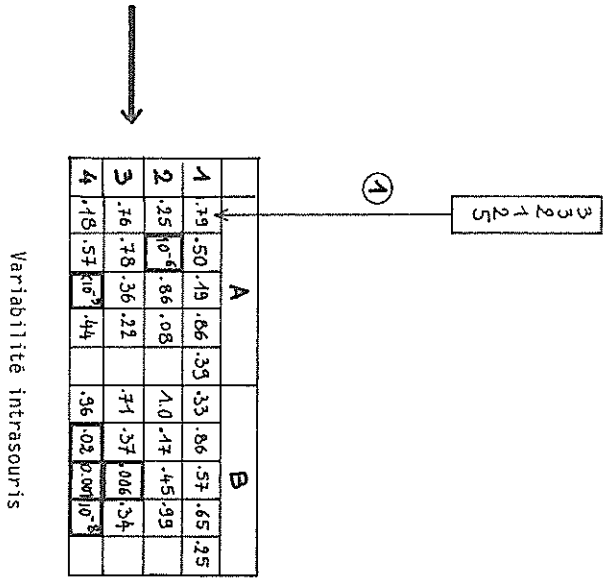
L'analyse statistique est résumée par la figure 1. Nous en décrivons ici les étapes numérotées de 1 à 5 sur cette figure.

* On peut remarquer l'utilisation ici d'une excitation perturbant le système et permettant ainsi son étude

A						B					
53	S8	S14	S22	S28	S38	S8	S14	S20	S28	S38	S48
3	2	1	1	1	3	3	2	1	1	1	3
3	3	0	1	1	3	4	4	4	4	4	4
2	3	3	2	1	3	4	4	2	4	1	4
1	1	1	1	1	3	5	5	3	4	4	4
5	1	2	1	2	3	5	5	3	4	4	4
53	S8	S14	S28	S38	S48	S8	S14	S20	S28	S38	S48
7	13	6	13	10	10	4	9	3	4	5	5
8	7	4	4	7	6	6	6	3	4	4	4
8	12	6	4	10	5	6	6	6	6	7	8
9	15	6	4	5	9	5	4	3	3	8	8
8	15	4	5	9	6	5	4	3	3	5	5
10	14	5	5	9	6	6	5	3	5	8	8
53	S8	S14	S28	S38	S48	S8	S14	S20	S28	S38	S48
1	2	6	7	7	4	4	12	10	9	9	9
5	3	3	7	7	4	6	11	9	9	11	11
5	4	5	7	7	4	6	10	8	4	10	10
3	3	3	3	8	6	6	6	8	4	8	8
3	3	1	9	6	6	6	12	9	9	8	8
4	6	7	5	5	6	3	3	2	10	10	10
53	S8	S22	S28	S38	S48	S8	S20	S28	S38	S48	S58
4	4	4	9	8	4	13	5	5	5	5	5
4	1	5	10	4	4	9	6	14	13	14	14
1	1	2	5	6	4	9	0	13	7	7	7
4	2	5	13	12	6	9	2	12	12	12	12
4	4	5	12	6	9	6	13	6	2	15	15
6	3	12	0	9	6	5	6	2	15	15	15

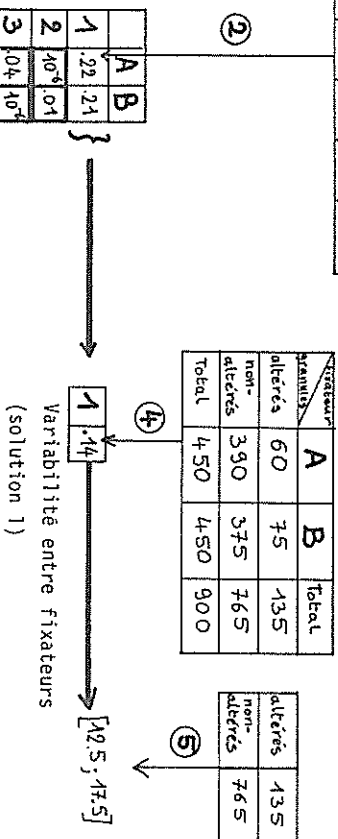
Données expérimentales

Exemple de S3, traitement 1/A.



Exemple du traitement 1/A

	souris		non-altérés		altérés		Total
Altérés	n°3	n°8	n°14	n°22	n°28	n°38	Total
Altérés	16	10	9	9	16	60	
Non-altérés	74	80	81	81	74	390	
Total	90	90	90	90	90	450	



Exemple du traitement 4/A

	souris		non-altérés		altérés		Total
Altérés	n°3	n°8	n°14	n°22	n°28	n°38	Total
Altérés	4	4	4	4	9	9	
Non-altérés	1	1	2	2	5	6	
Total	4	4	5	5	13	12	
	6	6	2	2	12	6	
	3	3	12	0	9	9	

Figure 1 - ANALYSE HIERARCHIQUE DE LA VARIABILITE.

Les résultats des analyses statistiques sont indiqués par leur niveau critique (probabilité de se tromper si l'hypothèse nulle est rejetée). Les cases encadrées correspondent aux rejets (niveau critique inférieur à 0.05). La structure des données analysées est dans chaque cas illustrée par un exemple.

1. Le modèle de référence est : le nombre de granules altérés dans une souris suit une loi binomiale. Dans ce cas, la loi de $\sum Y_i^2$, sachant $\sum Y_i$ est connue où Y_i est le nombre de granules altérés parmi les 15 granules observés dans la cellule n° i. Une trop forte valeur de $\sum Y_i^2$ détectera donc une hétérogénéité entre cellules d'une même souris. Sur le tableau, "variabilité intra souris", on voit que pour les 3 premières lignes, le modèle pourra être accepté (au plus 1 résultat significatif parmi 8), par contre, il sera rejeté pour la 4ième solution d'incubation (4ième ligne) (4 résultats significatifs parmi 8).

2. Le modèle binomial pour les granules d'une même souris ayant été accepté pour les 3 premières solutions, on peut sommer les nombres de granules altérés d'une même souris et construire ainsi des tableaux (dits tables de contingence comme celui "exemple traitement 1A") où, pour chaque souris, on note le nombre de granules altérés et le nombre de granules non altérés. Le calcul d'un χ^2 sur ces tableaux permet alors de tester le modèle binomial entre souris soumises au même traitement. On remarque sur le tableau "variabilité entre souris, sol.1.2.3" que pour la solution 1, le modèle est accepté, quel que soit le fixateur, par contre il est rejeté pour les deux autres solutions, les niveaux critiques indiquant un effet croisé entre solution et fixateur (effet ayant d'ailleurs une interprétation biologique).

3. Pour la quatrième solution, on ne peut pas sommer les résultats des cellules d'une même souris, deux analyses de variances non paramétriques seront effectuées montrant l'existence dans les deux cas d'un effet souris (cf. tableau variabilité entre souris - solution 4).*

4. Le modèle binomial ayant été accepté entre souris pour la solution 1 pour un même fixateur, alors les résultats sont sommés pour toutes les souris comme en 2. et un calcul du χ^2 permet de tester le modèle binomial. Ce modèle est accepté.

5. L'étude 4 permet une nouvelle sommation et une estimation globale du pourcentage de granules altérés.

* On pourrait remarquer que cette même analyse aurait pu être effectuée sur les 3 premières lignes mais au prix d'une perte importante de puissance.

L'analyse de systèmes par l'utilisation de plusieurs techniques statistiques organisées de manière hiérarchique (le choix d'une méthode dépend du résultat des analyses précédentes) à partir d'un plan d'échantillonnage spécialement conçu est particulièrement fructueuse dans de nombreux domaines biologiques ou écologiques (il faut cependant remarquer que l'utilisation de tests en cascade rend impossible une évaluation exacte des risques d'erreur). Elle montre clairement qu'une analyse de système commence à la conception du plan expérimental et non après l'acquisition des résultats.

II.1-2 Analyse de données

L'analyse de fichiers de données expérimentales importants est souvent impossible sans une réduction préalable de ces données. Cette réduction peut d'ailleurs fournir les résultats recherchés en mettant en évidence certaines structures du fichier. La démarche est donc à rapprocher de celle de l'exemple 2, cependant nous parlerons ici de méthodes spécifiquement adaptées à la réduction de masses importantes de données. L'exemple 4 décrit la démarche utilisée.

EXEMPLE 4 : FREQUENCE DES CODONS DANS LES GENES

Le cadre biologique est le même que celui de l'exemple 2 mais on ne s'intéressera pas ici à la succession des bases le long du gène mais au nombre de présences de chaque codon.^{*} Ainsi, de ce point de vue, un gène est caractérisé par 61 valeurs. On connaît actuellement 52 gènes obtenus par l'étude de la séquence de macromolécules^{**} les ARN messagers. Les données se présentent donc comme un tableau 52 x 61 contenant 3172 valeurs. La taille du tableau de données justifie pleinement l'utilisation d'une méthode d'analyse de données "automatique". Nous avons utilisé une analyse des correspondances qui fournit, classiquement, une carte où sont projetées les lignes du tableau de données (ici les gènes) et ses colonnes (ici les codons type) (figure 2). Nous disposons maintenant d'un

* Un codon est un ensemble de 3 bases consécutives (GCU, AGC...). Comme il existe 4 bases, il existe à priori $4^3 = 64$ codons types, cependant 3 d'entre eux jouent un rôle particulier et ne peuvent se trouver dans un gène ; ainsi nous ne parlerons dans la suite que de 61 codons types.

** Au 1/2/79

résumé des données : malgré sa complexité, la figure obtenue est plus lisible que le tableau des données initiales. Cependant, il serait très dangereux de tenter de lire ce résumé indépendamment de la technique utilisée. Il est clair que la recherche d'un résumé de données n'a pas de sens en soi : il convient de préciser avant tout quel type d'information est essentiel et doit être conservé au mieux et quel type peut être perdu. Ici, nous avons considéré le tableau des données comme une table de contingence : chacun des 11 147 codons utilisés dans l'analyse appartient d'une part à un des 52 messagers, d'autre part à l'un des 61 types de codons.

Comme nous cherchons ici à détecter et interpréter un usage différent des codons types par chacun des gènes, certaines des informations contenues dans ce tableau ne nous intéressent pas. Ce sont, en particulier, le nombre total de codons de chaque gène ainsi que la fréquence globale de chaque type de codons. L'analyse des correspondances répond alors à notre problème (traitement des tables de contingences conditionnées par les marges). Cependant, un exemple simple va montrer que la lecture du résumé nécessite un examen plus approfondi : la représentation sur une même figure des codons types et des gènes laisse supposer une relation entre un gène et un codon type de même position. Par exemple, UCA est proche de PAMPR et éloigné de G4E ; on s'attend donc à ce que UCA soit particulièrement fréquent dans PAMPR et rare dans G4E, le tableau ci-dessous contredit formellement cette interprétation :

Fréquence globale de UCA	Fréquence de UCA dans PAMPR	Fréquence de UCA dans G4E
0.013	0.008	0.063

Notre propos ne peut être ici de développer la théorie de l'analyse des correspondances [LEBART (1978), BENZECRI (1973), HILL (1974)] ; nous en donnerons cependant une vue géométrique simplifiée, permettant l'utilisation de la figure. Pour représenter la similarité entre deux gènes, on définit une distance entre eux par :

$$(1) d_{i_1 i_2} = \sum_{j=1}^{61} \frac{1}{n \cdot j} \left(\frac{n_{i_1 j}}{n_{i_1 \cdot}} - \frac{n_{i_2 j}}{n_{i_2 \cdot}} \right)^2$$

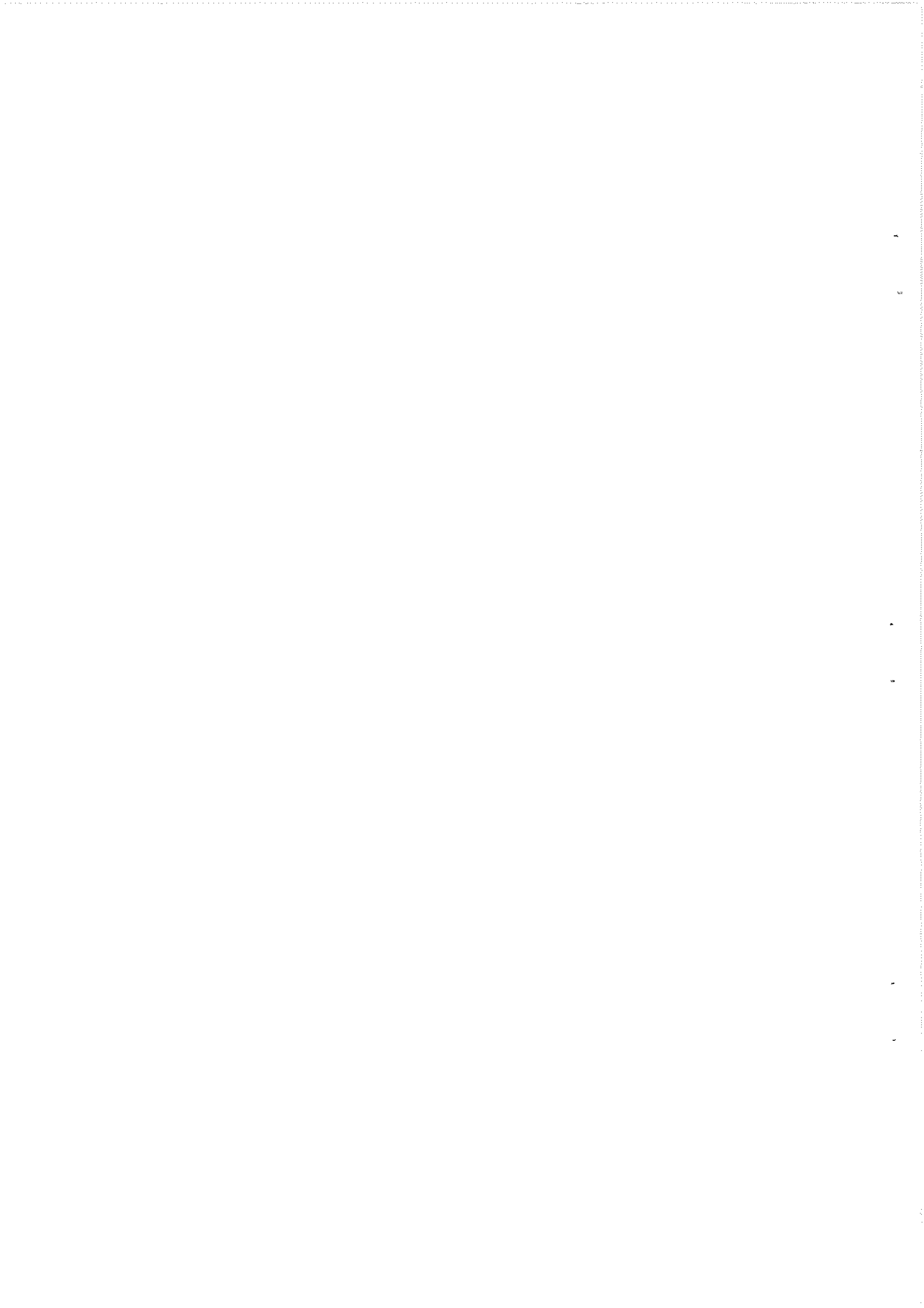
où i_1 et i_2 représentent les numéros des deux gènes.

n_{ij} est nombre de codons du type j présents dans le i^{e} gène

et
$$n_{.j} = \sum_{i=1}^{52} n_{ij} ; n_{i.} = \sum_{j=1}^{61} n_{ij}$$

Le terme $\left(\frac{n_{i_1 j}}{n_{i_1 .}} - \frac{n_{i_2 j}}{n_{i_2 .}} \right)^2$ mesure l'écart entre les fréquences du codon type j dans les gènes i_1 et i_2 ; le terme $\frac{1}{n_{.j}}$ pondère cet écart de manière à favoriser les codons types rares (ce qui n'a de sens que dans le modèle d'une table de contingence). On montre que l'on peut alors placer ces 52 gènes dans R^{51} de manière à ce que la distance entre deux gènes soit celle de la formule (1). On affecte ensuite à chaque gène un poids égal à sa longueur (ce qui n'est logique que par référence à une table de contingence) et on cherche le plan d'inertie minimale du nuage. La position des gènes de la figure 2 est obtenue par projection orthogonale sur ce plan. On obtient ainsi la représentation plane la moins déformée du nuage dans R^{51} (il faut remarquer que l'on n'est pas tenu à se limiter à une projection plane et que de nombreuses applications demandent l'utilisation d'un espace ayant plus de deux dimensions). De par le rôle symétrique joué par les lignes et les colonnes dans une table de contingence, la même analyse peut être réalisée en permutant les rôles joués par les gènes et par les codons types. On obtient ainsi une représentation plane des codons types. Bien entendu, il faut tenir compte dans l'interprétation des figures obtenues des déformations qui sont certainement intervenues lors des projections : deux points ayant même projection peuvent être en fait éloignés sur des axes orthogonaux au plan (ceci peut se vérifier par l'examen d'axes supplémentaires ou par l'examen de la distance à l'origine). La pondération des points dans R^{51} entraîne d'autre part que les gènes courts risquent d'être plus mal représentés (de même que les codons types rares), ce qui, dans le cas présent, est peut-être discutable.

Ces remarques étant faites, la figure 2 montre, pour le biologiste, une forte organisation. D'une part les gènes issus d'un même organisme (ou d'organismes voisins) sont en général proches les uns des autres, ce qui fournit un argument pour l'existence d'une contrainte globale sur les gènes d'une espèce. D'autre part les codons se terminant par la même base se trouvent



rapprochés, ce qui peut également s'interpréter biologiquement.

Il nous reste à justifier la superposition des deux représentations planes et à montrer son apport dans l'analyse des données. Il nous faut, pour cela, évoquer un autre aspect de l'analyse des correspondances. Chacun des 11 157 codons a été affecté de deux valeurs par l'analyse : d'une part, l'abscisse du gène qui le contient dans la représentation plane des gènes, et d'autre part l'abscisse du codon type dans la représentation des codons types. On peut ainsi construire un nuage de 11 157 points dans \mathbb{R}^2 . On montre que l'analyse des correspondances maximise la corrélation de ce nuage (on peut remarquer encore l'importance de la notion de table de contingence dans l'interprétation de ce nouveau nuage). Si cette corrélation, dite corrélation canonique, est forte, cela signifie clairement qu'un gène contient essentiellement des codons types ayant une "valeur" voisine de la sienne. De manière symétrique, un codon type sera surtout présent dans les gènes ayant une valeur voisine. En résumé, un codon et un gène voisin dans la figure 2 sont tels que la fréquence de ce codon type est particulièrement élevée dans ce gène, du moins si la corrélation canonique est élevée. Plus la corrélation canonique sera faible, moins la correspondance entre les deux figures sera bonne et ceci sera d'autant plus vrai pour les points situés au centre du dessin dans la mesure où ils jouent un moindre rôle dans le calcul de cette corrélation*. Ces remarques expliquent l'exemple du codon UCA présenté plus haut. S'il est parfois dangereux d'utiliser la correspondance point par point, il est par contre souvent utile de la considérer globalement. Ici, par exemple, on remarque que l'ensemble (ou presque) des codons se terminant par G ou C se trouvent dans la moitié inférieure de la figure, ce qui permet une interprétation simple de l'axe des ordonnées pour les gènes de la figure 2. : elle doit être corrélée au pourcentage de G+C en dernière position des codons du gène.

Effectivement, cette corrélation est de 0.94 ; cette variable est donc responsable en grande partie de la structure chez les gènes.

Sur ces mêmes données, une méthode de classification automatique a également été utilisée [R. FAGES (1978)]. Nous tenons seulement à souligner

* L'analyse de correspondance a bien d'autres facettes dont certaines seraient sans doute mieux adaptées à l'interprétation de ce problème mais nécessiteraient des développements mathématiques hors de propos ici.

l'intérêt de l'utilisation conjointe de ces deux méthodes qui fournissent des résumés complémentaires.

Les méthodes d'analyse de données sont très nombreuses. Mais la présentation de méthodes sophistiquées ne doit pas faire oublier l'efficacité dans d'autres contextes de méthodes graphiques simples (graphes, histogrammes...). Une critique fréquente de ce type d'analyse est que l'on peut démontrer ce que l'on veut ; cette critique nous paraît injustifiée si le modèle sous-jacent n'est pas oublié. Ainsi, l'utilisation d'une analyse des correspondances sur une table de contingence peut difficilement conduire à des interprétations fausses (moyennant les remarques faites dans l'exemple 4 sur le principe de correspondance), mais cette même analyse est fréquemment utilisée sur des données d'un autre type (présence-absence, classification multiple) l'interprétation correcte des figures obtenues demande alors une très bonne connaissance de la technique mathématique sur laquelle repose l'analyse.

II.1-3 Modèles propres au phénomène étudié

Outre les modèles généraux d'analyse des données, des modèles adaptés à des systèmes particuliers sont actuellement proposés au plan de la morphologie (description de la forme d'un organisme ou d'un organe) ou de la morphogénétique (genèse de cette forme) ; il est clair que le deuxième point fait intervenir la variable temps, il sera abordé dans le cadre du chapitre suivant.

Les outils de la géométrie analytique et différentielle conduisent à quelques résultats dans la mesure où certaines courbes, surfaces ou volumes pouvaient représenter la forme de certains systèmes biologiques ; on pourra se référer à ce sujet aux travaux de Y. BOULIGAND (1969) . Signalons, par exemple, que les spirales sont une représentation acceptable d'organismes animaux ou végétaux, ou de certaines structures observées dans ces organismes (escargots, certains coquillages). Des molécules telles les acides nucléiques ont des géométries bien définies (hélice ou double hélice). Certains organismes ont des formes assez simples pour être repré-

sentés par des surfaces ou volumes élémentaires de la géométrie analytique (éllipsoïdes, sphères) ou par des combinaisons simples de ces éléments.

A part les outils "classiques" de la géométrie, il apparaît actuellement des tentatives intéressantes de modélisation de structure spatiale utilisant la théorie des langages, avec un objectif de réalisation d'outils pour la reconnaissance des formes. Nous prendrons l'exemple, simplifié, de la représentation de certains chromosomes [LEDLEY et al (1965) et FU (1974)].

EXEMPLE 5 : MODELISATION DES CHROMOSOMES

Depuis de nombreuses années, le problème de la reconnaissance automatique des formes se pose dans le domaine de l'analyse médicale, particulièrement pour l'établissement de cartes chromosomiques (ou caryotype). De nombreuses tentatives ont été faites pour proposer des algorithmes de reconnaissance automatique des chromosomes humains sans avoir, il faut l'avouer, encore débouché sur une utilisation effective au niveau du laboratoire d'analyse. Pour aborder ce problème, certains auteurs (principalement LEDLEY et al, FU et LEE & FU (1972) , cités ci-dessus, proposent des modèles dérivés de la théorie des langages (suivant la classification de CHOMSKY) car pour certains d'entre eux, (langage de KLEENE ou langages hors contextes), on sait construire des automates accepteurs, même pour une extension probabiliste de ce langage (cf. [FU (1974)]). Le principe est le suivant :

Un langage est un ensemble de chaînes de symboles vérifiant certaines règles de grammaire. On associe à chaque forme "élémentaire" d'un chromosome un symbole (ou caractère). Un chromosome sera alors représenté par une chaîne de symboles appartenant à un langage de grammaire aussi simple que possible. Un automate associé à cette grammaire pourra reconnaître, ou non, une chaîne quelconque, et ainsi décider si la forme proposée est le chromosome cherché (normal) ou un autre chromosome (anormal).

Signalons que des descriptions morphologiques peuvent être tentées en utilisant les systèmes de LINDENMAYER (1974) , qui s'apparentent

à la théorie des langages et qui sont particulièrement adaptés à la modélisation du développement des organismes vivants (nous y reviendrons dans le chapitre suivant). Une autre direction, actuellement encore dans l'enfance, pourra peut-être fournir un cadre intéressant de modélisation des formes : il s'agit des objets fractals [MANDELBRROT (1978)], notamment pour la représentation de systèmes ramifiés. Enfin la Théorie des Catastrophes [THOM (1973)] est un cadre important pour l'étude théorique des problèmes morphologiques et également ceux de la morphogénèse (ainsi la limitation du nombre de catas-trophes possibles, et de là du nombre de formes élémentaires, suivant la di-mension de l'espace considéré, nous semble un résultat fondamental).

II.2 Analyse dynamique

Dans ce cadre, on s'intéresse généralement au comportement d'un système biologique dans le temps ; la prise en compte du caractère spécifique de la variable temps distingue cette analyse de la précédente (analyse non dynamique). Ainsi, l'analyse dynamique des processus biologiques permet d'appréhender le fonctionnement des systèmes vivants et également de détecter des mécanismes élémentaires mis en oeuvre dans ces systèmes. Il est à signaler que ce deuxième point a longtemps prévalu et n'a pas toujours nécessité le support d'un modèle mathématique. Ainsi, J. MONOD dans sa thèse (1942) a proposé un modèle efficace de la croissance bactérienne. Cependant, des observations qualitatives de la croissance de E. Coli sur milieu complexe ont eu un impact au moins aussi important (point de départ des recherches sur les mécanismes de la synthèse protéique).

Dans le même ordre d'idée, l'étude simultanée en fonction du temps de 2 ou plusieurs variables internes d'un système et leur représentation dans l'espace de ces variables ont permis de détecter certains phénomènes importants. Ainsi, la notion d'allométrie introduite par TEISSIER (1948) revient à l'étude simultanée en fonction du temps de l'évolution de deux variables d'un système représentant des grandeurs comme longueur et poids de 2 organes d'un même organisme. Ainsi, l'étude de croissances relatives a permis de détecter des "crises" dans ce processus généralement associées à des modifications du métabolisme.

Ces exemples illustrent l'efficacité de modèles, et de méthodologie, simples. Il apparaît néanmoins nécessaire de faire appel à des modèles plus sophistiqués.

Un modèle est généralement choisi en fonction de 3 critères principaux :

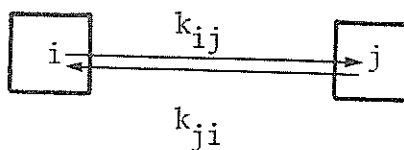
1. L'objectif de l'analyse
2. La "structure" du système biologique à étudier
3. Les possibilités d'expérimentation.

Dans ce rapport, nous présenterons les modèles d'analyse dynamique classés suivant leur expression mathématique. Nous avons préféré ce type d'exposé classique qui a l'avantage de la simplicité à une présentation, peut-être plus utilitaire, en fonction des critères définis ci-dessus.

Ainsi nous distinguerons d'une part les modèles en temps discret et en temps continu et d'autre part les modèles déterministes et stochastiques (ou probabilistes) ; cependant, il est clair qu'un même système pourra être modélisé de plusieurs façons.

De plus, il nous semble important de signaler que les modèles stochastiques sont souvent plus "réalistes" dans la mesure où les variabilités des différentes étapes d'un phénomène peuvent être précisément prises en compte, alors que les modèles déterministes représentent souvent globalement (ou en "moyenne") ces mêmes phénomènes. Dans certains cas, les liens entre modèles probabilistes et modèles déterministes d'un même phénomène ont pu être établis ; c'est le cas notamment d'une classe importante de modèles utilisés en biologie : les modèles à compartiments. Ces modèles sont essentiellement utilisés pour étudier des phénomènes de transports entre certaines structures de systèmes vivants (organes, cellules, structures cellulaires).

Un modèle à compartiment est défini de la façon suivante : à chaque variable interne du système, on associe un compartiment et on représente les interactions entre les variables par des flux entre compartiments :



où k_{ij} est un paramètre définissant le transfert entre le compartiment i et le compartiment j .

Dans ce schéma, les transferts peuvent être déterministes ou probabilistes. Une discussion très complète de cette distinction peut être trouvée dans [MATIS et GERALD (1977)].

D'une façon générale, un modèle probabiliste sera préféré à un modèle déterministe toutes les fois que la variabilité d'une ou plusieurs étapes du phénomène sera importante. Par contre, un modèle déterministe peut être choisi dans un souci de simplicité si les contraintes spécifiées ci-dessus sont "négligeables", par exemple, si la variance propre du phénomène est petite devant l'incertitude des mesures. Il faut cependant remarquer que les modèles probabilistes conduisent parfois à des impasses mathématiques (utilisation des modèles pour la commande). On se ramène alors au cas déterministe.

11.2-1 Modèles déterministes en temps continu

Cette classe de modèles est aujourd'hui la plus étudiée ; la dynamique d'un système est représentée par un ensemble d'équations différentielles ordinaires ou par un ensemble d'équations aux dérivées partielles. La première approche est la plus communément employée dans la mesure où les aspects théoriques et numériques sont mieux connus. Cependant, les développements récents de la théorie des équations aux dérivées partielles en font des outils qui pourront, vraisemblablement, être de plus en plus utilisés.

a) Modèles régis par des équations différentielles ordinaires :

Dans la pratique, pour un système biologique donné, on définit un certain nombre de variables internes (appelées variables d'état) qui caractérisent l'état du système à chaque instant et dont l'évolution est décrite par un système d'équations différentielles du type : $\frac{dX}{dt} = f(X, t)$, où X est le vecteur des variables d'état.

Lorsque le système reçoit des excitations extérieures, dont

on est plus ou moins maître, le système d'équations s'écrit :

$$\frac{dX}{dt} = f(X, t, U)$$

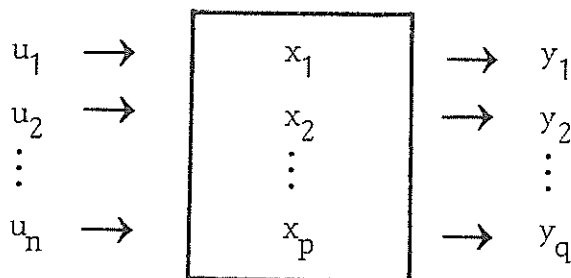
où l'on distingue :

X : le vecteur des variables d'état (ou variables internes) du système.

U : le vecteur des variables de commande (excitations).

Si, de plus, chacune des variables d'état n'est pas directement accessible à la mesure, mais l'est seulement par l'intermédiaire d'autres grandeurs fonction des variables d'état, on devra adjoindre aux équations précédentes un système d'équations dit d'observation : $y = g(X, t, U)$ où y est le vecteur des variables de sortie, c'est-à-dire des variables accessibles à l'expérimentateur, celles qu'il peut mesurer.

Ainsi, un système peut se représenter par le schéma suivant :



Cette deuxième formulation, faisant intervenir la commande et l'observation, n'est pas très habituelle dans les sciences biologiques, elle nous semble cependant très importante à considérer puisqu'elle a des implications non seulement au niveau modélisation mais également au niveau de la conception d'expériences (cf. section 3.5 pour les modèles à compartiments).

Ces modèles à base d'équations différentielles sont largement utilisés dans les sciences biologiques, notamment dans l'analyse compartimentale qui permet d'aborder les problèmes des transferts des particules dans les

systemes vivants (modèles déterministes). On trouvera dans les ouvrages de SHEPPARD (1962) et de JACQUEZ (1972) un exposé détaillé de cette approche, avec de très nombreux exemples. Nous présenterons ici un exemple tiré de la biologie cellulaire que nous avons eu à traiter.

EXEMPLE 6 : LE TRANSPORT INTRACELLULAIRE DE LA FIBROÏNE DANS LA GLANDE SERICIGENE DE BOMBYX MORI [P. COUBLE, J.C. PRUDHOMME et J.DAILLIE (1977)], [A. PAVE (1977)]

L'étude de ce problème entre dans le cadre plus général des études moléculaires sur le fonctionnement des glandes séricigènes. En vérité, il s'agit de l'étude d'un matériel biologique où les problèmes de la biosynthèse de protéines chez les organismes évolués (eukaryotes) sont simplifiés. En effet, l'activité des cellules est limitée en partie à la synthèse en grande quantité d'une protéine : la fibroïne (environ 70 % de la synthèse totale).

La transport intracellulaire de la fibroïne et sa sécrétion ont été suivis par autoradiographie après marquage radioactif d'un acide aminé constitutif de cette protéine.

La représentation au niveau cellulaire de ces phénomènes est donnée par la figure ci-dessous.

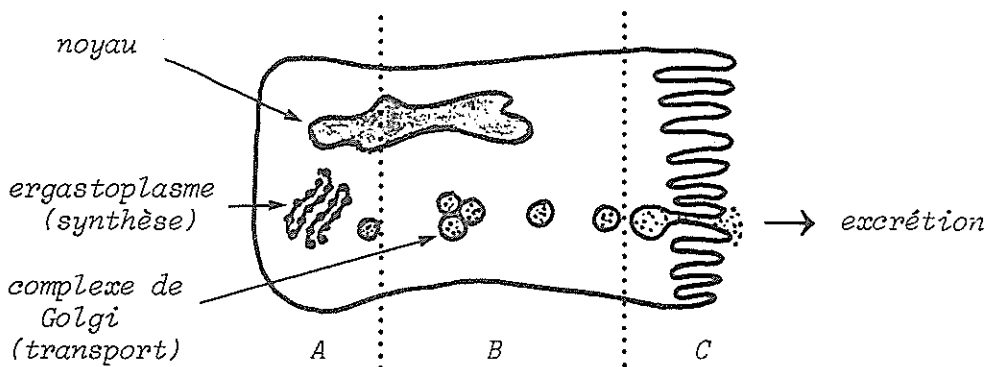
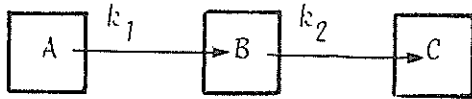


Fig. : Schéma d'une cellule sécrétrice de fibroïne

On peut alors définir 3 compartiments :

- le compartiment A où s'effectue l'incorporation du marqueur,
- le compartiment B correspondant au transport intracellulaire,
- le compartiment C où la fibroïne est excrétée.



Au plan expérimental, le nombre de molécules dans les compartiments A, B et C a été mesuré. Le modèle mathématique correspondant au schéma précédent est :

$$\frac{dN_A}{dt} = -k_1 N_A$$

$$\frac{dN_B}{dt} = k_1 N_A - k_2 N_B$$

$$\frac{dN_C}{dt} = k_3 N_C$$

qui admet une solution analytique non développée ici. Cependant, il s'avère que pour le premier compartiment la comparaison de la solution donnée par le modèle avec les résultats expérimentaux n'était pas satisfaisante, comme le montre la figure ci-dessous.

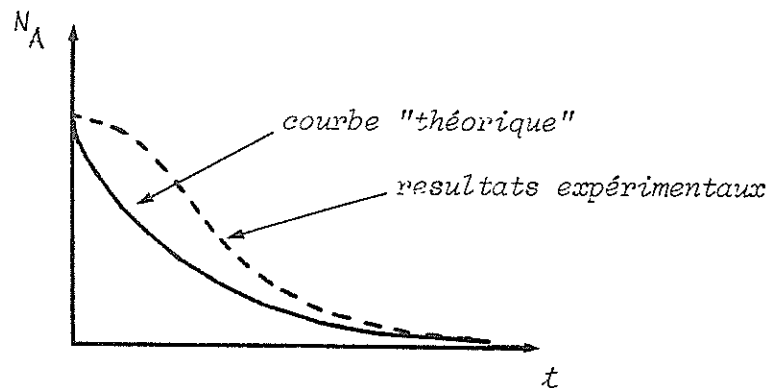
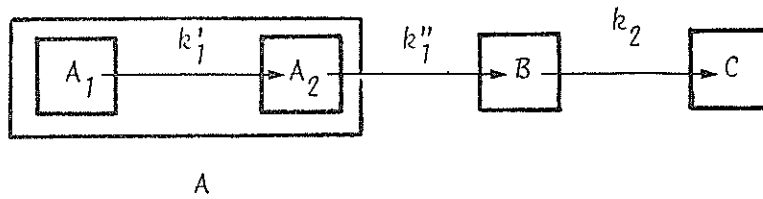


Fig. : Evolution du nombre de molécules marquées dans le compartiment A

En réalité, dans ce compartiment, les phénomènes sont plus complexes (entre autre processus de synthèse), ce qui nous a conduit à scinder ce compartiment en deux sous-compartiments. Le modèle devient alors :



$$\frac{dN_{A_1}}{dt} = -k_1 N_{A_1}$$

$$\frac{dN_{A_2}}{dt} = k'_{11} N_{A_1} - k''_{11} N_{A_2}$$

$$\frac{dN_B}{dt} = k''_{11} N_{A_2} - k_2 N_B$$

$$\frac{dN_C}{dt} = k_2 N_B$$

$$\text{ou } N_A = N_{A_1} + N_{A_2}$$

Dans ces conditions, on obtient une bonne concordance avec les résultats expérimentaux.

Dans un tel modèle, les constantes k'_{11} , k''_{11} , k_2 peuvent s'interpréter comme étant l'inverse des temps de séjour moyens dans chaque compartiment et on peut aisément démontrer, avec ce type d'équations, qu'il y a additivité de ce temps de séjour moyen. Ainsi :

$$\tau_1 = \frac{1}{k'_{11}} + \frac{1}{k''_{11}}$$

peut être considéré comme le temps de séjour dans le compartiment A (assimilable essentiellement au temps de synthèse).

$$\tau_2 = \frac{1}{k_2}$$

représente le temps moyen de transit des molécules du lieu de synthèse à

l'extérieur de la cellule.

La comparaison du taux de synthèse (i.e. du nombre d'acides aminés incorporés par seconde) estimé dans cette expérience avec ceux estimés pour d'autres systèmes protéosynthétiques chez les eukaryotes montre qu'il semble y avoir une relative constante de ce paramètre. Par contre, l'estimation du temps de transport n'a pu être comparé avec d'autres données.

Aussi, un modèle relativement simple a permis d'obtenir une estimation de paramètres difficilement accessibles à l'expérience.

Cependant, signalons que :

. Ce processus de synthèse ne peut être qu'approché par un système linéaire dont les paramètres (ou du moins la somme de leur inverse) ont une interprétation biologique.

. On considère que l'assimilation par la cellule de l'acide aminé marqué est très rapide ; son rôle est donc uniquement de fixer la valeur initiale de N_{A_1} , aussi elle n'intervient pas comme variable de commande.

Il est clair que les conclusions tirées d'une telle approche doivent tenir compte de ces hypothèses simplificatrices.

Bien que la majorité des systèmes étudiés par l'analyse compartimentale se limitent à des systèmes linéaires à paramètres constants, on peut néanmoins introduire des non-linéarités dans les transferts entre compartiments [JACQUEZ (1972)].

En outre, l'utilisation de modèles à équations différentielles ordinaires ne conduit pas nécessairement à la définition de compartiment.

EXEMPLE 7 : ANALYSE DU SYSTEME COAGULOLYTIQUE

Cet exemple présente l'élaboration d'un modèle pour le système coagulolytique. Il s'agit d'un système biologique complexe comprenant un ensem-

bile de réactions enzymatiques et physicochimiques couplées entre elles : les unes ayant pour fonction la coagulation du sang, les autres la lyse du caillot de sang. De l'équilibre entre ces mécanismes (coagulation et lyse) dépend la fluidité du sang circulant. Ce modèle a précisément été établi en vue de l'étude de l'équilibre du système, équilibre dont les propriétés, comme la stabilité, sont fondamentales d'un point de vue médical. Les objectifs de la modélisation sont donc d'une part de rendre compte du comportement dynamique du système coagulolytique et d'autre part de tester a posteriori la validité des hypothèses biologiques utilisées pour la modélisation.

Pour mener à bien une telle modélisation, les problèmes à résoudre se situaient au niveau :

- . de la délimitation du système et de son environnement,
- . de la définition des hypothèses biologiques qui ne sont souvent pas suffisamment précises pour être formulées,
- . de la mise en équation (en effet, il s'agit des cascades des réactions enzymatiques *in vivo*).

Pour la délimitation, nous avons défini les réactions clés du système et considéré que les mécanismes situés en amont et en aval de ces réactions appartenaient à l'environnement. Ces mécanismes constituent en fait les mécanismes sources des réactifs et les mécanismes éliminateurs des produits du système coagulolytique.

Pour le second point, nous avons fait préciser les hypothèses biologiques de manière à pouvoir les synthétiser dans un schéma fonctionnel.

Quant au troisième problème, la mise en équation, nous avons établi le modèle en écrivant une équation de bilan pour chacun des composés, dont les concentrations sont considérées comme variables d'état du système. Nous aboutissons ainsi à un système de 2 équations algébriques et de 9 équations différentielles non-linéaires du type :

$$x_i = u_i - C_{1i}$$

$$\frac{dx_i}{dt} = \frac{C_{2i} x_i}{1 + C_{3i} x_j + C_{4i} x_k + C_{5i} x_l} - C_{6i} x_i$$

où x_i, x_j, x_k, x_l sont des variables d'état

u_i est une action sur la variable d'état x_i

C_{1i}, \dots, C_{6i} sont des constantes cinétiques relatives.

Le modèle ainsi obtenu, qui traduit mathématiquement les hypothèses biologiques concernant le système coagulolytique, a été confronté avec les données et observations biomédicales. Ayant été établi en vue de rendre compte des propriétés de l'équilibre coagulolytique, le modèle a été utilisé pour définir mathématiquement cet équilibre et mettre en évidence ses propriétés (existence, stabilité, sensibilité aux paramètres...).

Cette modélisation donne un exemple de modèle utilisant la notion de variable d'état sans qu'elle soit associée à une notion de compartiment et de plus, les équations différentielles utilisées sont non linéaires.

En ce qui concerne les non-linéarités, on citera les modèles bilinéaires qui permettent de tenir compte d'interactions du type multiplicatif entre variables du système [A. CHERUY (1974)] et dont le principal intérêt réside d'une part dans la disponibilité d'outils mathématiques efficaces et d'autre part dans le fait qu'ils fournissent de bonnes approximations de modèles plus complexes.

EXEMPLE 8 : UNE APPROCHE DE LA DYNAMIQUE DES MACROMOLECULES : Variation de la quantité de RNA total dans la glande séricigène du ver à soie [PAVÉ (1978)] - Un essai d'interprétation d'un mécanisme moléculaire de la sélection.

Chez le ver à soie, au cours du dernier âge larvaire, la synthèse de la soie prend une importance considérable. Cette synthèse s'arrêtera après filage du cocon (ce qui correspond à un arrêt de la synthèse protéique au niveau des cellules de la glande séricigène). Dans le processus de la synthèse des protéines, certaines molécules jouent un rôle clé, notamment les acides ribonucléiques (RNA). L'évolution de la quantité de RNA total peut être suivie

pendant toute cette période [A. FOURNIER (1974), PRUDHOMME (1976)]. On obtient un graphe du type :

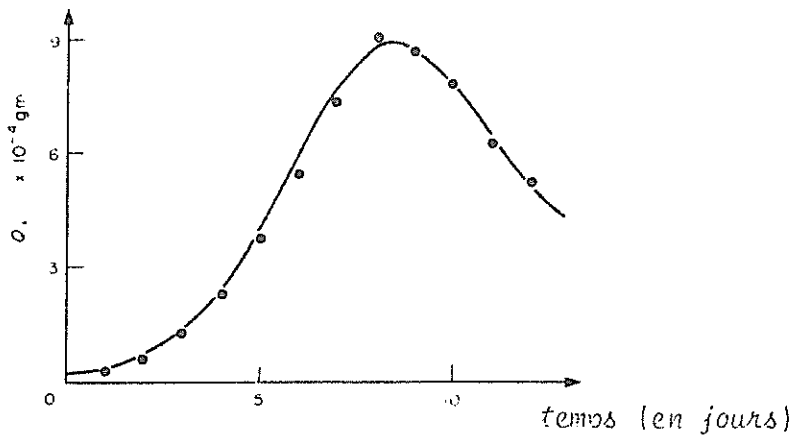


FIGURE 5

Evolution de la quantité de RNA total
dans la glande séricigène

La forme de cette courbe peut s'expliquer par la présence simultanée d'un processus de synthèse et d'un processus de dégradation. Un modèle simple traduisant une telle évolution peut être déduit du modèle de VOLTERRA-KOSTITZIN, [KOSTITZIN (1937)] utilisé d'abord pour représenter la croissance de populations cellulaires, de populations de microorganismes, ou la croissance embryonnaire :

$$\frac{dq}{dt} = aq - bq^2 - cqF$$

$$\frac{dF}{dt} = q$$

où q représente la quantité de RNA, F la quantité d'un facteur de dégradation (une enzyme : la RNAase), le paramètre a peut s'interpréter comme un taux de synthèse, c comme un taux de dégradation et b comme un paramètre de limitation de la synthèse (effet stérique ou répressif).

a) Un tel modèle peut être "schématisé" en utilisant une représentation équivalente à celle des réactions chimiques :

$$q \xrightarrow[\pm]{a} 2q \quad (\text{"autoreproduction" limitée})$$

$$q \xrightarrow{c_1} q + F + \dots \quad (\text{production d'un facteur de dégradation})$$

$$(c = c_1 \times c_2)$$

$$q + F \xrightarrow{c_2} F + \dots \quad (\text{dégradation du RNA})$$

Une telle représentation permet d'interpréter plus aisément les paramètres [PAVÉ, PAGNOTTE (1977)]. Notons que cette démarche qui consiste à proposer un schéma fonctionnel à partir d'un modèle mathématique n'est pas classique puisque généralement on suit le chemin inverse (cf. par exemple les modèles à compartiments).

b) Nous avons utilisé ce modèle pour étudier chez différentes souches de ver à soie, l'évolution de la quantité de RNA et sa corrélation avec la production de soie. En effet, il est bien connu que le ver à soie est sélectionné depuis plusieurs millénaires en vue de la production de soie, aussi on dispose actuellement d'une assez grande variété de souches se distinguant, en particulier, par leur production de soie.

A partir de données expérimentales obtenues d'une part au Japon et d'autre part à Lyon, sur l'évolution du RNA, nous avons fait des estimations de paramètres a , b et c pour plusieurs souches. Ainsi, nous avons pu mettre en évidence le pouvoir discriminant du paramètre c entre ces souches ainsi que la corrélation négative entre ce paramètre et la production de soie, alors que les paramètres a et b sont peu discriminants. Nous avons donc été amenés à proposer que l'effet principal d'une sélection en vue de la production de soie se produisait non pas en augmentant les taux de synthèse, mais plutôt en limitant le processus de dégradation de la machinerie cellulaire. Il va sans dire qu'il ne s'agit que d'une hypothèse induite par l'analyse du modèle, qui demande confirmation, notamment au plan expérimental.

En conclusion, pour tous ces modèles régis par des équations différentielles ordinaires, on dispose de nombreux outils mathématiques (en particulier dans le cas linéaire) qui expliquent leur large utilisation aussi bien pour l'analyse de mécanisme et le test d'hypothèses biologiques

que pour la commande de systèmes et la prédiction...

b) Modèles régis par des équations aux dérivées partielles

Ce type de modèle permet de rendre compte de l'évolution d'un système où interviennent des phénomènes de diffusion et propagation. On les écrit généralement sous la forme :

$$\frac{\partial^p X}{\partial t^p} = f(x, t, \frac{\partial X}{\partial x}, \frac{\partial^2 X}{\partial x^2}, \dots)$$

où t représente le temps, x le vecteur de variables d'état du système et x , un vecteur de l'espace dans lequel s'effectue la diffusion ou la propagation (par exemple R , R^2 ou R^3).

On peut citer les applications suivantes :

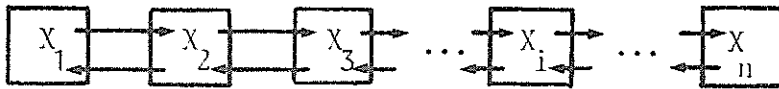
- . Transfert de chaleur dans les systèmes vivants,
- . Diffusion de fluides dans l'organisme (par exemple les échanges gazeux dans les poumons : [CHERRUAULT et LORIDAN (1977)]),
- . Diffusion de molécules en biologie cellulaire.

Bien que de nombreuses recherches théoriques soient développées sur ce type d'équations, [LIONS (1973)], actuellement, l'étude de tels phénomènes est souvent traitée par une représentation approchée utilisant des équations différentielles ordinaires.

Par exemple, certains processus de diffusion peuvent s'approcher par un système à compartiments. Ainsi, le modèle classique de la diffusion d'un produit suivant une direction x s'écrit :

$$\frac{\partial X}{\partial t} = k \frac{\partial^2 X}{\partial x^2}$$

Ce modèle peut être représenté par un système à compartiments :



où les variables X_i représentent les quantités de produit dans des sections élémentaires $x_i, x_i + \Delta x$. C'est-à-dire par un modèle mathématique :

$$\frac{dx}{dt} = Ax \text{ où } x = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \text{ et } A \text{ est une matrice tridiagonale.}$$

La solution de ce système tend vers la solution de l'équation aux dérivées partielles quand $n \rightarrow \infty$.

Pratiquement, il s'avère que pour un faible nombre de compartiments, l'approximation est satisfaisante. Ainsi, BOISVIEUX (1977) a réalisé une étude par simulation et montré que pour $n = 3$, on obtenait une bonne approche de la solution.

Au niveau de la démarche on peut noter qu'il est toujours intéressant d'essayer de comparer le modèle le plus "exact" possible avec le modèle simplifié que l'on veut (ou que l'on peut) utiliser, car les niveaux d'approximation sont mieux cernés.

II.2-1 Modèles déterministes en temps discret

L'hypothèse implicite qui préside à une telle approche est l'occurrence d'un ensemble de phénomènes à des temps discrets, généralement séparés par des intervalles constants. Aussi le terme "horloge du processus" est souvent employé. Les équations de récurrence caractérisent ces modèles :

$$x_t = f(x_{t-1}, x_{t-2}, \dots, x_{t-p})$$

Le processus est entièrement déterminé par la connaissance des p premières valeurs des variables d'état x , ce qui explique l'appellation de "modèles à mémoire".

Un cas important est celui où $p = 1$ et f est une fonction linéaire. Alors, on a :

$$X_t = M X_{t-1} \text{ où } M \text{ est une matrice } n \times n,$$

si n est le nombre de variables d'état du système.

Ces modèles sont très intéressants lorsque l'on veut simuler un phénomène car ils permettent souvent d'éviter l'utilisation de procédures d'intégration longues et coûteuses pour les systèmes continus.

Cependant, si dans le cas linéaire on connaît bien les méthodes de discrétisation, il n'en est pas de même pour des équations non linéaires.

Enfin, dans le cas linéaire, ces modèles présentent de nombreux points communs avec les modèles discrets stochastiques (évolution moyenne).

EXEMPLE 9 : FONCTIONS DE CROISSANCE DANS LES "SYSTEMES" DE LINDENMAVER

Comme nous l'avons déjà signalé, les "systèmes" de LINDENMAVER constituent une approche théorique possible des problèmes de morphogénèse, liée à la théorie des langages. Ils manipulent des symboles auxquels on peut attacher des significations biologiques. On peut trouver un développement important de cette approche dans l'ouvrage de HERMAN et ROZENBERG (1975). L'aspect quantitatif a été développé en particulier par VITANYI (1974).

Il n'est pas question de donner ici une définition générale de ces modèles mais plutôt de citer un cas particulier orienté vers l'aspect quantitatif de la croissance d'une population (population cellulaire, par exemple).

Considérons une population de cellules constituée de deux types d'individus notés a et b , associés à des classes d'âges : a est une cellule jeune qui parvient à maturation à un temps donné et se transforme alors en cellule b , cette cellule b se divise et donne une cellule jeune a et une cellule mature b . Ce processus peut se schématiser par :

$$\begin{aligned} a &\rightarrow b \\ &\quad (P) \\ b &\rightarrow ab \end{aligned}$$

Supposons qu'à l'instant $t = 0$, on parte d'une seule cellule a et que l'on applique indépendamment à chaque symbole de la chaîne obtenue à l'instant $t - 1$ les règles de production décrites ci-dessus ; la lignée cellulaire s'écrit :

$$\begin{array}{ll} t = 0 & a \\ t = 1 & b \\ t = 2 & ab \\ t = 3 & bab \\ t = 4 & abbab \end{array}$$

(On suppose qu'il n'y a pas d'interaction entre cellules)

Considérons les variables d'état suivantes : N_a , nombre de cellules à l'état a et N_b , nombre de cellules à l'état b ; alors, il est aisé de montrer que le processus de croissance s'écrit :

$$\begin{pmatrix} N_a \\ N_b \end{pmatrix}_t = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} N_a \\ N_b \end{pmatrix}_{t-1}$$

Si on note $N(t)$ le nombre total de cellules au temps t (i.e. $N(t) = N_a(t) + N_b(t)$), alors, on trouve immédiatement :

$$N(t) = N(t-2) + N(t-1)$$

Il s'agit tout simplement d'une suite de FIBONACCI (on remarquera en outre que N_a et N_b suivent eux-mêmes une suite de FIBONACCI).

Ces suites permettent de rendre compte d'un certain nombre de processus de croissance "sub-exponentiels", on peut trouver des discussions sur ce point dans les articles de LEGAY (1968) et de VITANYI (1974) .

L'avantage de la présentation proposée dans le cadre des systèmes de LINDENMAVER est d'inciter à la formulation d'hypothèses sur la présence d'états cellulaires ou de classes d'âges (du type mature-immature). Il va sans dire qu'une grande variété de modèles peut être obtenue dans ce cadre, par exemple, en augmentant le nombre d'états, ou encore en modifiant les productions.

II.2-3 Modèles stochastiques en temps continu

On s'intéresse à l'évolution d'un phénomène aléatoire au cours du temps considéré comme une variable déterministe continue (on parle de processus permanent). On rencontre un certain nombre de difficultés dans la définition de tels processus qui conduisent à de nécessaires simplifications ; nous ne détaillerons pas ces difficultés techniques, un exposé concis peut être trouvé dans l'ouvrage de GIRAULT (1965) . On peut distinguer grossièrement le cas où l'ensemble des états possibles du système est continu (processus continu) de celui où l'ensemble des états est discret (processus discret).

- Les processus continus : citons par exemple les processus laplaciens, le mouvement brownien ... On peut ainsi avoir une approche stochastique des phénomènes de diffusion.

Ainsi, considérons le problème du déplacement des tRNA (acides ribonucléiques de transfert), lors de la synthèse des protéines, peut-on considérer qu'un processus de diffusion est suffisant pour expliquer les taux de synthèses de protéines (estimés par ailleurs) ou faut-il supposer l'existence d'un processus actif de transport de ces macromolécules ? Pour répondre à cette question, on peut envisager de tester l'hypothèse "nulle" (transport non actif gouverné par un mouvement brownien) et de comparer les taux de synthèse protéiques estimés sous cette hypothèse à ceux déterminés expérimentalement. Cependant, cette approche, non encore résolue, se heurte à des difficultés d'ordre mathématique (mouvement browniens dans un espace à trois dimensions, non isotrope ; cependant, au moins dans un premier temps, l'hypothèse d'isotropie pourrait être retenue.

- Le processus discret le plus connu est celui de POISSON. Un exemple d'utilisation de ce processus se trouve dans la construction du modèle dit de la "cible" qui permet d'explorer les phénomènes de mutations induits par les irradiations (UV, X ou γ).

EXEMPLE 10 : LE MODELE DE LA CIBLE : APPLICATION A L'ETUDE DE LA SYNTHÈSE DES CHLOROPLASTES [J. ESTEVE, P. NICOLAS et V. NIGON (1978)]

Le matériel biologique est une algue verte unicellulaire : Euglena gracilis, qui contient normalement une dizaine de chloroplastes par cellule. Ces chloroplastes, contenant de la chlorophylle, sont responsables de la couleur verte de cet organisme. Ils contiennent, en outre, des molécules de DNA (acide désoxyribonucléique) et il semble que ce DNA extranucléaire intervienne de façon déterminante dans la synthèse chlorophyllienne. Après irradiation U.V., il est possible de créer des mutants "blancs" où la mutation apparaît portée par les molécules de DNA extranucléaires. Le problème posé est celui de l'estimation du nombre de "cibles", que l'on peut associer à ce DNA dont l'atteinte a pour conséquence le blanchissement des cellules (i.e. pas de synthèse de la chlorophylle).

Ainsi, on suppose que chaque cellule contient c cibles ; chacune d'entre elles pouvant être inactivée par irradiation de façon identique et indépendante les unes des autres. L'inactivation de chaque cible suit alors un processus de Poisson indépendant pour chacune d'elles.

Alors, la probabilité qu'une cible ne soit pas atteinte après une irradiation d'une durée t est :

$$p_t(0) = e^{-at}$$

Inversement, la probabilité que cette même cible soit atteinte au moins une fois est :

$$p_t(\bar{0}) = 1 - e^{-at}$$

Et que les c cibles soient touchées :

$$p_{t,c}(\bar{0}) = (1 - e^{-at})^c$$

où a représente la sensibilité (ou l'affinité) d'une cible au rayonnement U.V.

On suppose enfin qu'une cellule maintient sa capacité de synthèse, i.e. sa couleur verte, si au moins une cible est sauvegardée. La probabilité d'un tel événement est alors :

$$p_{t,c}(V) = 1 - p_{t,c}(\bar{0}) = 1 - (1 - e^{-at})^c$$

probabilité qui peut être estimée par voie expérimentale en déterminant la fréquence du nombre de cellules vertes dans la population totale des cellules (il est clair que le modèle a été construit en fonction de la mesure accessible).

A partir des résultats expérimentaux, il est possible d'estimer numériquement les paramètres a et c et d'envisager la comparaison entre des variétés ou des espèces différentes de cellules. De même, à partir des hypothèses de base sur le processus poissonien, des modèles différents peuvent être construits, en supposant par exemple que deux événements sont nécessaires pour inactiver une cible.

En fait, il apparaît pratiquement que les conclusions que l'on pouvait espérer obtenir sont limitées. La forte corrélation entre les paramètres dans un même modèle rend délicate la comparaison entre expériences différentes. Ainsi, bien que l'on ait pu mettre en évidence une différence de sensibilité entre deux variétés d'euglènes, la comparaison entre les nombres de cibles obtenues par différents auteurs (avec des techniques d'estimation généralement différentes) ne permet pas de décider si la variabilité des résultats est due aux différentes techniques utilisées ou à une variabilité génétique propre aux souches étudiées.

On notera que cet exemple illustre la difficulté d'emploi des modèles ne dépendant pas linéairement des paramètres et les réserves à faire dans les conclusions, notamment lorsque des techniques d'inférence statistique sont utilisées pour décider de la différence entre des estimations de paramètres et de l'interprétation de ces différences.

Nous citerons, enfin, un exemple qui, en plus de son importance dans le domaine considéré (cinétique de croissance) nous paraît avoir un grand intérêt méthodologique. En effet, la technique utilisée ici dans le cadre d'un modèle dynamique est plutôt habituellement considérée comme propre à l'analyse de données.

EXEMPLE 11 : ANALYSE DES COURBES DE CROISSANCES D'INDIVIDUS OU DE POPULATIONS [J. ESTEVE, E. SCHIFFLERS (1976)]

Les auteurs présentent d'une part une revue bibliographique critique des méthodes d'analyse des courbes de croissance et d'autre part un nouvel outil de modélisation. Cet outil est une excellente démonstration du fait qu'une solution technique ne peut jamais être associée de manière rigide à un type de données ou à un modèle.

Les données utilisées sont des mesures du poids de N souris répétées chaque jour sur chacune des souris. La modélisation proposée consiste en une décomposition du processus multidimensionnel $Y(t)$ (les composantes de $Y(t)$ sont les poids des N souris au temps t) suivant la formule :

$$Y(t) = \mu(t) + \sum_{k=1}^d \alpha_k V_k(t) + \varepsilon(t)$$

avec :

- $\mu(t)$: moyenne de $Y(t)$
- $V_k(t)$: famille de fonctions du temps ($d < N$)
- α_k : variables normales indépendantes de variances unité
- $\varepsilon(t)$: une "erreur" aléatoire normale centrée.

Cette méthode est, dans son principe, relativement voisine de la

méthode classique utilisant une approximation par une somme de polynômes orthogonaux : il s'agit là aussi d'approcher un processus dans R^N par une combinaison de d fonctions du temps ($d < N$). L'originalité méthodologique réside dans la méthode d'estimation des V_k qui conduit à réaliser une analyse en composantes principales sur les N souris repérées chacune par leurs poids aux différents instants considérés. Ainsi, une technique souvent considérée comme propre à l'analyse de données "non dynamique" est utilisée ici dans le cadre d'un modèle dynamique. Les qualités propres de cette technique exploitées en analyse de données permettent ici une interprétation relativement simple des fonctions V_k (facteurs de l'analyse en composantes principales).

II.2-4 Modèles stochastiques en temps discret

Comme pour les modèles déterministes en temps discret, on considère qu'une horloge est associée au processus, c'est-à-dire que les événements se produisant à des temps discrets régulièrement espacés : par exemple les générations successives d'une lignée cellulaire (cette notion apparaît "naturellement" dans les lignées synchrones).

Les processus utilisés sont essentiellement les processus de Markov et les processus de ramifications. La distinction entre eux est d'ailleurs arbitraire dans la mesure où la deuxième classe de modèles est construite sur la base de processus MARKOVIENS individuels. Dans les deux cas, des théorèmes limites permettent de passer en temps continu.

i) Processus de MARKOV

On considère qu'un système peut se trouver dans un certain nombre d'états discrets $E_1, E_2 \dots E_j \dots E_n$, on connaît les probabilités de passage P_{ij} de l'état E_j à l'état E_i pendant l'intervalle de temps $t, t+1$. Ainsi, si la distribution des probabilités $(x_1, x_2 \dots x_i \dots x_n)_t$ des différents états à l'instant t est connue, on peut calculer cette distribution à l'instant $t+1$ (ceci caractérise l'aspect markovien du processus).

$$X_{t+1} = P \cdot X_t$$

où P est la matrice des probabilités p_{ij} .

Il est clair que :

$$X_{t_0+n} = P^n X_{t_0}$$

Les propriétés asymptotiques de ce type de processus peuvent se déduire de la matrice P. Ainsi, ce type de modélisation est-il particulièrement adapté à la prédiction de la distribution des états quand n devient grand.

La théorie et des applications de ces processus peuvent être trouvées dans l'ouvrage de KARLIN (1966) . Une introduction à l'utilisation de cette approche en biochimie et en biologie cellulaire et moléculaire, avec passage en temps continu, est exposée par JOLIVET (1978) .

ii) Processus de ramification

Cette classe de modèles permet d'analyser la croissance de populations, notamment lorsque celles-ci sont structurées en âge. Essentiellement utilisées en dynamique des populations d'organismes supérieurs, les méthodes et les techniques développées sont cependant susceptibles d'un plus vaste champ d'applications, notamment dans l'étude de populations cellulaires (on pourra se référer à l'ouvrage de JAGERS (1975), qui présente à la fois un développement théorique important et des applications).

Les matrices de LESLIE, classe d'opérateurs proposée par cet auteur dès 1945, ont été replacées plus tard dans le cadre plus général des processus de ramification (cf. par exemple HARRIS (1963)) : Elles conduisent à une modélisation particulièrement efficace pour l'analyse de la taille d'une population. Nous allons détailler maintenant ce point :

On s'intéresse à l'évolution du nombre d'individus d'une popula-

tion d'organismes, de cellules, voire de molécules .

. Les individus sont groupés en n classes d'âges successives de même durée choisies comme l'unité de temps.

. Ces individus sont caractérisés par deux paramètres : leur mortalité p et leur fécondité f qui ne dépendent que de l'âge.

Soit N_t vecteur moyen des effectifs des différentes classes d'âges au temps t ; il peut s'exprimer en fonction de N_{t-1} par l'expression suivante :

$$N_t = A N_{t-1}$$

$$\text{où } A = \begin{pmatrix} f_1 & f_2 & \dots & f_n \\ p_1 & 0 & \dots & 0 \\ 0 & p_2 & & \\ 0 & 0 & p_{n-1} & 0 \end{pmatrix}$$

Ce type de modélisation s'adapte bien à la fois à l'étude du passé d'un processus et à la prédiction de l'évolution future d'une population. Dans certaines conditions, où les f_i sont non tous nuls, on peut obtenir une structure stable des âges et ainsi définir un taux de croissance λ . L'étude de la sensibilité de ce taux λ à des paramètres "démographiques" est alors un aspect important de cette analyse.

On trouvera un exposé concis de ces problèmes dans LEBRETON (1973) . Notons également que la relation entre processus aléatoires et théorie des langages peut aisément être établie, ce qui permet une représentation schématique des processus aléatoires (cf. FU (1974) pour la classification de CHOMSKY, PAVÉ (1977) pour les systèmes de LINDENMAYER).

II.3 Conclusion

Comme nous venons de le voir, il existe un nombre important de types de modèles susceptibles d'être utilisés pour l'analyse d'un système biologique. Le choix d'un type de modélisation dépend avant tout de l'objectif de cette modélisation et notamment de l'utilisation que l'on souhaite faire du modèle.

Cependant, d'autres considérations doivent être prises en compte :

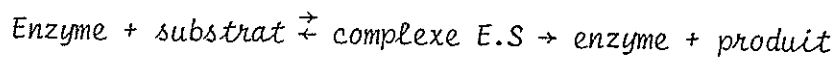
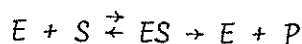
- . Les possibilités d'expérimentation,
- . La nature des données expérimentales,
- . Les problèmes techniques liés au modèle (facilités d'écriture et d'utilisation...)

III - PROBLEMES LIES A LA CONSTRUCTION DES MODELES

Historiquement, lorsque l'on a cherché à construire des modèles pour l'analyse des systèmes biologiques, 3 approches principales ont été utilisées :

i) La plus ancienne, illustrée par exemple par les travaux de RASHEVSKY (1961) , a été essentiellement une tentative d'application des théories biophysiques (dérivées de la physique) ou biochimiques (dérivées de la chimie) à l'étude des systèmes biologiques.

Par exemple, les lois de la cinétique enzymatique proposées par Michaelis Menten découlent de l'application des lois de la cinétique chimique à une réaction enzymatique représentée par le schéma de réaction suivant :



ii) Une autre approche consiste à proposer des lois phénoménologiques tirées de l'expérimentation et à construire des modèles sur cette base : par exemple, la classe des modèles à compartiments peut être rattachée à une telle démarche ainsi que la classe des modèles dérivés des travaux de [MONOD (1942)] sur la croissance bactérienne.

En effet, Monod ayant à rendre compte de la croissance de bactéries à partir d'un milieu nutritif (S), a eu l'idée d'exprimer le taux de croissance des bactéries en fonction de la quantité d'éléments nutritifs qui suit une loi analogue à celle donnant la vitesse d'une réaction enzymatique en fonction de la quantité de substrat.

$$\text{Taux de croissance} = \frac{dB}{Bdt} = \frac{\mu_m S}{K+S}$$

B = biomasse

S = éléments nutritifs

Les paramètres ainsi introduits (μ_m et K) gardent dans ce contexte une signification biologique (μ_m = taux max. de croissance, K = constante d'affinité de la biomasse pour les éléments nutritifs).

Cette loi, du fait de sa simplicité mathématique et de la signification biologique de ses paramètres, est très utilisée, en particulier pour rendre compte des phénomènes de croissance de microorganismes dans les procédés de fermentation (cf. par exemple [LUKASIK (1974)]).

On pourra remarquer, à ce propos, les similarités entre des lois qui semblent être valides à différents niveaux d'organisation [VON BERTALANFFY (1973)]. Ainsi, la loi dite "d'action de masse" et ses équivalents paraît avoir un caractère de généralité qui dépasse les phénomènes chimiques ou biochimiques (cf, par exemple, GARFINKEL (1968) (1977)).

iii) La 3ème approche fait appel à un modèle de représentation. Contrairement aux 2 approches précédentes, pour lesquelles la construction de modèle s'appuyait sur des hypothèses de mécanismes, cette dernière ne vise qu'une bonne description de l'évolution des variables observées.

Prenons l'exemple d'un dosage biologique, ou biochimique. On cherche à mettre en relation une variable y représentant la quantité, ou la concentration d'un produit avec une variable x représentant la mesure (par exemple une densité optique). En général, on cherche à établir la relation :

$$y = f(x)$$

en procédant à un étalonnage.

Même si un modèle "théorique" peut être construit sur la base de mécanismes (réaction chimique, diffusion ou absorption de lumière par l'échantillon...), il n'est pas certain qu'il fournisse la meilleure estimation de y connaissant x . En effet, il peut avoir des caractéristiques qui rendent son utilisation difficile et entraînent des erreurs d'interpolation (par exemple, des non-linéarités en fonction des paramètres). Il pourra donc être plus judicieux d'utiliser une représentation qui assure de bonnes propriétés d'interpolation, ou encore de modèles linéaires ou paraboliques construits sur une base statistique (qui permettent notamment d'associer une précision à l'estimation de y).

Il apparaît donc qu'un même phénomène biologique puisse être modélisé suivant différentes approches en fonction du problème à résoudre.

Quelle que soit l'approche utilisée pour la construction d'un modèle, des problèmes de nature variée (mathématique, numérique, expérimentale...) sont rencontrés au cours des différentes étapes de la modélisation. Pour des raisons de simplicité, nous avons choisi de les présenter étape par étape. Ainsi, nous distinguerons dans la modélisation les 3 étapes suivantes :

- la délimitation et la caractérisation du système incluant la mise en équation
- la simulation et l'identification des paramètres incluant la validation
- l'expérimentation.

III.1 Problèmes liés à la délimitation et à la caractérisation du système

III.1-1 Délimitation du système

C'est la première étape de la modélisation, elle consiste d'abord à "isoler" le système biologique à étudier de son environnement, puis à préciser ses relations avec cet environnement. La nature même de l'organisation des systèmes vivants entraîne que l'étude à un niveau donné (population, organisme, cellule, molécule) n'est jamais totalement indépendante des autres niveaux. Il faut donc juger de l'importance et de la nature de ces relations.

Cette étape de délimitation est essentielle mais très délicate ; aucune règle universelle ne peut être suivie ; chaque exemple est un cas d'espèce. On peut seulement remarquer que la prise en compte des objectifs et des connaissances biologiques préalables doit servir de base.

Ces problèmes de délimitation ont été soulevés dans l'exemple 7 à propos de la modélisation du système coagulolytique (cf. p. 25).

III.1-2 Caractérisation du système

Le système ayant été délimité de son environnement, l'étape suivante consiste à définir les variables caractéristiques du système, c'est-à-dire celles qui définissent entièrement l'aspect du système auquel on s'intéresse. On distinguera parmi ces variables: les variables internes ou d'état, les variables d'entrée ou d'excitation, et les variables de sortie accessibles à la mesure. Il n'existe pas forcément un ensemble unique de variables caractéristiques ; le choix tiendra compte des variables directement mesurables ou excitables et nécessitera le plus souvent une expérimentation préliminaire.

Après cela, il s'agit de préciser les interactions entre les variables caractéristiques de manière à définir un schéma fonctionnel du système à étudier. Ce schéma fonctionnel traduit les principales hypothèses biologi-

ques et sert de base à l'écriture du modèle mathématique. Un exemple typique est le schéma d'un système à compartiment.

Comme dans l'étape précédente, les difficultés rencontrées ici sont de nature biologique. On peut remarquer que, dans certains cas, la représentation du système par un schéma fonctionnel apporte une réponse suffisante aux biologistes.

Enfin, la dernière opération dans cette étape consiste à mettre le système en équation. Il est important de souligner que le choix de la formulation mathématique dépend à la fois des possibilités d'expériences et des objectifs de la modélisation. C'est de la considération simultanée de ces 2 critères que résulte le choix d'une classe de modèles. Par exemple, il est évident que si l'on construit un modèle pour faire de la commande, cela exclut pratiquement l'utilisation des modèles stochastiques mal adaptés aux traitements mathématiques de ce problème.

En outre, la nature même des systèmes biologiques conduit la plupart du temps à considérer des modèles incluant des non-linéarités venant de la complexité des interactions. Or, l'analyse mathématique des systèmes dynamiques non linéaires est très limitée, ce qui conduit, dans certains cas, à linéariser le modèle quitte à limiter son domaine de validité.

III.2 Problèmes liés à la simulation et à l'identification des paramètres

Nous allons examiner successivement les problèmes de simulation, d'estimation de paramètres à partir de données expérimentales et de validation du modèle. Il est clair que les problèmes de simulation doivent être résolus pour procéder à l'estimation des paramètres. De la même façon, il est nécessaire de connaître des valeurs numériques des paramètres pour valider le modèle.

III.2-1 Simulation

En simulation, pour les modèles déterministes, on rencontre des difficultés essentiellement dans les procédures d'intégration numérique alors que pour les modèles stochastiques, la difficulté majeure vient de la qualité des générateurs de nombres aléatoires. A ce propos, il faut remarquer qu'une mauvaise utilisation d'un générateur donne souvent lieu à des déviations indétectables par rapport au résultat cherché souvent dues à la présence de pseudo périodes dans la suite des nombres aléatoires obtenus. On pourra trouver dans CHASSÉ, DEBOUZIE (1974) une étude systématique d'un certain nombre de générateurs et de leurs propriétés statistiques.

Pour les modèles déterministes, les difficultés numériques d'intégration proviennent d'une part des non-linéarités des systèmes et d'autre part de l'occurrence simultanée de processus lents et rapides. Dans ce dernier cas, l'utilisation d'une méthode d'intégration du type prédiction-corrrection (par exemple méthode de Gear) est nettement préférable à celle d'une méthode classique (Runge-Kutta) tant au plan de la précision qu'à celui du temps de calcul. Ces difficultés d'intégration peuvent être évitées par l'utilisation de modèles discrets. Cependant, en dehors des systèmes linéaires, il n'existe pas de méthode générale de discrétisation et de plus, il est souvent difficile de construire directement un modèle discret d'un système biologique.

Un autre point de vue à considérer pour la simulation est l'accessibilité à des outils de calcul plus ou moins puissants. En effet, il existe des langages, adaptés à la simulation et très simples d'utilisation comme Dynamo, CSMP ... qui peuvent être de bonnes solutions pour répondre aux besoins des biologistes [MEYER (1978)] en leur facilitant la programmation. Cependant, ils masquent souvent des difficultés numériques liées à l'intégration.

On notera également l'existence de langages spécifiques à l'étude de certains systèmes comme le langage développé par [GARFINKEL (1968)] pour l'étude des cinétiques chimiques et enzymatiques.

III.2-2 Estimation des paramètres

L'estimation des paramètres consiste à déterminer les valeurs numériques de ces paramètres à partir de données expérimentales quantitatives, ainsi que les propriétés de ces estimations, comme par exemple la précision. Deux méthodes sont les plus utilisées:

. La méthode du maximum de vraisemblance qui consiste à trouver un ensemble de valeurs de paramètres tel que la fonction de vraisemblance soit maximum. Maximiser cette fonction revient à considérer que les données expérimentales utilisées sont les plus probables, sinon un tel échantillon de données n'aurait pu être observé.

. La méthode des moindres carrés qui revient à minimiser une distance entre les données expérimentales et les mêmes grandeurs calculées avec le modèle. Cette méthode est la plus couramment utilisée, elle est équivalente à la précédente quand les erreurs expérimentales sont gaussiennes, ce qui est fréquemment le cas.

Evidemment, on peut trouver les valeurs des paramètres si le problème de minimisation a au moins une solution. Dans le cas linéaire (linéarité par rapport aux paramètres), la recherche du minimum revient à résoudre un système d'équations algébriques linéaires. Dans le cas non linéaire, on utilise des méthodes numériques itératives (cf. algorithmes de la librairie HARWELL (1973)), qui très souvent, conduisent à des difficultés numériques (problèmes de convergence et de stabilité des solutions) et qui, par conséquent, peuvent devenir très onéreuses en temps calcul lorsque les équations sont mal conditionnées. De plus, dans le cas non linéaire, ces méthodes d'estimation ne conduisent pas forcément à une solution unique et il n'est généralement pas possible de démontrer l'unicité ou la non unicité des solutions sinon par une exploration numérique.

Quant aux propriétés statistiques des estimations, elles sont souvent caractérisées par la matrice de covariance des paramètres qui peut souvent être obtenue à la fin d'une procédure de minimisation [SEBER (1973), RAO (1973)]. Il est alors possible de fournir des domaines de confiance associés aux paramètres et de comparer les valeurs des paramètres obtenues à partir de diverses

expériences. Il est important d'étudier la corrélation entre paramètres et, si besoin est, de reparamétriser le modèle afin d'améliorer les propriétés numériques et statistiques de l'estimation [ROSS (1970) (1975)].

III.2-3 Validation

Valider un modèle consiste à accepter ou à refuser ce modèle comme étant une bonne représentation des résultats expérimentaux. Les critères utilisés peuvent être qualitatifs ou quantitatifs.

. Si un critère qualitatif (ou critère de forme) est pris en considération, la validation du modèle se fondera sur la ressemblance entre la réponse du modèle et les résultats expérimentaux. On peut trouver 2 types d'exemples : les modèles qui visent à simuler des formes (morphogénèse) et les modèles avec lesquels on cherche à reproduire uniquement l'allure des résultats expérimentaux (courbes croissantes, réponses oscillantes...). Dans ce dernier cas, il est souvent possible de trouver plusieurs fonctions donnant une même forme, aussi une grande prudence s'impose dans l'interprétation des résultats.

. Quant aux critères quantitatifs, ils sont exprimés en terme de distance entre la réponse du modèle après estimation des paramètres et les résultats expérimentaux. Dans ce cas, des tests statistiques (test χ^2 , test de linéarité...) sont utilisés pour la validation ; en général, ces tests sont construits sur la base de "l'hypothèse nulle" (on suppose que les différences entre données et réponses du modèle sont distribuées suivant une loi de probabilité de moyenne nulle). Cependant, il est important de vérifier que les conditions d'application du test sont vérifiées, par exemple : beaucoup de tests supposent l'indépendance des mesures entre elles. A ce sujet, on pourra se référer aux synthèses de WRITHT (1972) et de STEINHORST (1977) , où l'on trouvera également quelques indications quant au choix des tests suivant leur puissance. Enfin, signalons que les résultats de ces tests ne sont valables que pour le domaine expérimental exploré et n'offre aucune garantie lorsque le modèle est utilisé en prédiction.

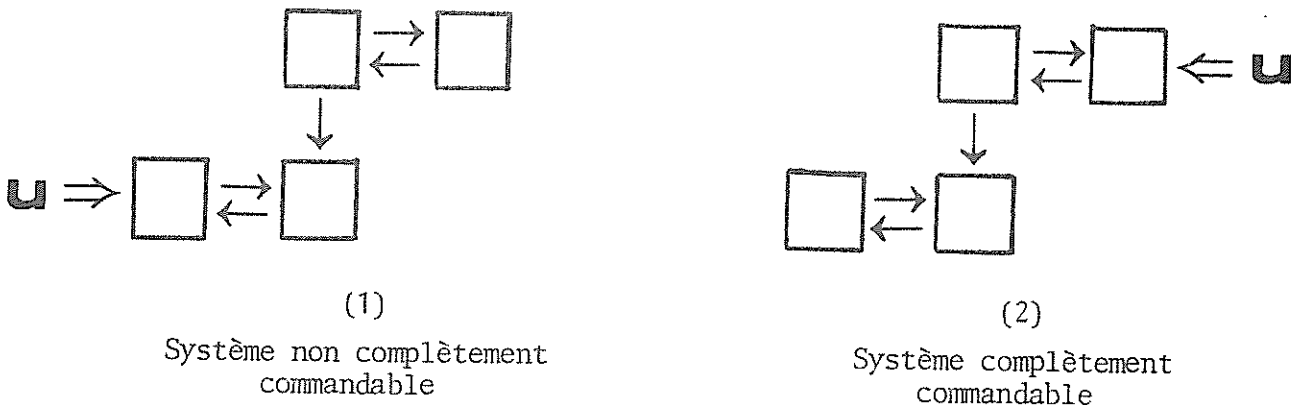
Dans la pratique, il est bien rare que le modèle soit retenu après

une 1ère identification et souvent l'amélioration de l'identification passe par la révision du protocole expérimental.

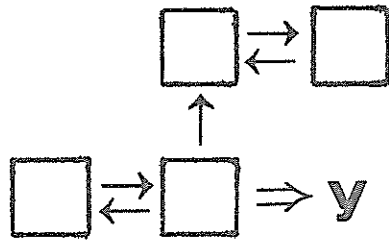
III.2-4 Identifiabilité des systèmes

Lorsque les problèmes d'identification numérique et de validation sont résolus, alors le système est considéré comme identifié, c'est-à-dire que le modèle mathématique est une bonne représentation du système biologique dans le domaine où il a été expérimenté. Pour certaines classes de modèles, il est possible de vérifier a priori si cette identification a un sens, il s'agit alors de résoudre le problème "d'identifiabilité". Par exemple, pour les modèles régis par un ensemble d'équations différentielles linéaires, on peut étudier l'identifiabilité du système s'il est à la fois complètement "observable" et complètement "commandable". Nous allons illustrer ces notions de commandabilité et d'observabilité sur les systèmes à compartiments, régis par des équations différentielles linéaires que nous avons déjà introduites à propos d'exemples (cf. exemple 6).

Un tel système sera complètement commandable (C.C) s'il existe au moins un chemin reliant tout compartiment commandé à un compartiment qui ne l'est pas.

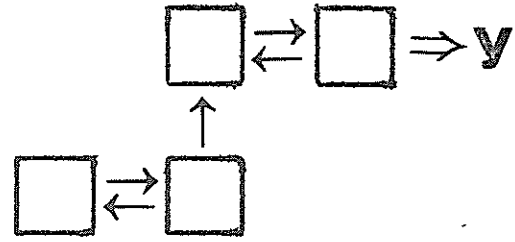


De la même façon, un système est complètement observable si tout compartiment non directement observé est relié par au moins un chemin à un compartiment qui l'est.



(1)

Systeme non complètement observable



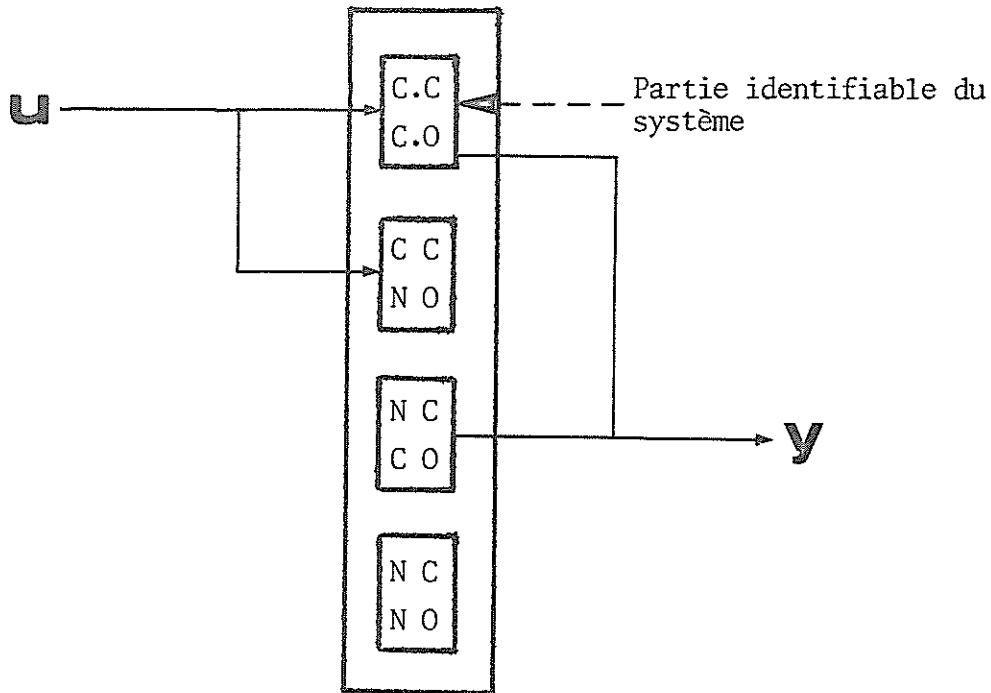
(2)

Systeme complètement observable

Remarques

1) On ne sait étudier les conditions d'identifiabilité que pour des systèmes d'équations différentielles linéaires.

2) Un système linéaire peut toujours être décomposé de la façon suivante, relativement aux propriétés de commandabilité et d'observabilité des sous-systèmes qui le composent :



3) La notion de commandabilité n'apparaît pas toujours explicitement lors de l'étude d'un système biologique car on maîtrise mal ou indirectement les excitations u du système. Par contre, la notion d'observabilité

est évidemment essentielle et prend le même sens que celui défini ci-dessus.

4) En ce qui concerne les systèmes à compartiments, surtout utilisés pour suivre la dynamique de ce système via un traceur (colorant, marqueur radioactif...), la notion de commande est explicite : le (ou les) compartiment(s) commandé(s) recevra le traceur en question.. On pourra trouver un développement plus détaillé des procédures à suivre pour vérifier *a priori* l'identifiabilité dans les articles de BELLMAN et ÅSTRÖM (1970) et de COBELLI et ROMANIN-JACUR (1975) (1976).

III.3 Problèmes liés à l'expérimentation

L'élaboration d'un modèle mathématique s'appuie toujours sur une base expérimentale. Aussi est-il nécessaire de bien connaître les possibilités d'expérimentation sur le système biologique à étudier afin de pouvoir l'adapter en vue, par exemple, d'une bonne identification. A ce propos, notons que l'expérimentation peut avoir recours à des méthodes qui permettent d'optimiser les protocoles expérimentaux : par exemple, des plans d'expériences peuvent être élaborés pour une analyse statistique ou pour une analyse dynamique particulières.

Ici, nous nous limiterons à citer les principales sources de difficultés en expérimentation :

- . La complexité des systèmes biologiques,
- . La mauvaise reproductibilité des expériences (due à la variabilité et à la non stationnarité),
- . Les expériences sont souvent faites *in vitro* et la transposition *in vivo* est délicate,
- . Le manque de capteur conduit souvent à des mesures indirectes : par exemple, la thermogénèse ou la densité optique sont souvent utilisées

pour estimer la taille d'une population microbienne. De même, pour doser un composé biologique, on lui associe une réaction enzymatique dont le produit possède des caractéristiques physiques mesurables (production de chaleur, fluorescence, colorimétrie, ...).

. les variables biologiques sont rarement mesurées de façon continue, la mesure nécessite souvent la prise d'échantillon,

. Beaucoup d'expériences sont coûteuses en matériel, temps et produits,

. Les possibilités d'excitation des systèmes biologiques sont limitées.

En conclusion, l'analyse des systèmes biologiques par la modélisation, si elle s'avère une voie prometteuse, se distingue néanmoins de l'analyse des systèmes physiques ou même chimiques par la multiplicité et la complexité des problèmes rencontrés. Comme on a pu le voir, ces problèmes concernent différents domaines de l'activité scientifique et il n'est pas toujours possible de leur apporter une solution, même approchée.

IV - CONTRIBUTION DE LA MODELISATION A L'ETUDE DES SYSTEMES BIOLOGIQUES

Dans les chapitres précédents, pour chaque présentation d'exemples, nous avons pris soin de préciser les objectifs de la modélisation dont il était question, et, pour chaque type de modèle, nous avons souvent fait des remarques sur leurs possibilités d'utilisation. Aussi, dans ce chapitre, nous chercherons avant tout à dégager la contribution de la modélisation à l'étude des systèmes biologiques.

En nous appuyant sur les exemples déjà cités; nous allons montrer comment l'utilisation de modèles mathématiques permet ou contribue à :

- l'analyse de structures biologiques,
- l'analyse et le test de mécanismes,
- la prédiction et la simulation de systèmes biologiques
- l'estimation de paramètres non accessibles,

- l'optimisation de protocoles expérimentaux,
- le contrôle de procédés biologiques.

IV.1 Analyse des structures biologiques

Cette contribution relève essentiellement de l'analyse non dynamique telle que nous l'avons décrite au paragraphe II.1. Ainsi, dans l'exemple 1 de ce paragraphe, l'auteur a montré la symétrie existant entre les glandes séricigènes gauche et droite d'un ver à soie dans une population particulière. On peut indiquer que d'autres expériences de cet auteur ont montré la généralité de cette propriété l'autorisant ainsi à considérer globalement les deux glandes simplifiant donc le plan expérimental (cf. IV.5 - Optimisation de protocole expérimental). Cependant, la correspondance entre l'analyse "non dynamique" d'un système et l'étude d'une structure biologique n'est en rien naturelle.

Dans l'exemple 3, la principale conclusion biologique est l'existence de contraintes s'exerçant, au sein d'un individu, sur les gènes indépendamment de la protéine codée. Ceci suggère donc l'existence d'une évolution des gènes non réductible à celle des protéines : l'information apportée porte sur l'aspect dynamique d'un mécanisme biologique alors que l'analyse était purement non dynamique. De même, bien que nous n'ayons pas été confrontés personnellement avec cette situation, soulignons qu'une analyse dynamique peut être utilisée à la mise en évidence d'une structure biologique particulière. Il apparaît une nouvelle fois ici que la modélisation ne peut résulter d'une démarche purement mécanique.

IV.2 Analyse et test de mécanismes

Très souvent en biologie, les modèles sont élaborés et utilisés pour faire du test d'hypothèses ou de mécanismes, ce qui correspond à la démarche suivante :

1. A partir de la connaissance que l'on a a priori sur le système biologique à étudier, on est à même de formuler certaines hypo-

thèses sur son fonctionnement ou sa structure.

2. On construit alors un modèle qui prend en compte ces hypothèses et les traduit mathématiquement.
3. On utilise ce modèle pour simuler le "comportement" du système biologique, en particulier dans des conditions où l'on peut l'expérimenter.
4. On compare résultats expérimentaux et résultats simulés afin d'apprécier la validité des hypothèses de départ.

Il est très important de souligner qu'un tel test d'hypothèse n'est vraiment concluant que s'il est négatif, c'est-à-dire si l'hypothèse conduit à des résultats incompatibles avec les données expérimentales. Dans un tel cas, l'hypothèse doit être rejetée alors que dans le cas contraire, on peut simplement dire que l'hypothèse est compatible avec les données expérimentales considérées. Mais, l'on ne peut absolument pas garantir d'une part qu'il n'existe pas d'autre hypothèse permettant la même compatibilité et, d'autre part, que l'hypothèse en question serait compatible avec des données expérimentales obtenues dans d'autres conditions.

Il existe de nombreux exemples illustrant une telle démarche. Parmi les exemples déjà cités, nous relèverons entre autres l'exemple 7 (analyse du système coagulolytique) où un modèle a été construit pour tester si certaines hypothèses proposées par les biologistes étaient compatibles avec les propriétés d'équilibre du système. De plus, nous citerons un nouvel exemple très typique de cette démarche et qui concerne la détermination du mécanisme de réaction d'une enzyme à partir de l'analyse de ses cinétiques.

EXEMPLE 12 : DETERMINATION DU MECANISME DE REACTION D'UNE ENZYME
A PARTIR DE SES CINETIQUES [A. CHERUY (1974)]

Il s'agit ici de déterminer le mécanisme de réaction d'une enzyme c'est-à-dire l'enchaînement des réactions élémentaires de l'enzyme avec chacun

de ses substrats, inhibiteurs et produits. Par exemple, pour une enzyme à 2 substrats S_1 et S_2 , il s'agit de savoir si le substrat réagit d'abord avec l'enzyme indépendamment du substrat S_2 ou si S_1 et S_2 se fixent simultanément sur l'enzyme pour qu'un 1er produit de réaction soit libéré ; et il s'agit également de déterminer la dynamique de chacune des étapes élémentaires de la réaction. Pour ce faire, on dispose en général d'expériences donnant la vitesse globale de réaction en fonction de différentes concentrations en substrats, inhibiteurs et produits (cinétiques à l'état stationnaire).

La méthode proposée dans [CHERUY (1974)] est la suivante :

1. On considère un mécanisme de réaction a priori à partir des connaissances que l'on peut avoir.
2. On écrit les équations cinétiques correspondantes et on déduit l'expression de la vitesse de réaction en fonction des concentrations en substrats, produits et inhibiteurs.
3. On teste d'abord qualitativement l'adéquation avec les données expérimentales en comparant par exemple l'allure des courbes de vitesse théoriques et expérimentales. Ensuite, si le test qualitatif ne révèle pas d'incompatibilité, on procède à un test quantitatif en estimant à partir des données expérimentales les paramètres introduits dans l'équation de vitesse.
4. Si l'un de ces tests n'est pas satisfait, un nouveau mécanisme de réaction est considéré et l'on est ramené à 2). Par contre, si le test quantitatif est satisfaisant avant de retenir le mécanisme de réaction, de nouvelles expériences sont effectuées pour tester sa validité dans d'autres conditions expérimentales intéressantes.

Cette méthode a été mise en oeuvre sur l'enzyme nucléoside diphosphate kinase et a permis de proposer, pour cette enzyme, un mécanisme prenant en compte l'action des 2 substrats.

A propos d'analyse de mécanisme, nous noterons également l'exemple 8 (une approche de la dynamique des macromolécules) où une analyse, à l'aide d'un modèle, de l'évolution du RNA dans la glande séricigène du ver à soie a conduit à faire des propositions sur le mécanisme de la sélection des vers à soie en vue d'améliorer la production de soie. Alors que "logiquement", l'on pouvait penser qu'une telle sélection devrait viser à augmenter les taux de synthèse, il s'avère qu'elle consiste plutôt à limiter le processus de dégradation.

IV.3 Prédiction et simulation de systèmes biologiques

Même s'ils n'ont pas été construits pour un objectif de simulation, les modèles sont souvent utilisés pour simuler le système biologique. Mais, pour qu'une telle simulation soit valable, les conditions de validité du modèle doivent être respectées. Cette possibilité de simulation est particulièrement intéressante en biologie où il est rare que l'ensemble des variables internes du système soient accessibles. Le principal intérêt de la simulation est qu'elle permet d'avoir l'évolution simultanée de toutes les variables internes du système.

Pour illustrer ce propos, reprenons l'exemple 7 sur le système coagulolytique. Ayant étudié les propriétés d'équilibre in vivo de ce système, un autre volet de l'étude a consisté à rechercher quelle information sur cet équilibre pouvait être apportée par des expériences in vitro sur du sang prélevé.

Ainsi, des conditions d'expérimentation sur des prélèvements sanguins ont été définies et l'on a alors établi le modèle permettant de rendre compte des résultats expérimentaux. Alors qu'expérimentalement on ne mesure que globalement la formation ou la lyse du caillot de sang, par simulation, pour chaque expérience, le modèle permet d'avoir l'évolution simultanée de toutes les variables internes. On a pu ainsi mettre en évidence que telle variable interne était très sensible à telle condition expérimentale et même situer l'impact de certaines perturbations, par exemple des

drogues, ce qui présente un grand intérêt du point de vue biomédical.

Ainsi, en s'appuyant sur des exemples de la littérature, on peut montrer clairement que de nombreuses contributions de la modélisation à l'analyse des systèmes biologiques viennent du fait qu'un modèle permet d'effectuer des simulations.

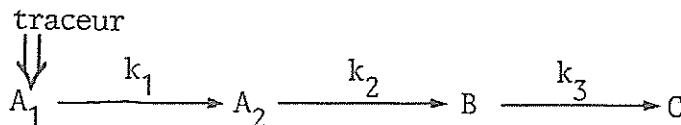
Quant à l'utilisation des modèles pour la prédiction, elle est relativement courante mais à notre avis souvent criticable. Nous ne présenterons pas d'exemple mais nous soulignerons cependant les difficultés de ce type d'analyse pour les systèmes biologiques, difficultés dues entre autres à la non-linéarité des systèmes.

Cependant, nous citerons à ce propos les études de sélection. En effet, la mise en place de plans de sélection en vue de l'amélioration économique de plantes ou d'animaux, relève de raisonnements de type prédictif.

IV.4 Estimation de paramètres non directement accessibles

Dans les modèles utilisés en analyse dynamique, apparaissent un certain nombre de paramètres qui peuvent avoir une signification biologique précise et la contribution de la modélisation peut être l'estimation de ces paramètres associés à la quantification d'un phénomène biologique.

Ainsi, reprenons le modèle à compartiment de l'exemple 6. Le but de cette modélisation était d'estimer le temps de synthèse d'une protéine (la fibroïne en l'occurrence), et le temps de transfert intracellulaire de cette molécule du lieu de la synthèse à l'extérieur de la cellule. La structure choisie est :



La plupart du temps, l'estimation de ces paramètres est faite sur la base de représentations graphiques des résultats expérimentaux, donc im-

précise et variable car les critères retenus dépendent des auteurs.

Or, à partir du modèle à compartiment défini ci-dessus, il est possible de déduire aisément la loi de probabilité du temps d'attente d'une molécule dans un compartiment quelconque, et on montre que le temps de transfert moyen t_n d'une molécule dans le nième compartiment (i.e. temps de transfert = temps d'attente) est donné par l'expression :

$$t_n = \frac{1}{k_n}$$

(sauf pour le dernier compartiment où ce temps est infini).

Donc, pour estimer ce paramètre, il suffit d'estimer les paramètres k_1 , k_2 et k_3 par les techniques classiques d'estimation.

En outre, le temps de passage dans deux compartiments est la somme des temps de passage dans chacun d'eux (i.e. additivité des temps de transfert). Ainsi, le temps de transfert dans le compartiment A peut être estimé par :

$$t_A = t_{A_1} + t_{A_2} = \frac{1}{k_1} + \frac{1}{k_2}$$

Ce temps de transfert peut être associé au temps de synthèse. Ainsi a-t-il été possible, connaissant la composition moyenne et le poids moléculaire de la fibroïne, d'estimer le taux d'incorporation moyen des acides aminés (au moins son ordre de grandeur : environ 2 acides aminés par seconde). Ce résultat est comparable au taux d'incorporation obtenu sur d'autres systèmes protéosynthétiques chez les eukaryotes (notamment la synthèse du collagène et la synthèse des protéines exogènes du pancréas).

Cet exemple illustre la complémentarité du modèle par rapport à l'expérience. En effet, si l'estimation du taux d'incorporation des acides aminés dans la chaîne protéique est possible par voie expérimentale, il s'avère que l'expérimentation suivie dans notre exemple est plus simple et permet d'obtenir plus d'information sur les processus cellulaires (notamment l'estimation du temps de transit du lieu de synthèse à la périphé-

rie cellulaire).

Cependant, il y a lieu de considérer de telles approches à leur juste valeur. Dans l'exemple que nous venons de discuter, des ordres de grandeur étaient suffisants et nous n'avons pas eu à nous interroger sur la précision des estimations. Par contre, l'exemple 10 (modèle de la cible) incite à plus de prudence, les auteurs de ce travail ont été en fait conduits à une discussion des procédures d'estimation, assez largement utilisées, plutôt qu'à répondre à l'objectif premier d'estimation de ce nombre de cibles. En effet, cette estimation s'avère trop variable pour que des conclusions raisonnables puissent être tirées, même quant à l'origine précise de cette variabilité.

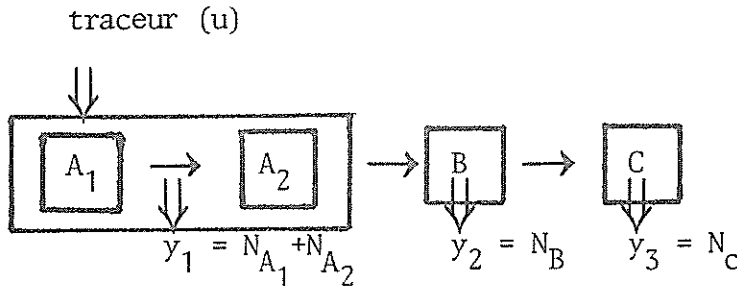
IV.5 Optimisation des protocoles expérimentaux

Un autre type de contribution de la modélisation à l'analyse des systèmes biologiques est la conception et l'optimisation du protocole expérimental en vue de la minimisation d'un critère (coût) ou de l'utilisation d'un modèle adapté à l'objectif.

Ainsi, les modèles de la statistique conduisent à la notion de plans d'expérience. La construction d'un tel plan tient compte du modèle sous-jacent et de la technique mathématique utilisée pour l'analyse statistique (par exemple, le modèle linéaire dans l'analyse de la variance). En statistique, toute estimation et tout test d'hypothèse dépend de la procédure expérimentale (il est souhaitable de choisir avant l'expérience cette procédure, sinon l'existence d'un traitement des résultats n'est pas assurée). On pourra remarquer qu'une part importante de l'effort théorique dans cette discipline a concerné la conception de plans expérimentaux conférant une grande puissance à la méthode d'analyse [Cochran, Cox (1957); Coll. Int. CNRS (1961)]. Cet aspect développé au départ pour répondre aux besoins des agronomes est de plus en plus utilisé par les biologistes.

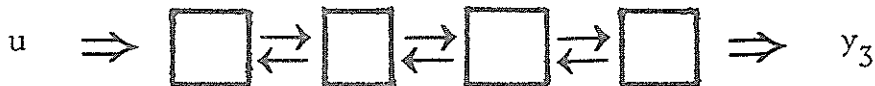
L'optimisation de protocole peut s'envisager dans le cadre d'autres modèles. Nous prendrons ici l'exemple des modèles à compartiments,

pour lesquels nous avons souligné l'importance des notions de commandabilité et d'observabilité. Reprenons l'exemple 6 de la synthèse de la fibroïne.



Dans l'expérience réalisée, le traceur (acide aminé marqué) est incorporé au niveau du premier compartiment (c'est en fait une contrainte due à la nature du système : une incorporation dans B ou C de ce traceur n'a aucun sens au plan biologique). Il est clair que ce système est C.C. et C.O.. A partir des résultats de BELLMAN et ÅSTRÖM (1970) et de COBELLI et ROMANIN-JACUR (1976) , il est possible de montrer que l'observation de y_3 suffit à identifier ce système (a fortiori l'observation de y_1 , y_2 et y_3).

Par contre, supposons que le système suivant ait été retenu :



Il est C.C. et C.O. mais on peut démontrer que la seule observation de y_3 ne permet pas de l'identifier (cf. BELLMAN & ÅSTRÖM (1970)).

Ainsi, une étude préalable d'un système, dans le cadre d'une classe de modèles choisie a priori, peut conduire à définir précisément un protocole expérimental, qui non seulement peut optimiser l'expérience envisagée mais assure que l'utilisation du modèle ne sera pas impossible ou inadaptée par manque d'information.

IV.6 Contrôle de procédés biologiques

Les modèles peuvent être utilisés afin d'analyser les propriétés de régulation propres aux systèmes biologiques ou encore afin d'élaborer une commande de procédés utilisant du matériel biologique.

Parmi les systèmes biologiques dont on a étudié les mécanismes de régulation, citons comme exemples le système de thermorégulation chez les homéothermes [DUNCAM & al. (1972)] et certains systèmes de régulation hormonale. [YATES & al. (1968)].

La démarche généralement suivie consiste à faire des analogies avec les systèmes de régulation physique. En effet, les régulations biologiques, comme les régulations physiques, procèdent généralement par rétroaction ; cependant, l'analyse des écarts avec ces modèles de référence a permis de mettre en évidence la grande complexité des régulations biologiques (non-linéarités et multiplicité des interactions) et leur possibilité d'adaptation aux conditions expérimentales et à l'environnement. Dans de telles analyses, les modèles utilisés sont les modèles d'analyse dynamique régis par des équations différentielles, c'est-à-dire ceux qu'utilisent les automaticiens pour concevoir des régulations.

En génie biologique, on est souvent confronté avec un problème de commande de procédés utilisant du matériel biologique. Citons, par exemple, les fermentations où les procédés mettent en jeu des microorganismes dont on utilise les propriétés de biosynthèse pour produire des substances comme des protéines, ou des antibiotiques. Pour concevoir toute commande de ces procédés, un modèle est nécessaire. Ce modèle doit permettre de rendre compte du comportement dynamique du système, de sa réponse, lorsque ce dernier est soumis à certaines excitations (variables de commandes). Aussi a-t-on à faire à des modèles où l'on s'attache plutôt à représenter le comportement entrée-sortie du système qu'à décrire l'évolution de ses variables internes. Les modèles les plus utilisés dans ce domaine sont ceux régis par des équations différentielles (si possible linéaires à coefficients constants) car ce sont les seuls avec lesquels actuellement on sache déterminer sans ambiguïté une commande optimale.

Comme exemple d'application, citons la commande d'un procédé de fermentation produisant un antibiotique [BOURDAUD (1974), CHERUY, DURAND (1978)]. Le but de commande était de maximiser la production d'antibiotique. Le 1er problème à résoudre dans une telle étude est le choix des variables de commande. Ces variables doivent d'une part être techniquement "maîtrisables", c'est-à-dire qu'on doit pouvoir les actionner facilement et, d'autre part, elles doivent bien "exciter" le système, c'est-à-dire que toute variation de ces commandes doit entraîner une variation sensible sur la réponse du système (ici la production d'antibiotique). En général, le 1er point est souvent limitant en génie biologique ; par exemple, dans notre cas, nous avons choisi comme variables de commande la température et le pH du fermenteur qui sont 2 grandeurs physiques facilement maîtrisables. Or, il était évident que de telles variables n'étaient pas les plus "efficaces" pour agir sur la production d'antibiotique -une variable "efficace" étant, par exemple, la quantité de nutriment apportée au fermenteur-. Cependant, le contrôle de la température et du pH, s'il n'a pas permis d'augmenter sensiblement la production, a permis de diminuer de 30 à 40 % la durée de la fermentation, ce qui correspond à une augmentation d'autant de la productivité du procédé. La seconde étape, dans une telle étude de commande, est l'élaboration du modèle et la conception des expériences permettant l'identification. Dans cette étape, les possibilités de choix du modèle sont assez limitées, étant donnée l'utilisation qui en sera faite. Cependant, il peut être intéressant d'élaborer un modèle plus complet, prenant en compte, par exemple, les connaissances biologiques que l'on a sur le système quitte à le simplifier ensuite pour l'utiliser en commande. Dans cette seconde étape, une source importante de difficultés vient du fait que l'on dispose en général d'un nombre très limité de capteurs, c'est-à-dire que les possibilités de mesures sont réduites. Quant à la 3ième étape, qui concerne la commande proprement dite, il est bien évident que ses performances dépendront de la "qualité" du modèle.

Comme autres exemples de l'utilisation des modèles à des fins de commande en biologie, nous noterons la définition de posologies optimales en pharmacologie. Il s'agit là de définir une prise de médicament au cours du temps pour obtenir une bonne efficacité tout en limitant les contraintes d'absorption [BOISVIEUX (1977)].

De manière générale, on retiendra l'importance de la notion de commande en génie biologique.

V - CONCLUSION

Notre exposé a été limité à la considération de modélisations se fondant sur des outils mathématiques. Il est clair, notamment dans les Sciences biologiques, que d'autres directions ont été explorées telles les approches analogiques ou physiques, et sont même plus usitées comme les modèles discursifs.

Cette limitation s'explique par le caractère souvent très spécifique des modèles physiques. Quant au modèle discursif, nous avons admis qu'il relevait essentiellement de la logique de la pensée et que nous étions incompétents pour donner des conseils au lecteur dans ce domaine (il pourra cependant remarquer que les étapes de construction et de validation présentées ici se retrouvent en fait dans leur principe dans toutes les formes de modélisations). Nous avons, par contre, inclus l'ensemble des modèles mathématiques, la ségrégation faite souvent à l'égard des modèles de la statistique [SOLOMON (1977)] ne nous est pas apparue fondée sur des différences essentielles.

Ce point de vue est, sans doute, issu de la confrontation que ce travail a permise entre les approches Biométrique et Automatique de la modélisation. Si cette confrontation a nécessité certaines explications réciproques, notamment au niveau du vocabulaire des techniques et des objectifs, elle a montré une grande unité dans les concepts. En particulier, dans tous les domaines confrontés, la notion de modèle elle-même, l'importance accordée aux objectifs ainsi que des points plus techniques tels l'estimation et la validation sont apparus identiques.

Nous pensons que des tentatives de synthèse de ce type sont extrêmement profitables dans la mesure où, outre l'adaptation et la formulation des idées, elles conduisent à fondre ensembles des concepts ou des

méthodes souvent cantonnés aux domaines propres des compétences (citons par exemple certains aspects du raisonnement "statistique" classiques pour les biométriciens et des notions fondamentales telles celles de commande et d'observation habituelles aux automaticiens).

On peut considérer que la modélisation, en vue de l'analyse d'un système biologique, repose sur trois critères essentiels : l'objectif choisi, la nature des données et les contraintes expérimentales, et enfin, les possibilités techniques offertes par la classe de modèles retenue. Ainsi, l'un des écueils auquel se heurtent les tentatives de modélisation est la limitation drastique des observations, d'ailleurs souvent très bruitées, et des actions sur le système, il se situe donc au niveau expérimental. Enfin, signalons que l'aspect le plus souvent négligé concerne les possibilités techniques des modèles qui pourront influencer sur le choix des expérimentations futures et même conduire à une limitation des objectifs. En effet, on constate que trop souvent la démarche suivie consiste à accumuler d'abord des données, souvent au prix d'expérimentations longues et coûteuses, et de chercher ensuite les possibilités de traitement (qui ne sont pas assurées a priori). Alors qu'il serait plus judicieux de passer par une phase exploratoire pour cerner le phénomène étudié, puis à concevoir un plan expérimental en fonction du traitement auquel les données seront soumises (pensons notamment à l'efficacité des "plans expérimentaux" en agronomie).

S'il est clair que l'approche d'un problème biologique par une voie formelle comme la modélisation mathématique n'est pas toujours possible, ni même souhaitable (trop grande complexité de l'outil, par exemple), il n'en est pas moins vrai que cette démarche peut apporter des réponses intéressantes dans de nombreuses situations. Ainsi, comme le montre le chapitre 4, des outils efficaces permettent de contribuer à la résolution de problèmes biologiques (Analyse de structure, analyse de mécanismes...). Cependant, on retiendra l'aspect essentiellement méthodologique de l'apport de la modélisation.

Ainsi, on peut remarquer que, contrairement à d'autres domaines, la modélisation en biologie ne dérive pas, à l'heure actuelle, d'une théorie

globale. La seule exception à cette remarque paraît être une partie de la génétique qui utilise ce que l'on désigne habituellement par "génétique mathématique" où une formalisation propre a été développée [MALECOT (1966), KIMURA (1971)]. Ainsi, la modélisation procède essentiellement de théories "à objectifs" (statistique, analyse...).

Il nous semble que l'évolution, dans ce domaine, sera dominée par la part croissante prise par la construction de modèles et de techniques propres à la question biologique posée (on peut penser par exemple aux possibilités offertes par la statistique non paramétrique dans la construction de statistiques puissantes contre deux alternatives précises). Cependant, il faut souligner qu'une accession croissante des biologistes aux techniques de modélisation et leur concertation accrue avec les "biométriciens" est une condition essentielle à ce développement.

L'importance des modèles en biologie cellulaire et moléculaire ne semble pas niable actuellement ; cependant nous pensons que ce rôle est appelé à connaître un essor très important. Cette "conviction" s'appuie sur le récent développement des connaissances à ces deux niveaux. Par exemple, l'accumulation des données concernant les mécanismes de transcription et de traduction intervenant dans la biosynthèse des protéines a sensibilisé les biologistes à l'importance d'une modélisation synthétique de ces processus. Cette modélisation paraît d'autant plus essentielle que les techniques biologiques offrent la possibilité, nouvelle dans ce domaine, d'agir sur ce système par des "commandes" (traduction in vitro avec contrôle de certaines populations moléculaires : tRNA, facteurs...). D'autre part, l'utilisation croissante de l'introduction de gènes étrangers dans *Echerichia Coli* à des fins économiques (production d'insuline...) ou de recherche plus fondamentale (production de messagers purs pour en déterminer la séquence) pose clairement le problème de l'optimisation de cette production aussi bien au niveau de l'expression de ce gène que de la croissance de la souche ainsi constituée. Ainsi, il nous semble qu'outre les techniques purement biologiques ou biochimiques, qui interviennent dans cette activité croissante qu'est l'"ingénierie biologique", il y aura lieu d'étudier et de mettre au point des outils formels, répondant, par exemple, aux impératifs de production. On peut enfin remarquer que les expériences biologiques que nous venons de résumer

permettront sans doute une nouvelle approche par la modélisation des relations entre niveaux moléculaire et cellulaire [(A. Chang (1978)], car l'incorporation d'un gène apparaîtra comme une perturbation au niveau moléculaire dont le résultat sera une modification phénotypique (niveau cellulaire).

REFERENCES

- Bellman, R. & Aström, K.J. (1970) On structural identifiability *Math. Biosciences* 7, 329-339
- Benzecri, J.P. (1973) *L'analyse des données* (2 tomes). Dunod, Paris
- Blaineau-Peyretti, S. (1976) Etude physiologique des granules spécifiques du myocarde *Thèse 3^e cycle LYON*
- Boisvieux, J.F. (1977) Modélisation et commande des processus biologiques. Aspects théoriques et mise en oeuvre *Thèse PARIS VI*
- Bouligand, Y. (1969) Sur l'existence de pseudomorphes cholestériques chez divers être vivants *Journ. Phys.* 31, 1041-1054
- Bourdaud, (1974) Contribution à l'identification et à l'optimisation de procédés de fermentation discontinus. *Thèse 3^e cycle GRENOBLE*
- Chang, A. et al (1978) Phenotypic expression in E. Coli of a DNA sequence coding for mouse dihydrofolate reductase. *Nature*, 5681, 617-624
- Chassé, J.L. & Debouzie, D. (1974) Utilisation des tests de Kiveliovich et Vialar pour l'étude de quelques générateurs de nombres pseudo aléatoires *Rev. Stat. Appl.* 22, 83-90
- Chassé, J.L., Legay, J.M. & Pavé, A. (1977) Le modèle de Volterra-Kostitzin en dynamique des populations. Ajustement et interprétation des paramètres. *Ann. Zool. Ecol. Anim.*, 9, 3, 425-441
- Cherruault, T. & Loridan, P. (1977) *Modélisation et méthodes mathématiques en Biomédecine*. Masson, PARIS
- Cheruy, A. (1974) Analyse cinétique d'un biosystème: transphosphorylation catalysée par l'enzyme nucléoside diphosphate kinase *Thèse, GRENOBLE*
- Cheruy, A. (1975) Analyse d'une classe importante de biosystèmes: les réactions enzymatiques *RAIRO* 1, 37-69
- Cheruy, A. & Durand, A. (1978) Optimization of the erythromycin biosynthesis by controlling pH and temperature: theoretical and practical aspects *2nd Intern. Cong. on computer appl. in fermentation techn. PHILADELPHY, august 1978*
- Chessel, D. & Gautier, C. (1977) Des statistiques non paramétriques pour l'analyse des données binaires *Rev. Stat. Appl.* 25, 57-73

- Cobelli, C. & Romanin-Jacur, G. (1975) Structural identifiability of strongly connected biological compartmental systems *Medic. and Bio. Eng.* 13, 831-838
- Cobelli, C. & Romanin-Jacur, G. (1976) Controllability, observability and structural identifiability of multi-input and multi-output biological compartmental systems *IEEE trans. on Biom. Engin.* BME-23, 2, 93-100
- Cochran, W.G. & Cox, G.M. (1957) *Experimental Designs*, J. Wiley and Sons, NEW YORK
- Coll. Intern. du C.N.R.S. (1961) Le plan d'expérience Ed. CNRS PARIS
- Couble, P., Prudhomme, J.C. & Daillie, J. (1977) The biosynthesis of fibroin III: The golgi transport and the secretion of fibroin studied by electron microscope autoradiography *Exp. Cell Res.* 109, 138-148
- Duncam, M., Atkins, A.R. & Wyndham, C.H. (1972) Mathematical and physical models of thermoregulation *Essays in temperature regulation Bligh et Moore éditeurs AMSTERDAM* 37-54
- Estève, J., Nicolas, P. & Nigon, V. (1978) Fitting survival curves to theoretical models: goodness of fit tests and parameter estimation, application to the clonal survival of chlorella and the chloroplastic survival of Euglena Gracilis after irradiation *Mathem. Biosc.* 42, 3/4, 279-298.
- Estève, J. & Schiffers, E. (1976) Discussion et illustration de quelques méthodes d'analyse longitudinale *Proceed. of the 9th Int. Biometric Conf.* BOSTON 343-480
- Fages, R. (1978) La notion de dispersion en classification automatique *Journée de Class. Autom. NICE Mai 1978*
- Fayard, J.M. (1976) Etude biométrique des glandes séricigènes en rapport avec la production de soie chez Bombyx MoriLL. *Thèse 3^e cycle LYON*
- Fayard, J.M. (1978) Introduction à une analyse génétique et embryologique des populations cellulaires des glandes séricigènes chez le vers à soie Bombyx Mori *Ann. Genet. Sel. Anim.* sous presse
- Fisher, R.A. (1950) Review of the theory of inbreeding *Ann. J. Human Genet.* 2, 97-100
- Fletcher, R. (1970) A new approach to variable metric algorithm *Comp. Journ.* 13, 317-322
- Fu, K.S. (1974) *Syntactic methods in pattern recognition Acad. Press. NEW YORK*

- Gautier, C. & Grantham, R. (1976) Evolution moléculaire *Rapp. annuel du GLRBC LYON* 54-77
- Garfinkel, P., Rutledge, J.D. & Higgins, J.J. (1961) Simulation and analysis of biochemical systems. I Representation of chemical kinetics *Comm. ACM* 4 559-562
- Garfinkel, D. (1968) A machine independent language for the simulation of complex chemical and biochemical systems. *Comput. in Biomed. research* 2 31-44
- Garfinkel, D. (1977) Simulating biochemical activity in physiological systems *Simulation* 28, 3, 193-196
- Girault, M. (1965) *Processus aléatoire*. Dunod PARIS
- Harris, T.E. (1963) *The theory of branching processes*. Springer Verlag BERLIN
- Harwell (1973) *Harwell subroutine library. Theoretical Physics Division U. K. , A.E.E. Research Group*
- Hemon, D., Angeller, N. & Bouteille, M. (1978) Analyse des sources de variabilité en autoradiographie à haute résolution — Application à l'étude de la transcription nucléolaire *Biométrie et Biologie cellulaire* (Legay, J.M. et Tomassone, R. éditeurs) Soc. F. Biom.
- Herman, G.T. & Rozenberg, G. (1975) Developmental systems and languages *North Holl. Pub. , ELSEVIER*
- Hill, M.O. (1974) Correspondence analysis : a neglected multivariate method *Appl. Stat.* 23,3, 340-354
- Iliadis, A., Cheruy, A., Daver, J. & Desnoyers, P. (1977) Modélisation du système coagulolytique *C. R. Acad. Sci.* 284, 2411-2414
- Iliadis, A., Cheruy, A., Daver, J. & Desnoyers, P. (1978) Modelling of the coagulolytic system. I Methodology. II. Contribution to system analysis. III. Experimental results 4th *Intern. Conf. on synthetic fibrinolytic trombolytic agents KARLOVY-VARY (Tchécoslovaquie)*
- Jacquez, J.A. (1972) *Compartmental analysis in biology and medicine*. ELSEVIER, NEW YORK

- Jagers, P. (1975) *Branching processes with biological applications*. J. Wiley and Sons LONDON
- Jolivet, E. (1978) Introduction aux modèles aléatoires en biochimie *Biométrie et Biologie cellulaire* Legay, J.M. et Tomassone, R. ed., Soc. F. Biom.
- Karlin, S. (1969) *Initiation aux processus aléatoires*. Dunod PARIS
- Kimura, M. & Ohta, T. (1971) *Theoretical aspect of population genetic*. PRINCETON univ. Press
- Kostitzin, V.A. (1937) *Biologie mathématique*. Armand Colin PARIS
- Lebart, L., Morineau, A. & Tabard, N. (1977) *Techniques de la description statistique*. Dunod PARIS
- Lebreton, J.D. (1973) Introduction aux modèles mathématiques de la dynamique des populations *Informatique et Biosphère* 75-116
- Lebreton, J.D. (1978) Un modèle probabiliste de la dynamique des populations de cigognes blanches (*Ciconia ciconia* L.) en Europe Occidentale *Biométrie et Ecologie* (Legay, J.M. et Tomassone, R. editeurs) 277-344
- Ledley, R.S., Rotolo, L.S., Golab, T.J., Jacobsen, J.D., Ginsburg, M.D. & Wilson, J.B. (1965) FIDAC: Film input to digital automatic computer and associated syntax directed pattern recognition programming system *Optical and Electro-optical Inform. Proc.*, MIT Press, Cambridge Mass. 591-614
- Lee, H.C. & Fu, K.S. (1972) A syntactic pattern recognition system with learning capability *Int. Symp. Comput. Inform. Sci.* 47th MIAMI (12/72)
- Legay, J.M. (1968) Elements d'une théorie générale de la croissance d'une population *Bull. Math. Bioph.* 30, 33-46
- Legay, J.M. (1973) La méthode des modèles, état actuel de la méthode expérimentale *Informatique et Biosphère* 5-73
- Lehman, E.L. (1959) *Testing statistical hypothesis* J. Wiley and Sons NEW YORK
- Lindenmayer, A. (1971) Developmental systems without cellular interaction, their languages and grammars *J. Theor. Biol.* 30, 455-484
- Lukasic, A. (1974) Sur l'identification de procédés de fermentation discontinus *Thèse 3^e cycle GRENOBLE*
- Lukasic, A., Foulard, C. & Cheruy, A. (1975) Identification de procédés de fermentation discontinus: application au procédé de fabrication de l'Erythromycine. *Journées AFCET "Automatisation des processus de fermentation"*
- Malecot, G. (1966) *Probabilités et hérédité* INED cahier n°47 P.U.F. PARIS

- Mandelbrot, B. (1978) Les objets fractals *La Recherche* 9, 5-13
- Mathis, J.H., Gerald, K.B. (1977) Compartmental models analysis in Ecology
Intern. Stat. Ecol. Program, Aout 1977 BERKELEY
- Meyer, J.A. (1978) Les langages de simulation, CSMP GPSS et SIMSCRIPT,
illustrés par divers exemples en Biologie cellulaire *Biométrie et
Biologie cellulaire (Legay, J.M. et Tomassone, R. éditeurs) Soc. F. Biom.*
- Monod, J. (1942) Recherches sur la croissance de cultures bactériennes.
Thèse, Hermann PARIS
- Pavé, A. (1977) Estimation of tracer waiting time in morphological structures
using a catenary irreversible compartmental system *Exp. Cell. Res.* 109,
149-150
- Pavé, A. (1977) Description languages. A tool for mathematical modelling.
Biological and ecological applications *First world Conf. on Math. at the
service of Man BARCELONA, Juillet 1977* 27p
- Pavé, A. (1979) Dynamics of macromolecular populations : the changes of RNA
quantities in the silkgland during the last larval instar. *Biochimie*, 61, 263-273.
- Pavé, A. (1978) Introduction aux modèles morphologiques et morphogénétiques
dérivés de la théorie des langages. *Biométrie et Biologie cellulaire (Legay,
J.M. et Tomassone, R. éditeurs) Soc. F. Biom.*
- Pavé, A. & Pagnotte, Y. (1977) An approach to computer aided design a
tool for mathematical modelling in biology and ecology. *Comp. Biol. Med.* 7,
301-310
- Rao, C.R. (1973) *Linear statistical inference and its applications. J. Wiley
and Sons, NEW YORK*
- Rashevsky, N. (1960) *Mathematical biophysics. Physicomathematical foundations
of biology. Dover Publish., NEW YORK*
- Ross, G.J.S. (1970) The efficient use of function minimization in non linear
maximum likelihood estimation. *J. R. Stat. Soc. C* . 19, 205-221
- Ross, G.J.S. (1975) Simple non linear modelling for the general user. *I.S.I.*
1-9, Varsovie

- Seber, G.A.F. (1973) *The estimation of animal abundance*. Griffin, LONDON
- Sheppard, C.W. (1962) *Basic principles of the tracer method*. J. Wiley, NEW YORK
- Solomon, D.L. (1977) On a paradigm for mathematical modeling. *Int. Stat. Ecol..Program, Juillet 1977, COLLEGE STATION*
- Steinhorst, K.R. (1977) Validation and sensitivity analysis of large system models. *Int. Stast. Ecol. Program, Aout 1977, BERKELEY*
- Thom, R. (1973) A global dynamical scheme for vertebrate embryology. *Lect. on Math. in the life science (American Math. Society, PROVIDENCE)*. 5, 3-95
- Thom, R. (1974) *Modèles mathématiques de la morphogénèse*. Union G. Ed. (70/18), PARIS
- Teissier, G. (1948) La relation d'allométrie, sa signification statistique et biologique. *Biometrics*, 4, 14-53
- Vitanyi, C.M.B. (1974) Growths of strings in context dependant L. systems. *Lect. Notes in comput. Sciences*, 15, 104-126
- Von Bertalanffy, L. (1973) *Theorie générale des systèmes*. Dunod, PARIS
- Wright, R.D. (1972) Validating dynamic models, an evaluation of test of predictive power. *Proceed. of Summer Comput. Sim. Conf.*, 1286-1294
- Yates, S.E., Rennam, R.D., Urquhart, J., Dallman, M.S., Lee, C.C. & Halpern, N. (1968) Continuous systems model of adrenocortical functions. *System theory and biologie (Mesarovic, M.D. éditeur) Springer Verlag BERLIN*

